**Investigation of genes involved in zinc uptake, storage, and sequestration at identified *Brassica oleracea* quantitative trait loci**

Katie Gwathmey & Chris Polo

**Abstract**

Zinc is an essential nutrient for plants and animals. In plants, zinc concentrations are tightly regulated by a variety of proteins to ensure proper growth and to avoid the buildup of toxic concentrations. Therefore, the identification of zinc regulatory genes in *Brassica oleracea* would be beneficial for general plant health and growth. In the newly assembled genome sequence of *B. oleracea*, four QTLs for zinc were identified. Here, we determined that nine potential zinc regulating genes were located within two million base pairs of three of the four QTLs. Our results suggest that the *B. oleracea* genome is reputable, and future research can verify and validate these findings.

**Introduction**

 Zinc is an essential micronutrient for the proper growth of plants (Sommer and Lipman, 1926). It serves as a cofactor for over 300 enzymes and approximately 1230 proteins interact with, bind or transport zinc within the cell (Marschner, 1995; Ricachenevsky *et al.,* 2013). Zinc is a necessary component in a diverse array of cellular processes including transcription, translation, protein degradation, and detoxification of reactive oxidative species within the plant (Marschner, 1995; Song *et al.,* 2010). Song *et al.* (year) and Grotz *et* al. (year) reported that plants need around 20 to 100 µg of zinc in tissues while less than 15 to 20 µg in leaves lead to deficiency. However, excessive concentrations of zinc are toxic to plants and can lead to impaired growth, chlorosis and interference with cellular processes due to incorrect cofactor interactions (Marschner, 1995; Ricchenevsky *et al.* 2013). Therefore, plants tightly regulate zinc homeostasis in order to keep a balanced concentration for proper growth while avoiding toxicity (Claus *et al.,* 2012). Song *et al.* (year) reviewed that plants control zinc concentrations by tightly regulating its influx and efflux from cells, intracellularly binding zinc to chaperone chelators, and storing it in vacuoles. Storage and sequestration is imperative to this process, and although zinc is utilized in many organelles, most of the articles I read cited that zinc is primarily stored in the vacuole (Eide, 2006; Ricchenevsky *et al.*, 2013).

Previous researchers have identified a number of proteins involved in zinc uptake, transport, and storage in *Arabidopsis thaliana*. Excreted chelators and rhizosphere acidification digest zinc in the soil as it can only be absorbed into the plant as the cation Zn2+ (Clemens *et al.*, 2002). Unable to simply diffuse across the membrane, members of the ZIP protein family (ZIP1 and ZIP3) and certain iron-regulated transporters (IRT1, IRT2, IRT3) transport Zn2+ from the soil into the root cell cytosol (Grotz *et al.* 1998). Certain ZIP transcription factors have also been identified and knockouts of these genes - *bZIP1, bZIP19, bZIP32 -* result in reduced plant growth and zinc uptake (Assuncao *et al.* 2010).

When inside the cell, zinc must be transported across the root cells into the xylem. As mentioned above, Zn2+ is usually bound to certain chelator when it moves through the cytosol, which regulates its molecular interactions (Clemens *et al.*, 2002; Hofman, 2012). Zinc moves through the root cells via apoplastic or symplatic pathway to the xylem in order to be transported to the rest of the plant’s cells (Claus *et al.*, 2012). In a review of zinc homeostasis, HMA2 and HMA4 were cited as important for movement from root to shoot and the general efflux of Zn2+ from cellular cytoplasm (Grotz and Guerinot, 2006). Heavy metal ATPase’s are a P1B-type ATPase, and there are eight proteins in the protein family present throughout the plant cells and organelles all of which generally move Zn2+ from the cytoplasm (Grotz and Guerinot, 2006). Further studies have shown that plant cadmium resistance 2 (PCR2) pumps Zn2+ into the xylem and out of the root cells when there is a toxic zinc concentration (Song *et al.,* 2010).

Zinc storage and sequestration is just as important as its transport as maintaining beneficial concentrations are crucial for plant health. Within the cell, zinc transporting proteins have been found on the chloroplast and vacuole (Grotz *et al.* 1998). ZIP4, which is thought to be localized to the chloroplasts, moves Zn2+ into the plastid for use. Conversely, HMA1 has been shown to move zinc out of the chloroplast when there cells experience excessive zinc concentrations (Kim *et al.*, 2009). Although zinc is transported into the chloroplast, most articles cite that the vacuole stores most of supplemental zinc. HMA3, has been implicated in transporting zinc and other heavy metal into the vacuole (Morel *et al.*, 2009). Furthermore, metal tolerance proteins (MTPs), a class of cation diffusion facilitators, are critical for zinc vacuolar storage (Ricachenevsky *et al.,* 2013*)* MTPs that are specific to zinc transport are also referred to zinc transport proteins or ZATs (Grotz and Guerinot, 2006). Ricachenevsky *et al.* reviews thatMTP1 and MTP3 has been shown to move excessive zinc into the vacuole in both *A. thaliana* and *A. halleri*, a species of *Arabidopsis* able to hyper accumulate zinc. Furthermore, in hyperaccumulators MTP1 is overexpressed highlighting its importance in zinc detoxification through vacuolar sequestration (Ricachenevsky *et al.*, 2013).

Humans also require zinc to function with zinc deficiency leading to disturbances in brain function, immune system, and physical development ([www.harvestzinc.org](http://www.harvestzinc.org)). Unfortunately, around thirty percent of agricultural soils across the globe are zinc-deficient due to changes in pH, excessive water, and over cropping (Mertens and Smolders, 1990). Meanwhile, Claus *et al.* (year) reported that industrial and mining waste has contaminated many other soils, which now contain excessive amounts of heavy metal, including zinc. Therefore, it would be beneficial for plant and human health to identify genes involved in the intake, transport, and sequestration of zinc. This would enable the development of genetically modified crops that would be capable of growing in various soil conditions around the world, providing nutritionally accessible zinc in human populations

 *Brassica oleracea*, more commonly known as broccoli, contains a large amount of important nutrients for humans (Brown, 2014). *B. oleracea* recently diverged from *A. thaliana* resulting in relevant genetic sequence overlap between the two species (Brown, 2014). Dr. Allan Brown’s lab in Kannapolis, NC recently sequenced and mapped the *B. oleracea* genome and identified QTLs for various minerals across its nine chromosomes. These QTLs were identified in samples from the flowering part of the plant. Four QTLs for zinc were determined. Here, we display that a variety of zinc-related regulatory genes are located near three of the four QTLs in the constructed *B. oleracea* genome.

**Methods**

|  |  |  |
| --- | --- | --- |
| **Chromosome** | **Nucleotide**  | **Shared with:** |
| 6 | 9,847,363 | Boron |
| 8 | 34,578,133 |  |
| 9 | 3,558,943 | Iron |
| 9 | 41,992,687 | Boron and Sodium |

 We examined previous research and identified proteins involved in zinc uptake, transport, and storage in *Arabidopsis thaliana*. We compiled a list of *A. thaliana* orthologs for proteins of interest, and used NCBI BLAST to identify the amino acid sequence of each target protein **(Appendix B)**. If proteins had multiple accession numbers listed, we either confirmed that the amino acid sequence was the same or noted where there was a difference. Our four Zinc QTLs on chromosomes 6, 8, and 9 of *B. oleracea* genome were provided by Dr. David’s lab **(Table 1)**. On blueberry and broccoli genome database, dev.vaccinium.org, we performed a tBLASTn on the compiled *B. oleracea* genome for each candidate protein. We recorded the nucleotide coverage for every hit with an E-value less than 10-6 located two million base pairs away from one of our four QTLs. We visualized our results with graphs.

**Table 1.** *B. oleracea* zinc quantitative trait loci

 We used the IGB genome browser to see if our BLAST hits were present and annotated and to identify any other proteins potentially involved in zinc regulation within one million base pairs of each of our QTLs. Using manual browsing, we examined the location of all BLAST hits within two million base pairs of a QTL. We aimed to establish if our protein of interest was there, if another protein was there, if our protein was mislabeled, or if it was simply not annotated. If missing, we highlighted and obtained the nucleotide sequence where it should have been present and blasted against *A. thaliana* genome on NCBI.

To determine if there were any other proteins of interest, we manually scrolled through the program one million base pairs in each direction from our QTL. When we found a protein of interest we used the “Selection Info” tab to find and record its IGB accession number, name, and nucleotide coverage. We further utilized the “Advanced Search” tool in IGB, using terms such as ‘zinc’ or ‘zinc transporter’ to quickly scan around our QTLs, recording all zinc-finger and zinc-binding proteins found within our specified nucleotide range. For a protein that seemed to be involved in zinc transport and regulation, we generated their amino acid sequence by highlighting the protein and selecting to view the sequence in genomic sequence viewer. From here, the sequence viewer generated the cDNA (upper right corner) and under the drop tab “Show”, we selected +1 Reading Frame. This gave us a translated amino acid sequence that we then blasted in NCBI *Arabidopsis thaliana* protein blast in order to validate the protein name and function. Many of the resulting proteins varied in name and function highlighting the importance of checking the information stated on IGB.

We used the TAIR database (REF) to identify the protein’s cellular location, function and developmental expression in *A. thaliana.* We used a list of simple sequence repeats, which was provided for us by Dr. Brown’s lab, to choose the best tri-nucleotide SSR primers located within 100,000 base pairs of our best candidate proteins.

**Results**

 Our BLAST and IGB searches produced in list of 9 gene candidates: two for our QTL on chromosome 6, five for our QTL on chromosome 8, and two for our first QTL on chromosome 9. All of our identified proteins were different from the exact types we found in our research and blasted. However, some were a part of protein families previously studied. On TAIR, we confirmed that all proteins were found in flowering cell types and expressed in the petal expansion and differentiation stage. We recorded candidate genes’ location, location from the QTL, and TAIR accession number **(Table 2).**

**Table 2.** Candidate gene locations on chromosome, their distances from the QTL, and TAIR accession number

|  |  |  |  |
| --- | --- | --- | --- |
| **Protein** | **Location** | **Distance from QTL (bp)** | **TAIR Accession Number** |
| IAR1 | C6: 9,022,635-9,025,010 | 824,728  | AT1G68100.1 |
| PCR12 | C6: 9,434,940-9,435,744 | 412,423  | AT1G68630.1 |
| MTPA1 | C8: 33,235,433-33,235,797 | 1,342,700 | AT3G61940.1 |
| H(+)-ATPase 5 | C8: 34,442,724-34,443,770 | 135,409 | AT2G24520.1 |
| H(+)-ATPase 1/2 | C8: 34,653,982-34,653,210 | 75,077 | AT4G30190.2 |
| HMA1 | C8: 34,921,138 – 34,921,306 | 656,827 | AT4G37270.1 |
| Ca2+ ATPase 7 | C8: 35,036,530-35,040,779 | 757,562 | AT2G22950.1 |
| Ca2+ ATPase 2 | C9: 514,143-514,673 | 3,044,270  | AT4G37640.1 |
| COPT1 | C9: 4,068,689-4,069,617 | 510,674 | AT5G59030.1 |

For chromosome 6, we found IAA-alanine resistance protein 1 (IAR1) and plant cadmium resistance 12 (PCR12) downstream from our QTL **(Figure 1A)**. We identified IAR1 through IGB alone, while a PCR2 BLAST hit aligned with PCR12 protein presumably because they share a homologous domain. Both proteins on IGB were incorrectly annotated, but blasting the amino acid sequence resulted in the correct protein identification.

Chromosome 8 had a number of zinc transport related proteins including a member of the MTP family (MTPA1), proton ATPase 5 and 1 or 2 (AHA5 and AHA1/2), heavy metal ATPase 1 (HMA1), and calcium ATPase 7 (ACA7) **(Figure 1B)**. When we BLASTed the IGB proton ATPase amino acid sequence in NCBI, it showed significant similarity to both AHA1 and AHA2 so we could not verify its exact nature. Most of our BLAST hits aligned with general ion transporter proteins, but bZIP1 could not be identified on IGB or through a subsequent nucleotide blast. We did locate and verify MTPA1 at the site of MTP1 blast hit, which was unlabeled in IGB **(Appendix A).** Finally, we found HMA1 through IGB alone, which was also mislabeled. Unlike all the other examples, Ca2+ ATPase 7 was more exclusively expressed in the flower and pollen cell type and implicated as important for pollen tube development.

For the first QTL on chromosome 9, we identified calcium ATPase 2 (ACA2) and a copper ion transporter 1 (COPT1) **(Figure 1C)**. Our HMA2/4 blast hit aligned with another calcium ATPase while an unrelated transcription factor was located at the bZIP1 blast hit. We identified COPT1 on IGB. Unfortunately, we were unable to identify any zinc regulatory proteins for the second QTL on chromosome 9 **(Figure 1C).** No proteins could be found at either blast location through IGB or a nucleotide blast, and most proteins annotated on IGB had nonspecific names. We found three SSR primers for each protein of interest, which are listed in **Appendix C**.

**Discussion**

Our results present an overview of all proteins we found that were remotely related. There are some better candidates than others. Proteins found on chromosome 6 were not well characterized or extensively examined in previous literature; however, our two candidates are similar to ZIP and PCR proteins. IAR1 shares a functionally similar shape to ZIP proteins and contains the same His residues and transmembrane domains (Laswell *et al.,* 2000*)*. Current publications on IAR1 focus on its potential role in auxin homeostasis, but they also postulate that it may transport heavy metals, including zinc, out of the endoplasmic reticulum and suggest that it is located there as well (Laswell *et al.,* 2000). Similarly, PCR12’s function and location are unknown but it is assumed to be involved in heavy metal transport and located in the plasma membrane based on its shape on Uniprot. Previous research had shown that both ZIP proteins and PCR2 are critical for zinc intake, transport and plant’s detoxification making these two good candidates for this QTL. This QTL was also a QTL for boron, but we do not share any of the same candidate genes.

 The QTL on chromosome 8 was the only QTL we did not share with any other mineral and the one where we found the best results. We included the two proton ATPases because they were very close to our QTL and both were stated to transport cations by creating the electrochemical gradients. We found that AHA5 was located in the plasma membrane, while AHA1 and AHA2 are located in different organelle membranes. A recent review has proposed that certain H(+)-ATPases may play a more direct role in heavy metal transport and uptake in root cells (Palmgren, 2001). The auto-regulated calcium ATPase, ACA7, is more directly involved in calcium ion transport and located in the plasma membrane. However, on TAIR, it is said to play a role in zinc ion homeostasis but we could



**A**

****

**B**

****

**C**

**Figure 1.** A visual representation of candidate proteins involved in zinc uptake, transport and storage located within two million base pairs offour *B. oleracea* QTL’s. (A) Proteins on chromosome 6. (B) Proteins on chromosome 8. (C) Proteins near two QTLs on chromosome 9. Red lines signify QTL location. Dotted lines signal unlabeled proteins on IGB that are present. Protein hits found using NCBI blast are listed and marked under “BLAST” while those found on IGB under “IGB”. Proteins in quotations are those that are misnamed on IGB and its correct name labeled after the equal sign. (These diagrams were inspired by Dr. Campbell’s, Eric Sawyer’s, and Hannah Itell’s figures.)

not find the original publication on this particular function. These proteins are not specific to zinc transport, but are thought to be involved with cation transport and are closely located to our QTLs. They are not our best candidates, but should be noted.

At this QTL, we also found two proteins that were apart of some of our researched protein families: HMAs and MTPs. As stated in our introduction, previous research has shown that HMA1 plays a role in the movement of zinc out of the chloroplasts when zinc is present in toxic concentrations (Kim *et al,* 2009). Although MTPA1 has not been as well studied or characterized, it is a part of the zinc transporter protein, and MTP1 is known to move zinc into the vacuoles for storage (Ricachenevsky *et al.*, 2013). These two proteins are within two million base pairs of our QTLs and are great zinc specific candidates.

 For chromosome 9, we have the fewest gene candidates and none are directly related to zinc uptake, transport, or storage. The calcium ATPase, ACA2, is involved in calcium ion transmembrane transport but since it is relatively unstudied, it may play a role in zinc transport similar to ACA7. Unfortunately, it is three million base pairs away from the QTL. We thought that a copper transporter may transport zinc as well since zinc transporters, like HMA1, 2 and 4, additionally bind to and move copper ions (Grotz and Guerinot, 2006). However, studies show that COPT1 is a specific copper ion transporter and has no known involvement in significant zinc uptake (Sancenon *et al.*, 2004). Therefore, these two protein candidates are our weakest so far. As stated above, we found no significant results on our second QTL for chromosome 9.

**Conclusion**

 Throughout our research, it became clear that a use of both BLAST and IGB were necessary to obtain significant and reliable results. BLAST hits usually aligned with proteins of interest while IGB was useful for confirming protein locations and identifying other proteins previously un-researched. In terms of our results, I propose that we found the most significant results around the QTL on chromosome 8 because it was not a QTL for other minerals and because it was well annotated on IGB, resulting in relatively few unnamed proteins. Conversely, both of our QTLs on chromosome 9 were QTLs for other minerals and I suspect that a more general heavy metal transporter is accounting for these joint QTLs. Furthermore, chromosome 9 had many more unidentifiable proteins in comparison to the annotations on the other two chromosomes, which also may have contributed to our results.

As this project continues, IGB should be updated with the correct names of the proteins we identified. Furthermore, MTPA1 should be added to the annotation on chromosome 8. Finally, future research in a wet lab should examine if these candidate genes are contributing to increased zinc uptake and storage or zinc detoxification.

**Acknowledgement**

 This paper is the culmination of research performed at Davidson College in the course *Laboratory Methods in Genomics*. I want to thank Dr. Campbell for his advice, guidance and enthusiasm. This project would never have been completed without him. I want to thank my partner Chris Polo who contributed to these results and who was resilient researcher even in the face of extreme illness. Finally I want to thank Dr. Allan Brown and Dr. Charles David who provided access to all the tools and information needed to complete this project, as well as guidance and information.

**References**

1. Assuncao, A., Herrero E. & Aarts, M. *Arabidopsis thaliana*transcription factors bZIP18 and bZIP23 regulate the adaptation to zinc deficiency. *Proc Natl Acad***107(22)**, 10296-10301 (2010).

2. Claus, J., Bohmann, A. & Chavarria-Krauser, A. Zinc uptake and radial transport in roots of *Arabidopsis thaliana*: a modeling approach to understand accumulation. *Ann Bot***112 (2)**, 369-380 (2012).

3. Clemens, S., Palmgren, M. & Kramer, U. A long way ahead: understanding and engineering plant metal accumulation. *Trends in Plant Science***7(7)**, 309-315 (2002).

4. Eide, D. Zinc transporters and the cellular trafficking of zinc. *Biochimica et Biophysica Acta***1763**, 711-722 (2006).

5. Grotz, N.*et al*. Identification of a family of zinc transporter genes from *Arabidopsis* that respond to zinc deficiency. *Proc. Natl. Acad. Sci.***2**, 7220-7224 (1998).

6. Grotz, N. & Guerinot, M. Molecular aspects of Cu, Fe and Zn homeostasis in plants. *Biochimica et Biophysica Acta***1763(7)**, 585-608 (2006).

7. Hoffman, N. Nicotianamine in Zinc and Iron Homeostasis. *The Plant Cell***24(2)**, 373 (2012).

8. Kim, Y.*et al*. AtHMA1 contributes to the detoxification of excess Zn(II) in Arabidopsis. *Plant J.***58(5)**, 737-53 (2009).

9. Lasswell, J., Rogg, L., Nelson, D., Rongey, C. & Bartell, B. Cloning and Characterization of *IAR1,*a Gene Required for Auxin Conjugate Sensitivity in Arabidopisis. *American Society of Plant Biologists***12(12)**, 2395-2408 (2000).

10. Marschner, H. in *Mineral Nutrition of Higher Plants*(Academic Press, Boston: MA, 1995).

11. Mertens, J. & Smolders, E. in *Heavy Metals in Soils*(ed Alloway, B.) 465-490 (Springer, New York, 1990).

12. Morel, M.*et al*. AtHMA3, a P1B-ATPase Allowing Cd/Zn/Co/Pb Vacuolar Storage in Arabidopsis. *Plant Physiol.***148(2)**, 894-904 (2009).

13. Palmgren, M. Plant Plasma Membrane H+-ATPases: Powerhouses for Nutrient UPtake. *Annual Review of Plant Physiology and Plant Molecular Biology***52**, 817-845 (2001).

14. Ricachenevsky, F., Menguer, P., Sperotto, R., Williams, L. & Fett, J. Roles of plant metal tolerance proteins (MTP) in metal storage and potential use in biofortification strategies. *Front. Plant Sci.***4**, 02/24/14 (2013).

15. Sancenon, V.*et al*. The *Arabidopsis*Copper Transporter COPT1 Functions in Root Elongation and Pollen Development. *The Journal of Biological Chemistry***279 (15)**, 15348-55 (2004).

16. Sommer, A. L. & Lipman, C. B. Evidence on the indispensible nature of zinc and boron for higher green plants. *Plant Physiology***1**, 231-249 (1926).

17. Song, W. & et al. *Arabidopsis*PCR2 is a Zinc Exporeter Involved in Both Zinc Extrusion and Long-Distance Zinc transport. *American Society of Plant Biologists***22 (7)**, 2237-2252 (2010).

**Appendices A – Verifying that MTPA1 is present in *B. oleracea genome***

1. On IGB there was no protein labeled at the MTP1 blast hit (33,235,433 – 33,235,797)

****

1. Highlighted and copied nucleotide sequence 33,232,019 to 33,235,843 base pairs and performed a nucleotide blast on NCBI *Arabidopsis* genome

Sequence: GGATACCAAATATACTTTAGATAACCGCACTTTAGTAACGTTTTTTTTTTTCAGATCCTTCTAATATATAGTATTATATTAGGCTAAGAAGAATAAGAGAGTATGTTGGCTCATGCATTGTAAAGATACCAGTATTACCGTACTATGTACGGGCACTAATATAAAATATGTTTATTTTGAACATACTATTAGAATTTATAATTAAGATAACATTAAAATTTTGATGTTTATAATTTACAACAGTTCAATATAGACACAAAATATATAATGTTAATATTGTTTCAAAACCAATATAATCATTAAATATGATCGAACAACGAAAATATATATCTTACAAAAGTACAATTTGTCAACTTAGTTTTGAGGTTATTATATTTTGTAGTCATTTTCTTATTCTCTCACTCATGTTTACTTTTTTGGTCATATTCTTTTATATTTCTTTTTATACAACGATGCTATTTGCTTCGACATTTCTTTTTGAGTGAATCTACAAATTCAATAAAACTATTTAAGGAAAAAAAATCAGAAGTGTATATTGAATAAAAAATAAGATATCACCACATTTTATAGATTATAAAAAGTTAAACACACACAAGAAAAATATGGAATACTTACCGAGTCTAAATAATAGACATAAGTCTTTTTTCTCCTACAACCCGAATTTCTTATGCCCTATTTTTTATTTCTAATTATCACAACATTTTTCCTTTCTTCTTTTCTTCTTGTTGTACTTCTCAAGTTTTCTAGGGGCCATGGAAGTGTGGTTTAAGGTGAGCCTCTTACTCCTTCCCTGAATTATCACCAAAACAAATAAACGTGTTCCATAAATTGATAATTAGTTTAGTCTAGTTTAGTCCATTATGAAAATTAATTGACCTATTTTGTTTACGTACCTTTTTCTTCTGTTCTTCGGCATCTAAAATAATAAAAGAAAAGAAATATATAGTATCAGTGTTTTTTTCCTTTGAAATGCTAAAATAGATGAAAAAAAAGCAAAGCGAGAAATTTTCTAAAATTAAACCATAGAAACAGATTTGTAAAATACTTGTTTTTGTACAATAGAGTTAGGGATGGTTATTGTTACCTTTGTTGCTGAAAAAACTTTGATTTGTCATTCATTTATTGGAGAATTAAGCTTTCAAAAGCTTTGAAACATAAACCATATGTTTTTTTTGCTACAGTCAAGGACTCCGTCGTCGTAAGAAAGAAGAATATGAAACAATAGAAAAGCAAAAGGAGAGAGGAAGGAGAGAGGAGGAGTTTAATAGAGAGTAAAACTGTTTTGTCAAACTTATATAAGAACAAAACAGTTTCAAATCGTAACTGTTATTTTAGAACGTGAAGTTTTTTTTTGTCAAGAAGTGGAAGTGATAATGTGAGAAATTTTAGGGTTTAGTTATTTCGAATTTTAGAAATTTTAGGATCTAACATTATTTTTTGTTGCATTTTAATTTTAATTTTTTAAACAGATGTCATGTGTCAAATTTTGATTCGCTAGGTGACTTGTGCTTTAGTAAATAAGGGATAGGTGTAAATTTCTACATTTTTTTTTCGTCTAAAGAATGTAAATTCACATTTACACAAATTAAATATGAATTCTTCATCATACTCAAATACTATAATTATTTTTACTTCAAAATGTGAAAGATACACTAGTATATAAACGATATGCGGTATATGTACTCATTCTTCCACATGGTATTAAGGTCCAGCTAGTAAATTTAAACAAACGTAACACAATTTAAAATAACATAAGATTTGGTTTTTAAATATAAGTTGAAGAACGATCATGTCAAAGGGAATCATAATGTGTGAACCTTTCACGAACGAGAAATTTTAAATGGATTTCTCAATTATTTGAGACACTTTACTACTTAATAAAATACTAATGATTTTGTTTTCACACTTAACTCAGATAACAATACCAATAAATTAATGATGCTCTAAAATCTCATATCATAAATATAAAAGACAAATACTAATATATACTCCATCTGTTTCATATTGAGTGTCGTTGTAGAGAAATTTTTTCGTTACAAAATAAGTGTCGTTTTCGATTTTCAATGCAAAATTTATTAATTTTATACAGTAATTTATTTTTCTATTGGTTGAAATATGGTTAGGTGTATAGTTAATTGTATTTTTATATAGAAAATATACAAAATTAATTATTTTCTTAATTTGTATGCACAAGATCAAAATGACATTTAATATGAAACAGAGGGAGTATAAAAGAGACAAAGAGACGTAAGTAATATAAGTTGAGTTATATTTTCATAGTTTTAAATAATTATCAGTTTTTAAATTCACTTTAAACATTTCATCTTTGAGGATATCTCAAAGAATAAACTATTACATACAAACTTCAAATCCAATCGATAACTTTTAAACTCGAAACTTCACAATCGTCTCTGGATGAGTAATATAAAATATATCTTTTAAAATCGAGGTCGATGTCACGTGAAGTTAATAGTTAAATACATTTTCCACGATATTCTAGTTTCGGTCAAAGTCAAATATCGTTAGCGAAGACTTATACGTGGAAAGTGGAAACAACGGCAGCACCAGCCTAAACATTTCCTTGGCCCAAAAATATACGTAGTTCATTAACTACTTTTGAAAATGGGTCTGTTTAAGAAAACTAACAAGTAACTAAATCTATGTTTTACGTTATGTATCGGAGTTATTAGATTATATTAGTCAACAAATGTCTATCTTATAGACAAAAAAAAAATGTAAAACCCGTGAAGCGCATGAAAACTAATATTTGGAAACATTTGAAGAGTAAGCCAATATTATCCTAGCATGTTTACATTATTTATATTCCATTTGTGTATCAAATTTGAATTATGAAAGTTTACGTGATGTTTTTTTTTACATTATATGATGTTAGTAATATGTTAAAAGCTAATCACTATATATATGAATCTAGGAAAATTGCTTAAAATACCACATTCCTAATATCAGTTATAACTCTAGAATTAACTAATCTAGATTTAATATTTAGAGTTGAGGGGGTGGATAGCGTTTTGGAATGAGAATTTAGGATTCTAATAACTATATAAATAAATACTTAAAAAGTTAATTAAAAAAATAGTTTCAAAAATAATTTTTGAACTTCAAAAAAAAATTTGGAAAAAAAAGATTATATAAAAGTTCGAATTTAAAAACATATAATTCAAAAAAAATTATTACTTTTATTTAATTTTTTTATTATATATAAAAAAAAAGGATACAAGAGTCTTTTGTTTTTTAATAAAGAAAATATTTGTGAAAATGTCCTTTTATTGAAAATAAACATGAATAAATGATATGCTAATGCTGTAAATATGAAAATTTCCAATGAATCTGTAGTCTTGTATCAAACACAATTCATAGAATTTTTTTAGTGTCCAAAGTTCACATTGTGTTTCTTTGGCCGATGGATTCAACAAGAAATAAAGTCTGCGGAGAAACATCTTGTGTGTTTTCAACTTCTACAAGCGATACAAAGGAACGTTCTGCTTCGATGCGAAAGCTCGCTATCGCTATCGTTGTCGTGCTATGTATTTTGTTCATGACCATAGAAGTTGTTGGTGGCATTAAGGCAAATAGTTTAGCTATACTAACCGATGAAGCTCATTTGAAGAAGTGGTGTCTGTTCATGAGCTTCACATTTGGGCAATAACAGTTGGAAAATTCCTGTTTTCTTGTCATGTCAAAATCAAACAAGAATCTGATGATGCTATGGTGCTTAACAAAGTGATTGATTACATCCGGAGAGAGTACCGTATCAGTCACGTTACCATACAAATCGAGCGTTAGAACAACAAAGCAAGCCGGAGACGGTTAATATATCTGGTTCTCTGTCATTTTCATGATCAAATCTCAATCTTCTTTTTTGCTTTCTCGTGTTCTTTCCTTTACATTGTTCTTTGCTATGTCATTGTTTTTGTTTATGAAGTTTTTTTTCCTTGTTTTATCTTAAAATGGGTATCTGAAAGGACATAGCTTTATTTGAGATTAGGTCTCTTTGTGTTATGTTACTATAGAAGAATCTGAATCTTTAGGGTTTATCTCTGATAACATGCCTATCAATAAGAGTGGGGGCATGTTTGTCCTTAGAGCAAGTAACCATAGCTCGTGTGTGACAGTAGCAAACCAGATCACAGTACACACACACACGATAAAAAAAAAAAGAATTCAAACTTTCTTTGGAGTTTTTAGTTTTATAGTTGATAAGGTTTCTAAGGTCTTCATTAACAATGCAAATGCTGATTCCTTGTTTCTTTGTTAAGGATGTTGATTGTTGACTTGAGAGACTGGCTATGAATTTTACCCATTGACTGAGATTGTAGACAAAACTGGGACATTTTATAGTACAACACTATTCAATGCCGTTGAAGTCAAGGCATCAATAATACATAAGATTCTTTAGTGTCACATTATCTATAGATTCAAGATATTAGACAAAAACCACACATAAGAACATTATAAAATGATGGGGGCAGCGTAGTTCTGTATAAAGGCTTTGATTTTTGAGCTTTGAAGAAAAGGAATGGTGTGGTGTGTAATGTAATGATGTATAATGAAGAAAGGATGATTGAGACATCATGAGAAAGATGACAGAAACCCAATACAAGCTCTCTTACCCATGACTTGTCTCTATTTACCTTCTTCTTATGTGTAAATTAGAATGTCCCCACAAAGTCTCTGTCTCCACTTCAAGACAAAACTGCCTCCACCAATAATCAATCTTTCTCTCCCTCTCTCCTCTCCTATGCTCTGCTCAGCCTCACCAGCTCTTCTTACAGTACCTAAATAGATTTGTTTTCTTACCTCTCTGGTTTTTTACAGCTCCACTGTTTTCTTCTCTATTCAATAGCTACTAAAGAGTCACTGCTCTCTATATAAAGCTAACAGTTACAAGTCCAGAAAGACTCAGAATCAGAAGAGAGTACATGAGTTTCCTATTCTGCTTCTAAGCATTTATTGTTACAC

1. There was one significant result on the chromosome 3 *A. thaliana* that shared an 80% identity with zinc transporter protein. The E-value was 6e-120



****

1. The protein linked to NCBI page that cited the protein being called MTPA1, and linked to TAIR with Accession number AT3G61940.1

NCBI Link: <http://www.ncbi.nlm.nih.gov/nucleotide/240255695?report=gbwithparts&from=22937445&to=22938449&RID=MSW1Y49N014>

TAIR Link: <http://www.arabidopsis.org/servlets/TairObject?type=locus&name=AT3G61940>

**Appendix B – Protein Blast Candidates**

**Table 1.** Previously identified zinc regulation related proteins, their amino acid length, accession numbers and website links to NCBI page with amino acid sequence.

|  |  |  |  |
| --- | --- | --- | --- |
| **Protein** | **Amino Acid Length** | **Accession Number** | **NCBI Citation** |
| ZIP1 | 355 | NP\_187881 | http://www.ncbi.nlm.nih.gov/protein/NP\_187881.1 |
| bZIP1 | 609 | AAF22906 | http://www.ncbi.nlm.nih.gov/protein/AAF22906.1 |
| ZIP2 | 353 | NP\_200760 | http://www.ncbi.nlm.nih.gov/protein/NP\_200760.1 |
| ZIP3 | 339 | NP\_180786 | http://www.ncbi.nlm.nih.gov/protein/NP\_180786.1 |
| ZIP4 | 374 | AAB65480 | http://www.ncbi.nlm.nih.gov/protein/AAB65480.1 |
| bZIP19 | 261 | NP\_567974 | http://www.ncbi.nlm.nih.gov/protein/NP\_567974.1 |
| bZIP23 | 249 | NP\_179268 | http://www.ncbi.nlm.nih.gov/protein/NP\_179268.2 |
| IRT1 | 347 | NP\_567590 | http://www.ncbi.nlm.nih.gov/protein/NP\_567590.3 |
| IRT2 | 350 | NP\_001031670 | http://www.ncbi.nlm.nih.gov/protein/NP\_001031670.1 |
| IRT3 | 425 | NP\_564766 | http://www.ncbi.nlm.nih.gov/protein/NP\_564766.1 |
| HMA2 | 951 | NP\_194740 | http://www.ncbi.nlm.nih.gov/protein/NP\_194740.1 |
| HMA3 | 542 | NP\_194741 | http://www.ncbi.nlm.nih.gov/protein/NP\_194741.2 |
|   | 760 | P0CW78 | www.ncbi.nlm.nih.gov/protein/P0CW78.1 |
| HMA4 | 1172 | NP\_179501 | http://www.ncbi.nlm.nih.gov/protein/NP\_179501.1 |
|   | 95 | 2KKH\_A | http://www.ncbi.nlm.nih.gov/protein/2KKH\_A |
| PCR2 | 152 | NP\_172940 | http://www.ncbi.nlm.nih.gov/protein/NP\_172940.1 |
| Mtp1/ ZAT1  | 398 | NP\_850459 | http://www.ncbi.nlm.nih.gov/protein/NP\_850459.1 |

**Appendix C - SSR Primer Pairs for Potential Genes of Interest**

**Chromosome 6**

IAR1 located at 9,022,635-9,024,010 bp

1)

For Primer AAAACATTCAAAGACAACAC

Rev Primer GTTTGTGAGAGAGAGAATCA

Repeats (cct) x5PCR product = 363 bp & start at base 9,017,802

2)

For Primer ACATGTGTTAATGACTTGGT

Rev Primer AAAATCAAAGTAAAGGGAAT

Repeats (aat) x5PCR product = 277 bp & start at base 8,976,499

3)

For Primer TCTTAAAATATGTCTTTCCG

Rev Primer CTTCGAAATTTGTCATGTAT

Repeats (tat) x5PCR product = 386 bp & start at base 8,991,499

PCR12 located at 9,434,940-9,435,744 bp

1)

For Primer AATCCTAATCCTACGTTCTT

Rev Primer ACCCATCTCTCTAATAAACC

Repeats (tgt) x5PCR product = 290 bp & start at base 9,424,677

2)

For Primer AATACCAAAACACTCAAAGA

Rev Primer AACTGCAAACAAATAAAGAG

Repeats (tat) x5PCR product = 330 bp & start at base 9,431,851

3)

For Primer CTCCTCCTTCAACTTCAC

Rev Primer GTTGTTTTGAGGTGTATTGT

Repeats (cca) x5PCR product = 368 bp & start at base 9,447,692

**Chromosome 8**

MTPA1 located at 33,235,433-33,235,797

1)

For Primer GTCTGAGTATTGAGCAACAT

Rev Primer TCAACTCTCTCTAACGGTAA

Repeats (gaa) x22PCR product = 273 bp & start at base 34,303,669

2)

For Primer AGCAACAATTTACTTTTCTG

Rev Primer AGAGAGTACAGAGGAAGGAG

Repeats (ttc) x5PCR product = 189 bp & start at base 34,275,721

3)

For Primer GGGAAAATTAGATAAAAACC

Rev Primer TTCGCCTACATATTAAATTC

Repeats (cac) x6PCR product = 263 bp & start at base 34,322,402

H(+) –ATPase 5 at 34,442,724-34,443,770

1)

For Primer AGTTGTTTTCAGAGATGATG

Rev Primer CGAATATAGCTGTCTCATTC

Repeats (aga) x10PCR product = 242 bp & start at base 34,349,188

2)

For Primer ATAATTTCTGATGTTGGTTG

Rev Primer TGTCAATAGAAGGATTTGAC

Repeats (atc) x5PCR product = 298 bp & start at base 34,403,291

3)

For Primer TACAAAACAATTTTAGAGCC

Rev Primer CTGAGCTAAATGGAAGAATA

Repeats (gaa) x5PCR product = 286 bp & start at base 34,340,088

H(+) –ATPase 1/2 at 34,653,982-34,653,210

1)

For Primer GAAGTGAAAGACATCTGGTA

Rev Primer TTGCTCATAAAAGAAAAGAC

Repeats (atc) x7PCR product = 194 bp & start at base 34,638,263

2)

For Primer TATATGAATGGAGACTGGAG

Rev Primer ACTAGTGCTTTCATCATCAC

Repeats (tga) x7PCR product = 280 bp & start at base 34,681,263

3)

For Primer GATTCCTAGACTTGTTGTCA

Rev Primer GGAAAGAGATTGGACTAGAT

Repeats (ttg) x5PCR product = 257 bp & start at base 34,642,029

HMA1 located at 34,921,138 – 34,921,306 bp

1)

For Primer TTTCCCATAAACACAGATAC

Rev Primer TCAACTCTCTCTAACGGTAA

Repeats (aag) x7PCR product = 265 bp & start at base 34,923,615

2)

For Primer ACCCTCTCTCTTTCATAAGT

Rev Primer GATTGATTTACCGATCTTCT

Repeats (caa) x6PCR product = 314 bp & start at base 35,025,000

3)

For Primer AATGTTTGGTCTTGACTATG

Rev Primer TTTTCGGATCTATTTTGTAG

Repeats (gtt) x5PCR product = 185 bp & start at base 34,859,373

Ca2+ ATPase 7 at 35,036,530-35,040,779

1)

For Primer ACCCTCTCTCTTTCATAAGT

Rev Primer GATTGATTTACCGATCTTCT

Repeats (caa) x6PCR product = 314 bp & start at base 35,026,000

2)

For Primer TTTCCCATAAACACAGATAC

Rev Primer CTTACATCAAACCTCTTCAG

Repeats (aag) x7PCR product = 265 bp & start at base 34,923,615

3)

For Primer AAAATCATTTGTATTCTGGA

Rev Primer GAAGCTGTTCATTTATTTGT

Repeats (att) x8PCR product = 254 bp & start at base 35,181,253

**Chromosome 9.1**

Ca2+ ATPase 2 located at 514,143-514,673

1)

For Primer ATTATCGTCATCATCTTCAG

Rev Primer TGAGCTTCCTAAGTTAACAG

Repeats (tca) x6PCR product = 293 bp & start at base 536,627

2)

For Primer CAAAGATGGATTTTCTAGTG

Rev Primer TTGAATCTTGTTTGAAATCT

Repeats (gaa) x5PCR product = 267 bp & start at base 510,857

3)

For Primer TTCTTTTGTCTATTAGCTGG

Rev Primer AGCTGTTTCATTGTATCCTA

Repeats (gtg) x5PCR product = 246 bp & start at base 534,769

COPT1 at 4,068,689 – 4,069,617

1)

For Primer GGATAGTAGTTTCCTGAGGT

Rev Primer GTATCCTTCAATTTGTTTTG

Repeats (aat) x8PCR product = 148 bp & start at base 4,150,652

2)

For Primer ATTTGCATGATAAATGCTAT

Rev Primer TGTGATATTTTAAGTGGACC

Repeats (tta) x5PCR product = 277 bp & start at base 3,993,909

3)

For Primer TCCTTTTAACCAGAGTATCA

Rev Primer GATTACCTTCAACCTAACCT

Repeats (aga) x5PCR product = 340 bp & start at base 4,136,602

**Figure 2**. Results for Chromosome 8