

AtSNP_TATAdb: a knowledge base of uniform guess of how promoter proximal SNPs alter Arabidopsis gene expression

Zolotareva K.^{1*}, Chadaeva I.¹, Khandaev B.^{1, 2}, Podkolodnyy N.^{1, 3}, Tverdokhleby N.¹, Ponomarenko P.¹, Sharypova E.¹, Filonov S.¹, Kondratyuk E.^{1, 4, 5}, Zemlyanskaya E.^{1, 2}, Rasskazov D.¹, Bogomolov A.^{1, 2}, Savinkova L.¹, Podkolodnaya O.¹, Kolchanov N.^{1, 2}, Ponomarenko M.¹

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia

⁴ Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

⁵ Siberian Federal Scientific Centre of Agro-BioTechnologies, RAS, Krasnoobsk, Novosibirsk region, Russia

* ka125699ri@yandex.ru

Key words: plant; gene; promoter; TATA box; TATA-binding protein; SNP; expression alteration; genome-wide analysis; *in silico* estimate; *in vivo* test; handmade annotation; target-assisted breeding

Motivation and Aim: The mainstream of the post-genome target-assisted plant breeding is biofortification as high-throughput phenotyping along with genome-based selection.

Methods and Algorithms: Therefore, in this work, we used our previously developed Web-service Plant_SNP_TATA_Z-tester [1] to run a uniform *in silico* analysis of the transcription alterations of 54,013 protein-coding transcripts from 32,833 *Arabidopsis thaliana* L. genes caused by 871,707 single nucleotide polymorphisms (SNPs) within the 90 bp proximal promoter regions. We took their DNA sequences and SNPs from the databases Ensembl Plant [2] and TAIR [3], respectively, as depicted in Fig. 1.

Results: The analysis has identified 54,993 SNPs, each of which can significantly either upregulate or downregulate a proper Arabidopsis gene by means of an alteration in TATA-binding protein (TBP) binding affinity to the promoters carrying this SNP. The existence of these SNPs in highly conserved proximal promoters may be explained as intraspecific diversity kept by the stabilizing natural selection. To support this, using the database PubMed [4] we hand-annotated papers on some of the Arabidopsis genes possessing these SNPs or on their orthologs in other plant species and demonstrated the effects of changes in these gene expressions on plant vital traits. We integrated *in silico* estimates of the TBP-promoter affinity in the AtSNP_TATAdb knowledge base [5] (https://www.sysbio.ru/AtSNP_TATAdb/) and showed their significant correlations with independent *in vivo* experimental data, as exemplified in Fig. 2.

Conclusion: These correlations appeared to be robust to variations in statistical criteria, genomic environment of TATA box regions, plants species and growing conditions.

Funding: The study is supported by the Government Budget Project FWNR-2022-0020.

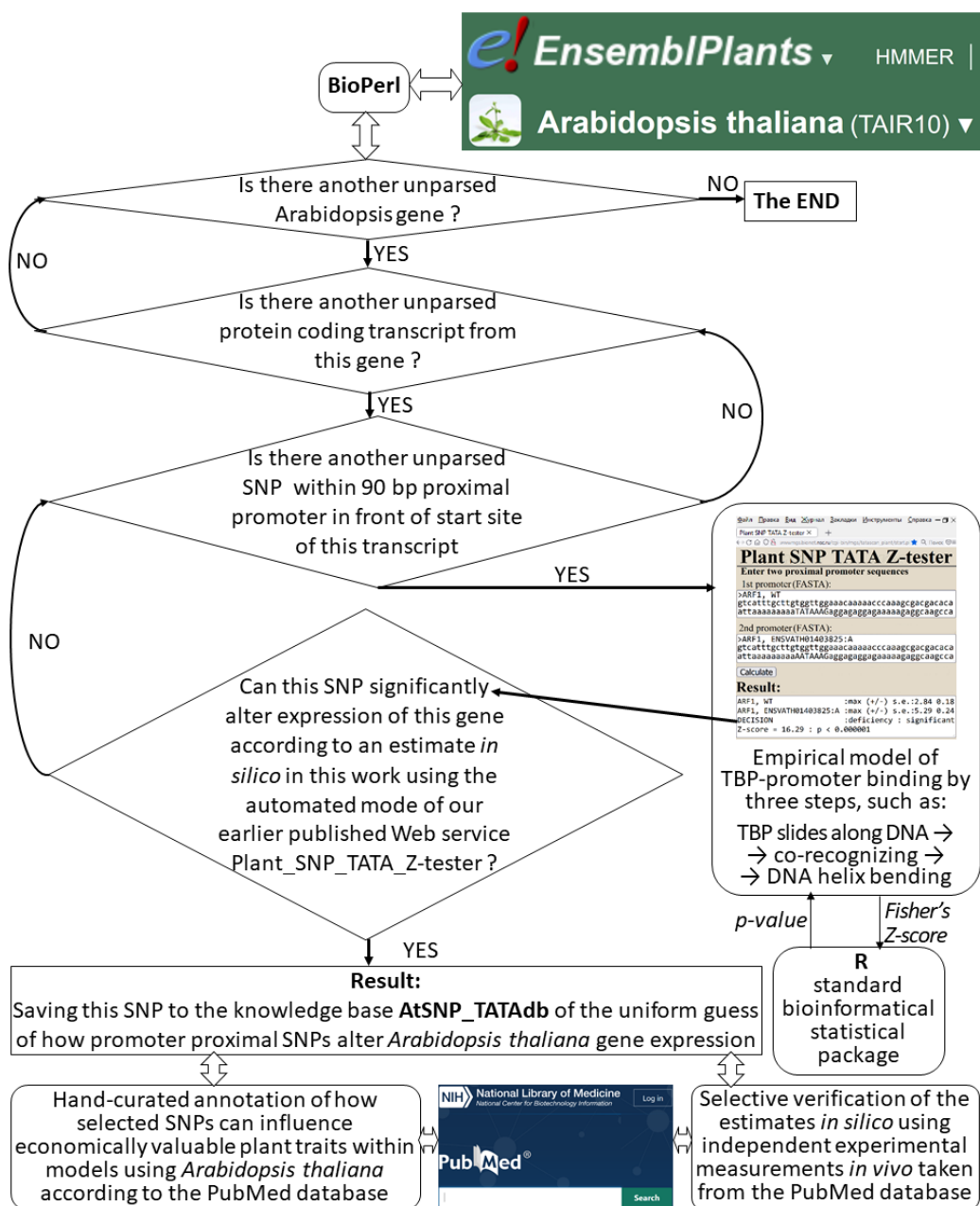


Fig. 1. Flowchart of AtSNP_TATAdb knowledge base [5] development by processing genome-wide information from both database Ensembl Plant [2] and TAIR [3] by first using Plant_SNP_TATA_Z-tester [1] and, after that, selectively annotating some of the SNPs by a search in the PubMed database [4] using information about where, when and under what conditions changes in expression of these genes or their homolog genes were observed in agricultural studies. Finally, we selectively verified estimates *in silico* using independent experimental data *in vivo* taken from the PubMed database [4]



AtSNP_TATAdb

START (HOMOLOG)

START (USUAL)



DOWNLOAD DB

Show information

Information block

☒ Information about the SNP (minor allele)

☒ Information about influence plant gene of its homolog to the feature

Single Nucleotide Polymorphism:

☐ A → C
 ☐ A → G
 ☒ A → T
 ☒ C → A
 ☐ C → G
 ☐ C → T
 ☐ G → A
 ☒ G → C
 ☐ G → T
 ☒ T → A
 ☐ T → C
 ☒ T → G

Find gene:

Find feature:

Chromosome:

Page Number:

ARF1

Feature

1

<< 1 / 3 >> OK

Gene	dtSNP ID	proximal promoter DNA sequence: ancestral allele minor allele	TSS SNP	KdtSEM, nm, ancestral allele	KdtSEM, nm, minor allele	Z-score	P-value	Minor allele expression change (Expression)	Gene-homolog	Expression change	Feature	Effects of expression	Health sign	PMID			
ARF1	FNSVATH01403825	#gtgcgcacacacttataaaaaaaTataaaggaggaggagagaaagagcgaaccca agtcgcgcacacacttataaaaaaaTataaaggaggaggagagaaagagcgaaccca	-33	T>A	2.8440	18	5.2940	24	16.20	0.000001	deficiency	this gene	deficiency	somatic embryogenic response	within plant somatic embryogenesis models using Arabidopsis thaliana (ecotype Columbia) treated with auxin and, next, tested with qPCR: ARF1 deficit is a molecular marker for somatic embryogenesis as well as mutational ARF1-downregulation can impair somatic embryogenic response	worsened	28255787
ARF1	ENSVATH01403825	#gtgcgcacacacacttataaaaaaaAataaaggaggaggagagaaagagcgaaccca agtcgcgcacacacacttataaaaaaaTataaaggaggaggagagaaagagcgaaccca	-35	A>T	2.8440	18	1.6540	11	11.94	0.000001	excess	this gene	excess	chlorophyll content	within plant development models using Cymbidium as an economically important genus of flowering orchids cultivated in China: ARF1 overexpression in leaves is a molecular marker for both leaf senescence and decreased chlorophyll content	worsened	35291484
ARF6	FNSVATH01403825	#gtgcgcacacacttataaaaaaaTataaaggaggaggagagaaagagcgaaccca agtcgcgcacacacttataaaaaaaTataaaggaggaggagagaaagagcgaaccca	-33	T>A	2.8440	18	5.2940	24	16.20	0.000001	deficiency	ARF6	deficiency	lower buds opening	within plant development models using arf6 double loss-of-function mutant Arabidopsis lines: arrested flowers just before wild-type flower buds normally	worsened	17021043

Fig. 2. A sample entry in the AtSNP_TATAdb documents [5] on two SNPs ENSVATH01403825:A and ENSVATH01403824:T of the *ARF1* (*Auxin Response Factor 1*) gene, which can significantly downregulate and upregulate its expression, as calculated in this work together with their annotation using the database PubMed [4] (PMIDs: 28255787, 35291484, 17021043)

References

1. Rasskazov D., Chadaeva I., Sharypova E. et al. Plant_SNP_TATA_Z-Tester: A Web service that unequivocally estimates the impact of proximal promoter mutations on plant gene expression. *Int J Mol Sci.* 2022;23(15):8684. doi 10.3390/ijms23158684
2. Contreras-Moreira B., Naamati G., Rosello M. et al. Scripting analyses of genomes in Ensembl Plants. *Methods Mol Biol.* 2022;2443:27-55. doi 10.1007/978-1-0716-2067-0_2
3. Berardini T.Z., Reiser L., Li D., Mezheritsky Y., Muller R., Strait E., Huala E. The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis.* 2015;53(8):474-485. doi 10.1002/dvg.22877
4. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford).* 2011;2011:baq036. doi 10.1093/database/baq036
5. Bogomolov A., Zolotareva K., Filonov S. et al. AtSNP_TATAdb: candidate molecular markers of plant advantages related to single nucleotide polymorphisms within proximal promoters of *Arabidopsis thaliana* L. *Int J Mol Sci.* 2024;25(1):607. doi 10.3390/ijms25010607