

Supplemental Results

AGRIS (AtTFDB) and DATF comparisons

There are several reasons for this, as enumerated below. In several cases, different names have been chosen to describe the same family (e.g. REM family in AtTFDB is called B3 family in DATF). Taking these cases into account, the two databases share 31 TF families in common. Another common cause of differences is alternative groupings of transcription factors into families. For example, the CCAAT family in DATF is described in four families in AtTFDB, reflecting subfamilies (CCAAT-DR1, CCAAT-HAP2, CCAAT-HAP3, CCAAT-HAP5), while the SRS family in DATF is included in the C2H2 family of AtTFDB.

There are also families unique to one database or the other. In AtTFDB, there are three unique families that include TFs not found elsewhere in DATF. These include the MYB-related family with two genes (Kirik and Baumlein, 1996) ; curated by Dr. Bernd Weisshaar and Dr. Ralf Stracke Bielefeld University), the BBR-BPC family of seven genes (Frustaci et al., 1995; Meister et al., 2004; Kooiker et al., 2005); curated by Dr. Dierk Wanke University of Tübingen), and the Orphan family. In AtTFDB, the Orphan family includes two TFs for which no other family members have been identified in Arabidopsis. This includes the *LFY* [*At5g61850*; (Weigel et al., 1992)] gene and the *STO* [*At1g06040*; (Lippuner et al., 1996; Nagaoka and Takano, 2003)] gene. *LFY* is included as a family in DATF; however, this family also includes two other loci, *At5g61860* and *At5g61870*, which do not appear as loci in TAIR and seem to encode identical genes to *At5g61850* according to DATF sequences (data not shown). DATF does not have *STO* in the database. DATF includes fourteen families not found in AtTFDB. These can be categorized into three classes: those including genes that likely encode transcriptional co-factors, but the encoded proteins have not been shown to have DNA binding activity or contain a domain shown to have DNA binding activity in another system (AS2, AUX-IAA, GIF, LIM, LUG, MBF1, NZZ, SAP, TAZ, and ULT); those with non-specific DNA binding activity (HMG, PcG, and PLATZ); and those with little information on function, but no real proof they act as DNA binding TFs (FHA). The genes found in these families do not meet our criteria for inclusion as TFs, and close inspection of the information

found in DATF also supports the fact that these proteins have not been shown to be TFs by our definition.

In addition to differences in families, there are differences in overall numbers of TFs found in certain families between the two databases. At least ten families currently have the same number of TFs in both databases (Alfin-like, BZR/BES, CPP, E2F-DP1, GeBP, GRF, Whirly, SBP, VOZ, and WRKY). Other families vary in number, with some having more TFs in DATF, and some having more in AtTFDB. Some of the differences between some families again depends on the different definitions of what a TF is. For example the JUMONJI family in DATF consists of thirteen genes, two of which *EARLY FLOWERING6 (ELF6; At5g04240)* and *RELATIVE OF EARLY FLOWERING6 [REF6; At3g48430; (Noh et al., 2004)]*, are included in the C2H2 family in AtTFDB. The other eleven genes can be broken down into two classes: those encoding proteins with a jmjN domain, a jmj C domain, and a C5HC2 DNA binding domains (*At1g08620, At1g30810, At2g34880, At2g38950, At4g20400*), which are most likely TFs and are included in AtTFDB in the JUMONJI family; and those encoding proteins with only a jmjC domain and no known DNA binding domain (*At1g63490, At3g20810, At4g00990, At4g21430, At5g46910, and At5g63080*), which are not included in AtTFDB.

Literature Cited

- Frustaci JM, Sangwan I, O'Brian MR** (1995) *gsa1* is a universal tetrapyrrole synthesis gene in soybean and is regulated by a GAGA element. *J Biol Chem* **270**: 7387-7393
- Kirik V, Baumlein H** (1996) A novel leaf-specific myb-related protein with a single binding repeat. *Gene* **183**: 109-113
- Kooiker M, Airoidi CA, Losa A, Manzotti PS, Finzi L, Kater MM, Colombo L** (2005) BASIC PENTACYSSTEINE1, a GA binding protein that induces conformational changes in the regulatory region of the homeotic Arabidopsis gene SEEDSTICK. *Plant Cell* **17**: 722-729
- Lippuner V, Cyert MS, Gasser CS** (1996) Two classes of plant cDNA clones differentially complement yeast calcineurin mutants and increase salt tolerance of wild-type yeast. *J Biol Chem* **271**: 12859-12866
- Meister RJ, Williams LA, Monfared MM, Gallagher TL, Kraft EA, Nelson CG, Gasser CS** (2004) Definition and interactions of a positive regulatory element of the Arabidopsis INNER NO OUTER promoter. *Plant J* **37**: 426-438

- Nagaoka S, Takano T** (2003) Salt tolerance-related protein STO binds to a Myb transcription factor homologue and confers salt tolerance in Arabidopsis. *J Exp Bot* **54**: 2231-2237
- Noh B, Lee SH, Kim HJ, Yi G, Shin EA, Lee M, Jung KJ, Doyle MR, Amasino RM, Noh YS** (2004) Divergent roles of a pair of homologous jumonji/zinc-finger-class transcription factor proteins in the regulation of Arabidopsis flowering time. *Plant Cell* **16**: 2601-2613
- Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM** (1992) LEAFY controls floral meristem identity in Arabidopsis. *Cell* **69**: 843-859