



## An *in silico* comparative genomic report on transcription factors in three *Arabidopsis* species Un reporte genómico comparativo *in silico* sobre los factores de transcripción en tres especies de *Arabidopsis*

Seyed Mehdi Jazayeri<sup>1\*</sup>, Mahtab Pooralinaghi<sup>2</sup>, Ronald Oswaldo Villamar Torres<sup>3,4</sup>, Luz Cecilia García Cruzatty<sup>5</sup>

<sup>1</sup> Universidad Nacional de Colombia, Bogotá, Colombia

<sup>2</sup> Universidad de Payame Noor, Karaj, Irán

<sup>3</sup> Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador (SENESCYT), Ecuador

<sup>4</sup> Université de Montpellier, Montpellier, Francia

<sup>5</sup> Facultad de Agronomía, Universidad Técnica de Manabí, Campus La Teodomira, Portoviejo, Manabí 130105, Ecuador

\*Corresponding author: [smjazayeri@unal.edu.co](mailto:smjazayeri@unal.edu.co)

Rec.: 19.02.2018. Acept.: 10.05.2018

Publicado el 2 de Julio de 2018

### Abstract

Transcription factors (TF) are the elements, which regulate gene expression. Regulatory function of TFs play an important role in plant biological processes and mechanisms. They may interconnect with other transcription factors or functional genes to modulate their expression in response to an internal/external factor like life cycle stage, growth, development and stress. *Arabidopsis* is the well-known and the most used model organism. Transcription factors of three *Arabidopsis* species including *A. halleri*, *A. lyrata* and *A. thaliana*, were compared. basic/helix-loop-helix (bHLH) with 220 TFs was the most abundant family among three *Arabidopsis* species while MYB and MYB related families considering as a whole group were more than bHLH with 308 TFs. No STERILE APETALA (SAP) TF homolog was found for *A. halleri*. The common transcription factors among three species were 4,172 grouped in 1,212 clusters. The species-specific clustered TFs were 12, 30 and 58 for *A. halleri*, *A. lyrata* and *A. thaliana* respectively. Eight hundred ninety two single-copy gene clusters those have one gene copy from each species, i.e. 2,676 genes, were listed. Four hundred forty five TF singletons were not clustered and are unique among three species. For clustered TF belonging to each species, GO terms and SwissProt hits showed that *A. halleri* has two species-specific TFs involved in heavy metal response including Zinc finger protein AZF2 and two-component response regulator ARR11 while for *A. lyrata* specific TFs are involved in stress response and plant development. *A. thaliana* specific clustered TFs work on plant flower development and acclimation.

**Keywords:** *Arabidopsis*, species, bHLH, ZF2, ARR11.

### Resumen

Los factores de transcripción (FTs) son los elementos que regulan la expresión génica. La función reguladora de los FTs desempeña un papel importante en los procesos y mecanismos biológicos de las plantas. Los FTs pueden interconectarse con otros FTs o genes funcionales para modular su expresión en respuesta a un factor interno / externo como la etapa del ciclo de vida, el crecimiento, el desarrollo y el estrés. *Arabidopsis* es el organismo modelo más conocido y más utilizado en investigaciones biológicas en plantas. Se compararon los FTs de tres especies de *Arabidopsis*, incluyendo *A. halleri*, *A. lyrata* y *A. thaliana*. La familia basic / helix-loop-helix (bHLH) con 220 FTs fue la más abundante entre las tres especies de *Arabidopsis*, mientras que, las familias MYB y MYB-related, considerandolas como un grupo completo, tuvieron una mayor representación comparadas con bHLH que presentó 308 FTs. No se encontró homología para el factor de transcripción SAP en *A. halleri*. Los FTs comunes entre las tres especies fueron 4,172 agrupados en 1,212 conglomerados (clusters). Los FTs agrupados como específicos de especie fueron 12, 30 y 58 para *A. halleri*, *A. lyrata* y *A. thaliana*, respectivamente. Se enumeraron 892 agrupamientos de genes de una sola copia (single-copy) en cada especie, es decir, 2,676 genes. Cuatrocientos cuarenta y cinco FTs con una sola copia (singletons) no fueron agrupados y son únicos entre las tres especies. Para los FTs agrupados y perteneciente a cada especie, los términos GO y SwissProt mostraron que *A. halleri* presentó dos FTs específicos implicados en la respuesta de metales pesados, incluyendo la proteína con dedos de zinc (zinc finger) AZF2 y el regulador de respuesta de dos componentes (two-component response regulator) ARR11, mientras que, para *A. lyrata* están implicados los FTs específicos en la respuesta al estrés y el desarrollo de la planta. Para *A. thaliana* los FTs específicos agrupados trabajan en el desarrollo de la flor y la aclimatación.

**Palabras clave:** *Arabidopsis*, especies, bHLH, ZF2, ARR11.

## Introduction

Transcription factors are regulators, which modulate genes whose expression is activated or suppressed regarding to time, space and condition. They possess normally a DNA-binding region, an oligomerization site, a transcription-regulation domain, and a nuclear localization signal. However, there are exceptions for this typical structure (Liu *et al.*, 1999). The DNA-binding regions are normally found in cis-regulatory elements; i.e. specific sequences in the promoters of their target genes (Franco-Zorrilla *et al.*, 2014). They can switch on and off genes and link genes in downstream and upstream of a biological pathway acting to regulate desired reactions. In *A. thaliana*, 6-10% of all genes encode TFs while in other organism is less as seen only 3% in *Drosophila melanogaster* and 5% in humans. In *A. lyrata* and *A. halleri* TFs form around 5-6% of all genes. The importance of TFs has been disclosed via studying mutants affected in spatio-temporal transcription regulation (Pireyre and Burow, 2015). As plants are sessile, this relatively high amount of TFs may allow them to react to their changing environment and adapt to undesirable condition rapidly via regulation of gene expression, RNA transcription and protein translation. The TF involvement in various biological processes and mechanisms in plants and their regulatory function make it mandatory to study them to disclose their interaction with other TFs and genes.

There are various plant TF databases. Plant TFs are distributed in 58 families according to PlantTFDB (Jin *et al.*, 2017) (<http://planttfdb.cbi.pku.edu.cn/>) or 63 families according to PlnTFDB (Pérez-Rodríguez *et al.*, 2010) (<http://plntfdb.bio.uni-potsdam.de/v3.0/>). In other database they are listed in 70 groups (<http://itak.feilab.net/>) (Zheng *et al.*, 2016). The basic helix-loop-helix (bHLH) family is the biggest and SPOROCTELESS/NOZZLE (NZZ/SPL) is the smallest TF in terms of the number of family members.

*Arabidopsis* is the well-known and mostly ever used plant model, particularly *A. thaliana*. *A. thaliana* has highly contributed in providing the information and data for almost all of genomes even human and animals, as it is the first ever-published living organism and the most annotated among plants. Its genome has restructured biology in terms of Omics, mainly genomics and transcriptomics and revealed many mechanisms those had not previously been discovered. *Arabidopsis* has 11 accepted species names (<http://www.theplantlist.org/browse/A/Brassicaceae/Arabidopsis/>) with different chromosome number as five (*A. thaliana*), eight (*A. lyrata* and *A. halleri*), thirteen (*A. suecica*) and sixteen (*A. arenosa*). The chromosome number has changed during evolutionary events by duplication, deletion and insertion among the different species starting around 13 MYA from

a common ancestor (Beilstein *et al.*, 2010). It seems that *A. thaliana* is separated from the other species of *Arabidopsis* and each species has acquired somewhat characters differing from the others like heavy metal tolerance adaptation seen in *A. halleri* (Bomblies and Weigel, 2010). Among *Arabidopsis* species, only three genomes including *A. thaliana*, *A. lyrata* and *A. halleri* have been sequenced.

As the most contributing organism, there is a need for more details on their structure and genomic information on one of the most important gene groups, namely TFs. To our knowledge, no comparative report is available on *Arabidopsis* genus detailing TFs among three genome-published species. In this article, we compare the TF datasets of three genome-published *Arabidopsis* species to disclose similarities and differences among them. The common and differed TFs among three species are discussed.

## Materials and methods

The TF protein sequences of three *Arabidopsis* species including *A. halleri*, *A. lyrata* and *A. thaliana* were downloaded from PlantTFDB (Jin *et al.*, 2017) website (<http://planttfdb.cbi.pku.edu.cn/>), as it includes both TF number and loci while the other available databases do not. The OrthoVenn was used to compare and cluster three TF datasets. The  $p$ value  $\leq 1e-03$  was used to form the clusters related to GO terms. This program is a web platform that cluster orthologous gene datasets and generates Venn diagram and relevant annotation data outputting SwissProt hit and GO terms (Wang *et al.*, 2015). MEME version 4.8.1 was employed for finding protein motif sequences (Bailey and Elkan, 1994) (<http://meme-suite.org/>). Blast+ were used to find homologs (Camacho *et al.*, 2009).

## Results and discussion

The info about TFs of the three species have been disclosed in terms of quantity and loci (Table 1) and the number of each TF family (Table 2). *A. thaliana* possesses more TFs than the other two species although it has a smaller genome size. Interestingly, *A. lyrata* has most TF loci because of its bigger genome while *A. thaliana* with the smallest genome size stands in the second place. Considering the same ancestor from whom *Arabidopsis* species diverged, the speciation events such as deletion and duplication have differed for *A. thaliana* whose loci possess various TFs from *A. lyrata* that each locus present one TF. Further study for each TF family related to loci is required to reveal how TFs are distributed on loci in *Arabidopsis* genus.

**Table 1. TF loci and gene numbers of three *Arabidopsis* species. The information compiled from PlantTFDB.**

Species	TFs	Loci
<i>A. thaliana</i>	2,296	1,717
<i>A. lyrata</i>	1,759	1,759
<i>A. halleri</i>	1,586	1,446

**Table 2. TF families and their member number of three *Arabidopsis* species. The info compiled from PlantTFDB.**

TF	Total	Aha	Aly	Ath	TF	Total	Aha	Aly	Ath
AP2	66	16	20	30	LBD	130	33	47	50
ARF	75	18	20	37	LFY	3	1	1	1
ARR-B	46	13	12	21	LSD	20	5	3	12
B3	230	72	81	77	MIKC_MADS	154	39	39	76
BBR-BPC	34	9	8	17	M--type_MADS	161	51	40	70
BES1	31	8	9	14	MYB	441	124	149	168
bHLH	522	142	155	225	MYB_related	233	72	64	97
bZIP	278	71	80	127	NAC_	359	99	122	138
C2H2	317	97	104	116	NF-X1	8	3	3	2
C3H	181	61	54	66	NF-YA	47	14	12	21
CAMTA	22	6	6	10	NF-YB	58	15	16	27
CO-like	49	10	17	22	NF-YC	47	10	16	21
CPP	23	7	7	9	Nin-like	53	20	16	17
DBB	27	6	7	14	NZZ/SPL	3	1	1	1
Dof	107	26	34	47	RAV	16	3	6	7
E2F/DP	34	11	7	16	S1Fa-like	8	1	3	4
EIL	19	7	6	6	SAP	2	0	1	1
ERF	338	75	124	139	SBP	65	20	15	30
FAR1	55	24	5	26	SRS	35	9	10	16
G2-like	173	55	54	64	STAT	9	2	3	4
GATA	92	23	28	41	TALE	75	21	21	33
GeBP	64	19	22	23	TCP	75	17	25	33
GRAS	101	27	37	37	Trihelix	88	24	30	34
GRF	27	9	9	9	VOZ	9	3	3	3
HB-other	27	8	8	11	Whirly	11	3	4	4
HB-PHD	7	2	2	3	WOX	48	14	16	18
HD-ZIP	148	55	35	58	WRKY	229	60	79	90
HRT-like	5	1	2	2	YABBY	21	7	6	8
HSF	91	28	38	25	ZF-HD	44	9	17	18

Three smallest families were LFY, NZZ/SPL and SAP. STERILE APETALA (SAP) has no TF in *A. halleri* while LFY and NZZ/SPL possess only one TF for each species. The same TF number for these three families were seen in other plant TF databases including iTAK and PlnTFDB (Pérez-Rodríguez *et al.*, 2010; Zheng *et al.*, 2016). SAP plays a role in female reproductive organs; inflorescence, flower and ovule development. It is required for flower recognition in a manner similar to APETALA1 and modulates processes and gene expression related to organ size by targeting the KIX-PPD repressor complex for degradation (Byzova *et al.*, 1999). SAP functions in development of flower and its mutants have shown severe aberrations in such developmental process related to inflorescence, ovule and flower. SAP and APETALA2 (AP2) are involved in the initiation of female gametophyte development (Byzova *et al.*, 1999). In addition, it can interact with AGAMOUS to maintain floral identity acting in a manner similar to APETALA1.

The plant-specific LEAFY (LFY) protein is also involved in switching vegetative stage to reproductive stage. It controls the process of flower production and development and is known as a meristem identity gene (William *et al.*, 2004). It also interact with AGAMOUS and APETALA as floral homeotic genes in flower morphogenesis like what seen in SAP. Another rare TF family is the “floral organ-building” gene SPOROCTELESS/NOZZLE (SPL/NZZ). It also plays a central role in male flower development regulating anther cell differentiation (Liu *et al.*, 2009). Interestingly, it interact with AGAMOUS and AP2 involving in stamen identity. These TFs may work as a network sharing common interacting among various TFs in flower development. However, they exist in the least number in *Arabidopsis* and this makes them interesting to be studied further.

As seen generally in plants and reported in PlantTFDB, bHLH family with totally 522 TFs is the most abundant family in three *Arabidopsis* species (Table 2). Previously the MYB

TF family has been reported as the biggest TF family among plants and particularly *Arabidopsis* (Toledo-Ortiz *et al.*, 2003). Reported by Pireyre and Burow, a large proportion of the TFs in *Arabidopsis* were MYB with 339 and bHLH with 162 members (Pireyre and Burow, 2015). However, considering two MYB and MYB-like families together with 674 TFs, they are bigger than bHLH family. These two TF families forms different MYB/bHLH complexes regulate distinct cellular processes such as cell wall synthesis, cell death, circadian clock, responses to abiotic and biotic stress, hormone signaling, and the biosynthesis of specialized metabolites

(Lindemose *et al.*, 2013; Pireyre and Burow, 2015; Seo and Mas, 2014; Stracke *et al.*, 2007, 2001; Vaillau *et al.*, 2002). In addition, bHLHs and MYB can physically interact with each other to success their function.

In total 220 bHLH (Table 3) were common among three species while MYB and its related TFs were 308 (Supplementary File). bHLH35 and bHLH115 in *A. thaliana* and *A. lyrata* respectively were the species-specific bHLHs. bHLH115 is found among the three but they might be other isoforms found in *A. lyrata* as specific TFs. No specie-specific bHLH was found for *A. halleri*.

**Table 3. The clustered bHLH genes among three *Arabidopsis*.**

ID	# Proteins	Swiss-Prot Hit	ID	# Proteins	Swiss-Prot Hit
cluster15	7	bHLH69	cluster704	3	bHLH55
cluster47	6	BHLH {ECO:0000313 EMBL:CAE09167.1}	cluster756	3	bHLH36
cluster86	5	bHLH76	cluster766	3	bHLH28
cluster98	5	bHLH13	cluster776	3	bHLH81
cluster116	4	bHLH61	cluster799	3	bHLH30
cluster152	4	bHLH23	cluster827	3	bHLH101
cluster169	4	bHLH100	cluster831	3	bHLH18
cluster172	4	bHLH112	cluster842	3	bHLH14
cluster207	4	bHLH7	cluster857	3	bHLH62
cluster213	4	bHLH155	cluster871	3	bHLH10
cluster222	4	bHLH115 {ECO:0000313 EMBL:AEE32619.1}	cluster882	3	bHLH131
cluster224	4	bHLH104	cluster887	3	bHLH84
cluster235	4	bHLH137	cluster913	3	bHLH19
cluster243	4	bHLH27	cluster932	3	bHLH27
cluster263	4	bHLH129 {ECO:0000313 EMBL:AEC10214.1}	cluster943	3	bHLH131
cluster266	4	bHLH148	cluster952	3	bHLH81
cluster268	4	bHLH109	cluster963	3	bHLH27
cluster298	4	bHLH18	cluster1005	3	bHLH10
cluster304	4	bHLH89	cluster1016	3	bHLH101
cluster312	4	bHLH144	cluster1042	3	bHLH3
cluster318	4	bHLH119	cluster1054	3	bHLH23
cluster556	3	bHLH114	cluster1062	3	bHLH137
cluster559	3	bHLH85	cluster1065	3	bHLH119
cluster616	3	bHLH30	cluster1092	3	bHLH23
cluster617	3	bHLH92	cluster1093	3	bHLH60
cluster618	3	bHLH10	cluster1097	3	bHLH118
cluster641	3	bHLH131	cluster1098	3	bHLH80
cluster648	3	bHLH139	cluster1100	3	bHLH82
cluster670	3	bHLH28	cluster1114	3	Putative bHLH086
cluster684	3	bHLH30	cluster1199	3	bHLH79
cluster690	3	bHLH30	cluster1204	3	bHLH117
cluster1277	3	Putative bHLH086	cluster1273	3	bHLH137

All TFs were 5,641 for three species and classified in 4,503 clusters (Table 4, Supplementary File). Singletons those are species unique genes and not clustered were 445 in total. 892 single-copy gene clusters including 2,676 TFs were generated, i.e. the clusters containing one TF from one species summing 3 TFs. *A. thaliana* showed most numbers of TFs and subsequently clusters

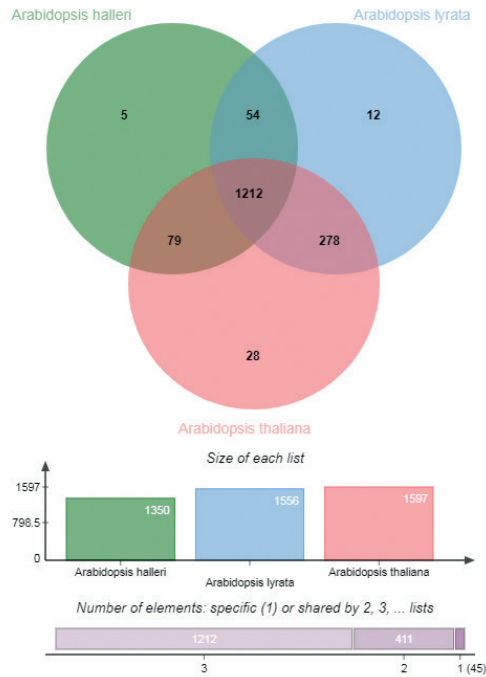
and singletons followed by *A. lyrata* and *A. halleri* (Table 4). This might be due to well assembled genome and information availability on *A. thaliana* although it has less chromosome number and its variation is less than other species (Hu *et al.*, 2011). However, the number of genes and GO distribution is well similarly distributed among the species

**Table 4. The proteins, clusters and singletons for three *Arabidopsis* species.**

Species	TF Proteins	Clusters	Singletons
<i>A. halleri</i>	1586	1350	102
<i>A. lyrata</i>	1759	1556	117
<i>A. thaliana</i>	2296	1597	226
Total	5641	4503	445

As shown in Venn diagram (Figure 1) and Table 5, there were 1,212 clusters containing 4,172 genes (74%) those are common among three species. Four hundred eleven clusters containing 924 genes (16.38%) were pairwise interspecific TFs, i.e. common between two species. Common interspecific TFs between *A. thaliana* and *A. lyrata* are more than those of *A. halleri* and *A. lyrata*. This more similarity between *A. thaliana* and *A. lyrata* suggest that these species

might be closer to each other than *A. halleri* in terms of TF families. Considering singletons and species-specific clustered TFs, *A. thaliana* showed more species-specific TFs as 2.53 % of all its TFs while for *A. lyrata* and *A. halleri* it was 1.7 and 0.75% respectively. However, further comparative analyses are required to study each TF family to reveal the speciation and evolutionary relation between three *Arabidopsis* species.



**Figure 1 .Venn diagram comparing three *Arabidopsis* TFs. The numbers are clusters.**

**Table 5. The number of TF clusters and genes among three *Arabidopsis***

Comparison	No. of cluster	No. of gene
Three species	1,212	4,172
<i>A.halleri</i> vs <i>A.lyrata</i>	54	111
<i>A.halleri</i> vs <i>A.thaliana</i>	79	197
<i>A.lyrata</i> vs <i>A.thaliana</i>	278	616
<i>A.halleri</i>	5	12
<i>A.lyrata</i>	12	30
<i>A.thaliana</i>	28	58

Homeobox-leucine zipper protein HAT22/ATHB-15 is the first cluster with most proteins containing eighteen TFs distributed among three species with eight in *A. halleri*, seven in *A. lyrata* and nine in *A. thaliana* (Supplementary File). It is unique to plants and belongs to HD-Zip class II that has nine members in *Arabidopsis* (Ariel *et al.*, 2007) or class III (Prigge *et al.*, 2005). The TFs of the family are involved in light response, shade avoidance, auxin signaling, cellular redox status perception (Elhiti

and Stasolla, 2009), lateral organ formation, formation or maintenance of vascular development in inflorescence stems (Prigge *et al.*, 2005; Turchi *et al.*, 2015). Its more highly conserved homeodomain motif is shown in Figure 2. It contains 41 sites from which 33 are constant in all three species. 5 sites show two alternatives for the site. Only 3 sites show substitution of 3 amino acids. Its regular expression is CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA.



Name	Start	p-value	Sites
A_thaliana AT4G32880.1	49	3.29e-53	SMRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA MNKLLMEEND
A_thaliana AT1G52150.2	51	3.29e-53	SIRRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA MNKLLMEEND
A_thaliana AT1G52150.1	51	3.29e-53	SIRRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA MNKLLMEEND
A_thaliana AT1G52150.3	51	3.29e-53	SIRRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA MNKLLMEEND
A_lyrata 314555	51	3.29e-53	SIRRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA MNKLLMEEND
A_lyrata 491408	48	3.29e-53	SMRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA MNKLLMEEND
A_halleri Araha.2499s0002.1.p	51	3.29e-53	SIRRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA MNKLLMEEND
A_halleri Araha.17232s0001.1.p	49	3.29e-53	SMRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA MNKLLMEEND
A_halleri Araha.17232s0001.2.p	49	3.29e-53	SMRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEASRLQAVNRKLTSA MNKLLMEEND
A_thaliana AT2G34710.1	59	3.46e-52	SIRRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEAARLQTVNRKLTSA MNKLLMEEND
A_lyrata 902559	66	3.46e-52	SIRRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEAARLQTVNRKLTSA MNKLLMEEND
A_halleri Araha.9962s0002.1.p	54	3.46e-52	SIRRRQQLIRE CPILSNIEPKQIKVWFQNRRCREKQRKEAARLQTVNRKLTSA MNKLLMEEND
A_thaliana AT1G30490.1	55	7.19e-50	SIRRRQQLIRE CPILCNIEPRQIKVWFQNRRCREKQRKEASRLQTVNRKLTSA MNKLLMEEND
A_lyrata 473277	54	7.19e-50	SIRRRQQLIRE CPILSNIEPRQIKVWFQNRRCREKQRKEASRLQTVNRKLTSA MNKLLMEEND
A_halleri Araha.2050s0004.1.p	54	7.19e-50	SIRRRQQLIRE CPILCNIEPRQIKVWFQNRRCREKQRKEASRLQTVNRKLTSA MNKLLMEEND
A_thaliana AT5G60690.1	59	3.57e-49	SIRRRQQLIRE CSILANIEPKQIKVWFQNRRCRDKQRKEASRLQSVNRKLTSA MNKLLMEEND
A_lyrata 496218	59	3.57e-49	SIRRRQQLIRE CSILANIEPKQIKVWFQNRRCRDKQRKEASRLQSVNRKLTSA MNKLLMEEND
A_halleri Araha.0447s0009.1.p	59	3.57e-49	SIRRRQQLIRE CSILANIEPKQIKVWFQNRRCRDKQRKEASRLQSVNRKLTSA MNKLLMEEND

**Figure 2.** HAT22/ATHB-15 highly conserved motif among 18 TFs of three *Arabidopsis* species. Above) Motif sequence, Below) Comparison of motif among 18 members of TF.

Four hundred forty five singletons were found those were not classified into clusters (Table 4, Supplementary File). These genes will be analyzed in a further study. The specific clustered TFs of each species were subject to analyze using their GO and the SwissProt hit provided for each cluster. GO terms related to transcription were not considered and only those with unique functionality were kept and mined. Thus, the specific TFs of each species were shortlisted having unique SwissProt and GO terms. As there were similar TFs those might belong to the subfamilies or isoforms of TFs, we limited the unique TFs those particularly found in each species.

*A. halleri* species-specific TFs were BEL1-like homeodomain protein 2, zinc finger protein AZF2 and two-component response regulator ARR11 (Supplementary File). BEL1-like homeodomain protein 2 under GO:0009965 “leaf morphogenesis” contributes in leaf shape by repressing growth in specific subdomains of the leaf. It is a negative regulator of

knox homeobox gene KNAT1/BP expression that functions in shoot apical meristem (SAM) and carpel development (Scofield *et al.*, 2008). Zinc finger protein AZF2 as a specific TF in *A. halleri* is classified under GO:0042538 hyperosmotic salinity response, GO:0009737 response to abscisic acid GO:0010200 response to chitin and GO:0009414 response to water deprivation. Intriguingly, it shows 80% similarity to ZAT6 that positively regulate cadmium tolerance (Chen *et al.*, 2016). Taking into account this high similarity between ZAT6 and ZF2, it may be involved in heavy metal tolerance as well as in tolerance to other abiotic stresses such as drought, salinity and cold (Kodaira *et al.*, 2011).

Two-component response regulator ARR11 has been identified as DNA motif AGAT(T/C) containing TF that play role in cytokinin related processes (D’Agostino *et al.*, 2000). The extended DNA motif (AAGAT[T/C]TT) has been reported as well (Raines *et al.*, 2016). Interestingly, cytokinin can enhance

plant tolerance to heavy metals and in response to heavy metal (Gangwar *et al.*, 2010). This TF is involved in His-to-Asp phosphorelay signal transduction with the ability to acquire a phosphoryl group from a histidine-containing phosphotransfer intermediate (AHP) improving cytokinin related process (Imamura *et al.*, 2003). On the other hand, RRs in bacteria regulate heavy metal response and expression of the Cu (I)/Ag (I) efflux system to extrude them (Fung *et al.*, 2016). There might exist a similar process of responding to heavy metals and resisting them by interaction between cytokinin and two-component response regulator ARR family in *A. halleri* as it is a metal hyperaccumulator and pseudometallophyte (Sarret *et al.*, 2009) and can tolerate excess of heavy metals. These two TFs might be involved in *A. halleri* resistance to heavy metals. The 4 genes of these two TF clusters are Araha.3286s0004.2.p and Araha.3286s0004.4.p belonging to two-component response regulator ARR11 and Araha.9992s0007.1.p and Araha.9992s0008.1.p belonging to Zinc finger protein AZF2. *A. lyrata* species-specific TFs (Supplementary File) were heat shock protein 90-2 functioning in ATPase activity, defense response to bacterium and protein folding; DnaJ protein homolog involved in protein folding; heat stress transcription factor A-4a, ethylene-responsive transcription factor ERF026 and ERF043 those participate in response to chitin and heat. The other *A. lyrata* specific TFs were scarecrow-like protein 23 engaging in leaf development and nuclear transcription factor Y subunit B-9 involved in positive regulation of fatty acid biosynthetic process. These TFs function mostly in different stress responses like heat for ERF026 and ERF043 and are as known stress responsive TF and development processes. For *A. thaliana*, according to their GO, mainly they are involved in plant development as flower development, negative regulation of vernalization response, double fertilization forming a zygote and endosperm and regulation of photoperiodism, for Agamous-like, MADS-box, bZIP, MADS affecting flowering5 respectively. In addition, MYB48 and dehydration-responsive element-binding protein 1B function in response to salicylic acid and cold acclimation. Zinc finger CCCH domain-containing protein 24 has RNA methyltransferase activity (Supplementary File). These specific TFs of *A. thaliana* participate in development processes and stress responses

### Conclusions

The importance of TFs make their study necessary in order to reveal how gene expression and function change in response to internal and external factors. *Arabidopsis* as the well-known model has revealed the function of TFs but it lacks more details of TF among its three genome-published species. This study was performed to compare TFs of these three *Arabidopsis* species in silico. Three *Arabidopsis* species showed 90.34% common TFs in all three or at least two species. There were 545 species-specific TFs for *A. halleri*, *A. lyrata*, and *A.*

*thaliana*. *A. halleri* has a known metallophytic character and two specific TFs of this species are proposed to be involved in response to heavy metals, namely AZF2 and two-component response regulator ARR11. Our results revealed comparatively similarities and differences in genomic level among three *Arabidopsis* TF families and species-specific TFs those can be used in subsequent studies of this plant model to disclose more details on biological processes like that of heavy metal, stress response and plant development. However, as other TF databases are available, another study will be executed to disclose more details on the TFs of *Arabidopsis* genus.

The Supplementary File is accessible from <https://doi.org/10.13140/RG.2.2.33662.08008>.

### References

- Ariel, F.D., Manavella, P.A., Dezar, C.A., Chan, R.L., 2007. The true story of the HD-Zip family. *Trends Plant Sci.* 12, 419–26. doi:10.1016/j.tplants.2007.08.003
- Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
- Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., Mathews, S., 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 107, 18724–8. doi:10.1073/pnas.0909766107
- Bombly, K., Weigel, D., 2010. *Arabidopsis* and relatives as models for the study of genetic and genomic incompatibilities. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365, 1815–23. doi:10.1098/rstb.2009.0304
- Byzova, M. V, Franken, J., Aarts, M.G., de Almeida-Engler, J., Engler, G., Mariani, C., Van Lookeren Campagne, M.M., Angenent, G.C., 1999. *Arabidopsis* STERILE APETALA, a multifunctional gene regulating inflorescence, flower, and ovule development. *Genes Dev.* 13, 1002–14.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421
- Chen, J., Yang, L., Yan, X., Liu, Y., Wang, R., Fan, T., Ren, Y., Tang, X., Xiao, F., Liu, Y., Cao, S., 2016. Zinc-finger transcription factor ZAT6 positively regulates cadmium tolerance through the Glutathione-dependent pathway in *Arabidopsis*. *Plant Physiol.* 171, 707–19. doi:10.1104/pp.15.01882
- D'Agostino, I.B., Deruère, J., Kieber, J.J., 2000. Characterization of the response of the *Arabidopsis* response regulator gene family to cytokinin. *Plant Physiol.* 124, 1706–17.
- Elhiti, M., Stasolla, C., 2009. Structure and function of homodomain-leucine zipper (HD-Zip) proteins. *Plant Signal. Behav.* 4, 86–88.

- Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., Solano, R., 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2367–72. doi:10.1073/pnas.1316278111
- Fung, D.K.C., Ma, Y., Xia, T., Luk, J.C.H., Yan, A., 2016. Signaling by the heavy-metal sensor CusS involves rearranged helical interactions in specific transmembrane regions. *Mol. Microbiol.* 100, 774–787. doi:10.1111/mmi.13348
- Gangwar, S., Singh, V.P., Prasad, S.M., Maurya, J.N., 2010. Modulation of manganese toxicity in *Pisum sativum* L. seedlings by kinetin. *Sci. Hortic. (Amsterdam)*. 126, 467–474. doi:10.1016/J.SCIENTA.2010.08.013
- Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.-F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J.D., Ossowski, S., Ottillar, R.P., Salamov, A.A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M.E., Bergelson, J., Carrington, J.C., Gaut, B.S., Schmutz, J., Mayer, K.F.X., Van de Peer, Y., Grigoriev, I. V, Nordborg, M., Weigel, D., Guo, Y.-L., 2011. The *Arabidopsis* lyrata genome sequence and the basis of rapid genome size change. *Nat. Genet.* 43, 476–481. doi:10.1038/ng.807
- Imamura, A., Kiba, T., Tajima, Y., Yamashino, T., Mizuno, T., Mizuno, T., 2003. In vivo and in vitro characterization of the ARR11 response regulator implicated in the His-to-Asp phosphorelay signal transduction in *Arabidopsis thaliana*. *Plant Cell Physiol.* 44, 122–31. doi:10.1093/PCP/PCG014
- Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., Gao, G., 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi:10.1093/nar/gkw982
- Kodaira, K.-S., Qin, F., Tran, L.-S.P., Maruyama, K., Kidokoro, S., Fujita, Y., Shinozaki, K., Yamaguchi-Shinozaki, K., 2011. *Arabidopsis* Cys2/His2 zinc-finger proteins AZF1 and AZF2 negatively regulate abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions. *Plant Physiol.* 157, 742–56. doi:10.1104/pp.111.182683
- Lindemose, S., O’Shea, C., Jensen, M.K., Skriver, K., 2013. Structure, function and networks of transcription factors involved in abiotic stress responses. *Int. J. Mol. Sci.* 14, 5842–5878.
- Liu, L., White, M.J., MacRae, T.H., 1999. Transcription factors and their genes in higher plants. Functional domains, evolution and regulation. *Eur. J. Biochem.* 262, 247–257. doi:10.1046/j.1432-1327.1999.00349.x
- Liu, X., Huang, J., Parameswaran, S., Ito, T., Seubert, B., Auer, M., Rymaszewski, A., Jia, G., Owen, H.A., Zhao, D., 2009. The SPOROCTELESS/NOZZLE Gene Is Involved in Controlling Stamen Identity in *Arabidopsis*. *PLANT Physiol.* 151, 1401–1411. doi:10.1104/pp.109.145896
- Pérez-Rodríguez, P., Riaño-Pachón, D.M., Corrêa, L.G.G., Rensing, S.A., Kersten, B., Mueller-Roeber, B., 2010. PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 38, D822–7. doi:10.1093/nar/gkp805
- Pireyre, M., Burow, M., 2015. Regulation of MYB and bHLH transcription factors: A glance at the protein level. *Mol. Plant* 8, 378–388. doi:10.1016/J.MOLP.2014.11.022
- Prigge, M.J., Otsuga, D., Alonso, J.M., Ecker, J.R., Drews, G.N., Clark, S.E., 2005. Class III Homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *PLANT CELL ONLINE* 17, 61–76. doi:10.1105/tpc.104.026161
- Raines, T., Blakley, I.C., Tsai, Y.-C., Worthen, J.M., Franco-Zorrilla, J.M., Solano, R., Schaller, G.E., Loraine, A.E., Kieber, J.J., 2016. Characterization of the cytokinin-responsive transcriptome in rice. *BMC Plant Biol.* 16, 260. doi:10.1186/s12870-016-0932-z
- Sarret, G., Willems, G., Isaure, M.-P., Marcus, M.A., Fakra, S.C., Frérot, H., Pairis, S., Geoffroy, N., Manceau, A., Saumitou-Laprade, P., 2009. Zinc distribution and speciation in *Arabidopsis halleri* × *Arabidopsis lyrata* progenies presenting various zinc accumulation capacities. *New Phytol.* 184, 581–595. doi:10.1111/j.1469-8137.2009.02996.x
- Scotfield, S., Dewitte, W., Murray, J.A., 2008. A model for *Arabidopsis* class-I KNOX gene function. *Plant Signal. Behav.* 3, 257–9.
- Seo, P.J., Mas, P., 2014. Multiple layers of posttranslational regulation refine circadian clock activity in *Arabidopsis*. *Plant Cell* 26, 79–87. doi:10.1105/tpc.113.119842
- Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrrens, F., Niehaus, K., Weisshaar, B., 2007. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* 50, 660–677. doi:10.1111/j.1365-313X.2007.03078.x
- Stracke, R., Werber, M., Weisshaar, B., 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* 4, 447–456. doi:10.1016/S1369-5266(00)00199-0
- Toledo-Ortiz, G., Huq, E., Quail, P.H., 2003. The *Arabidopsis* basic/helix-loop-helix transcription factor family. *Plant Cell* 15, 1749–70. doi:10.1105/TPC.013839
- Turchi, L., Baima, S., Morelli, G., Ruberti, I., 2015. Interplay of HD-Zip II and III transcription factors in auxin-regulated plant development. *J. Exp. Bot.* 66, 5043–5053. doi:10.1093/jxb/erv174
- Vailleau, F., Daniel, X., Tronchet, M., Montillet, J.-L., Triantaphylidès, C., Roby, D., 2002. A R2R3-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10179–84. doi:10.1073/pnas.152047199



- Wang, Y., Coleman-Derr, D., Chen, G., Gu, Y.Q., 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 43, W78-84. doi:10.1093/nar/gkv487
- William, D.A., Su, Y., Smith, M.R., Lu, M., Baldwin, D.A., Wagner, D., 2004. Genomic identification of direct target genes of LEAFY. *Proc. Natl. Acad. Sci.* 101, 1775–1780. doi:10.1073/pnas.0307842100
- Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin, G.B., Giovannoni, J.J., Zhao, P.X., Rhee, S.Y., Fei, Z., 2016. iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Mol. Plant.* doi:10.1016/j.molp.2016.09.014