Genome wide identification, characterization and expression profiles of heavy metal ATPase3 (HMA3) in plants

Ahmad Kabir
University of Rajshahi
Urmi Das
University of Rajshahi
AFM Haque
University of Rajshahi
Amit Dutta
University of Rajshahi
Gholamreza Gohari (gohari.gh@maragheh.ac.ir)
University of Maragheh
Md Reza
University of Rajshahi

Research Article

Keywords: ATPases family, phylogeny, conserved motif, gene interactions, sequence homology.

Posted Date: November 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-110370/v1

Abstract

HMA (heavy metal associated) is a member of the ATPases protein family involved in metal transport in plants. This study characterizes several HMA3 homologs and infers their molecular functions in different plant species. Arabidopsis AtHMA3 (AT4G30120) was used as a reference to retrieve 11 HMA3 homologs having 97-100% query cover, 535-542 residues, 56983 to 58642 (Da) molecular weight, and 5.74 to 8.16 pI value, 29.10 to 33.89 instability index, and 0.222 to 0.380 grand average of hydropathicity. Topological analyses showed 4 transmembrane domains in these HMA3 homologs positioned similarly in terms of cytoplasmic and non-cytoplasmic regions along with ~22-28% α-helices, ~22-28% extended strands, and ~50% random coils. HMA3 protein of Arabidopsis lyrata subsp. lyrata and Eutrema salsugineum are located at chromosome 2, while others are positioned at chromosome 4. All these HMA3 homologs are localized in the plasma membrane sharing a few common biological and molecular functions. Besides, these HMA3 genes contain 8-9 exons in which promoter positions are varied among the homologs. The cis-acting elements of HMA3 genes were projected to be involved with stress response, anaerobic induction, and light responsive regulation in plants. Three out of five motifs encode E1-E2_ATPase involved in proton-pumping in the plasma membrane. The Arabidopsis thaliana HMA3 protein clustered with Camelina sativa and Capsella rubella show a close phylogenetic relationship. Also, AtHMA3 exhibits a close association with AtHMA3 with MTPA2, ZAT, NRAMP3, IRT2, and NRAMP2 under the local network of AtHMA3 linked to metal transport. Further, AtHMA3 is most potentially expressed during senescence, germinating seed, seedlings, young rosette, bolting, and young flower. In addition, AtHMA3 showed a significant upregulation (>6.0 fold) under Fe-deficiency. These findings may provide essential background to perform wet-lab experiments to understand the role of HMA3 in metal homeostasis.

Introduction

Heavy metals are abundant in nature due to natural and anthropogenic causes. Heavy metals are taken up by humans through water and food-based meals and may cause serious health problems. Many of the metals, such as Fe, Cu, Zn, are essential for plants, but they need to be at an optimized level. In contrast, some of the heavy metals (Pb, Cd) are highly toxic to plants hampering photosynthesis, nutrient uptake, and yield in plants[1]. However, plants acclimatize different strategies consisting of uptake, sequestration, and chelation to regulate metal homeostasis in withstanding heavy metal toxicity[2]. The association of different metal transporters and their binding capacity play an integral part in the cellular detoxification and maintenance of metals in plants.

The ATPases (P-type adenosine triphosphatases) are the largest superfamily of integral membrane proteins involved in transporting transition metal cations in plants. Eight P-type IB ATPases are encoded in the genome of *Arabidopsis thaliana*[3]. The ATPases are clustered into two groups in plants: monovalent Cu/Ag ion transporting ATPases and divalent Zn/Cd/Co/Pb ion transporting ATPases[4]. Due to the distinctive N-terminal sequence, ATPases in plants are named HMA (heavy metal associated) protein. In Arabidopsis, HMA1-HMA4 and HMA5-HMA8 proteins are belonging to cluster 1 (Zn/Cd/Co/Pb) and 2

(Cu/Ag) groups[3]. In particular, HMA3 proteins participate in heavy metal ion transport and detoxification in plants. In *Arabidopsis thaliana*, AtHMA3 localized in tonoplast is involved in the vacuolar storage of Cd[5]. Furthermore, Arabidopsis overexpressed with *AtHMA3* showed increased tolerance to Cd, Zn, Pb, and Co, while AtHMA3-knockout mutant exhibited sensitivity to Cd and Zn[6]. Similarly, the overexpression of *SaHMA3h* in tobacco improved the Cd accumulation and tolerance of transgenic plants[7]. Further, *BjHMA3* is shown to be associated with the varied Cd accumulation in leaves of *Brassica rapa*[8]. Again, *OsHMA3* ectopic over-expression resulted in increased Cd tolerance and lower Cd concentration in leaves and grains but increased Cd concentration in rice roots[9,10].

Protein topology, such as transmembrane helices, domain recognition, binding sites is crucial features for metal binding capacity in plant system. As a result, the identification of metal sites along with the components at transcriptional and Posttranslational regulation eventually determines the functions of a protein in response to metals. One of the identified candidate genes in *Arabidopsis halleri* was AhHMA3, which is highly similar to HMA3 in *Arabidopsis* thaliana (AT4G30120)[11]. HMA3, located in the vacuolar membrane, participates in vacuolar sequestration of Zn, Cd, Co, and Pb in Arabidopsis[12]. However, the function of HMA3 in hyperaccumulators remains unclear in planta. Although the molecular functions of Arabidopsis HMA3 are relatively well established, the analysis of HMA3 homologs and interactions with other transporters/genes are barely studied.

The characterization of HMA3 possesses the immense potential to combat metal homeostasis in plants. The *in silico* characterization of HMA3 homologs may provide in-depth insight into these genes/proteins. In this study, we have searched for HMA3 homologs based on Arabidopsis heavy metal ATPase 3 (AtHMA3) referred to as AT4G30120 across different plant species. The CDS, mRNA, and protein sequences of these HMA3 homologs were taken into computational analysis with advanced bioinformatics software and online-based platforms.

Results

Retrieval of HMA3 transporter genes/proteins

Arabidopsis AtHMA3 was searched in the NCBI database to get the FASTA sequence of the protein (NP_194741.2) and mRNA (NM_119158.4). The blast analysis of AtHMA3 protein showed 11 homologs of the heavy metal atpase 3 family by filtering (E-value: 0.0, query cover: 97-100%, percentage identity: 71.73-100%) in 11 plant species, which include Arabidopsis thaliana, *Camelina sativa*, *Capsella rubella*, *Eutrema salsugineum*, *Brassica oleracea* var. oleracea, *Raphanus sativus*, *Brassica napus*, *Brassica rapa*, *Arabidopsis lyrata* subsp. lyrata, *Eutrema salsugineum* and *Tarenaya hassleriana* (Table 1).

Table 1. Physiochemical properties of HMA3 protein homologs.

Sl.	Protein Accession	Species	Protein length	MW (Da)	pI	Instability index	Grand average of hydropathicity (GRAVY)	α-helix	Extended strand	Random coil
1	NP_194741.2	*Arabidopsis thaliana*	542	58642.37	7.44	31.50	0.355	24.17%	26.38%	49.45%
2	NP_001289919.1	*Camelina sativa*	540	58517.12	6.72	29.10	0.380	23.70%	27.04%	49.26%
3	XP_006303912.1	*Capsella rubella*	539	58183.82	6.42	31.49	0.364	24.68%	25.05%	50.28%
4	XP_006412748.2	*Eutrema salsugineum*	539	58475.96	6.72	31.97	0.307	25.97%	24.49%	49.54%
5	XP_013591300.1	*Brassica oleracea var. oleracea*	540	58479.85	5.74	29.62	0.312	27.59%	23.70%	48.70%
6	XP_018480904.1	*Raphanus sativus*	540	58244.61	6.43	31.31	0.328	28.33%	25.00%	46.67%
7	XP_022562629.1	*Brassica napus*	540	58365.83	6.42	30.92	0.359	25.00%	26.11%	48.89%
8	XP_009137892.1	*Brassica rapa*	541	58293.66	6.98	33.89	0.325	28.47%	22.00%	49.54%
9	XP_020886284.1	*Arabidopsis lyrata subsp. lyrata*	528	57120.44	6.55	33.68	0.330	25.19%	26.70%	48.11%
10	XP_006409084.2	*Eutrema salsugineum*	525	56983.36	6.97	37.19	0.314	22.86%	28.57%	48.57%
11	XP_010548593.1	*Tarenaya hassleriana*	538	58204.12	8.16	32.42	0.222	24.16%	27.88%	47.96%

Physiochemical features and localization of HMA3 proteins

The 11 HMA3 protein homologs encoded a protein with residues of 525-542 amino acids having 56983.36 to 58642.37 (Da) molecular weight, and 5.74 to 8.16 pI value, 29.10 to 33.89 instability index, and 0.222 to 0.380 grand average of hydropathicity (Table 1). Topological prediction analyses of transmembrane (TM) domains of HMA3 protein homologs showed 4 transmembrane domains in protein representative from each of the plant species (Supplementary Fig. S.1). None of the HMA3 protein homologs contains signal peptide. These HMA3 proteins showed positioning similarity in terms of cytoplasmic and non-cytoplasmic regions (Supplementary Fig. S.1). In addition, secondary structure prediction showed that all HMA3 proteins contain above ~22-28% α-helices, ~22-28% extended strands, and ~50% random coils (Table 1).

Localization and functional annotation of HMA3 proteins

HMA3 protein of *Arabidopsis lyrata* subsp. lyrata (XP_020886284.1) and *Eutrema salsugineum* (XP_006409084.2) is located at chromosome 2; however, the rest of the HMA3 homologs positioned at chromosome 4 (Table 2). All of these HMA3 protein homologs are associated with E1-E2 ATPase (PF00122). The CELLO localization predictor showed that these HMA3 proteins are localized in the plasma membrane of roots in all 11 plant species (Table 2). Ontology analysis demonstrated that HMA3 proteins of *Arabidopsis thaliana*, *Camelina sativa*, *Capsella rubella*, *Eutrema salsugineum*, *Brassica oleracea* var. oleracea, *Raphanus sativus*, *Brassica napus* and *Brassica rapa* possess several cellular components, including vacuolar membrane, plasma membrane, and a membrane having involvement in the same biological process (cation transport, metal ion transport) and molecular function (nucleotide-binding, ATP binding, ATPase activity, hydrolase activity, metal ion binding). In addition, HMA3 *Arabidopsis lyrata* subsp. lyrata and *Eutrema salsugineum* showed the same cellular component

(vacuolar membrane, plasma membrane, membrane), biological process (transition metal ion transport, cation transport, zinc ion transport, cadmium ion transport, response to cadmium ion) and molecular function (nucleotide-binding, ATP binding, ATPase activity, hydrolase activity, metal ion binding, metal ion transmembrane transporter activity, cadmium-transporting ATPase activity). Lastly, HMA3 protein of *Tarenaya hassleriana* showed unique cellular components (plasma membrane, membrane, integral to membrane), biological process (ATP biosynthetic process, cation transport, metabolic process, metal ion transport, zinc ion homeostasis) but molecular function similar to *Arabidopsis lyrata* subsp. lyrata and *Eutrema salsugineum* (Table 2).

Table 2. Domain, localization, cellular component, biological and molecular function of HMA3 proteins.

Sl.	Protein Accession	Species	Domain	Localization	Cellular component	Biological Process	Molecular function
1	NP_194741.2	*Arabidopsis thaliana*	E1-E2 ATPase (PF00122)	Plasma Membrane	-vacuolar membrane -plasma membrane -membrane	-cation transport -metal ion transport	-nucleotide binding -ATP binding -ATPase activity -hydrolase activity, -metal ion binding
2	NP_001289919.1	*Camelina sativa*	As above	As above	As above	As above	As above
3	XP_006303912.1	*Capsella rubella*	As above	As above	As above	As above	As above
4	XP_006412748.2	*Eutrema salsugineum*	As above	As above	As above	As above	As above
5	XP_013591300.1	*Brassica oleracea var. oleracea*	As above	As above	As above	As above	As above
6	XP_018480904.1	*Raphanus sativus*	As above	As above	As above	As above	As above
7	XP_022562629.1	*Brassica napus*	As above	As above	As above	As above	As above
8	XP_009137892.1	*Brassica rapa*	As above	As above	As above	As above	As above
9	XP_020886284.1	*Arabidopsis lyrata subsp. lyrata*	As above	As above	-plasma membrane -plasmodesma -membrane -integral to membrane	-transition metal ion transport -cation transport -zinc ion transport - cadmium ion transport -response to cadmium ion	-nucleotide binding -ATP binding -ATPase activity -hydrolase activity, -metal ion binding - metal ion transmembrane transporter activity - cadmium-transporting ATPase activity
10	XP_006409084.2	*Eutrema salsugineum*	As above	As above	As above	As above	As above
11	XP_010548593.1	*Tarenaya hassleriana*	As above	As above	-plasma membrane -membrane -integral to membrane	-ATP biosynthetic process -cation transport -metabolic process -metal ion transport -zinc ion homeostasis	As above

Gene organization

ARAMEMNON analysis showed the presence of 8-9 exons among the HMA3 gene homologs located at different positions of gene ranged from 1-3535 base pairs (Fig. 1, Table 3). Promoter analysis showed marginal and highly like the prediction of promoter position in the gene sequence in different *HMA3* homologs across the 11 plant species. The Arabidopsis thaliana HMA3 showed three different positions of promoter marginal predicted at 200, 700, and 1800 bp (Table 3). Highly predicted position of promoters are located in 2200bp, 2100bp and 2500bp in XM_006303850.2 (*Capsella rubella*),

XM_006412685.2 (*Eutrema salsugineum*), XM_010550291.1 (*Tarenaya hassleriana*), respectively (Table 3). The position of TSS varied from 22-36 bp if found. Also, the PolA was positioned after the coding region in all HMA3 genes showing the position at 2369-4074 bp, if found (Table 3). The identified *cis*-acting elements were stress, hormone, and other responsive factors. Stress responsive, anaerobic induction, and light responsive regulators were found to be the height number and most common of cis-acting elements in *HMA3* genes. That projected their involvement of these activities (Table 4).

Table 3. Organization of HMA3 genes and position features.

No.	Gene Accession	Chromosome number	Coding region	Promoter position	Position of transcriptional start site (TSS)	PolA
1	NM_119158.4	4	127-1645, 1756-2408	200 Marginal prediction 700 Marginal prediction 1800 Marginal prediction	24	2516
2	NM_001302990.1	4	1-2286	1700 Marginal prediction	24	2516
3	XM_006303850.2	4	38-2302	2200 Highly likely prediction	-	2453
4	XM_006412685.2	4	6-2288	1100 Marginal prediction 2100 Highly likely prediction	-	2354
5	XM_013735846.1	4	1-2277	600 Marginal prediction 1100 Marginal prediction 1700 Marginal prediction	-	-
6	XM_018625402.1	4	71-2335	200 Marginal prediction 1000 Marginal prediction 1700 Marginal prediction	22	2376
7	XM_022706908.1	4	92-2341	200 Marginal prediction 1100 Marginal prediction	36	2369
8	XM_009139644.3	4	123-2417	1800 Marginal prediction	-	2681
9	XM_021030625.1	2	190-4023	1000 Marginal prediction 2700 Marginal prediction 3200 Marginal prediction 3600 Marginal prediction	-	4074
10	XM_006409021.2	2	135-4142	1100 Marginal prediction 2200 Marginal prediction 2600 Marginal prediction 3500 Marginal prediction 3900 Marginal prediction	-	4235
11	XM_010550291.1	4	155-516, 589-3535	2500 Highly likely prediction 3300 Marginal prediction	-	3788

Table 4. *Cis*-acting element analysis of *HMA3* gene promoters.

Gene Accession	*Cis*-acting elements																								
	MYC/Myc	MYB/myb	MRE	TC rich	GT1	G box/ABRE	W - box	TATC-Box	ERE	AREs	As-1/TGA	AuxRR-core	TGACG	CGTCA	TCA element	ACE	AE-Box	Box -4	GATA-Motif	GA-Motif	WUN3	WRE3	O2 Site	circadian	STRE	AT-rich element
NM_119158.4	1	4	1		2			1		2		1	1					4				1				1
NM_001302990.1	3	2					1			1	2		1	1	3			2			2	2		1		
XM_006303850.2		2								1	2			2		1	1	1		1	1					
XM_006412685.2	5	2		2	1	2			2	4		1				1		4		2			1			
XM_013735846.1	1					4			1	2	2				1			4			1					
XM_018625402.1	2	1				2	1			3		1	1		2			1			1				1	
XM_022706908.1	1	1	1			1	1		2	1		1	1					2								1
XM_009139644.3	2	4	1	1		1			1	1					1			6	2	1					2	1
XM_021030625.1	1	4							1	3	2			1				1	1						1	
XM_006409021.2	1	4	1		1					1	2			1			2		1					2		
XM_010550291.1	4			1	2			1		1	4		1	2		1		3					1		1	

Different *cis*- regulatory elements: MYC/Myc (Dehydration-responsive), MYB/Myb (drought-responsive), MBS (involved in drought induction), TC rich (defense and stress-responsive), GT1 (SALT-responsive), ABRE/G-box (Abscisic acid-responsive), W-box (defense responsive), TATC-Box (gibberellins responsive) ERE (ethylene-responsive), AREs (involved in anaerobic induction), As-1, and AuxRR-core, (auxin-responsive), TGACG and CGTCA (MeJA-responsive), TCA (salicylic acid-responsive), ACE, AE-box, Box-4, GATA-Motif, GA-motif, and G-Box (light-responsive), WUN3, and WRE3 (wound-responsive), O_2 site (Zein metabolism regulation), circadian (circadian control), STRE (Expression activator), and At-rich (DNA binding).

Conserved motif, Sequence similarities, and phylogenetic analysis

We have used the MEME tool to search for the five most conserved motifs in identified 11 HMA3 homologs (Fig. 2). All these 5 motifs are 50 residues long located at site 11. These motifs are as follows: motif 1 (INLNGYIKVKTTALARDCVVAKMTKLVEEAQKSQTKTQRFIDKCSRYYTP), motif 2 (HPMAAALIDYARSVSVEPKPDMVENFQNFPGEGVYGRIDGQDIYIGNKRI), motif 3 (NLSHWFHLALVVLVSGCPCGLILSTPVATFCALTKAATSGFLIKTGDCLE), motif 4 (KALNQARLEASVRPYGETSLKSQWPSPFAVVSGVLLALSFLKYFYSPLEW) and motif 5 (CMZDYTEAATIVFLFSVADWLESSAAHKASTVMSSLMSLAPRKAVIAETG). Motif 1, 3 and 5 encodes pfam_fs: E1-E2_ATPase. Motif 4 is linked to freq_pat:PKC_PHOSPHO_Site, while motif 2 shows no information (Fig. 2). The HMA3 protein homologs were aligned to check the similarities of sequence across the plant species. The MTP1 proteins showed 71.7% to 100% similarities among the different plant species, in which the consensus sequence ranged from 70%-100% (Supplementary Fig. S.2). The phylogenetic tree was clustered into four groups (A, B, C, D) based on tree topologies (Fig. 3). In cluster A, HMA3 of *Arabidopsis thaliana* formed a cluster with the *Camelina* sativa and *Capsella rubella*, while group B consist of HMA3 protein homologs of *Brassica oleracea* var. oleracea, *Raphanus sativus*, *Brassica napus* and *Brassica rapa*. The HMA3 of *Eurema salsugineum* clustered alone is located in the distance from *Arabidopsis thaliana homolog*. The cluster D consists of Arabidopsis lyrata subsp. Lyrata HMA4, *Eurema salsugineum* HMA4and *Tarenaya hassleriana* HMA3 (Fig. 3). In this phylogenetic tree,

HMA3 of *Arabidopsis thaliana*, *Brassica oleracea* var. oleracea, *Raphanus sativus*, *Brassica napus* and *Brassica rapa* showed the highest 100% bootstrap value (Fig. 3).

Predicted interaction partner analysis

Interactome analysis was performed for AtHMA3 (AT4G30120) on STRING server. STRING showed five closely associated putative interaction partners of AtHMA3. These include *MTPA2* (metal tolerance protein A2), *ZAT* (a member of the zinc transporter and cation diffusion facilitator), *NRAMP3* (natural-resistance-associated macrophage protein 3), *IRT1* (iron-regulated transporter 2) and *NRAMP4* (natural-resistance-associated macrophage protein 4) genes (Fig. 4). Further, the analysis showed four local network clusters, including CL:28166 (nickel transport and cation efflux protein), CL:28164 (manganese ion transport and nickel transport), CL:28176 (ion influx/efflux at the host-pathogen interface), CL:28126 (transition metal ion transmembrane transporter). Lastly, reactome pathways of *AtHMA3* include ion influx/efflux at host-pathogen interface, zinc efflux, and compartmentalization by the SLC30 family, peptide hormone metabolism, insulin processing, and metal ion SLC transporters (Fig. 4).

Expression profiles of HMA3

The genevestigator analysis against the Affymetrix Array Platforms showed expression potential and co-expression data of HMA3 in different anatomical parts, developmental stages, and perturbations. In the anatomical part, lateral roots, cauline leaf, silique inflorescence, and radicle in shoot apex seemed to be highly potential for HMA3 expression (Fig. 5a). Further, giant root cell and sperm cells in cell culture have the potential for HMA3 expression (Fig. 5a). Besides, HMA3 has expression potential during senescence, germinated seed, seedling, young rosette, bolting, and young flower stages of development (Fig. 5a). Among the selected stress, HMA3 only showed significant upregulation under Fe deficiency; while the expression did not notably vary in other stresses, such as anoxia, cold, drought, gamma irradiation, genotoxicity, heat, osmotic stress, salt stress, shift cold stress, submerge stress, wounding stress (Fig. 5c).

Co-expression analysis was filtered to five closely associated genes in different anatomical parts, developmental stages and perturbations (Fig. 6). In the anatomical part, the AtHMA3 gene is closely co-expressed with AT1G30560 (putative glycerol-3-phosphate transporter), AT1G63550 (cystein-rich repeat secretory protein 9), AT3G60270 (cupredoxin superfamily protein), AT5G43370 (probable inorganic phosphate transporter 1-2) and AT3G12900 (2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein). During the development stage, AT3G56891 (heavy metal transport/detoxification superfamily protein), AT5G19040 (adenylate isopentenyltransferase 5, chloroplastic), AT3G45410 (L-type lectin-domain containing receptor kinase I.3), AT3G29250 (short-chain dehydrogenase reductase 4) and AT2G25260 (unknown protein) co-expressed with AtHMA3 (Fig. 6). Under perturbations, the top five genes co-expressed with AtHMA3 are AT1G64480 (calcineurin B-like protein 8), AT5G17100 (cystatin/monellin superfamily protein), AT5G65980 (auxin efflux carrier family protein), AT2G28690 (protein of unknown function, DUF1635), and AT2G12190 (cytochrome P450 superfamily protein) (Fig. 6).

Discussion

In recent years, the characterization of membrane transporters involved in heavy metal scavenging in plants is emerging. Prior to the wet-lab experiment, the *in silico* analysis is of utmost interest to narrow down the target of studies. The role of *AtHMA3* in vacuolar storage of a few metals are documented, although the involvement of other metals and dynamic network of other associated genes and gene/protein properties are yet to extensively studies. This *in silico* characterization and expression profile *AtHMA3* and its homologs and closely associated genes unveil significant regulatory findings that can be essential contributors to the downstream genome-editing or biotechnological approach to heavy metal studies.

In this study, we selectively blasted the AtHMA3 sequences resulted in 11 different HMA3 protein homologs having 71.73-100% percentage identity. The similarities in protein size, pI, instability index and hydrophilicity suggest that these HMA3 proteins are biochemically relevant. Domain analysis further revealed the association of these HMA3 protein homologs with E1-E2 ATPase (PF00122) localized in the plasma membrane. The proton-pumping ATPase (H^+-ATPase) in the plasma membrane produces the proton motive force through the plasma membrane that is required to enable much of the transport of ions and metabolites[13]. Although HMA3 proteins showed diverse cellular component in HMA3 homologs, the unique feature of these limited to membrane, vacuolar membrane and plasma membrane. These cellular components are crucial in mineral absorption and metal homeostasis along with salt tolerance, intracellular pH regulation and cellular expansion in plants[13,14]. However, HMA3 proteins are predominantly associated with cation transport, metal ion transport, zinc ion transport and cadmium ion transport, as evident from our ontology analysis. *HMA3* gene is involved in cadmium and lead transport along with vacuolar sequestration potentiality in a heterologous system, but not in zinc transport. Vacuolar sequestration may have a detoxification function[15].

In predicting evolutionary relationships and functional genomics possibilities, the knowledge on the position and organization of the coding sequence of a gene is considered a critical factor. In this study, all of the identified sequences of HMA3 proteins demonstrated 4 transmembrane helices confirming similar hydropathy of these HMA3 protein homologs. The metal specificity of each subclade is determined by specific amino acids in the three transmembrane helices closest to C-terminus[16]. In this study, all *HMA3* gene homologs belonging to the 11 plant species showed 8-9 exons, suggesting that these HMA3 genes are evolutionarily closer to each other. Although promoter analysis predicted several promoter regions of each HMA3 gene, the highly likely prediction of the promoter was made at 2100-2500 bp in several plant species. Localization of exon and promoter plays an essential part in CRISPR-Cas9 and other genome editing studies in plant science. In addition, the identification of TSS and PolA in *HMA3* homologs will be crucial in understanding the transcriptional and translational genomics. Besides, promoter analysis reveals the involvement of *cis*-acting elements associated with stress response, hormone, anaerobic induction, and light responsive regulators in HMA3 genes.

Conserved motifs are identical sequences across species that are maintained by natural selection. A highly conserved sequence is of having functional roles in plants and can be a useful start point to start research on a particular topic of interest[17]. Out of the five motifs, three motifs are mainly matched with the E1-E2_ATPase associated with H^+ pumping. P-type proton ATPase is found in the plasma membranes of plants that in turn, drives secondary active transport processes across the membrane[18]. One of the motifs is also linked to the protein kinase C phosphorylation site that may play roles in controlling the catalytic activity, stability and intracellular localization of the enzyme[19]. Further, the phosphorylation site may be attributed to the release of Zn from intracellular stores leading to phosphorylation kinases and activation of signaling pathways[20]. The presence of common and long-preserved residues suggests that HMA3 homologs between species may have highly conserved structures. Additionally, for sequence-specific binding sites and transcription factor analysis, this information can be targeted. In phylogenetic analysis, HMA3 protein of *A. thaliana* positioned in the same cluster with *C. sativa* and *C. rubella*, suggesting its close relationship during the evolutionary trend. Consistently, HMA3 protein homologs of *Brassica* sp. and *Raphanus* sp. clustered within B, suggesting the close evolutionary emergence from a common ancestor within the Brassicaceae family. It appears that the HMA3 of *E. salsugineum* is relatively distantly related to *A. thaliana* over the evolutional trends. Thus, our results might infer a functional relationship of HMA3 sequences in metal uptake across different plant species.

The interaction network of a specific gene provides information of all physical associations that can occur among family members. Global gene co-expression analysis is an emerging tool to identify the tissues and the conditions in which significant interactions occur. The interactome map analyzed in String platform showed the most close association with *MTPA2*, *ZAT*, *NRAMP3*, *IRT2* and *NRAMP2*, mainly linked to metal transport in plants. Consistently, the local network of *AtHMA3* implies the involvement with metal transporter. As a result, these findings might be useful to characterize HMA3 and to interpret the interactions of multiple genes linked to particular stress of interest in plants. Studies reported that Zn homeostasis is closed associated with P-type ATPase heavy metal transporters (HMA). Again, both *HMA2* and *HMA4* were reported to be involved with Zn homeostasis in Arabidopsis[21]. Besides, AtHMA3 showed some reactome pathways, among which ion influx, zinc influx and metal ion SLC transporters may attribute to the metal transporter properties of this gene. Overall, this interactome finding might provide essential background for functional genomics studies of metal uptake and transport in plants.

The expression potential of a gene in different conditions is a crucial factor in determining the involvement in a particular trait. The *in silico* expression analysis in the Genevestigator platform showed interesting outputs concerning the expression of AtHMA3 (AT4G30120) in different anatomical, perturbations, and developmental stages. Being consistent with the AtHMA3 gene ontology, Genvestigator showed that root is the significant location where this gene showed expression potential. In a wet-lab experiment, root-specific expression of HMA3 was reported in rice[22]. Further, *AtHMA3* is most potentially expressed during senescence, but germinated seeds, seedlings, young rosette, bolting and young flower also possess significant potential for *AtHMA3* expression. Interestingly, *AtHMA3* showed a

significant upregulation (>6.0 fold) in response to Fe-deficiency. Till now, HMA3 is known to induce its expression subjected to heavy metals in several plant species[23,24]. Nevertheless, our results suggest that *AtHMA3* is a potential gene that could contribute to Fe-deficiency tolerance in plants.

Conclusion

This *in silico* work identifies and characterizes 11 HMA3 homologs from each plant species. The analysis showed similar physicochemical properties, gene organization, and conserved motifs related to metal transport. The identified *cis*-acting elements were linked to stress resoonse, hormone, and other responsive factors. Sequence homology and phylogenetic tree showed the closest evolutionary relationship of Arabidopsis HMA3 with *Camelina sativa* and *Capsella rubella*. In addition, the interactome map displayed some partner genes of *AtHMA3* involved in metal transport in plants. It was also predicted that *AtHMA3* is expressed in root tissue during senescence and was significantly upregulated in response to Fe-deficiency. These findings will provide basic theoretical knowledge for the downstream studies on HMA3 function and characterization related to metal homeostasis in various plants.

Materials And Methods

Retrieval of HMA3 genes/proteins

AtHMA3 gene named as AT4G30120 in Uniprort/Aramene database (protein accession: NP_194741.2 and gene accession: NM_119158.4) was obtained from NCBI to use as a reference for homology search[25]. The search is filtered to match records with expect value between 0 and 0. The corresponding FASTA sequences of gene and protein were retrieved from the NCBI database. During filtering, one accession for each species was selected for analysis.

Analyses of HMA3 genes/proteins

Physico-chemical features of HMA3 protein sequences were analyzed by the ProtParam tool (https://web.expasy.org/protparam) as previously instructed[26]. Chromosomal and exon position was detected by the ARAMEMNON database (http://aramemnon.uni-koeln.de/). The CELLO (http://cello.life.nctu.edu.tw) server predicted the subcellular localization of proteins[27]. Protein domain families were searched in the Pfam database (http://pfam.xfam.org), and functions were assessed by the Phytozome v12.1 database[28]. The structural organization of HMA3 genes was predicted by FGENESH online tool[29]. Promoter position was predicted by Promoter 2.0 Prediction Server (http://www.cbs.dtu.dk/services/Promoter). In addition, *in silico* promoter analysis was carried out for 1 kbp upstream of translation start site of each HMA3 genes from the respective databases. The PLACE[30] and PlantCare[31] were used for scanning of cis-elements present in promoter regions of these genes. Besides, transcriptional start site (TSS) and PolA site were predicted by TSSPlant[32] and FGENESH 2.6 [29], respectively.

Phylogenetic Relationships and Identification of Conserved Protein Motifs

Multiple sequence alignments of HMA3 proteins were performed to identify conserved residues by using Clustal Omega. Furthermore, the five conserved protein motifs of the proteins were characterized by MEME Suite 5.1.1 (http://meme-suite.org/tools/meme) with default parameters, but five maximum numbers of motifs to find[33]. Motifs were further scanned by MyHits (https://myhits.sib.swiss/cgi-bin/motif_scan) web tool to identify the matches with different domains[34]. The MEGA (V. 6.0) developed the phylogenetic tree with the maximum likelihood (ML) method for 1000 bootstraps using 11 HMA3 homologs from 11 plant species[35].

Interactions and co-expression of HMA3 protein

The interactome network of HMA3 protein was generated using the STRING server (http://string-db.org) visualized in Cytoscape[36]. Additionally, the expression data of Arabidopsis *HMA3* was retrieved from Genevestigator software. Expression and co-expression associations of HMA3 was analyzed in different anatomical, developmental, and perturbations based on the Affymetrix Array Platforms (AT_AFFY-ATH1-0).

Structural analysis of HMA3 proteins

Structural analysis, such as transmembrane domains, was constructed with Protter (http://wlab.ethz.ch/protter/start) tool[37]. Besides, a two-dimensional secondary structure of MTP1 proteins constructed GORIV (https://npsa-prabi.ibcp.fr/NPSA/npsa_gor4.html).

Declarations

Ethics approval and consent to participate

We confirm that our study does not involve human subjects.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interests

The author(s) declare no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions

All authors have contributed to carry out this research. AHK and GG designed the experimental setup. UD, AFMMH, AKD and MAR analyzed data and results and prepared the figures, while AHK and GG wrote the manuscript. All authors read and approved the final manuscript.

References

1. Shanker AK, Cervantes C, Loza-Tavera H, Avudainayagam S (2005) Chromium toxicity in plants. Environ Int 31:739– 753.
2. Tripathi DK, Singh VP, Kumar D, Chauhan DK (2012) Impact of exogenous silicon addition on chromium uptake, growth, mineral elements, oxidative stress, antioxidant capacity, and leaf and root structures in rice seedlings exposed to hexavalent chromium. Acta Physiol Plant 34:279-289.
3. Baxter, I., Tchieu, J., Sussman, M. R., Boutry, M., Palmgren, M. G., Gribskov, M., Harper, J. F., and Axelsen, K. B. (2003). Genomic comparison of P-Type ATPase ion pumps in Arabidopsis and rice. Plant Physiol., 132, 618–628.
4. Williams, L. E. and Mills, R. F. (2005). P(1B)-ATPases - an ancient family of transition metal pumps with diverse functions in plants. Trends Plant Sci., 10, 491–502.
5. Chao DY, Silva A, Baxter I, Huang YS, Nordborg M, Danku J, Lahner B, Yakubova E, Salt DE. 2012. Genome-wide association studies identify heavy metal ATPase3 as the primary determinant of natural variation in leaf cadmium in Arabidopsis thaliana. PLoS Genetics 8: e1002923.
6. Morel M, Crouzet J, Gravot A, Auroy P, Leonhardt N, Vavasseur A, Richaud P. 2009. AtHMA3, a P1B-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in Arabidopsis. Plant Physiology 149: 894–904.
7. Zhang J, Zhang M, Shohag MJ, Tian S, Song H, Feng Y, Yang X. 2016. Enhanced expression of SaHMA3 plays critical roles in Cd hyperaccumulation and hypertolerance in Cd hyperaccumulator Sedum alfredii Hance. Planta 243: 577–589.
8. Zhang L, Wu J, Tang Z, Huang XY, Wang X, Salt DE, Zhao FJ. Variation in the BrHMA3 coding region controls natural variation in cadmium accumulation in Brassica rapa vegetables. J Exp Bot. 2019 Oct;70(20) 5865-5878.
9. Sasaki, A., Yamaji, N., & Ma, J. F. (2014). Overexpression of OsHMA3 enhances Cd tolerance and expression of Zn transporter genes in rice. Journal of experimental botany, 65(20), 6013–6021.
10. Ueno, D., Yamaji, N., Kono, I., Huang, C. F., Ando, T., Yano, M., & Ma, J. F. (2010). Gene limiting cadmium accumulation in rice. Proceedings of the National Academy of Sciences of the United States of America, 107(38), 16500–16505.
11. Becher, M., Talke, I. N., Krall, L., and Krämer, U. (2004). Cross-species microarray transcript profiling reveals high constitutive expression of metal homeostasis genes in shoots of the zinc hyperaccumulator *Arabidopsis halleri*. *Plant J.* **37,** 251-268 (2004).
12. Morel, M. *et al.* AtHMA3, a P1B-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in Arabidopsis. *Plant. Physiol.* **149**, 894–904 (2009).

13. Morsomme, P., & Boutry, M. The plant plasma membrane H(+)-ATPase: structure, function and regulation. *Biochim. Biophys. Acta.* **1465,** 1–16 (2000).
14. Logan, H., Basset, M., Véry, A.-A., & Sentenac, H. Plasma membrane transport systems in higher plants: From black boxes to molecular physiology. *Physiol Plant.* **100,** 1-15 (1997).
15. Gravot, A. *et al.* AtHMA3, a plant P1B-ATPase, functions as a Cd/Pb transporter in yeast. *FEBS letters.* **561,** 22–28 (2004).
16. Raimunda, D., Subramanian, P., Stemmler, T., & Argüello, J. M. A tetrahedral coordination of zinc during transmembrane transport by P-type Zn2+-ATPases. *Biochim. Biophys. Acta*. **1818,** 1374–1377 (2012).
17. Wong, A., Gehring, C., & Irving, H.R. Conserved Functional Motifs and Homology Modeling to Predict Hidden Moonlighting Functional Sites. *Frontiers. Bioeng. Biotech.* **3,** 82 (2015).
18. Palmgren, M. G. Plant plasma membrane H^+-ATPases: Powerhouses for Nutrient Uptake. *Ann. Rev. Plant. Physiol. Plant. Mol. Biol.* **52,** 817-845 (2001).
19. Freeley, M., Kelleher, D., & Long, A. Regulation of Protein Kinase C function by phosphorylation on conserved and non-conserved sites. *Cell. Signal.* **23,** 753–762 (2011).
20. Thingholm, T. E., Rönnstrand, L., & Rosenberg PA. Why and how to investigate the role of protein phosphorylation in ZIP and ZnT zinc transporter activity and regulation. *Cell Mol Life Sci.* **77,** 3085–3102 (2020).
21. Hussain, D. *et al.* P-type ATPase heavy metal transporters with roles in essential zinc homeostasis in Arabidopsis. *Plant Cell.* **16,** 1327–1339 (2004).
22. Cai H. *et al.* Root-specific expression of rice OsHMA3 reduces shoot cadmium accumulation in transgenic tobacco. *Mol Breeding*. **39,** 49 (2009).
23. Yao, X., Cai, Y., Yu, D., & Liang, G. bHLH104 confers tolerance to cadmium stress in *Arabidopsis thaliana*. *J. Integr. Plant. Biol.* **60,** 691–702 (2018).
24. Colangelo, E. P., & Guerinot, M. L. The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. *Plant Cell.* **16,** 3400–3412 (2004).
25. Stephen, A. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic. Acids. Res.* **25,** 3389-3402 (1997).
26. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The proteomics protocols handbook. Humana, Louisville, p. 571–607 (2005).
27. Yu, C. S., Chen, Y. C., Lu, C. H., & Hwang, J. K. Prediction of protein subcellular localization. *Proteins.* **64,** 643–651)2006(.
28. El-Gebali, L. The Pfam protein families database in 2019. *Nucleic. Acids. Res.* **47,** D427–D432 (2019).
29. Solovyev, V., Kosarev, P,, Seledsov, I., & Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **7,** 1-10 (2006).

30. Higo, K., Ugawa, Y., Iwamoto, M., & Korenaga, T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic. Acids Res.* **27,** 297–300 (1999).
31. Lescot, M. *et al.* Plant CARE, a data base of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic. Acids. Res.* **30,** 325–327 (2002).
32. Shahmuradov, I, A,, Umarov, R, K., & Solovyev, V. V. TSSPlant: a new tool for prediction of plant Pol II promoters. *Acids. Res.* **45,** e65 (2017).
33. Timothy, L. B. *et al.* MEME SUITE: tools for motif discovery and searching. Nucleic *Nucleic. Acids. Res.* **37,** 202-208 (2009).
34. Sigrist, C. J. *et al.* ROSITE, a protein domain database for functional characterization and annotation. *Nucleic. Acids. Res.* **38,** D161-6 (2010).
35. Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30,** 12, 2725– 2729 (2013).
36. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic. Acids. Res.* **47,** D607-613 (2019).
37. Omasits, U., Ahrens, C. H., Müller, S., & Wollscheid, B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics.* 30, 884-886 (2014)

Figures

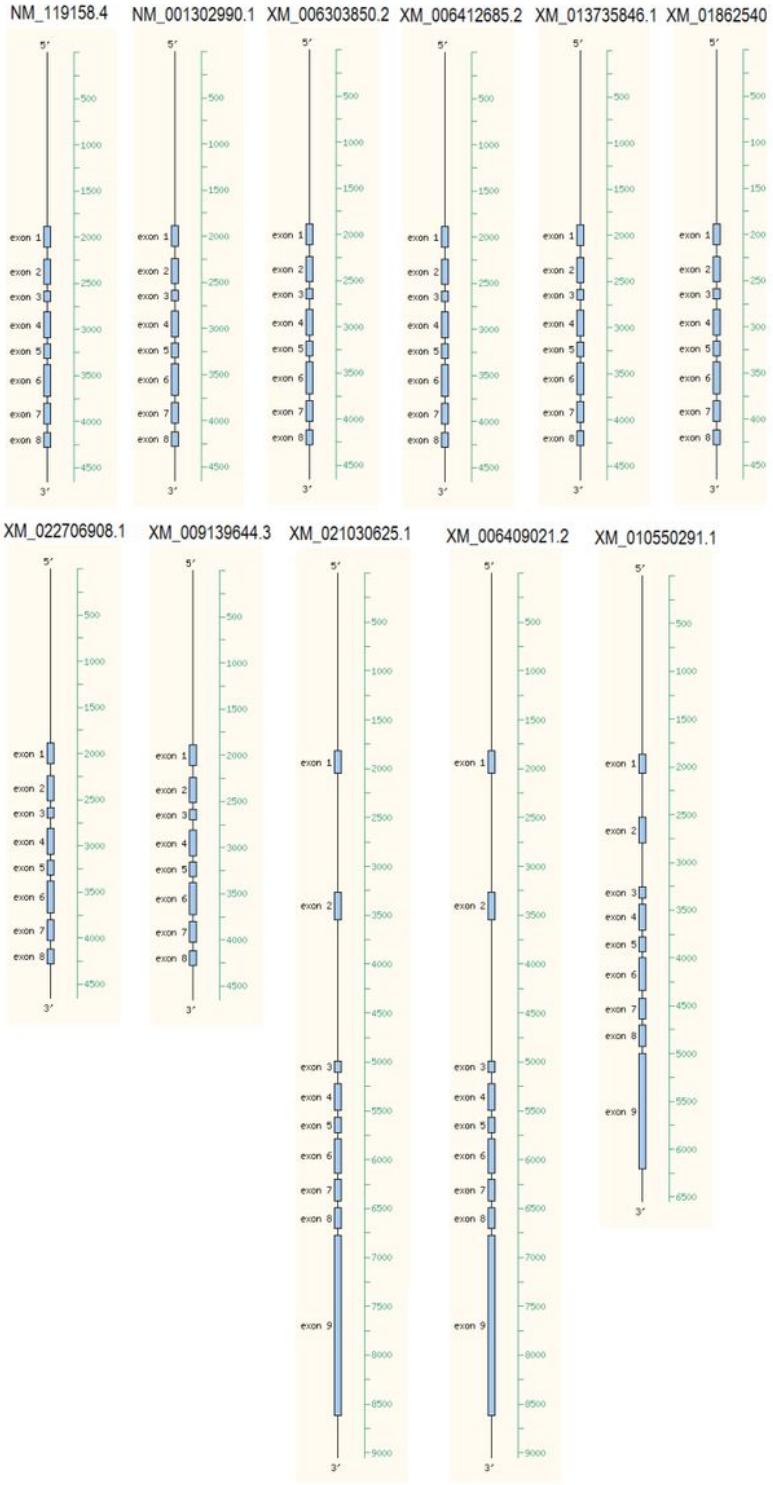


Figure 1

Gene organization of HMA3 homologs.

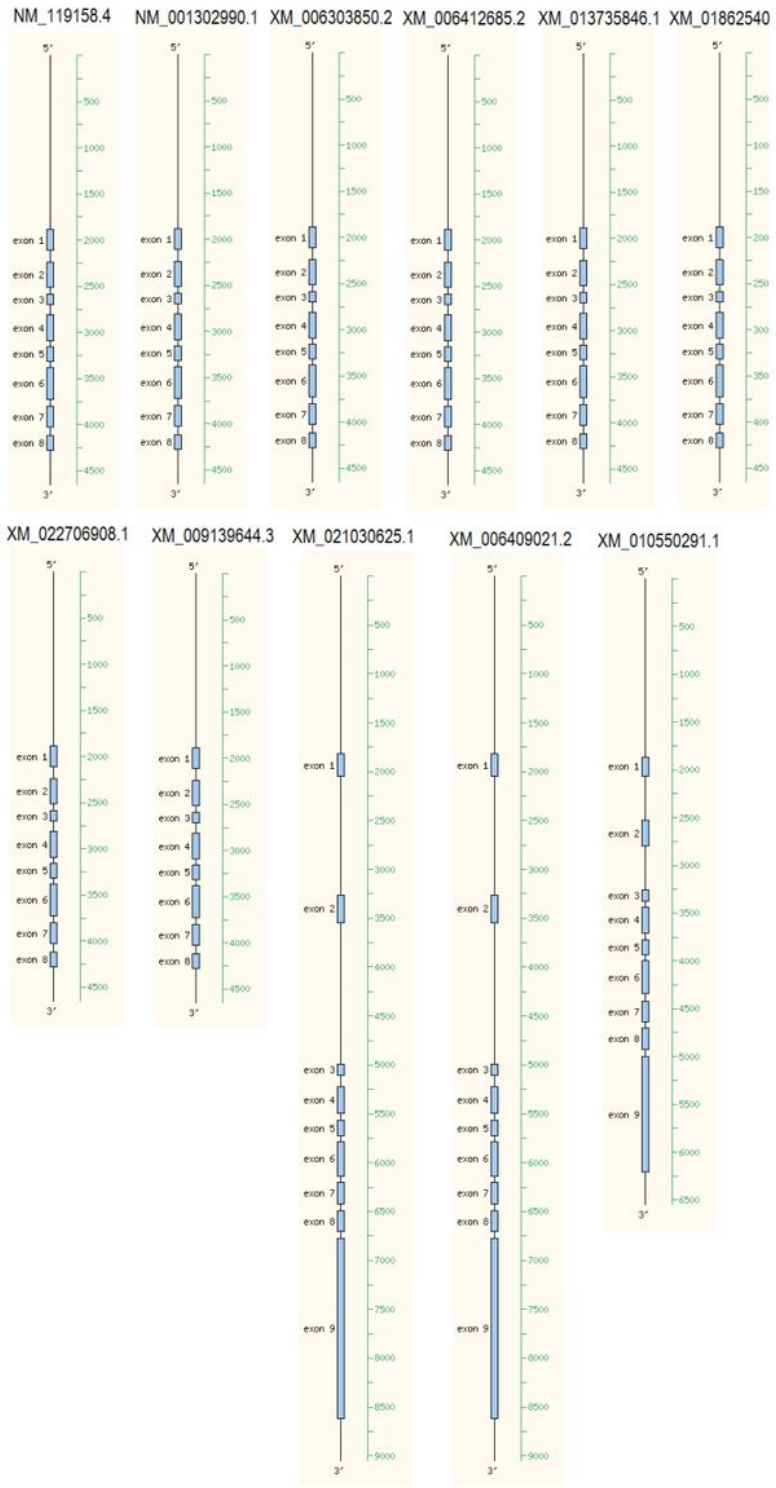


Figure 1

Gene organization of HMA3 homologs.

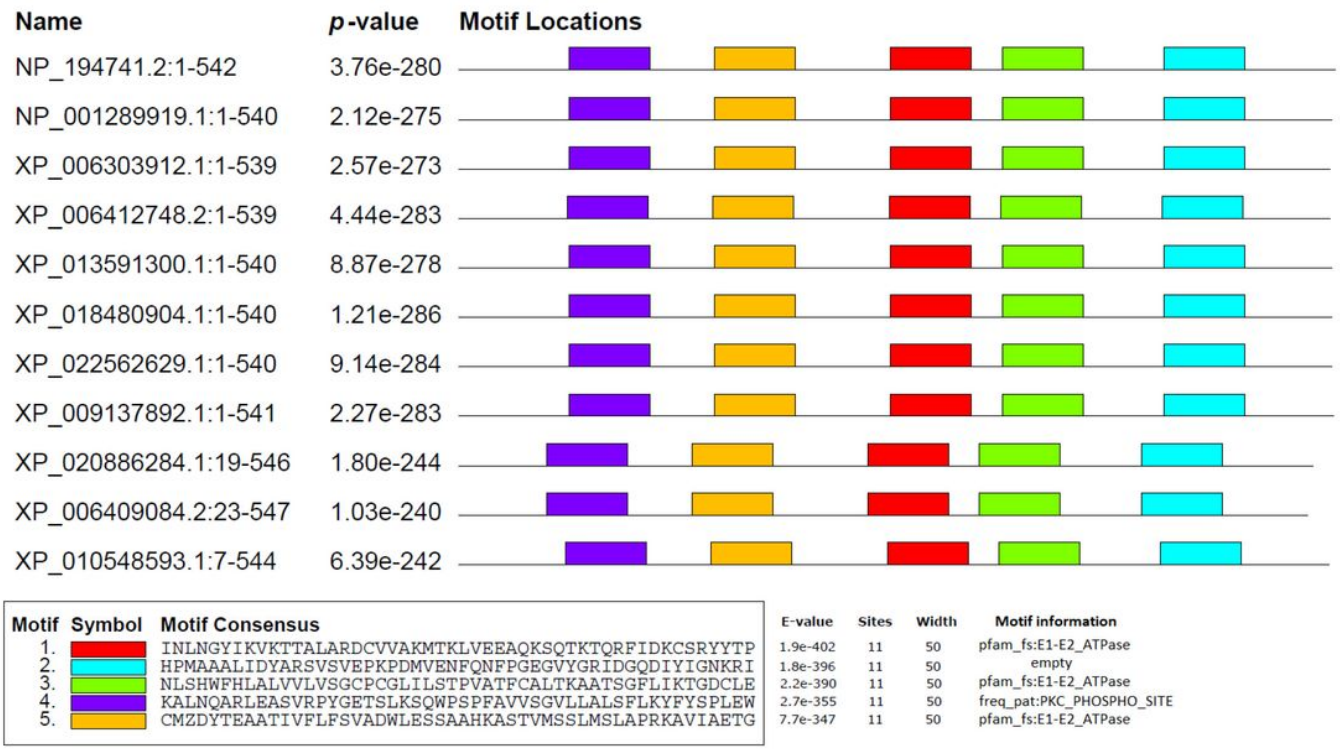


Figure 2

Schematic representation of the 5 conserved motifs in 11 HMA3 protein homologs across 11 plant species. Scale bar corresponds to 0.1 amino acid substitution per residue. Different motifs, numbered 1–5, are displayed in different colored boxes.

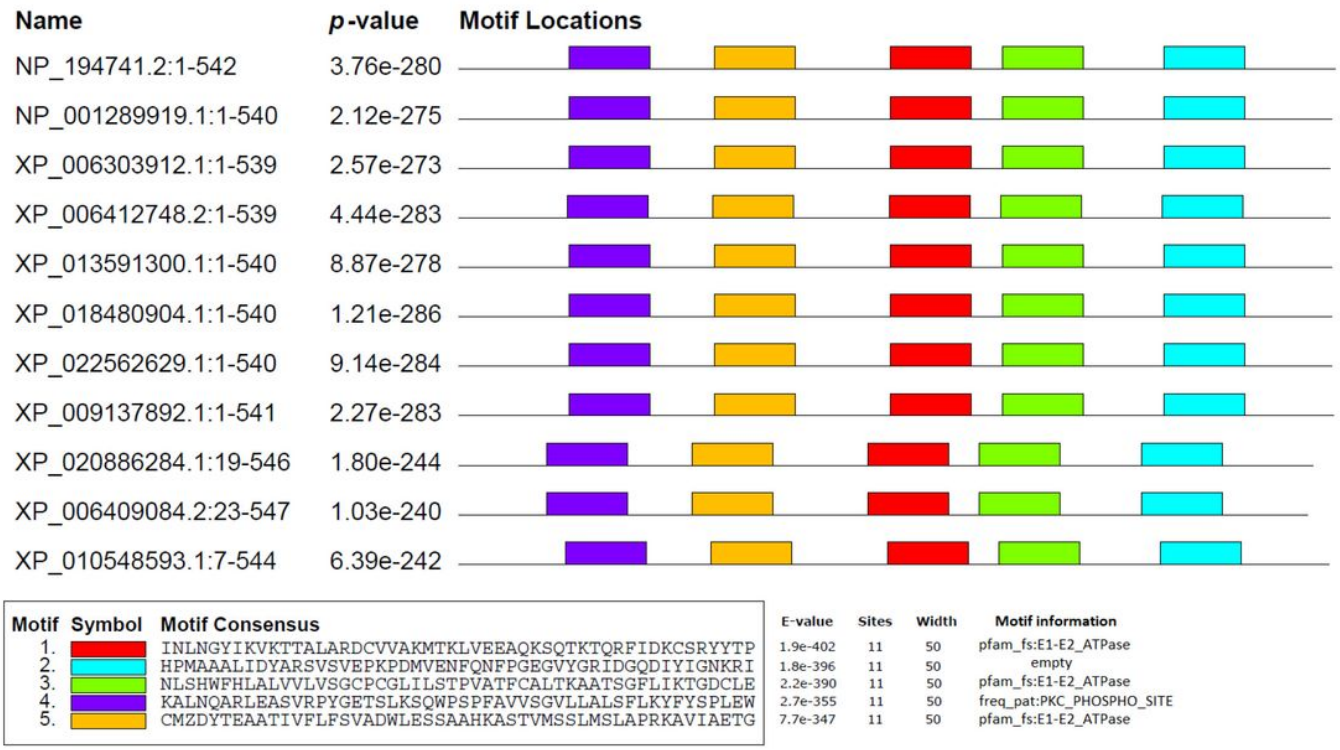


Figure 2

Schematic representation of the 5 conserved motifs in 11 HMA3 protein homologs across 11 plant species. Scale bar corresponds to 0.1 amino acid substitution per residue. Different motifs, numbered 1–5, are displayed in different colored boxes.

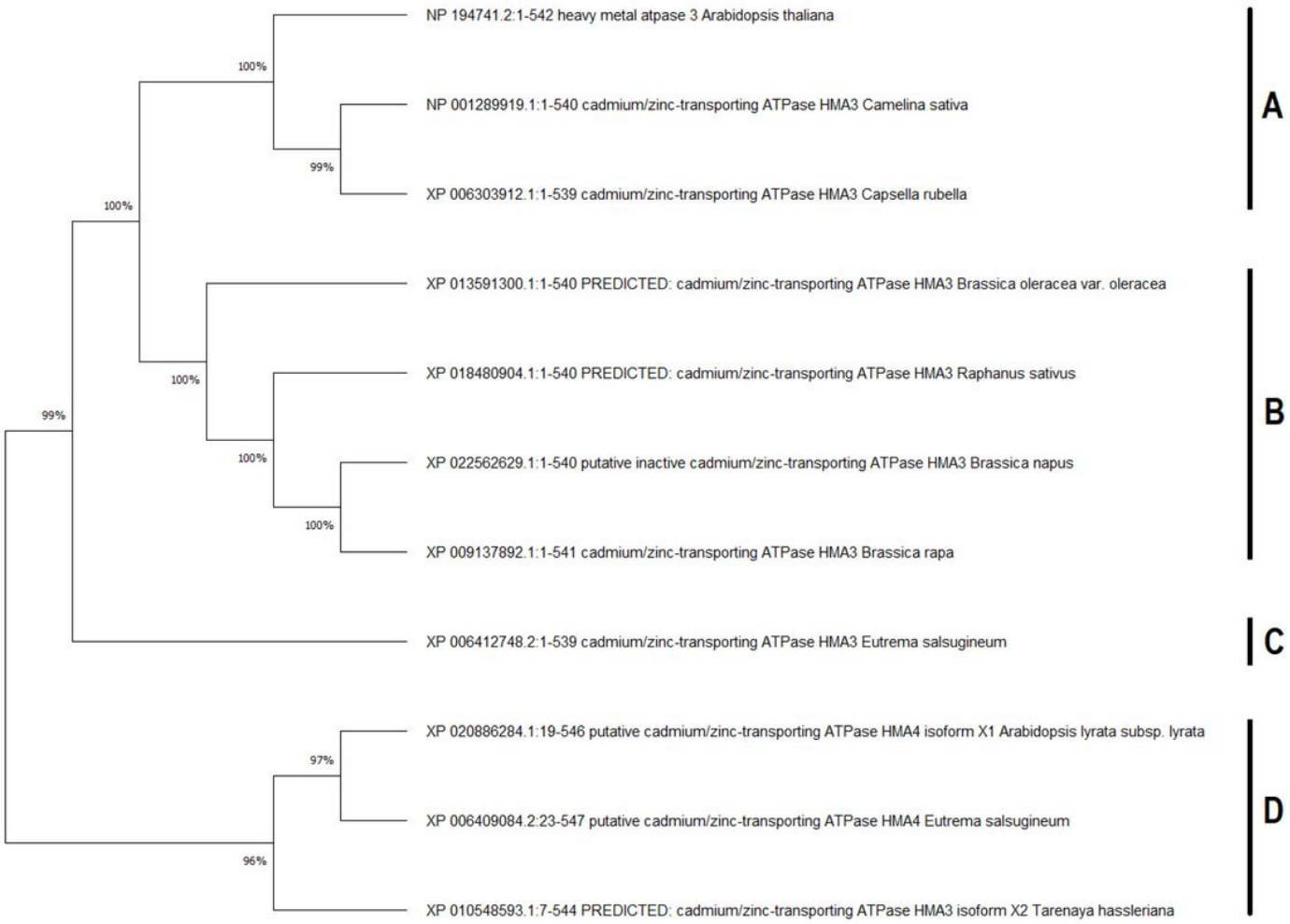


Figure 3

Phylogenetic trees of HMA3 protein homologs. Trees were constructed by MEGA 6 software with the maximum likelihood (ML) method for 1000 bootstrap values. Trees were used as benchmarks to analyze the clustering of 11 HMA3 sequences. A, B, C and D represents different clusters.

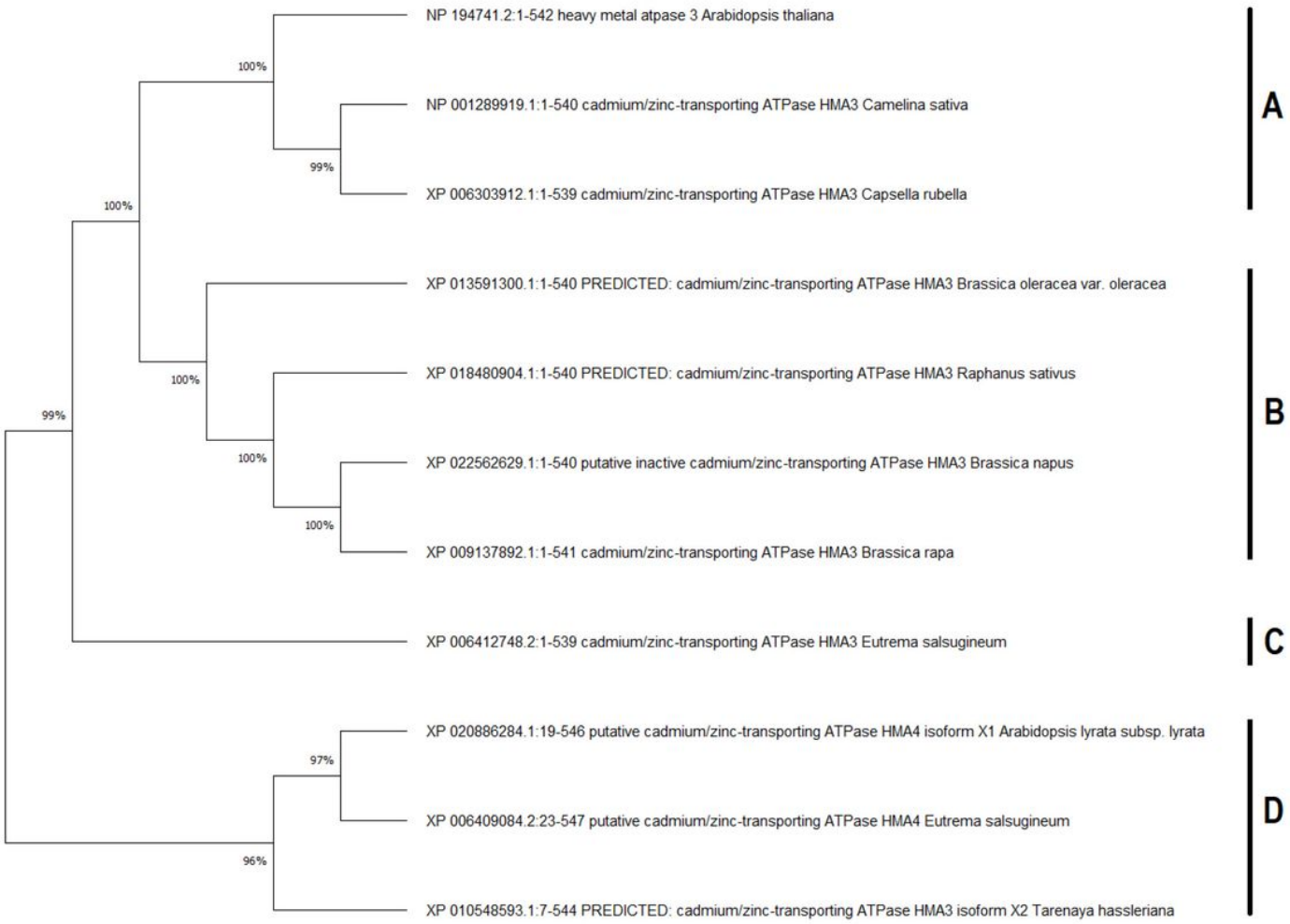
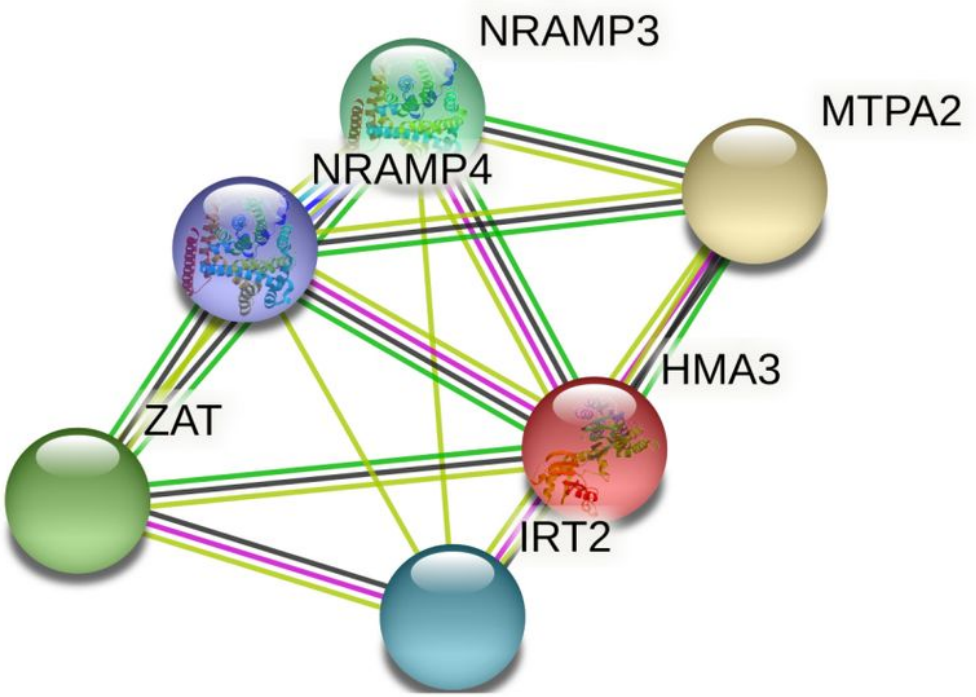


Figure 3

Phylogenetic trees of HMA3 protein homologs. Trees were constructed by MEGA 6 software with the maximum likelihood (ML) method for 1000 bootstrap values. Trees were used as benchmarks to analyze the clustering of 11 HMA3 sequences. A, B, C and D represents different clusters.



Your Input:

HMA3	*Putative inactive cadmium/zinc-transporting ATPase HMA3; Encodes a protein similar to Zn-ATPase, a P1B-type ATPases transport zinc (542 aa)*

Predicted Functional Partners:

MTPA2	*Metal tolerance protein A2; Member of Zinc transporter (ZAT) family. Contributes to basic cellular Zn tolerance and controls Z...*
ZAT	*Encodes a member of the zinc transporter (ZAT) and cation diffusion facilitator (CDF) families. It is expressed throughout the...*
NRAMP3	*Natural resistance-associated macrophage protein 3; Vacuolar metal transporter involved in intracellular metal homeostasis. ...*
IRT2	*Iron regulated transporter 2; High-affinity iron transporter that mediates under iron- deficiency the iron uptake from the rhizos...*
NRAMP4	*Natural resistance associated macrophage protein 4; Vacuolar metal transporter involved in intracellular metal homeostasis. ...*

local network cluster (STRING)

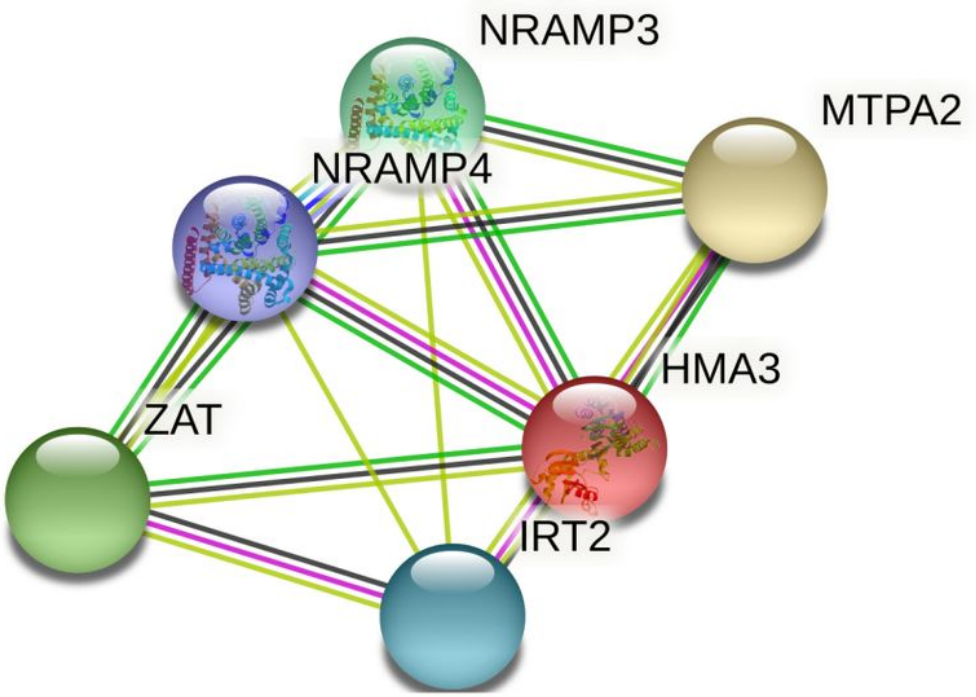
cluster	*description*	*count in network*	*strength*	*false discovery rate*
CL:28166	Nickel transport, and Cation efflux protein, cytoplasmic do...	3 of 5	3.44	4.56e-10
CL:28164	mixed, incl. manganese ion transport, and Nickel transport	5 of 11	3.32	2.84e-15
CL:28176	Ion influx/efflux at host-pathogen interface, and Solute carri...	2 of 5	3.26	1.07e-06
CL:28126	mixed, incl. transition metal ion transmembrane transporter...	6 of 84	2.51	7.39e-15

Reactome Pathways

pathway	*description*	*count in network*	*strength*	*false discovery rate*
ATH-6803544	Ion influx/efflux at host-pathogen interface	2 of 4	3.36	2.39e-06
ATH-435368	Zinc efflux and compartmentalization by the SLC30 family	2 of 4	3.36	2.39e-06
ATH-2980736	Peptide hormone metabolism	2 of 4	3.36	2.39e-06
ATH-264876	Insulin processing	2 of 4	3.36	2.39e-06
ATH-425410	Metal ion SLC transporters	4 of 11	3.22	1.39e-11

Figure 4

Gene interaction partners and gene network analysis of AtHMA3 and its homologs. Interactome was generated using Cytoscape for STRING data.



Your Input:

HMA3	*Putative inactive cadmium/zinc-transporting ATPase HMA3; Encodes a protein similar to Zn-ATPase, a P1B-type ATPases transport zinc (542 aa)*

Predicted Functional Partners:

MTPA2	*Metal tolerance protein A2; Member of Zinc transporter (ZAT) family. Contributes to basic cellular Zn tolerance and controls Z...*
ZAT	*Encodes a member of the zinc transporter (ZAT) and cation diffusion facilitator (CDF) families. It is expressed throughout the...*
NRAMP3	*Natural resistance-associated macrophage protein 3; Vacuolar metal transporter involved in intracellular metal homeostasis. ...*
IRT2	*Iron regulated transporter 2; High-affinity iron transporter that mediates under iron- deficiency the iron uptake from the rhizos...*
NRAMP4	*Natural resistance associated macrophage protein 4; Vacuolar metal transporter involved in intracellular metal homeostasis. ...*

local network cluster (STRING)

cluster	*description*	*count in network*	*strength*	*false discovery rate*
CL:28166	Nickel transport, and Cation efflux protein, cytoplasmic do...	3 of 5	3.44	4.56e-10
CL:28164	mixed, incl. manganese ion transport, and Nickel transport	5 of 11	3.32	2.84e-15
CL:28176	Ion influx/efflux at host-pathogen interface, and Solute carri...	2 of 5	3.26	1.07e-06
CL:28126	mixed, incl. transition metal ion transmembrane transporter...	6 of 84	2.51	7.39e-15

Reactome Pathways

pathway	*description*	*count in network*	*strength*	*false discovery rate*
ATH-6803544	Ion influx/efflux at host-pathogen interface	2 of 4	3.36	2.39e-06
ATH-435368	Zinc efflux and compartmentalization by the SLC30 family	2 of 4	3.36	2.39e-06
ATH-2980736	Peptide hormone metabolism	2 of 4	3.36	2.39e-06
ATH-264876	Insulin processing	2 of 4	3.36	2.39e-06
ATH-425410	Metal ion SLC transporters	4 of 11	3.22	1.39e-11

Figure 4

Gene interaction partners and gene network analysis of AtHMA3 and its homologs. Interactome was generated using Cytoscape for STRING data.

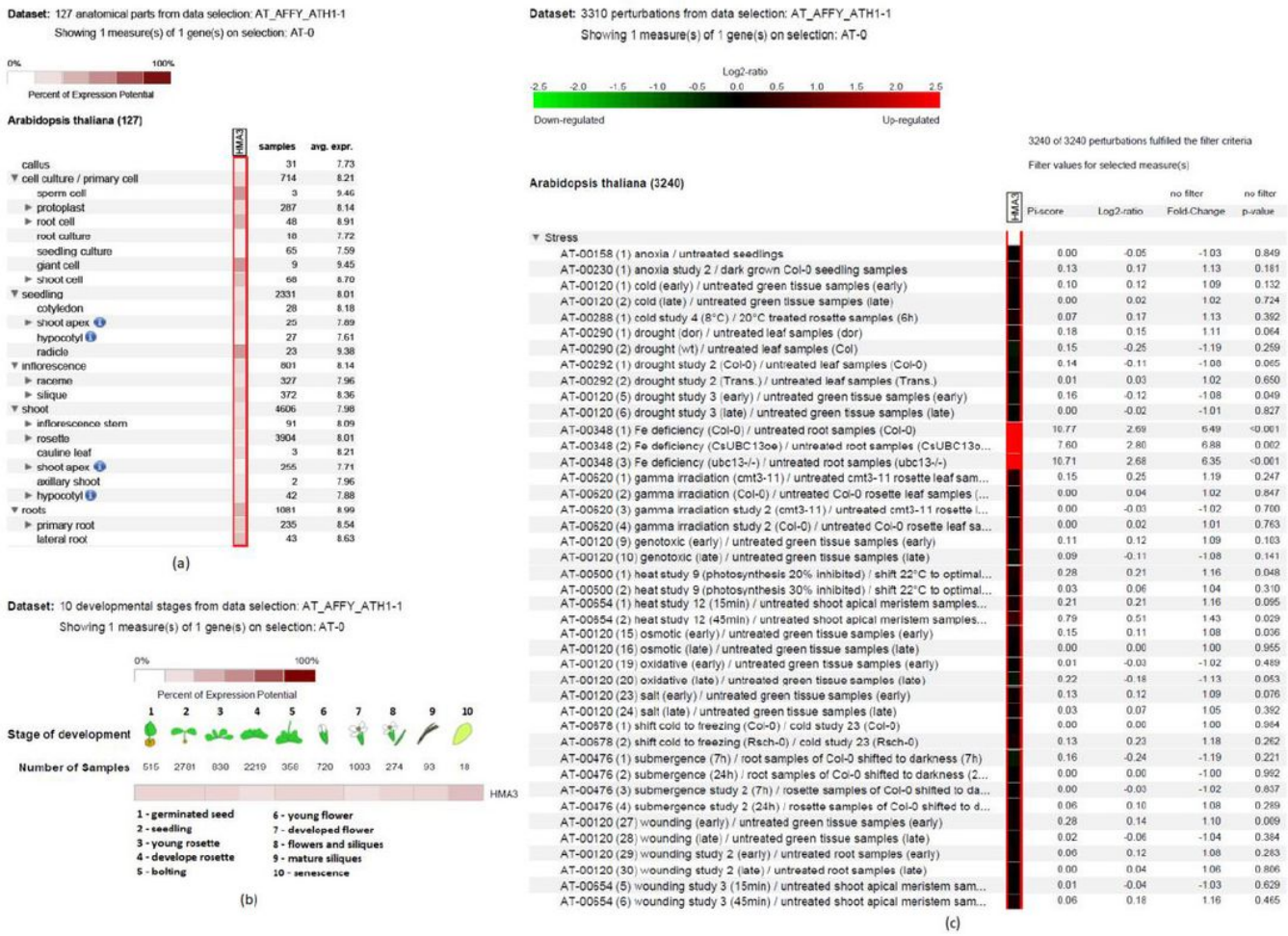


Figure 5

Expression profiles of AtHMA3 in different anatomical part, developmental stage and perturbations in Genevestigator Affymetrix platform.

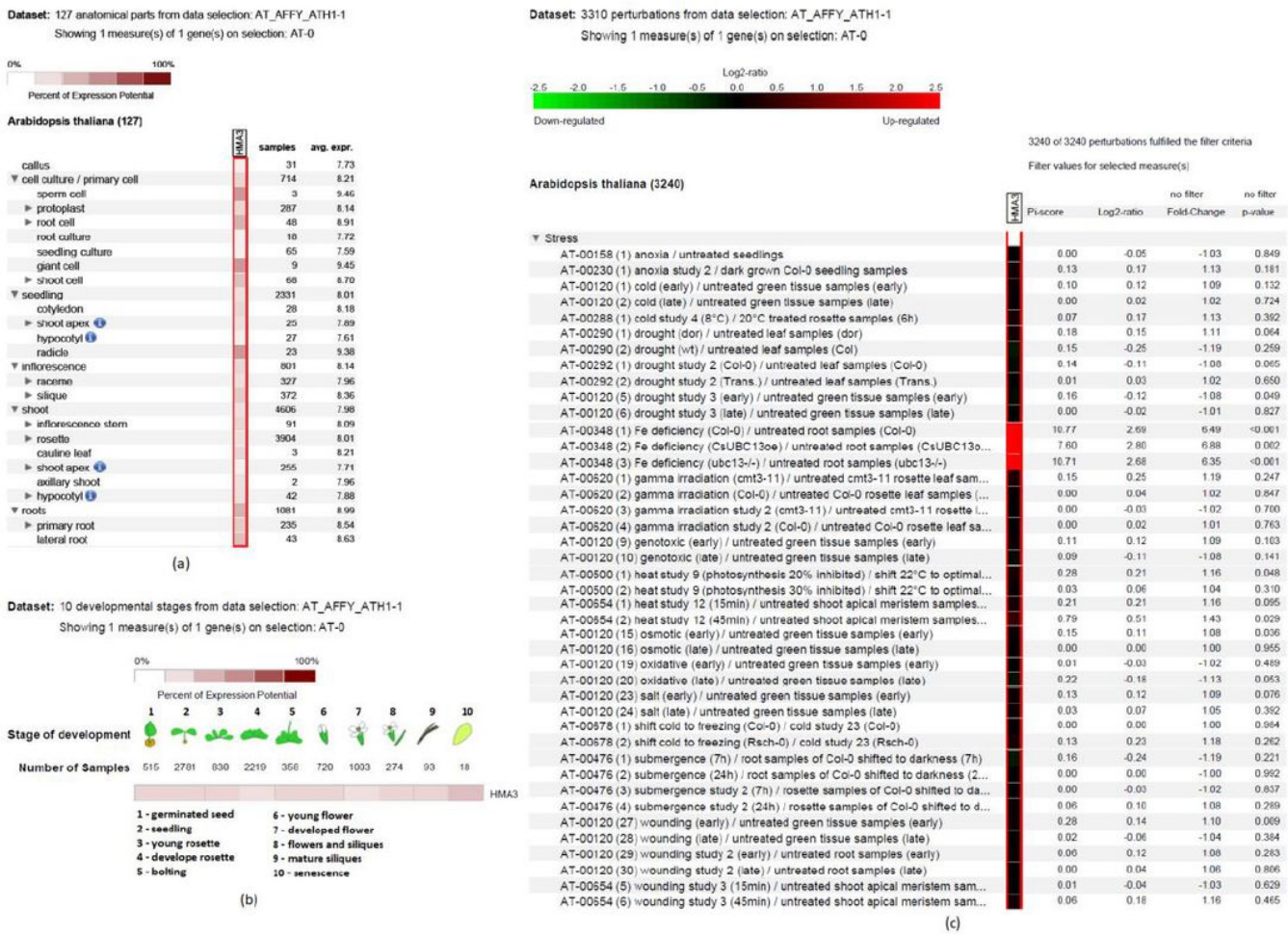


Figure 5

Expression profiles of AtHMA3 in different anatomical part, developmental stage and perturbations in Genevestigator Affymetrix platform.

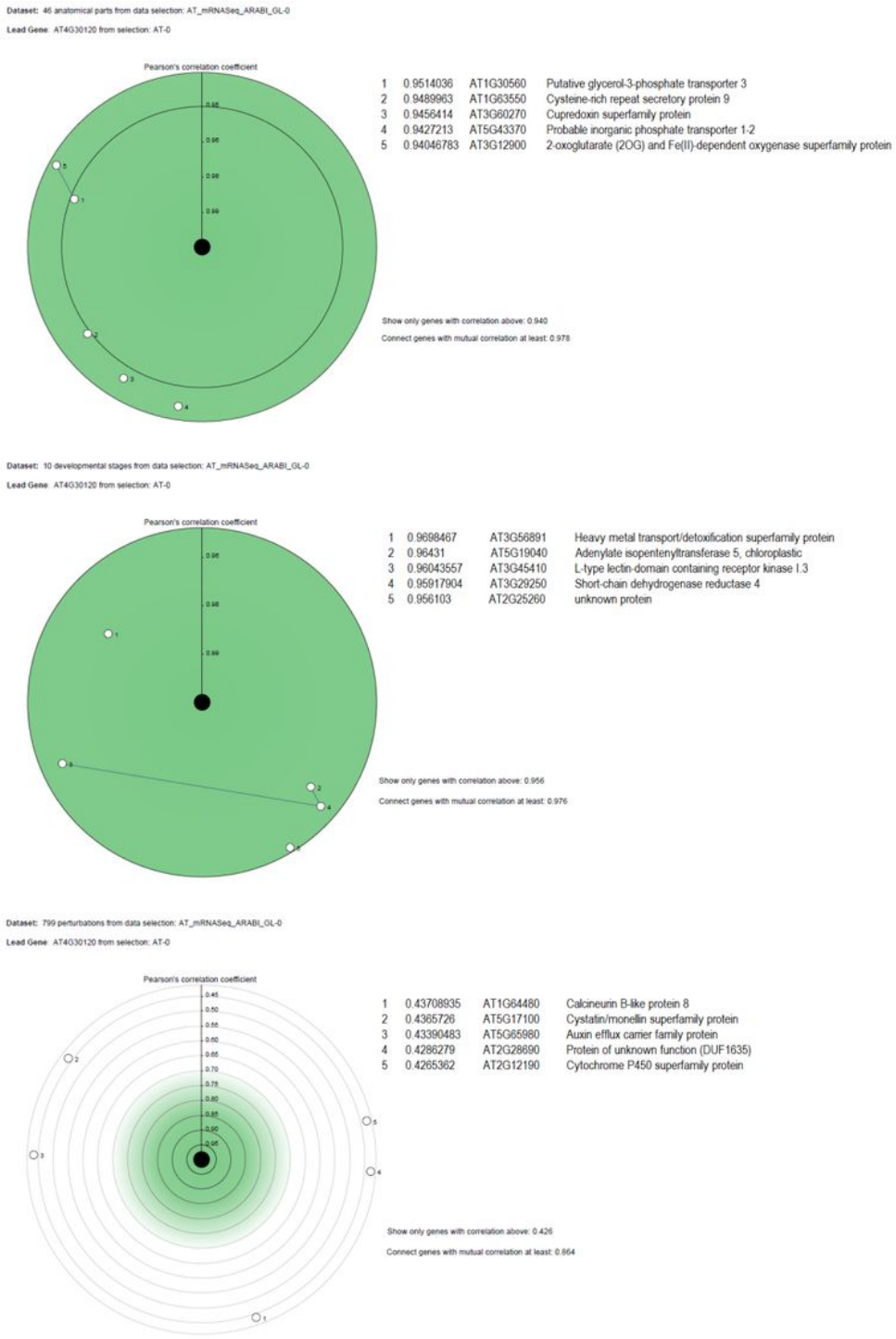


Figure 6

List of top 5 genes co-expressed with AtHMA3 in different anatomical parts and developmental stages of plants.

Figure 6

List of top 5 genes co-expressed with AtHMA3 in different anatomical parts and developmental stages of plants.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile.docx
- SupplementaryFile.docx
- Fig.S.1.jpg
- Fig.S.1.jpg
- Fig.S.2.jpg
- Fig.S.2.jpg