James Stewart and Nisha Crouser

Dr. Malcolm Campbell

BIO 343 Lab Methods in Genomics

8 May 2014

The Genomics of Calcium Uptake and Storage in Broccoli (*B. oleracea*)

Abstract

 Calcium is a vital nutrient in both plants and humans, providing the basis for many vital biological processes. This study attempts to identify genes that are important in calcium uptake and storage in Broccoli (*B. oleracea*) and match them with Quantitative Trait Loci (QTLs) provided to us by our NC State University collaborators. Using two different search tools, tBLASTn and the IGB genome browser, we mined the *B.* *oleracea* genome for candidate genes within 1,000,000 bp of the QTLs. We successfully identified genes near all seven QTLs. The proteins coded for by these genes belong to four different protein categories: ion channels, ATP-driven calcium pumps, H+-driven calcium pumps, and calcium binding proteins. Having identified likely candidate genes in all four protein categories in our results, we have provided evidence for a comprehensive model for calcium transport and storage in *B. oleracea* that match up well with the QTL list we were provided.

Introduction

 Calcium is a divalent cation that is an integral component in a vast number of biological systems in both plants and animals, particularly those related to growth and homeostasis. In plants, calcium provides strength and stability to the cell by forming pectate cross-links within the cell wall (Conn *et al.*, 2011), acts as an intracellular messenger in the cytosol in signal transduction events, and provides a basis for stress responses in the plant as a counter-cation exchanger (White *et al.*, 2003). Calcium is rarely lacking in most soils, but is can become scarce in an agricultural setting where calcium is often rapidly depleted from the soil. Calcium deficiency in plants results in weakened cell walls that can lead to a variety of horticultural diseases, such as blossom-end rot in tomato or tipburn in lettuce (White *et al.*, 2003).

 Uptake and transport of calcium in a plant occurs through two main pathways: the symplastic pathway (through cells) and the apoplastic pathway (between cells). Both are equally important pathways and provide distinct advantages and disadvantages to the plant. Apoplastic transport, which constitutes the majority of calcium transport in the plant (Conn *et al.*, 2011), keeps calcium from coming into contact with the cytosol of the cell, which must remain at a submicromolar level of Ca2+ so as not to trigger the many calcium-dependent enzymes that reside in the cell cytoplasm (White *et al.*, 2003). However, apoplastic transport is strongly affected by transpiration, which can lead to inconsistencies in the amount of calcium supplied to certain tissues like the shoot, and could ultimately lead to calcium disorders in the tissue (White *et al.*, 2003). Apoplastic transport is relatively non-selective in which divalent cations it transports, meaning toxic cations could also enter (White *et al.*, 2003). Conversely, symplastic transport often involves highly selective cation transporters, but is limited in how fast transport can occur, since rapid symplastic transport of Ca2+ would upset the carefully maintained submicromolar Ca2+ concentration in the cytoplasm (White *et al.*, 2003).

 Though a large amount of Ca2+ is stored in the cell wall of plants and in the apoloplast (Hirschi, 2004), it is also very often sequestered in other areas of the cell, including the vacuole, the endoplasmic reticulum, and the chloroplast, with the vacuole acting as the most prominent Ca2+ sink (Hirschi, 2004). These storage areas are essential for proper cellular function, as they allow intracellular Ca2+ to be removed from the cytosol and stored at high concentrations in areas that can be tapped rapidly when required by the cell for the calcium signaling events (Tuteja *et al.*, 2007).



Figure 1. Image outlining several different types of calcium transport proteins and how they function. 3 main types of proteins are shown: ions channels (*e.g.* GORK), ATP-driven calcium pumps (*e.g.* ECA1), and H+-driven calcium antiporter pumps (*e.g.* CAX). This figure also displays the major organelles involved in calcium storage, the vacuole and the endoplasmic reticulum. Image is from White et al. 2003.

There are a variety of protein-coding genes involved in calcium transport and storage within plants (Figure 1), but they can be grouped generally into one of four categories. The first category comprises various ion channels, which are often located within the plasma membrane. These integral proteins mediate passive entry of Ca2+ into the cytosol of the cell, and are driven by the large gradient differential between the apoplast and the cytosol (Hirschi, 2004). These ion channels are generally voltage-gated channels that are triggered by changes in cytosolic voltage (Hirschi, 2004). While these channels are extremely important in facilitating Ca2+ influx into the cytosol, it is not correct to refer to them as “Ca2+ channels;” rather, they are more correctly described as “Ca2+-permeable channels,” as they are very often non-specific to calcium and facilitate the influx of other divalent cations into the cytosol as well (Sanders *et al.*, 2002). There are similar voltage-gated channels located on the endomembrane of the calcium sequestering organelles, and allow for efflux of Ca2+ from the organelle into the cytosol. In addition to the majority of plasma membrane ion channels that are voltage-gated, there is also a smaller group of proteins that are non-gated and allow unregulated passive diffusion to a variety of different ions into the cytosol. These channels often discriminate poorly between monovalent and divalent cations (Sanders *et al.*, 2002). These channels provide a non-specific route for many ions into the cytosol, meaning they would likely serve a wide variety of cellular functions that are unrelated to calcium transport (Sanders *et al.*, 2002).

The second and third groups of calcium transport proteins are those responsible for calcium efflux from the cytosol, and are both calcium pump proteins. The first is composed of ATP-driven calcium pump proteins, endomembrane proteins that allow for active (ATP-driven) transport of Ca2+ ions into the sequestering organelles. These Ca2+ pumps are located primarily on the endoplasmic reticulum and on the chloroplast (Sanders *et al.*, 2002). The second group comprises the Ca2+/H+ exchangers, which facilitate active transport of Ca2+ across the gradient without direct ATP expenditure. These calcium “antiporters” pump Ca2+ out of the cytosol and into the sequestering organelle like the standard calcium pumps, but are driven by the downhill flow of H+ ions across the gradient into the cytosol rather than ATP. Although these calcium antiporters are located primarily in the vacuolar membrane, there is evidence that they may also exist in the plasma membrane (Sanders *et al.*, 2002).

The fourth and final group of important calcium proteins within plants includes the calcium binding proteins. These proteins are also split into two distinct groups: those that bind to free cytosolic Ca2+ in order to dampen cytosolic Ca2+ concentrations, and those that bind to Ca2+ to initiate one of the innumerable signal transduction events that occur in plant cells (Hirschi, 2004). The former group consists of integral membrane proteins found primarily on the endoplasmic reticulum (Roderick *et al.*, 2000). These proteins bind to calcium and store it in a relatively inactive state. The latter group, the proteins that initiate signal transduction, is extremely diverse in what the proteins eventually activate. Structurally, most of these calcium-binding proteins contain multiple “helix-loop-helix” structures (coined “EF hand”) that are each capable of binding to and holding a Ca2+ ion (White *et al.*, 2003). The binding of the Ca2+ allosterically modulates these binding proteins, activating them as signaling proteins to in turn activate something else.

In this paper, we investigate several protein-coding genes involved in calcium uptake and storage as they appear in the genome of broccoli (*Brassica oleracea*), with the goal being to locate genes that increase overall storage of calcium in *B. oleracea*, particularly in the florets. After looking through the primary literature, we compiled a list of candidate genes to search for in the *B. oleracea* genome based on the most noteworthy genes involved in calcium transport and storage in the model plant species, *Arabidopsis Thaliana*, that we predicted would also be present in *B. oleracea* (Table 1). From our collaborators at North Carolina State University, we received a chromosome map of quantitative trait loci (QTLs) in *B. oleracea* that outlined probable areas on the genome for calcium uptake genes. QTLs are stretches of a chromosome that have been identified as responsible for affecting certain quantitative traits, calcium content in his case. The goal of our study, then, was to determine which candidate genes identified in the primary literature best correlated to the QTL map provided to us.

|  |  |  |  |
| --- | --- | --- | --- |
| Protein Type | Gene Name | Location  | Details |
|   |  |   |   |
|   | Glutamate Receptor | Plasma Membrane |  Ligand-gated cation channel that is well documented in causing Ca2+ influx in plants. (Dennison *et al.*, 2000) |
| Ion Channels | GORK | Plasma Membrane | Voltage-gated K+ channels that influx Ca2+ at the same time as they efflux K+ (Ache *et al.*, 2000)  |
|   | MCA1 | Plasma Membrane | Ca2+-permeable mechanosensitive channel (Yamanaka *et al.*, 2010) |
|   |  |   |   |
|  | ECA1 | Endoplasmic Reticulum | Highly regulated Ca2+ pump that transports cytosolic Ca2+ in the Endoplasmic Reticulum (Sanders *et al.*, 2002) |
| ATP Ca pumps | ACA | Plasma membrane/ several organelles |  Family of highly regulated Ca2+ pumps that act in many membranes and are activated through Ca2+/Calmodulin binding (Sanders *et al.*, 2002) |
|   |  |   |   |
| Ca2+/H+ Antiporters | CAX2 | Vacuole |  Low affinity calcium transporter that uses diffusion of H+ into the cytosol to power Ca2+ transport to the vacuole. (Conn *et al.*, 2011) |
|  |  |  |   |
|   |  |  |  |
|   | Calmodulin | Cytoplasm | Calcium binding protein with 4 EF-Hands. Activates genes, including ACA, and can be a ligand for ligand-gated ion channels (Sanders *et al.*, 2002) (Luan *et al.*, 2002) |
|   | Calcineurin B | Cytoplasm | Calcium binding protein with 3 EF-Hands. Activates other genes (Luan *et al.*, 2002)  |
| Calcium-binding proteins | Annexin | Plasma Membrane | A non-mobile plasma membrane calcium-binding protein. Does not bind using EF hands (Tuteja *et al.*, 2002) (White *et al.*, 2002) |
|   | CDPK | Cytoplasm |  A superfamily of multifunctional cation-binding proteins. These genes are generally found in the cytoplasm and have 4 EF hands. (White *et al.*, 2002) |
|   | SOS3 | Cytoplasm | An EF Hand protein that activates SOS2, a protein kinase, during a salt-induced stress response. (Halfter *et al.*, 1999) |

Table 1. A list of candidate genes suspected to be present in *B oleracea* based on their existence and importance in *A. thaliana*. Genes are split into four categories, and include ion channel proteins, ATP-driven calcium pumps, H+-driven calcium antiporter pumps, and calcium binding proteins. Candidate genes are listed along with their location in the cell and a brief description of their function.

Methods

*tBLASTn against the B. oleracea genome*

 The first step to identifying candidate genes was to determine if any of the genes outlined in Table 1 appeared within 10,000,000 bp of the QTL marker on the *B. oleracea* genome. To determine this, we identified the specific amino acid sequences of each candidate gene in *A. thaliana* using the protein database on NCBI. We used the proteins in *A. thaliana* under the premise that it is one of the best studied plant species, and that it is relatively closely related to *B. oleracea*. Next, we did a tBLASTn protein alignment comparing the amino acid sequence for the target gene with the genome of *B. oleracea*, noting any candidate gene that appeared within 10,000,000 bp of one of the QTL markers on the genome. Access to the yet unpublished *B. oleracea* for tBLASTn analysis was provided by our collaborators through dev.vaccinium.org.

*Genome browsing using IGB*

 Our next goal was twofold: to both verify the existence of the candidate genes located through the tBLASTn method and to visually search for any additional candidate genes near the QTLs. To do this, we used IGB, a genome browsing tool, to search for candidate genes. IGB uses software to predict genes on the genome and automatically annotate them. For our purposes, we uploaded the genome of *B. oleracea* and searched through the annotations predicted by the program. We specifically targeted the genes identified previously using tBLASTn in order to verify their existence, but also searched for any candidate genes that might appear in IGB but had been missed by the tBLASTn search. Based on new information received at the time, we restricted our potential range around the QTL from 10,000,000 bp to 1,000,000 bp. We then went back to tBLAStn with the genes located on IGB and reanalyzed those sequences against the *B. oleracea* genome to reconfirm their existence in the genome. Our final list of candidate genes based on the QTL map provided by our collaborators therefore included all genes identified within 1,000,000 bp of the QTL using either the tBLASTn tool or the IGB genome browsing tool.

*Simple Sequence Repeat (SSR) selection*

 The final step of the process was to locate Simple Sequence Repeats (SSRs) for each candidate gene identified through tBLASTn and IGB browsing. These SSRs were either di- or tri-nucleotide repeats that occurred within 25,000 bp of the candidate gene on the genome. We used a list of SSRs provided by our collaborators at NC State University, and reported the three SSRs from the list that were closest to each candidate gene.

Results

 Figure 1 shows a visual representation of all candidate genes identified in the *B.* oleracea chromosomes within 1,000,000 bp of the QTLs. We mined all seven QTL-containing chromosomes and successfully located candidate genes on each chromosome within 1,000,000 bp of the QTL. All candidate genes identified using tBLASTn were successfully verified on the IGB genome browser across all chromosomes. However, we managed to locate important candidate genes using the IGB genome browser that did not find beforehand during our tBLASTn searching. These genes are shown in green in figure 1, while all genes that were located using both methods are shown in black.

In total, we located all of our initial genes of interest from table 1 within 1,000,000 bp of the QTLs, with the exception of genes from the ACA family. We found all three of the ion channel proteins (MCA1, GORK, and a glutamate receptor), an ATP-driven calcium pump (ECA1), a H+-driven calcium transporter (CAX2), and all five calcium-binding proteins (calmodulin, calcineurin-B, CDPK, annexin, and SOS3). Several of these candidate genes were located within 150,000 bp of the QTLs, and we marked these with a star to indicate that they are particularly strong candidates based on their QTL proximity. In addition to the genes from table 1, we also located a gene on chromosome 6 that was labeled “Calcium-binding EF-Hand.” We did not find this gene through tBLASTn, but did find it using IGB. We were not able to identify exactly what kind of EF-hand protein this was, but there is potential for it to be a significant candidate gene along with the other calcium-binding proteins. Also, note that this protein is 1,082,871 bp away from the QTL, which is just slightly over our 1,000,000 bp search range. All other identified genes were from the original candidate genes in table 1, and were all located within 1,000,000 bp.

 After compiling all of the good candidate genes that we had identified, we created a list of SSRs that were close to each of the candidate genes (Table 2). These SSRs were all either di- or tri-nucleotide repeats that were referenced from a list of SSRs provided by our collaborators at North Carolina State University. We chose the three closest SSRs to each candidate gene, and all SSRs selected were within 25,000 bp of the candidate genes. In Table 2, we provide both the forward and reverse primer sequences of the SSR, the nucleotide length of the SSR, and the starting nucleotide of the SSR.







Figure 2. *B. oleracea* chromosomes displaying candidate genes related to calcium uptake and storage located within 1,000,000 bp of each target QTL. C2 – C8 refer to the respective *B. oleracea* chromosomes. Each map displays the range of 1,000,000 bp above and below the QTL marker, with nucleotide number increasing from top to bottom. Each QTL marker is shown in red in the middle of each chromosome with its exact nucleotide location shown to the right. Candidate genes that were located using both IGB and tBLASTn are displayed in black, and those that were located through IGB exclusively are displayed in green. Gene names are to the left of the chromosome, and their exact nucleotide locations are on the right. A star indicates a particularly good candidate gene that appeared within 150,000 bp of the QTL.



Table 2. List of Simple Sequence Repeats (SSRs) located near candidate genes. Each candidate gene located within 1,000,000 bp of a QTL is displayed in the table along with its nucleotide location on the chromosome. Each gene is displayed with the three nearest SSRs from the SSR list provided by our collaborators. The size, starting location, and forward and reverse primer sequences for each SSR are shown to the right of the candidate gene. Candidate genes are organized by chromosome.

Discussion

 Based on the results we received from our tBLASTn searching and IGB browsing, we feel confident saying that we have located excellent candidate genes for all of the QTLs provided to us by our collaborators at NC State University. Using tBLASTn, we located almost all of our initial genes of interest from the literature within 1,000,000 bp of the QTLs. We are even more confident in the presence of these genes and their importance as candidate genes because we were successfully located each candidate gene using the IGB genome browser as a compliment to the tBLASTn results.

 We were very excited to identify and locate several genes coding for calcium pumps and ion channels, as these two types of proteins directly lead to transport and storage of Ca2+ in different parts of the cell. We located ECA1 on chromosomes 2, 3, and 7 and CAX2 very close to the QTL on chromosome 4. These two genes code for calcium pumps that sequester cytosolic Ca2+ into two important Ca2+ sinks, the endoplasmic reticulum and the vacuole of the cell. These two genes are potentially two excellent candidates, as transport into calcium sink organs is the final step of the journey in storing calcium in plants. Additionally, we identified a glutamate receptor gene on chromosome 5 and both the GORK gene and the MCA1 gene on chromosome 7. These three genes code for gated ion channels in the plasma membrane that allow for controlled influx of apoplastic calcium into the cytosol of the plant. These channels are also very important for achieving the final goal of calcium storage, as increasing the influx of calcium into the cytosol must occur first in order for the calcium pump genes, ECA1 and CAX2, to function. Together, these two types of genes work together to transport calcium from the apoplast to storage in the calcium sink organs.

 In addition to calcium pump and ion channel genes, we also identified several calcium binding genes. We located annexin on chromosome 2, calcium-dependent protein kinase (CDPK) on chromosomes 2, 4, 5, 6, 7, and 8, a calcineurin B-like gene on chromosome 7, an SOS3 gene on chromosome 7, and finally calmodulin genes on chromosomes 3, 4, 5, 7, and 8. We also located the unidentified EF-hand calcium binding protein on chromosome 6. While these genes do not code for calcium sequestering proteins that store away calcium, they do play a very important role in the grand scheme of calcium storage. Apart from annexin, all of these genes code for free floating cytosolic proteins that bind calcium molecules and act as second messengers to activate other genes. At first glance, it appears that these proteins would not be significant in the overall goal of calcium storage and accumulation. However, some of these proteins are actually responsible for activating other calcium storage proteins. For example, Sanders et al. report that calmodulin is responsible for directly activating ACA genes, a large family of calcium pump proteins (similar to ECA1) that are located in the plasma membrane, the vacuole, the endoplasmic reticulum, and the plastid (Sanders *et al.*, 2002). Calmodulin and the other calcium binding proteins could therefore prove important in indirectly increasing the amount of calcium transport and storage occurring in the cell.

 As a final step in assessing the likelihood of our candidate genes matching the QTLs, we referenced the TAIR database (www.arabidopsis.org), a database dedicated to research done on many aspects of *A. thaliana*. We used this database to verify that all of the abovementioned genes were not just expressed in the plant, but were specifically expressed in the florets of the plant, as this is the most important area of the plant when considering increasing stored calcium for human benefit. Using this tool, we verified that all candidate genes (except for ECA1) were indeed expressed in the florets of the plant, in addition to other areas. ECA1 is reported to be expressed primarily in the guard cells and the plant callus of plants, but we decided to report it in or results given that it’s a very significant gene in calcium transport regardless.

 Ultimately, there is much at stake in identifying the genes in *B. oleracea* that are most important to calcium uptake and storage. *B. oleracae* has previously been noted to be an excellent source of dietary calcium for humans (Heaney *et al.*, 1990) and shows potential, based on the data collected during this study, to be an accumulator of additional calcium to further supplement the human diet. Using both tBLASTn and IGB, we have compiled a comprehensive list of candidate genes in *B. oleracea* that could be targeted to increase calcium transport and storage. These candidate genes form a complete model for calcium transport and storage, including apoplastic Ca2+ efflux into the cytoplasm, cytoplasmic Ca2+ binding, signaling, and gene activation, and finally Ca2+ influx into calcium storing organelles. The SSRs that we have provided should also provide a more direct and quicker method for researchers to do additional research on these genes. Further research will need to be done to determine which of the candidate genes we have identified are correct matches to the different calcium related QTLs provided to us by our collaborators.

References

Ache, P., Becker, D., Ivashikina, N., Dietrich, P., Roelfsema, M., Rob, G., & Hedrich, R.,. (2000). GORK, a delayed outward rectifier expressed in guard cells of arabidopsis thaliana, is a K+-selective, K+-sensing ion channel.*FEBS Letters, 486*(2), 93-98.

Carnegie Institute for Science Department of Plant Biology. (2014). Tair. Retrieved May 7, 2014, Retrieved from www.arabidopsis.org

Conn, S. J., Gilliham, M., Athman, A., Schreiber, A. W., Baumann, U., Moller, I., . . . Leigh, R. A. (Jan 2011). Cell-specific vacuolar calcium storage mediated by CAX1 regulates apoplastic calcium concentration, gas exchange, and plant productivity in arabidopsis.*The Plant Cell, 23*(1), 240-257.

Dennison, K., & Spalding, E. (2000). Glutamate-gated calcium fluxes in arabidopsis.*Plant Physiology, 124*(4), 1511-4.

Genome database for vaccinium. (2014). Retrieved May 7, 2014, Retrieved from dev.vaccinium.org

 Halfter, U., Ishitani, M., & Zhu, J. (2000). The arabidopsis SOS2 protein kinase physically interacts with and is activated by the calcium-binding protein SOS3.*Proceedings of the National Academy of Sciences of the United States of America, 97*(7), 3735-3740.

Heaney, R., & Weaver, C. (1990). Calcium absorption from kale.*The American Journal of Clinical Nutrition, 51*(4), 656-7.

Hirschi, K. D. (2004). The calcium conundrum. both versatile nutrient and specific signal.*Plant Physiology, 136*(1), 2438-2442.

Luan, S., Kudla, J., Rodriguez-Concepcion, M., Yalovsky, S., & Gruissem, W. (2002). Calmodulins and calcineurin B–like proteins: Calcium sensors for specific signal response coupling in plants.*The Plant Cell, 14*, S389-S400.

Roderick, H. L., Lechleiter, J. D., & Camacho, P. (2000). Cytosolic phosphorylation of calnexin controls intracellular Ca2+ Oscillations via an interaction with SERCA2b.*The Journal of Cell Biology, 149*(6), 1235-1247.

Sanders, D., Pelloux, J., Brownlee, C., & Harper, J. F. (2002). Calcium at the crossroads of signaling.*The Plant Cell, 14*, 401-417.

Tuteja, N., & Mahajan, S. (2007). Calcium signaling network in plants: An overview.*Plant Signaling & Behavior, 2*(2), 79-85.

White, P., & Broadley, M. (2003). Calcium in plants.*Annals of Botany, 92*(4), 487-511.

Yamanaka, T., Nakagawa, Y., Mori, K., Nakano, M., Imamura, Tomomi, Kataoka, Hajime, Terashima, A., . . . Iida, H.,. (2010). MCA1 and MCA2 that mediate ca²⁺ uptake have distinct and overlapping roles in arabidopsis.*Plant Physiology, 152*(3), 1284-1296.