

AlfalfaTFDB: a database of computationally predicted alfalfa transcription factors

(<http://plantpathology.ba.ars.usda.gov/alfalfatfdb.html>)

Olga A. Postnikova, Jonathan Shao and Lev G. Nemchinov

United States Department of Agriculture, Beltsville Agricultural Research Center, Molecular Plant Pathology Laboratory, 10300 Baltimore Avenue, Beltsville, MD 20705.

Comprehensive plant TF databases do not contain any information on *Medicago sativa*, the most extensively cultivated forage legume in the world. In this work, we have performed an *in silico* analysis of transcriptome data generated in our lab and publicly available from other sources to reveal and systematize alfalfa transcription factors (1). We first used RNA-seq data from our previous study (2) to generate two transcriptomes with Velvet 1.2.03 and Oases 0.2.06 genome assemblers and *k*-mer values 35 and 45. They were combined into one synthetic assembly that was further merged with alfalfa transcriptome collections available at the Legume Information System. All redundancies were removed via CD-HIT tool. Transcription factors were predicted based on the best BLASTX hits against the comprehensive PlantTFDB and against the nr protein database maintained by NCBI. Further analyses and annotations were performed by InterProScan, a protein domain identifier and by CDD, a conserved domain search. Phylogenetic trees were deduced from domain alignments of *M. sativa* and *M. truncatula* TF families with Arabidopsis TFs. Transcriptome profiling enabled prediction of 983 putative TFs. Among them 617 TFs were shared with *M. truncatula* TFs and 366 TFs were traced to species other than *M. truncatula*, indicating their possible uniqueness to alfalfa. All data were assembled into a simple open-access database, AlfalfaTFDB. Database provides users with an overview of all annotated alfalfa TF families, a reference to the description of each family, identified transcript and protein sequences and predicted ORFs. Transcription factor datasets can be retrieved using a convenient search tool. We hope that integrated TF repertoires of *Medicago sativa* will serve as a useful resource for alfalfa community.

1. Postnikova OA, Shao J and Nemchinov LG. *In silico* identification of transcription factors in *Medicago sativa* using available transcriptomic resources. Mol Genet Genomics. 2014 Feb 21. [Epub ahead of print, PMID: 24556904].
2. Postnikova OA, Shao J and Nemchinov LG. Analysis of the alfalfa root transcriptome in response to salinity stress. Plant Cell Physiol. 2013, 54:1041-55.