Nature Environment & Pollution Technology

This is a peer-reviewed prepublished version of the paper

Identification and Functional Annotation of *Echium plantagineum* Metallothioneins for Reduction in Heavy Metals in Soil Using Molecular Docking

Y. S. Rasheed[1]†, M.S. AL-Janaby[2] and M. H. Abbas[2]
[1]Department of Biology, College of Science, University of Anbar, Iraq
[2]Department of Biotechnology, College of Science, University of Anbar, Iraq
†Corresponding author: Yousry Sadoon Rasheed; yousry.rasheed@uoanbar.edu.iq

Key Words	Metallothioneins, *Echium plantagineum*, Ctr copper transporter, Zinc/iron permease, Nicotianamine synthase
DOI	https://doi.org/10.46488/NEPT.2024.v24i01.D1671 (DOI will be active only after the final publication of the paper)
Citation of the Paper	Rasheed, Y. S., AL-Janaby, M. S. and Abbas, M. H., 2025. Identification and functional annotation of *Echium plantagineum* metallothioneins for reduction in heavy metals in soil using molecular docking. *Nature Environment and Pollution Technology*, 24(1), D1671. https://doi.org/10.46488/NEPT.2024.v24i01.D1671

ABSTRACT

Heavy metal contamination in soil poses a significant environmental challenge globally, affecting agricultural productivity and human health. Phytoremediation, using plants to extract and detoxify heavy metals, presents a promising solution. This study investigates the novel potential of *Echium plantagineum*, a metal-tolerant species, in phytostabilization and phytoremediation and explores the role of metallothioneins in heavy metal reduction. A comprehensive literature review identified known metallothioneins involved in heavy metal reduction across various plant species. Moreover, genome annotation and gene prediction of *Echium plantagineum* were performed, predicting a total of 39,520 proteins. This comprehensive protein list facilitates the identification of metallothioneins or other metal-related proteins with potential functional roles in heavy metal tolerance, suggesting new targets to improve the effectiveness of phytoremediation. The sequences of these proteins were utilized to construct a protein BLAST database, against which known metallothioneins protein sequences from other plant species were subjected to BLAST searches, resulting in 41 top hits. Subsequent 3D modeling, structural analysis, protein-metal virtual screening, and functional annotation of the proteins revealed novel high affinities of Ctr copper transporter, zinc/iron permease, and nicotianamine synthase proteins with nickel, zinc, and zinc ion, suggesting their unexplored roles in the uptake of aforementioned ligands. Notably, this study identifies novel metallothioneins proteins in *Echium plantagineum*, highlighting their role in metal tolerance and phytoremediation.

INTRODUCTION

Heavy metals are naturally occurring elements found in the crust of the earth, and some of them are essential for biological processes. However, human activities introduce these metals into aquatic environments (Al-Khashman 2004). The primary sources of heavy metals are industrial activities like mining, foundries, and smelters, as well as diffuse sources such as pipes, products, and the combustion of by-products. These metals can also come from traffic and other human activities (Guvenç et al. 2003). Heavy metals such as copper (Cu), iron (Fe), manganese (Mn), zinc (Zn), cadmium (Cd), arsenic (As), nickel (Ni), cobalt (Co), lead (Pb), aluminum (Al), and mercury (Hg) contamination threatens soil ecosystems, agricultural productivity, and global human health (Rashid et al. 2023).

For instance, Cu and Zn are commonly used in industrial and agricultural applications, while Cd is often found in industrial waste and fertilizers. Metal contamination can result from mining activities and arsenic pesticides, while Pb comes from paints and vehicle emissions (Zaccone et al. 2010). Fe and Mn originate from industrial activities and natural watering processes, whereas Ni and Co may stem from industrial emissions and mining activities (Santana et al. 2020). Coal combustion and mining are primary sources of Hg pollution, whereas Al causes environmental pollution through soil and water (Shinzato & Hypolito 2016, Streets et al. 2018).

The vague information on potential functions, structures, and properties of plant metallothioneins has resulted in sequence diversity (Freisinger 2009). Metallothioneins belong to the small, cystine-rich family and perform critical functions in regulating metal ions in plants. Similarly, these proteins act as metal chelators to maintain metal homeostasis through metal ion binding (Chatterjee et al. 2020, Joshi et al. 2016, Sharma et al. 2016).

Metallothioneins play a homeostatic role by preventing the accumulation of heavy metals through detoxification (Yaashikaa et al. 2022). Additionally, environmental factors such as oxidative stress and metal exposure to plant tissues are responsible for the higher expression of plant metallothionein genes, indicating their crucial role in metal tolerance (Ahn et al. 2012).

Due to their crucial role in heavy metal detoxification from soil, there is a dire need for the identification, characterization, and functional annotation of novel metallothioneins in plant species. Purple viper's bugloss, or *Echium plantagineum*, is a plant species with a high tolerance to contaminants in polluted soils. (Latini et al. 2022). *Echium plantagineum* tissues aid in soil decontamination by absorbing heavy metals and may have phytoremediation potential due to increased uptake of contaminants and soil microbes (Midhat et al. 2019, Skoneczny et al. 2019, Zine et al. 2020). In previous studies, the presence and role of metallothioneins as hyperaccumulators in *Echium plantagineum* have not been reported in the literature despite its promising potential for phytostabilization. These metallothioneins could play a crucial role in the uptake of heavy metals from the soil.

This study addresses this gap by identifying, characterizing, and functionally annotating novel metallothioneins in *Echium plantagineum*. Due to unreported hyperaccumulator genes, this study employs sequence homology to comprehend the tolerance mechanisms of this plant towards different heavy metal ions, including Cu, Fe, Mn, Zn, Cd, As, Ni, Co, Pb, Mo, Ca, Fe^{+2}, Fe^{+3}, Mn^{+2}, Zn^{+2}, Al, Hg, and Cr. This study also emphasizes the functional roles and interactions of these metallothioneins, demonstrating the potential of *Echium plantagineum* in tolerating heavy metals. These novel metallothionein genes may act as transgenic agents to facilitate phytoremediation in other plants capable of hyperaccumulating contaminants.

MATERIALS AND METHODS

Literature Review

An extensive literature review was conducted using Google Scholar and PubMed databases to identify the metallothionein genes across various hyperaccumulator plant species involved in heavy metal reduction in soil. The literature review was performed using both subject terms and free words such as phytoremediation, metal absorption, metal transport, sequestration, soil detoxification, heavy metals, genes, proteins, hyperaccumulators, plants, and metallothioneins.

Sequences and Genomic Annotation Data Retrieval

A genome annotation data file and protein sequences of the literature-identified genes were retrieved from the National Center for Biotechnology Information (NCBI) (Anon n.d.-a) and the UniProt database (Anon n.d.-c). Moreover, the *Arabidopsis thaliana* species was chosen to predict the genes from the genome FASTA file through the AUGUSTUS tool (Stanke 2003), resulting in a General Gesture Format (GFF3) file with predicted genes and their genomic coordinates. NCBI is a repository of biological information and data (Sayers et al. 2021, Pruitt & Sherry 2021), while

UniProt is a vast database of annotation data and protein sequences. Similarly, the AUGUSTUS tool is a popular gene prediction software (Keller et al. 2011).

The BLAST Sequence Analysis of *Echium plantagineum*

The BLAST sequence analysis was conducted to identify the homologous genes based on protein sequence similarities. The BLAST+ package was utilized to create a protein database for predicted *Echium plantagineum* proteins. Literature-identified proteins from other plant species with metal uptake features were used as a query for BLAST database searching. Subsequently, the "blastp" command was employed with parameters such as a maximum of 1 alignment and an e-value threshold of 1e-30. The E-value is a measure of sequence alignment quality, indicating the number of alignments with a score of S that can be found by chance in a database. The default value is 10, but increasing it increases the number of chance matches and makes the search more stringent. Protein-based homology search typically uses a threshold of $E \leq 1e\text{-}5$, with lower values indicating clearer homology, such as 1e-25 (Choudhuri 2014). A low E-value threshold (1e-30) was set to ensure only highly significant matches are reported, making it useful for applications like homologous protein identification or functional annotation. Similarly, limiting the number of alignments to 1 helps focus on the single best match for each query sequence, identifying the top hit for each protein. The results were organized into a tabular format, with hit identifiers stored separately in a text file. Moreover, the "blastdbcmd" command tool of the NCBI BLAST+ was employed to retrieve top hits from the protein database (Anon 2021). This analysis identified homologous sequences of metal uptake-related proteins from *Echium plantagineum* species.

Structural Prediction of Homologs and Functional Annotation

The next step involved predicting the 3D structures of the identified protein sequences. This was accomplished by utilizing Omegafold, a well-known 3D modeling tool for its accuracy in *ab initio* protein structure prediction (Wu et al. 2022). OmegaFold can handle protein sequences as long as 4096 residues when running on an NVIDIA A100 Graphics card with 80 GB of memory (Anon n.d.-b). However, due to limitations in computational resources, sequences larger in length (>1500 amino acids) were predicted through SWISS-MODEL (Waterhouse et al. 2018), a widely used homology modeling tool that leverages experimentally resolved homologous protein structures as templates. To functionally annotate the novel proteins, BLAST web-based search against other plant organisms was performed, whereas if no homologous sequences were found, InterPro database searching was performed to identify protein family or domain based on sequence search. InterPro is an extensive repository of protein families, superfamilies, and domains (Paysan-Lafosse et al. 2023).

Ligand Retrieval and Virtual Screening

Structures of ligands (Cu, Fe, Mn, Zn, Cd, As, Ni, Co, Pb, Mo, Ca, Fe^{+2}, Fe^{+3}, Mn^{+2}, Zn^{+2}, Al, Hg, and Cr) were retrieved from the PubChem database (PubChem n.d.), a chemical repository (Burley et al. 2023). The canonical SMILES for each ligand were retrieved and converted into *pdb.file* format through the *rdkit* library in Python. RDkit library is extensively utilized for molecular representation, substructure searching, and property prediction (Kunnakkattu et al. 2023). The virtual screening of the shortlisted proteins against heavy metals was performed using the GNINA tool, which utilizes an ensemble of convolutional neural networks (CNNs) as a scoring function (McNutt et al. 2021). Grid box coordinates were specified to guide the GINNA tool during molecular docking, facilitating the interaction analysis between proteins and heavy metal compounds. Moreover, the top complexes for each ligand were characterized by the best binding affinity scores. Additionally, PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) facilitated the visualization of the docked complexes (Yuan et al. 2017). The top 3 proteins were shortlisted for each ligand.

RESULTS AND DISCUSSION

Literature Review

An extensive literature review was performed to identify the genes involved in the metal uptake in different species. A total of 129 genes were identified that are involved in the metal uptake in 19 different species and two genera (*Arabidopsis* and *Pyrus*). Furthermore, the literature review revealed different species involved in the uptake of different metals, such as Cu, Fe, Mn, Zn, Cd, As, Ni, Co, Pb, Mo, Ca, Fe^{+2}, Fe^{+3}, Mn^{+2}, Zn^{+2}, AI, Hg, and Cr. The metallothionein genes and their metal uptake in their respective species are mentioned in Table 1.

Table 1: The metallothionein genes and their metal uptake in their respective species.

Genes	Organism/Genus	Metals
PgIREG1	*Psychotria gabriellae*	Ni
AtNRAMP1, MATE, FRD3	*Arabidopsis*	Mn, Fe, AI
AhZIP6, HMA4, NAS2	*Arabidopsis halleri*	Cd, Zn
COPT1, COPT5, AtMTP8, AtZIP1, AtMTP1, AtABCC3, AtVIT1, AtZIP1, AtCAX3, TpNRAMP5, IRT1, NRAMP1, PvACR3, AtZIP2, COPT1, CMT1, AtPIC1, HMA1, PAA1, YSL4, YSL6, PAA2, HMA8, PAM71, FRO3, MOT1, AtATM3, AtNRAMP6, RAN1, AtIRT2, AtECA3, MTP11, PAM71, ZTP29, AtIRT1, HMA2, HMA3, HMA4, AtHMA2, AtHMA4, AtABCC1, AtABBC2, AtPCR2, AtYSL1, AtYSL3, HMA4, NAS2	*Arabidopsis thaliana*	Mn, Zn, Cd, Fe, Co, As, Cu, Mo, Pb, Ca, Hg
BcIRT1, BcABCC1, BcABCC2	*Brassica chinensis*	Cd
NRAMP, BnMTP3	*Brassica napus*	Zn, Mn, Cd, Pb
CsMTP8.2	*Camellia sinensis*	Fe, Mn, Zn
bHLH, C2H2, ERF, bZIP, GRAS, MYB	*Cunninghamia lanceolata*	Al
ALS1, ART1, STOP, STAR1, STAR2, ALS3, FtFRDL1, FtFRDL2, MTPC2, PCSL, PCS	*Fagopyrum tataricum*	Al
PbMTP8.1	*Pyrus*	Mn
GmHMA8, GmHMA3, GmZIP1	*Glycine max*	Cd, Cu
HvHMA1, HvIRT1, HvZIP3, HvZIP5, HvZIP5	*Hordeum vulgare*	Zn, Cu, Fe^{+2}, Fe^{+3}, Mn^{+2}, Zn^{+2}

SOD, GST, IRT, CDFs, WRKY, ERFs, POD	*Medicago sativa*	Pb
MtVTL4, MtVTL8, TaVTL2, TaVTL3	*Medicago truncatula*	Fe, Co
NtMTP2, MTP1, AhNRAMP1, NtPIC1	*Nicotiana tabacum*	Co, Ni, Mn, Zn, Fe
IREG2	*Noccaea japonica*	Ni
OsHMA3, OsMTP1, OsABCC1, OsLCT1, OsHMA2, OsZIP3, OsZIP1, OsHMA5, OsNRAMP1, ARG1, MIT, OsIRT1, OsZIP4, OsNRAMP5, MAPK, YSL	*Oryza sativa*	Cd, Zn, As, Cu, Co, Ni, Fe, Mn, Cr, Fe^{+2}, Fe^{+3}, Mn^{+2}, Zn^{+2}
SaHMA3	*Sedum alfredii Hance*	Cd
SpHMA3	*Sedum plumbizincicola*	Cd
TaCT1	*Triticum aestivum*	Cu
ZmFDR4, ZmmCHAA1, ZmZLP1, YSL, ZmNAS4, ZmNAS2, ZmNAS9	*Zea mays*	Fe, Mn, Zn, Pb

Genomic Annotation Data for *Echium plantagineum*, Sequence Retrieval, and Sequence Alignment Through BLAST

The GFF3 files of *Echium plantagineum* were not available on the NCBI, Ensembl Plants, and Phytozome databases as the genomic annotation of this plant has not been done previously; therefore, genome sequences file of *Echium plantagineum* in FASTA format was retrieved from the NCBI and used in the AUGUSTUS tool for gene prediction of *Echium plantagineum*. AUGUSTUS predicted a total of 39,520 protein sequences given the *Echium plantagineum* genome FASTA file, while *Arabidopsis* was used as a model organism for the prediction.

Furthermore, the protein sequences of the identified genes from the literature review were retrieved using the UniProt database. A total of 70 sequences were retrieved out of 129 genes in different species from the UniProt database due to the unavailability of specific protein sequences. The literature-retrieved protein sequences were searched against the BLAST database for *Echium* *plantagineum*. The database search through BLAST resulted in 41 hits. The identifiers and gene names of aligned sequences are mentioned in Table 2.

Table 2: The genes of different species aligned with the gene sequences of the Echium database.

Species	Genes	Identifiers (*Echium plantagineum*)
Arabidopsis thaliana	*YSL6*	g10948
Arabidopsis thaliana	*YSL1*	g11126
Oryza sativa	*IRT1*	g11754
Arabidopsis thaliana	*CAX3*	g12062
Arabidopsis thaliana	ABCB25	g13956
Arabidopsis thaliana	HMA1	g14577
Brassica napus	RING-type E3 ubiquitin transferase	g14584
Cunninghamia lanceolata	*MYB1*	g14711
Oryza sativa	*HMA5*	g15210
Arabidopsis thaliana	*COPT1*	g18528
Arabidopsis thaliana	*ABCC1*	g18885
Arabidopsis thaliana	*MTP11*	g19065
Oryza sativa	*ZIP1*	g20449
Arabidopsis thaliana	*CMT1*	g21210
Arabidopsis thaliana	*ZTP29*	g21375
Oryza sativa	*OsNramp5*	g21607
Arabidopsis thaliana	*DTX43*	g22946
Arabidopsis thaliana	*ABCC3*	g24041
Oryza sativa	*ZIP3*	g24118
Arabidopsis thaliana	*MTPC3*	g25342
Arabidopsis thaliana	*TIC21*	g27043

Oryza sativa	*MTP1*	g28609
Arabidopsis thaliana	*NRAMP1*	g31075
Arabidopsis thaliana	*HMA2*	g33157
Arabidopsis thaliana	*MTP1*	g33971
Arabidopsis thaliana	*DTX42*	g34339
Oryza sativa	*ARG1*	g346
Arabidopsis thaliana	*NAS2*	g35440
Triticum aestivum	*CT1-5B*	g3625
Arabidopsis thaliana	*PAA2*	g36529
Arabidopsis thaliana	*ECA3*	g38544
Arabidopsis thaliana	*PAM71*	g38741
Arabidopsis thaliana	*COPT5*	g4534
Arabidopsis thaliana	*MOT1*	g4715
Medicago sativa	*WRKY33*	g4892
Oryza sativa	*ZIP4*	g4949
Arabidopsis thaliana	*PCR2*	g539
Arabidopsis thaliana	*VIT1*	g560
Arabidopsis thaliana	*FRO3*	g5731
Arabidopsis thaliana	*PAA1*	g7401
Arabidopsis thaliana	*RAN1*	g78

3D Structure Prediction of Homologous Sequences

OmegaFold and SWISS-MODEL were used for the 3D protein structure prediction of the 41 aligned sequences, while 24 structures were predicted using OmegaFold and 17 with SWISS-MODEL. The proteins such as g19065, g24041, g33157, g4949, and g27043, which were predicted using SWISS-MODEL, did not show full coverage with the templates, and their partial structures were retrieved after modeling. The protein structures predicted through SWISS-MODEL and OmegaFold are mentioned in Table 3.

Table 3: The protein sequences predicted using SWISS-MODEL, showing sequence coverage and sequence identity.

Proteins	Sequence Length	Coverage	Sequence Identity
g10948	646	4-646	82.06%
g11126	664	3-664	75.64%
g15210	989	1-989	75.94%
g19065	1239	839-1238	86.62%
g24041	1907	13-1329	70.99%
g33157	1078	7-691	70.45%
g36529	928	17-927	72.94%
g38544	959	1-959	84.15%
g4949	1146	130-666	11.38%
g7401	958	1-957	71.31%
g18885	1640	1-1639	73.70%
g13956	827	3-827	75.20%
g14584	839	1-839	70.11%
g21210	878	91-868	53.77%
g14577	762	1-760	72.16%
g27043	743	4-555	78.68%

g5731	647	1-647	70.70%

Functional Annotation of the Homologous Sequences

The 41 predicted gene structures of *Echium plantagineum* underwent analysis using web-based servers such as BLAST and InterPro to identify their respective gene or protein domain/family names. Among these, 21 gene names were successfully identified through BLAST searches, whereas sequences that did not yield matches via BLAST, an InterPro sequence search was conducted, leading to the identification of 20 protein family/domain names for these sequences. The gene and protein family/domain names are mentioned in Table 4.

Table 4: The gene and protein family/domain names of the predicted structures.

BLAST Search	
Predicted Genes Identifiers	Identified Gene Names
g10948	HvYS1
g11126	HvYS1
g13956	ABCB25
g14577	PAA1
g14584	UBC core domain-containing protein
g14711	WER
g15210	AHA2
g18885	ABCB25
g21210	MET2A
g21607	101781512
g24041	ABCB19
g31075	101781512
g33157	HMA4
g346	ARGAH1

g3625	STP10
g36529	PAA2
g38544	AHA2
g38741	DIR6
g4892	WRKY4
g7401	PAA1
g78	RABF2B
InterPro Sequence Search	
Predicted Genes Identifiers	**Identified Protein Family/Domains**
g33971	Cation efflux protein
g35440	Nicotianamine synthase
g12062	Calcium/proton exchanger
g27043	Na(+)/H(+) antiporter NhaD-like
g539	PLAC8 motif-containing protein
g18528	Ctr copper transporter
g4949	Meiosis specific protein Spo22/ZIP4/TEX11
g560	Ccc1 family
g21375	Zinc/iron permease
g28609	Cation efflux protein
g34339	Multi-antimicrobial extrusion protein
g4715	Molybdate transporter 1/2
g4534	Ctr copper transporter
g25342	Cation efflux protein
g20449	Zinc/iron permease
g19065	Cation efflux protein

g24118	Zinc/iron permease
g5731	Cytochrome b245, heavy chain
g11754	Zinc/iron permease
g22946	Multi-antimicrobial extrusion protein

Retrieval of the Ligands and Virtual Screening

The ligands (metals) identified through the literature review were retrieved using the PubChem database. A total of 18 ligands, including Cu, Fe, Mn, Zn, Cd, As, Ni, Co, Pb, Mo, Ca, Fe^{+2}, Fe^{+3}, Mn^{+2}, Zn^{+2}, Al, Hg, and Cr, were retrieved in SMILES format and converted to pdb format. The ligands, their PubChem IDs, and SMILES are mentioned in Table 5.

The virtual screening of the 41 predicted protein structures with 18 ligands was performed using the molecular docking tool Gnina. It resulted in 738 dockings with various binding affinities ranging from -1.05 kcal/mol to -1.91 kcal/mol. The top 5 best binding affinities were shown by the same protein, HvYS1, with Fe, Hg, Mo, As, and Mn, all showing the binding affinities of -1.91 kcal/mol. Notably, the Ctr copper transporter (g4534), Zinc/iron permease (g21375), and Nicotianamine synthase (g35440) proteins were shortlisted based on their specific interactions with the ligands Ni, Zn, and Zn^{+2}, with affinity scores of -1.7, -1.68, and -1.67, respectively. However, the remaining proteins did not exhibit any interaction with any of the nine poses of ligands. The interacting proteins, their corresponding ligands, and the affinity scores are listed in Table 6. Additionally, the zoomed view of the interacting proteins screened against the Ni, Zn, and Zn^{+2} ligands is illustrated in Fig. 1.

Protein-ligand interaction analysis was performed to gain insights into the interacting residues that contribute to the binding of shortlisted interacting proteins with Ni, Zn, and Zn^{+2}. The nicotianamine synthase protein exhibited binding interactions with Ni, Zn, and Zn^{+2} at the same residues: 181A (ALA), 179A (HIS), and 153A (ASP). Similarly, the zinc/iron permease exhibited binding at residues 52A (ALA), 53A (HIS), 47A (SER), and 50A (ASP) with all three ligands. Lastly, the Ctr copper transporter protein interacted with Ni at residues 20A (THR), 11A (VAL), and 9A (ARG). Moreover, the Ctr copper transporter protein also displayed interactions with Zn and Zn^{+2} at residues 100A (ALA), 101A (VAL), 105A (ASN), 104A (TYR), and 103A (SER). The protein-ligand interactions are illustrated in Fig. 2.

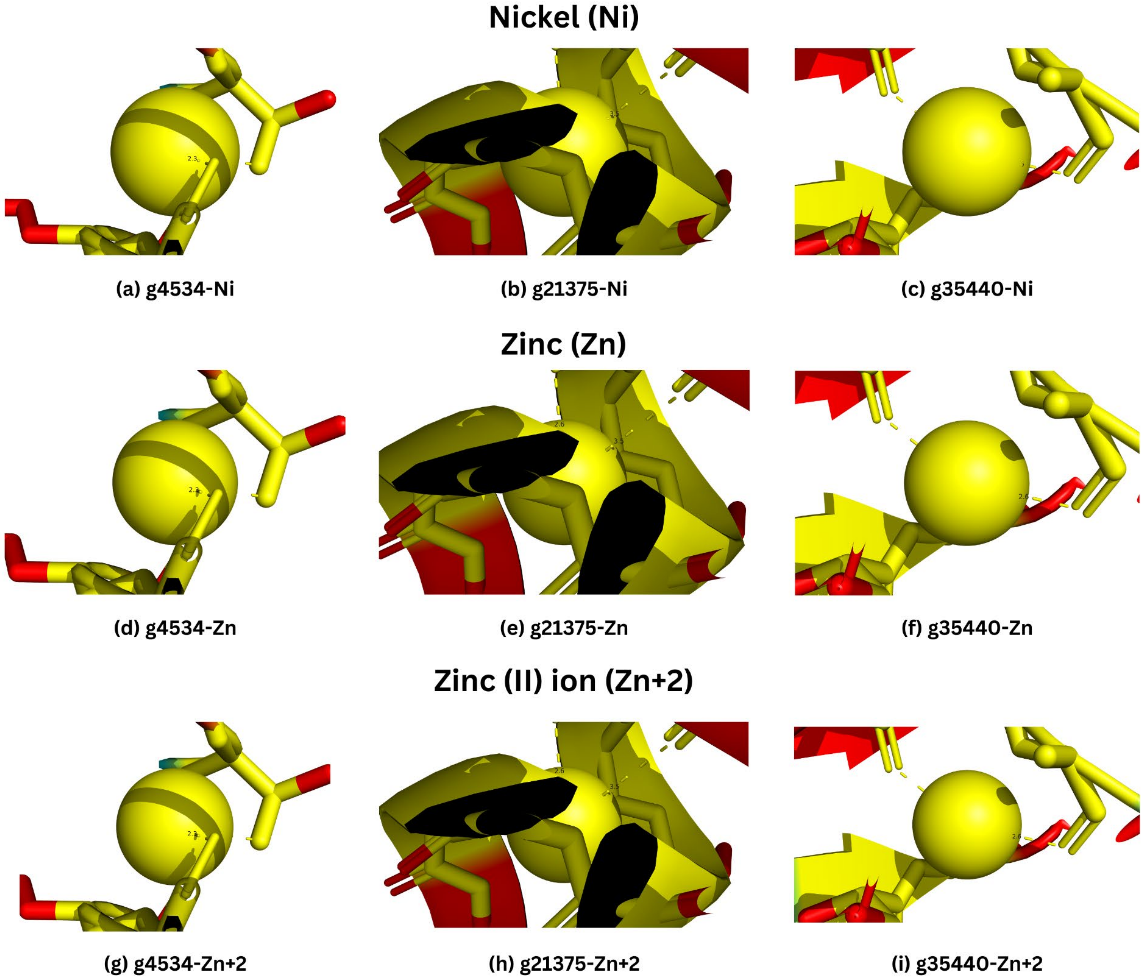
Table 5: The ligands, PubChem IDs and SMILES.

Ligands	PubChem IDs	SMILES
Cu	23978	[Cu]
Fe	23925	[Fe]
Mn	23930	[Mn]
Zn	23994	[Zn]
Cd	23973	[Cd]
As	5359596	[As]
Ni	935	[Ni]
Co	104730	[Co]
Pb	5352425	[Pb]
Mo	23932	[Mo]
Ca	5460341	[Ca]
Fe^{+2}	27284	[Fe+2]
Fe^{+3}	29936	[Fe+3]
Mn^{+2}	27854	[Mn+2]
Zn^{+2}	32051	[Zn+2]
AI	5359268	[Al]
Hg	23931	[Hg]
Cr	23976	[Cr]

Table 6: Binding affinities of interacting proteins against Ni, Zn, and $Zn^{+2.}$

Ligand	Protein	Affinity
Ni	Ctr copper transporter (g4534)	-1.7
	Zinc/iron permease (g21375)	-1.68
	Nicotianamine synthase (g35440)	-1.67
Zn	Ctr copper transporter	-1.7
	Zinc/iron permease	-1.68
	Nicotianamine synthase	-1.67
Zn+2	Ctr copper transporter	-1.7

	Zinc/iron permease	-1.68
	Nicotianamine synthase	-1.67



(a) g4534-Ni (b) g21375-Ni (c) g35440-Ni

(d) g4534-Zn (e) g21375-Zn (f) g35440-Zn

(g) g4534-Zn+2 (h) g21375-Zn+2 (i) g35440-Zn+2

Fig. 1: The zoomed view illustration of interacting protein structures screened against Ni, Zn, and $Zn^{+2.}$

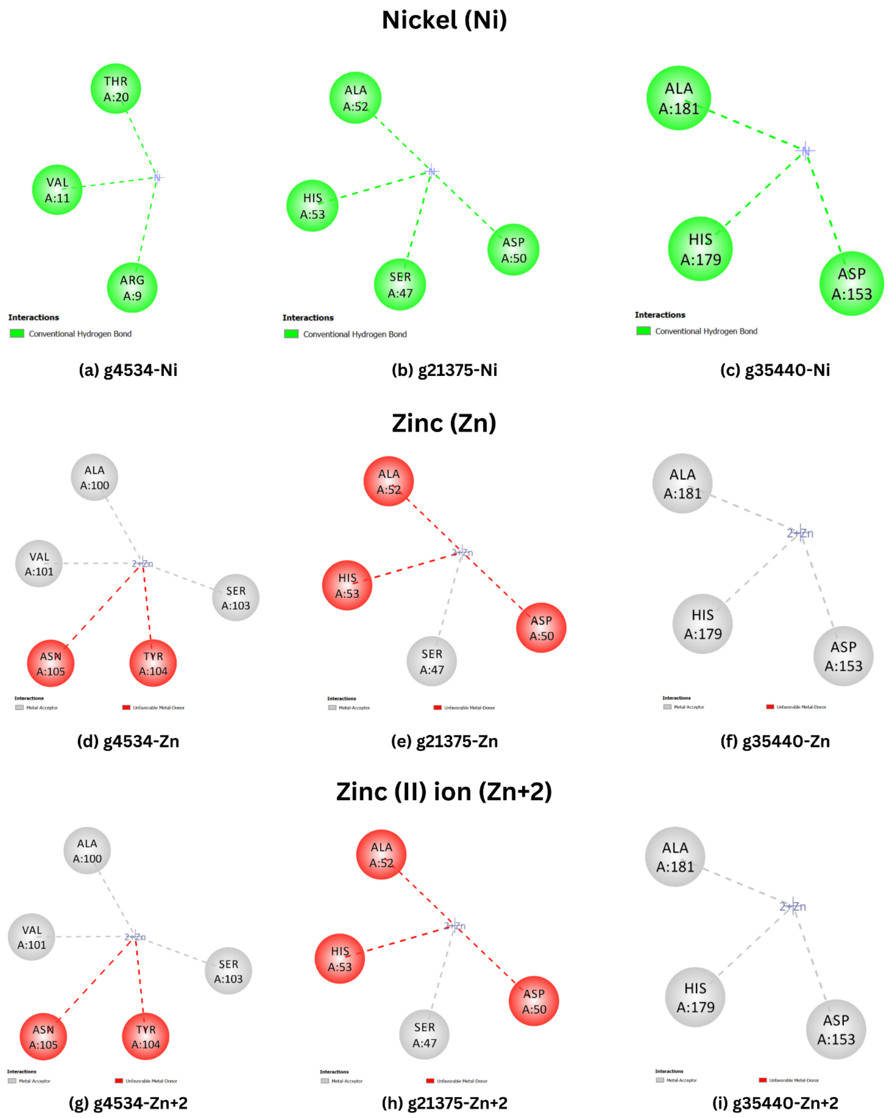


Fig. 2: 2D illustration of interactions of shortlisted interacting proteins with Ni, Zn, and Zn^{+2}.

Protein Domain Analysis and Functional Annotation

The shortlisted proteins were then functionally annotated through InterPro. It was observed that the g4534 protein has a Ctr copper transporter domain from 1-131 residues, whereas the g21375 protein has a Zinc/iron permease domain spanning 11-253. Contrary to this, it was observed that in the g35440 protein, almost the entirety of the protein length is covered by the Nicotianamine synthase domain. The domains are illustrated in Fig. 3. Furthermore, the functional annotation through Gene Ontology revealed key functional aspects of these proteins, such as the metal ion transmembrane transporter activity of g21375 protein, indicating a metal ion transport functionality of the protein, Fig. 3. This means that the protein is capable of transporting metal ions across cell membranes. In plants, metal ion transporters play crucial roles in regulating the uptake, distribution, and storage of essential metal ions such as iron, zinc, manganese, and copper. These transporters are essential for maintaining metal ion homeostasis and ensuring that plants supply these nutrients adequately without experiencing toxicity from excess metal ions.

Moreover, the g35440 protein has nicotianamine synthase activity. Nictioanamine is involved in plant metal detoxification processes, particularly in the chelation (binding) of heavy metals such as nickel, cadmium, and zinc. Plants can take up heavy metals from the soil, either through their roots or other mechanisms, and high concentrations of these metals can be toxic to the plant. Nictioanamine synthase activity allows the plant to produce nictioanamine, which can bind to heavy metals and form stable complexes, reducing their toxicity and facilitating their sequestration or removal from the plant's tissues. Contrary to this, the g4534 protein has a Ctr copper transmembrane transporter activity. This protein is capable of transporting copper ions across cell membranes. In plants, copper is an essential micronutrient, but excessive copper can be toxic. Copper transporters, like the one encoded by the g4534 gene, play a crucial role in maintaining copper homeostasis by regulating the uptake, distribution, and storage of copper ions within the plant.

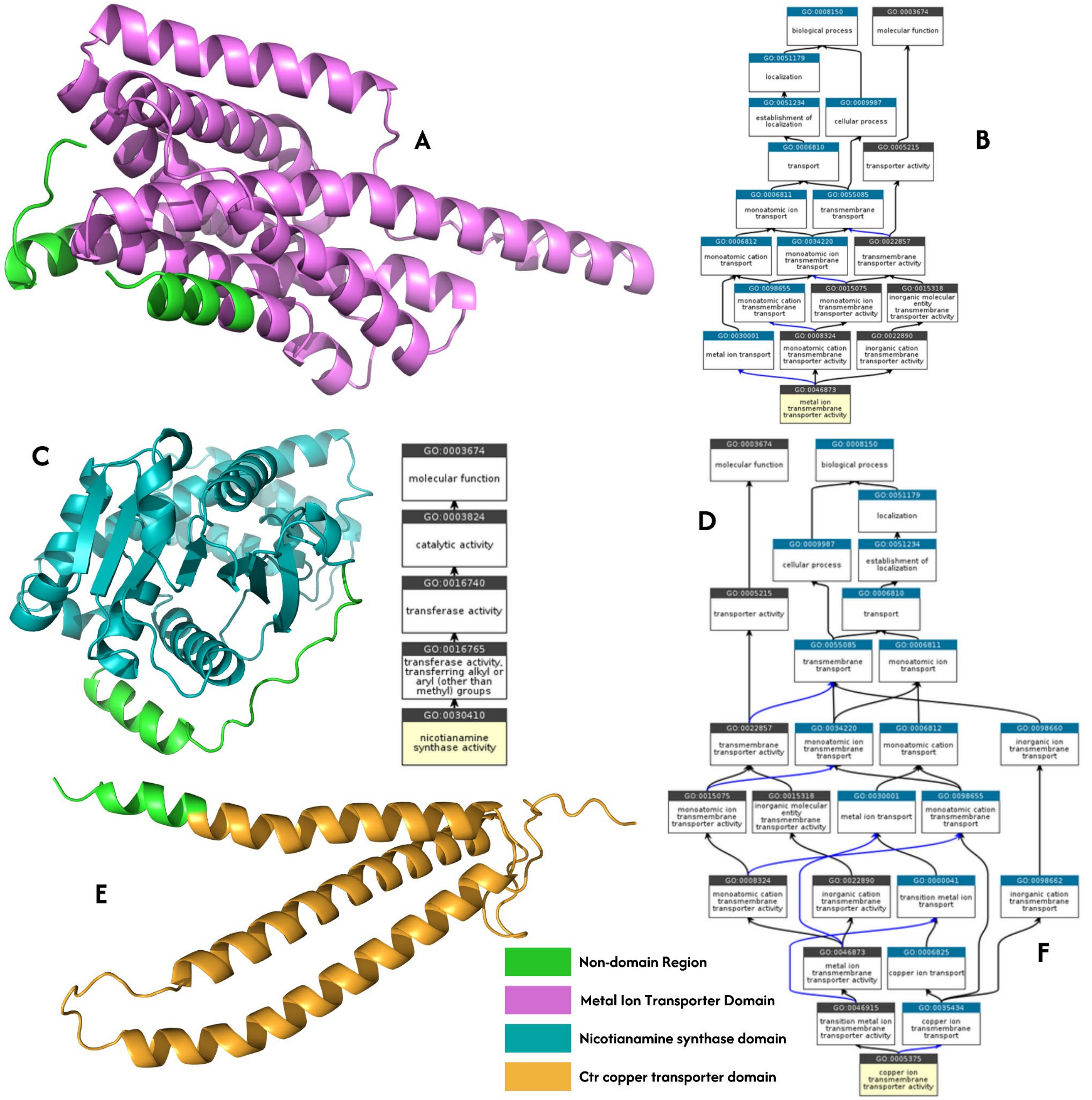


Fig. 3: Illustration of the identified domains with potential phytoremediation role and GO functions. (A) g21375 protein with its metal ion transporter domain (B) GO biological process and molecular functions of g21375 protein (C) g35440 protein with its nicotianamine synthase domain (D) GO biological process and molecular functions of g35440 protein (E) g4534 protein with its Ctr copper ion domain (F) GO biological process and molecular functions of g4534 protein.

Discussion

Soil heavy metal contamination has emerged as a global environmental challenge, drawing significant public attention due to worries about the safety of agricultural products. Heavy metals encompass metallic elements and metalloids known for their potential biological harm, including cadmium, mercury, arsenic, lead, and chromium (Mitra et al. 2022). Understanding the metabolic responses and adaptations of plants to toxic metal exposure is critical for advancing phytoremediation efforts in contaminated areas (Alsafran et al. 2022).

Phytoremediation has emerged as a promising alternative to traditional methods due to its cost-effectiveness, environmental friendliness, and aesthetic appeal. Thus far, approximately 500 plant species have been identified as hyperaccumulators of one or more metals, harnessing their natural extraction abilities (Li et al. 2019). However, further exploration is necessary, integrating biotechnological strategies with comprehensive multidisciplinary research to enhance plant tolerance and reduce the accumulation of toxic metals in soil.

Therefore, a comprehensive literature review was conducted to identify metallothioneins across diverse plant species (*Psychotria gabriellae, Arabidopsis halleri, Brassica chinensis, Camellia sinensis, Cunninghamia lanceolata, Fagopyrum tataricum, Pyrus, Hordeum vulgare, Medicago truncatula, Nicotiana tabacum, Noccaea japonica, Sedum alfredii, and Sedum plumbizincicola*) involved in the reduction of heavy metals, such as Ni, Mn/Mn^{+2}, Fe/Fe^{+2}/Fe^{+3}, Al, Cd, Zn/Zn^{+2}, Co, As, Cu, Mo, Pb, Ca, Hg, Cr in soil (Angulo-Bejarano et al. 2021, Jan & Parray 2016, Jogawat et al. 2021).

Notably, *Echium plantagineum*, a metal-tolerant species exhibiting a notably high Bioconcentration Factor (BCF) and low Translocation Factor (TF), holds the potential for phytostabilization purposes (El Berkaoui et al. 2021). Despite the promising potential of *Echium plantagineum* for phytostabilization, no previous study has investigated the presence or role of metallothioneins in this species, which could potentially be involved in the uptake of heavy metals from the soil.

However, the *Echium plantagineum* does not have enough proteome and genome resources available on public databases, a major limitation in identifying metallothioneins or metal uptake-related proteins from this plant. Consequently, this study highlights the novelty and significance of investigating the metallothioneins of *Echium plantagineum* by predicting the genes and proteins from the entire genome, resulting in 39,520 predicted protein sequences. However, *Arabidopsis* *thaliana* was used as the reference genome due to its fully sequenced genome and its regulatory mechanisms in genomics, transcriptomics, and proteomics comparable to other plant species (Di Silvestre et al. 2018). These predicted protein sequences were systematically compiled to construct a protein database. This database was subsequently employed for BLAST search against the 70 protein sequences of 129 genes identified from the literature review across different plant species, resulting in 41 top hits.

Moreover, the 41 protein structures of these top hits were subsequently predicted and subjected to screening against 15 ligands, identifying only Ctr copper transporter, zinc/iron permease, and nicotianamine synthase proteins exhibiting high affinity and interactions with Ni, Zn, and Zn+2 ligands. The copper homeostasis in plants and regulating processes (absorption, transportation, and growth) are maintained primarily by the copper transporter family (COPT/Ctr) (Nayeri et al. 2023, Wang et al. 2021). However, the functional characterization of these COPT genes encoding copper transporter proteins has been limited to *Arabidopsis* and *Oryza sativa* (Wang et al. 2018). Hence, this study reveals that the COPT/Ctr copper transporter protein has a novel high affinity with Ni, Zn, and $Zn^{+2 \text{ ligands}}$, suggesting potential roles beyond its traditional copper transport function.

Similarly, the zinc/iron permease, ZIP family, is crucial for metal ion transport and primarily facilitates the influx of transition metal ions such as zinc and iron. In addition to Zn^{+2}, ZIP transporters are known to facilitate the transportation of various other transition metal cations, including Mn^{+2}, Fe^{+2}, Cd^{+2}, and Ni^{+2} (Ajeesh Krishna et al. 2020, Zhao & Zhou 2020). *Thlaspi caerulescens* and *Arabidopsis halleri*, known as Zn hyperaccumulators, exhibit increased Zn uptake compared to non-hyperaccumulator species, correlating with elevated expression of certain ZIP family members in hyperaccumulators (Yan et al. 2020). The unreported literature on the role of zinc/iron permease in metal uptake, particularly for *Echium plantagineum*, underscores the novelty of our study, revealing its involvement in Ni, Zn, and Zn^{+2} uptake within this species.

Furthermore, nicotianamine (NA) serves as a primary metal chelator in plants and is synthesized catalytically by nicotianamine synthase. Nicotianamine synthase genes have been identified in various plant species, including *Oryza sativa, Zea mays, Hordeum vulgare, Triticum aestivum, Arabidopsis thaliana, Arabidopsis halleri, Solanum lycopersicum, Noccaea caerulescens,* and *Malus xiaojinensis*. Studies have demonstrated that overexpression of nicotianamine synthase in transgenic plants increases NA concentration, consequently elevating metal concentration. Specifically, the nicotianamine synthase gene from *N. caerulescens* enhances nickel accumulation in transgenic *Arabidopsis thaliana*, indicating its potential for improving divalent metal ion content and uptake, particularly in contaminated soil (Chen et al. 2019). Thus, the nicotianamine synthase can reduce the heavy metal toxicity from the contaminated soils (Nozoye 2018). Additionally, this study reveals that the nicotianamine synthase protein has a novel high affinity with ligands Ni, Zn, and Zn^{+2}, highlighting potential roles in phytoremediation.

Conclusively, this study identified the role of *Echium plantagineum* in heavy metal stress in contaminated soils. Metallothioneins essential for heavy metal reduction were identified through a comprehensive literature review and BLAST search. Key proteins such as Ctr copper transporter, zinc/iron permease, and nicotianamine synthase were found to have novel high affinities with Ni, Zn, and Zn+2 metal ions, indicating their role in metal uptake and transport. This study underscores the functional annotation of metallothioneins of *Echium plantagineum* for reduction in heavy metals in soil, paving the way for future research and environmental applications.

CONCLUSIONS

In conclusion, the intricate interplay between *Echium plantagineum* and the heavy metal stress in contaminated soils was identified in this study. Metallothioneins essential for heavy metal reduction in various plant species were identified through a literature review, and BLAST was searched against plant gene sequences, identifying 41 top hits. Notably, screening these hits against ligands revealed novel high affinities of key proteins such as Ctr copper transporter, zinc/iron permease, and nicotianamine synthase with Ni, Zn, and Zn^{+2}, implying their role in metal uptake and transport. Hence, this study identified for the first time the functional annotation of the metallothioneins in the *Echium plantagineum* involved in reducing heavy metals from the soil, paving the way for future research and environmental applications.

REFERENCES

Ahn, Y.O., Kim, S.H., Lee, J., Kim, H., Lee, H.-S. and Kwak, S.S., 2012. Three Brassica rapa metallothionein genes are differentially regulated under various stress conditions. *Molecular Biology Reports*, 39(3), pp.2059–2067. doi: 10.1007/s11033-011-0953-5.

Ajeesh Krishna, T.P.A., Maharajan, T., Roch, G.V., Ignacimuthu, S. and Ceasar, S.A., 2020. Structure, function, regulation, and phylogenetic relationship of ZIP family transporters of plants. *Frontiers in Plant Science*, 11, p.662. doi: 10.3389/fpls.2020.00662.

Al-Khashman, O.A., 2004. Heavy metal distribution in dust, street dust and soils from the workplace in Karak Industrial Estate, Jordan. *Atmospheric Environment*, 38(39), pp.6803–6812. doi: 10.1016/j.atmosenv.2004.09.011.

Alsafran, M., Usman, K., Ahmed, B., Rizwan, M., Saleem, M.H. and Al Jabri, H., 2022. Understanding the phytoremediation mechanisms of potentially toxic elements: A proteomic overview of recent advances. *Frontiers in Plant Science*, 13, p.881242. doi: 10.3389/fpls.2022.881242.

Angulo-Bejarano, P.I., Puente-Rivera, J. and Cruz-Ortega, R., 2021. Metal and metalloid toxicity in plants: An overview on molecular aspects. *Plants*, 10(4), p.635. doi: 10.3390/plants10040635.

Anon, 2021. Extracting data from BLAST databases with Blastdbcmd. in *BLAST® Command Line Applications User Manual* [online]. National Center for Biotechnology Information (US). Available at https://www.ncbi.nlm.nih.gov/sites/books/NBK569853/

Anon, n.d.-a. National Center for Biotechnology Information. Retrieved June 25, 2024. from: https://www.ncbi.nlm.nih.gov/.

Anon, n.d.-b. OmegaFold/README.md at main · HeliXonProtein/OmegaFold. GitHub. Retrieved June 25, 2024, from https://github.com/HeliXonProtein/OmegaFold/blob/main/README.md.

Anon, n.d.-c. UniProt. Retrieved June 25, 2024, from https://www.uniprot.org/.

Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chao, H., Chen, L., Craig, P.A., Crichlow, G.V., Dalenberg, K., Duarte, J.M., Dutta, S., Fayazi, M., Feng, Z., Flatt, J.W., Ganesan, S., Ghosh, S., Goodsell, D.S., Green, R.K., Guranovic, V., Henry, J., Hudson, B.P., Khokhriakov, I., Lawson, C.L., Liang, Y., Lowe, R., Peisach, E., Persikova, I., Piehl, D.W., Rose, Y., Sali, A., Segura, J., Sekharan, M., Shao, C., Vallat, B., Voigt, M., Webb, B., Westbrook, J.D., Whetstone, S., Young, J.Y., Zalevsky, A. and Zardecki, C., 2023. RCSB Protein Data Bank (RCSB.org): Delivery of experimentally determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Research*, 51(D1), pp.D488–D508. doi: 10.1093/nar/gkac1077.

Chatterjee, S., Kumari, S., Rath, S., Priyadarshanee, M. and Das, S., 2020. Diversity, structure, and regulation of microbial metallothionein: Metal resistance and possible applications in sequestration of toxic metals. *Metallomics*, 12(11), pp.1637–1655. Doi: 10.1039/d0mt00140f.

Chen, S., Zhang, M., Feng, Y., Sahito, Z.A., Tian, S. and Yang, X., 2019. Nicotianamine synthase gene 1 from the hyperaccumulator *Sedum alfredii* hance is associated with Cd/Zn Tolerance and Accumulation in Plants. *Plant and Soil*, 443(1), pp.413–427. doi: 10.1007/s11104-019-04233-4.

Choudhuri, S., 2014. *Bioinformatics for Beginners.* Academic Press, pp. 133–155.

Di Silvestre, D., Bergamaschi, A., Bellini, E. and Mauri, P.L., 2018. Large-scale proteomic data and network-based systems biology approaches to explore the plant world. *Proteomes*, 6(2), p.27. doi: 10.3390/proteomes6020027.

El Berkaoui, M., El Adnani, M., Hakkou, R., Ouhammou, A., Bendaou, N. and Smouni, A., 2021. Phytostabilization of phosphate mine wastes used as a store-and-release cover to control acid mine drainage in a semiarid climate. *Plants*, 10(5), p.900. doi: 10.3390/plants10050900.

Freisinger, E., 2009. Metallothioneins in Plants. In *Metallothioneins and Related Chelators* (pp. 107-153). RCB Publishing

Guvenç, N., Alagha, O. and Tuncel, G., 2003. Investigation of soil multi-element composition in Antalya, Turkey. *Environment International*, 29(5), pp.631–640. doi: 10.1016/S0160-4120(03)00046-1.

Jan, S. and Parray, J.A., 2016. *Heavy Metal Uptake in Plants*. Springer, pp.1-18.

Jogawat, A., Yadav, B., Chhaya, and Narayan, O.P., 2021. Metal Transporters in Organelles and Their Roles in Heavy Metal Transportation and Sequestration Mechanisms in Plants. *Physiologia Plantarum*, 173(1), pp.259–275. doi: 10.1111/ppl.13370.

Joshi, R., Pareek, A. and Singla-Pareek, S.L., 2016. *Plant Metal Interaction.* Elsevier, pp. 239–261.

Keller, O., Kollmar, M., Stanke, M. and Waack, S., 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, 27(6), pp.757–763. doi: 10.1093/bioinformatics/btr010.

Kunnakkattu, I.R., Choudhary, P., Pravda, L., Nadzirin, N., Smart, O.S., Yuan, Q., Anyango, S., Nair, S., Varadi, M. and Velankar, S., 2023. PDBe CCDUtils: An RDKit-based toolkit for handling and analyzing small molecules in the protein data bank. *Journal of Cheminformatics*, 15(1), p.117. doi: 10.1186/s13321-023-00786-w.

Latini, A., Papagni, I., Gatti, L., De Rossi, P., Campiotti, A., Giagnacovo, G., Mirabile Gattia, D. and Mariani, S., 2022. Echium vulgare and Echium plantagineum: A comparative study to evaluate their inclusion in Mediterranean urban green roofs. *Sustainability*, 14(15), p.9581. doi: 10.3390/su14159581.

Li, C., Zhou, K., Qin, W., Tian, C., Qi, M., Yan, X. and Han, W., 2019. A review on heavy metals contamination in soil: Effects, sources, and remediation techniques. *Soil and Sediment Contamination: An International Journal*, 28(4), pp.380–394. doi: 10.1080/15320383.2019.1592108.

McNutt, A.T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., Sunseri, J. and Koes, D.R., 2021. GNINA 1.0: Molecular docking with deep learning. *Journal of Cheminformatics*, 13(1), p.43. doi: 10.1186/s13321-021-00522-2.

Midhat, L., Ouazzani, N., Hejjaj, A., Ouhammou, A. and Mandi, L., 2019. Accumulation of heavy metals in metallophytes from three mining sites (Southern Centre Morocco) and evaluation of their phytoremediation potential. *Ecotoxicology and Environmental Safety*, 169, pp.150–160. doi: 10.1016/j.ecoenv.2018.11.009.

Mitra, S., Chakraborty, A.J., Tareq, A.M., Bin Emran, T., Nainu, F., Khusro, A., Idris, A.M., Khandaker, M.U., Osman, H., Alhumaydhi, F.A. and Simal-Gandara, J., 2022. Impact of heavy metals on the environment and human health: Novel therapeutic insights to counter the toxicity. *Journal of King Saud University - Science*, 34(3), p.101865. doi: 10.1016/j.jksus.2022.101865.

Nayeri, N., Li, P., Górecki, K., Lindkvist-Petersson, K. and Gourdon, P., 2023. Principles to recover copper-conducting CTR proteins for the purpose of structural and functional studies. *Protein Expression and Purification*, 203, p.106213. doi: 10.1016/j.pep.2022.106213.

Nozoye, T., 2018. The nicotianamine synthase gene is a useful candidate for improving the nutritional qualities and Fe-deficiency tolerance of various crops. *Frontiers in Plant Science*, 9. doi: 10.3389/fpls.2018.00340.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Lázaro Pinto, B., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D.H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H. and Bateman, A., 2023. InterPro in 2022. *Nucleic Acids Research*, 51(D1), pp.D418–D427. doi: 10.1093/nar/gkac993.

PubChem, n.d. PubChem. Retrieved June 25, 2024 from https://pubchem.ncbi.nlm.nih.gov/).

Rashid, A., Schutte, B.J., Ulery, A., Deyholos, M.K., Sanogo, S., Lehnhoff, E.A. and Beck, L., 2023. Heavy metal contamination in agricultural soil: Environmental pollutants affecting crop health. *Agronomy*, 13(6), p.1521. doi: 10.3390/agronomy13061521.

Santana, C.S., Montalván Olivares, D.M., Silva, V.H.C., Luzardo, F.H.M., Velasco, F.G. and de Jesus, R.M., 2020. Assessment of water resources pollution associated with mining activity in a semi-arid region. *Journal of Environmental Management*, 273, p.111148. doi: 10.1016/j.jenvman.2020.111148.

Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., Wang, J., Williams, R., Trawick, B.W., Pruitt, K.D. and Sherry, S.T., 2021. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 50(D1), pp.D20–D26. doi: 10.1093/nar/gkab1112.

Sharma, R., Bhardwaj, R., Handa, N., Gautam, V., Kohli, S.K., Bali, S., Kaur, P., Thukral, A.K., Arora, S., Ohri, P. and Vig, A.P., 2016. *Plant Metal Interaction*. Elsevier, pp. 263–283.

Shinzato, M.C. and Hypolito, R., 2016. Effect of disposal of aluminum recycling waste in soil and water bodies. *Environmental Earth Sciences*, 75(7), p.628. doi: 10.1007/s12665-016-5438-3.

Skoneczny, D., Zhu, X., Weston, P.A., Gurr, G.M., Callaway, R.M. and Weston, L.A., 2019. Production of pyrrolizidine alkaloids and shikonins in *Echium plantagineum* L. in response to various plant stressors. *Pest Management Science*, 75(9), pp.2530–2541. doi: 10.1002/ps.5540.

Stanke, M., 2003. The AUGUSTUS Gene Prediction Tool. Retrieved June 25, 2024, from https://bioinf.uni-greifswald.de/augustus/

Streets, D.G., Lu, Z., Levin, L., ter Schure, A.F.H. and Sunderland, E.M., 2018. Historical releases of mercury to air, land, and water from coal combustion. *Science of The Total Environment*, 615, pp.131–140. doi: 10.1016/j.scitotenv.2017.09.207.

Wang, H., Du, H., Li, H., Huang, Y., Ding, J., Liu, C., Wang, N., Lan, H. and Zhang, S., 2018. Identification and functional characterization of the ZmCOPT copper transporter family in maize. *PLOS ONE*, 13(7), p.e0199081. Doi: 10.1371/journal.pone.0199081.

Wang, Q., Wei, N., Jin, X., Min, X., Ma, Y. and Liu, W., 2021. Molecular characterization of the COPT/Ctr-type copper transporter family under heavy metal stress in alfalfa. *International Journal of Biological Macromolecules*, 181, pp.644–652. doi: 10.1016/j.ijbiomac.2021.03.173.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T., 2018. SWISS-MODEL: Homology modeling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), pp.W296–W303. doi: 10.1093/nar/gky427.

Wu, R., Ding, F., Wang, R., Shen, R., Zhang, X., Luo, S., Su, C., Wu, Z., Xie, Q., Berger, B., Ma, J. and Peng, J., 2022. High-resolution de novo structure prediction from primary sequence. 2022.07.21.500999.

Yaashikaa, P.R., Senthil Kumar, P., Jeevanantham, S. and Saravanan, R., 2022. A review on bioremediation approach for heavy metal detoxification and accumulation in plants. *Environmental Pollution*, 301, p.119035. doi: 10.1016/j.envpol.2022.119035.

Yan, A., Wang, Y., Tan, S.N., Ghosh, S. and Chen, Z., 2020. Phytoremediation: A promising approach for revegetation of heavy metal-polluted land. *Frontiers in Plant Science*, 11, p.359. doi: 10.3389/fpls.2020.00359.

Yuan, S., Chan, H. and Hu, Z., 2017. Using PyMOL as a platform for computational drug design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 7, p.e1298. Doi: 10.1002/wcms.1298.

Zaccone, C., Di Caterina, R., Rotunno, T. and Quinto, M., 2010. Soil–farming system – food – health: Effect of conventional and organic fertilizers on heavy metal (Cd, Cr, Cu, Ni, Pb, Zn) content in semolina samples. *Soil and Tillage Research*, 107(2), pp.97–105. doi: 10.1016/j.still.2010.02.004.

Zhao, M. and Zhou, B., 2020. A distinctive sequence motif in the fourth transmembrane domain confers ZIP13 iron function in *Drosophila melanogaster*. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1867(2), p.118607. doi: 10.1016/j.bbamcr.2019.118607.

Zine, H., Midhat, L., Hakkou, R., El Adnani, M. and Ouhammou, A., 2020. Guidelines for a phytomanagement plan by the phytostabilization of mining wastes. *Scientific African*, 10, p.e00654. doi: 10.1016/j.sciaf.2020.e00654.