

Applying TFmodeller to plant transcription factors: structural analysis and prediction of DNA motifs

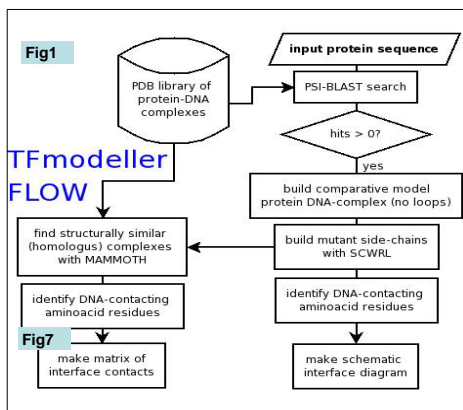


Bruno Contreras-Moreira
bcontreras@eead.csic.es
Estación Experimental de Aula Dei/CSIC
Av.Montañana 1005, 50059 Zaragoza (España)

Plant genomes house vast repertoires of transcription factors (TFs), What are their functions? For most TFs we know very little. However, experimental work on many other TFs show that they can be involved in important processes, such as plant growth and differentiation. Here I take TF collections for *Arabidopsis thaliana*, poplar, rice (var. indica & japonica) and *Vitis vinifera* in order to run the TFmodeller software (www.ccg.unam.mx/tfmodeller) and try to answer a few questions:

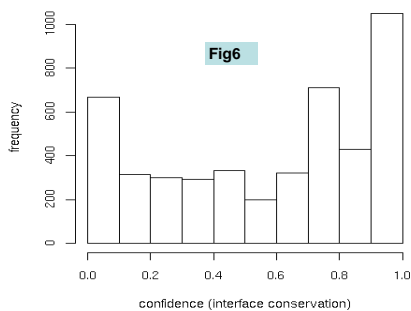
- 1) What are the most common folds of plant TFs? Which species provide most fold templates?
- 2) How many TFs can be modelled using the current knowledge about protein-DNA interfaces contained in the Protein Data Bank?
- 3) Can we predict DNA motifs with confidence?

TFmodeller example: *Arabidopsis* AT4G27950

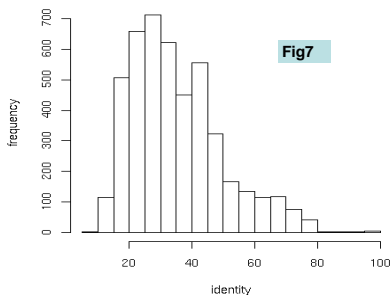


Benchmark on representative plant genomes

confidence (interface conservation) (4610 data points)



identity (4610 data points)



n_of_modelableTFs = 3925 (39%) n_of_nonmodelableTFs = 6220
n_of_templates used = 125

Protein Data Bank template	times used	source	fold classification (SCOP)
1h89_C	1716	HOMO SAPIENS	Homeodomain-like;
2t13_A	1007	MUS MUSCULUS	Zinc finger;
1gcc_A	899	ARABIDOPSIS THALIANA	DNA-binding domain;
1h8a_C	890	HOMO SAPIENS	Homeodomain-like;
1nkp_B	554	HOMO SAPIENS	HLH, helix-loop-helix DNA-binding domain;
[...]			
1n6j_A	421	HOMO SAPIENS	SRF-like;
1egw_A	420	HOMO SAPIENS	SRF-like;
1dh3_A	404	MUS MUSCULUS	Luciferase zipper domain;
1mnm_A	383	SACCHAROMYCES CEREVISIAE	Homeodomain-like;
1t2k_D	364	HOMO SAPIENS	Winged helix DNA-binding domain;
[...]			
4gat_A	166	EMERICELLA NIDULANS	Glucocorticoid receptor-like;
[...]			
2up1_A	88	HOMO SAPIENS	RNA-binding domain, RBD;
1jfi_B	88	HOMO SAPIENS	Histone-fold;
[...]			
1g0_D	38	HOMO SAPIENS	lambda repressor-like DNA-binding domains;
1kq_A	36	DROSOPHILA MELANOGASTER	ARID-like;
1j4w_A	36	HOMO SAPIENS	Eukaryotic type KH-domain (KH-domain type I);
1ubd_C	33	HOMO SAPIENS	C2H2 and C2HC zinc fingers;
1hry_A	33	HOMO SAPIENS	HMG-box;
[...]			
1gd2_E	24	SCHIZOSACCHAROMYCES POMBE	Translation proteins SH3-like domain;
[...]			
1a02_F	7	HOMO SAPIENS	p53-like transcription factors;

Table1

Figures 1-5 show how TFmodeller works, by using a transcription factor annotated at TAIR (<http://www.arabidopsis.org/>). **Table 1** shows a selection of the most commonly used modelling templates when working with plant transcription factors. Note that homeodomain folds and zinc fingers dominate. **Figure 6** shows the confidence of DNA motif predictions made by TFmodeller. Note that a large number of TFs in our benchmark have completely conserved protein-DNA interfaces, suggesting that their specifically recognized motifs are probably conserved as well. **Figure 7** shows that most modelling templates, the Protein Data Bank structures used by TFmodeller, fall in the 20-80 range of % sequence identity.

references and resources

- Contreras-Moreira, B., Branger, P.A. & Collado-Vides, J. (2007). TFmodeller: comparative modelling of protein-DNA complexes. *Bioinformatics*, doi: 10.1093/bioinformatics/btm148.
- Contreras-Moreira, B. & Collado-Vides, J. (2006). Comparative footprinting of DNA-binding proteins. *Bioinformatics*, 22(14): e74.
- Plant Transcription Factor Database (<http://planttfdb.cbi.pku.edu.cn>)