Genetic Analysis of the Function of Specific Small Heat Shock Proteins (sHsps) in Plants

An Honors Thesis

Presented by

Alyssa McQuillan

Completion Date: April 28th, 2017

Approved By:

____
Dr. Elizabeth Vierling, Biochemistry and Molecular Biology

____
Dr. Joohyun Lee, Biology

Genetic Analysis of the Function of Specific Small Heat Shock Proteins (sHsps) in Plants

By

Alyssa McQuillan

Approved for submittal to the Department of Biochemistry and Molecular Biology and Commonwealth Honors College for granting graduation with Departmental Honors:

Student
Ms. Alyssa McQuillan _______________________ Date ___________________________
(signature)

Committee Chair
Dr. Elizabeth Vierling _______________________ Date ___________________________
(signature)

Committee Member
Dr. Joohyun Lee _______________________ Date ___________________________
(signature)

Acknowledgments

My parents for providing me with never ending emotional support throughout each year of my undergraduate career at UMass Amherst.

Indu Santhanagoplan, Minsoo Kim, Thi Bui, and the members of the Vierling lab (past and present) for assisting me and answering all questions I had throughout, on top of being a great group of people to spend each day with.

Dr. Lee for providing insight and understanding into working with plants and how to maintain them.

Dr. Vierling for providing me with an incredible undergraduate research opportunity that has taught me how to excel not only in the lab, but also in life.

Table of Contents

Abstract ……………………………………………………………………………..….5

Introduction …………………………………………………………………..…..7

Chapter 1: A review of the Small Heat Shock Protein (sHsp) literature ………………....13

Introduction ……………………………………………………………………13

sHsp structure and ability to interact with other proteins.
Roles of sHsps in various forms of human disease.
Specific sHsps being analyzed in this project- Class I, Class II, Chloroplast-targeted, Mitochondrial-targeted.
Class I and Class II Cytosolic sHsps function and characteristics.
Chloroplast-targeted sHsps function and characteristics.
Mitochondrial-targeted sHsps function and characteristics.

Conclusion ……………………………………………………………………….21

Chapter 2: Generating the double knockout of class II sHsps.………………..……..…22

Materials and Methods
Results
Discussion
Conclusions

Chapter 3: Generating a triple mutant of all mitochondrial- or all chloroplast-targeted sHsps ……………………………………………………………………..…..37

Materials and Methods
Results
Discussion
Conclusions

References ……………………………………………………………..…………… 49

Abstract

Molecular chaperones are proteins that assist in the unfolding or refolding of other macromolecular structures. Small heat shock proteins (sHsps) are key chaperones that are found across many species and play a role in stress tolerance by preventing the irreversible aggregation of proteins that have become misfolded. When the protein quality control network is either overwhelmed with damaged proteins or is somehow defective, it often leads to various disease states. Understanding the mechanism of sHsps and their interactions with other chaperones has wide-ranging implications, including fully recognizing the roles these proteins actually play in cellular stress, as well as in disease processes. The goal of this project is to analyze the phenotypes of *Arabidopsis thaliana* plants carrying mutations in genes that code for specific sHsps. The first mutant will have both class II sHsps, Hsp17.6 and Hsp17.7, knocked out. Both genes will be knocked out via the CRISPR/Cas9 mechanism, or by starting with a T-DNA knockout of Hsp17.6 combined with a CRISPR/Cas9 mutation of Hsp17.7. The other mutants will eliminate the organelle-targeted sHsps in the mitochondria (Hsp26.5_MT, Hsp23.5_C/MT and Hsp23.6_C/MT) or chloroplasts (Hsp25.3_CP, Hsp23.5_C/MT and Hsp23.6_C/MT). These mutants will be created by crossing plants carrying single or double gene knockouts already available in the lab in order to create plants that are triple knockouts for all chloroplast or all mitochondrion-targeted sHsps, knocking out each entire class of sHsps. Understanding how these mutants behave and handle different stresses by generating null mutants of the genes encoding these proteins will provide a key to understanding how important sHsps are in these plants and in other organisms, including humans.

Abstract

Molecular chaperones are proteins that assist in the unfolding or refolding of other macromolecular structures. Small heat shock proteins (sHsps) are key chaperones that are found across many species and play a large role in stress tolerance by preventing the irreversible aggregation of proteins that have become misfolded. When the protein quality control network is either overwhelmed with damaged proteins or is somehow defective, it often leads to various disease states. Understanding the mechanism of sHsps and the interactions with other chaperones has wide ranging implications, including fully recognizing the roles these proteins actually play in cellular stress, as well as in disease processes. The goal of this project is to analyze the phenotypes of *Arabidopsis thaliana* plants carrying mutations in genes that code for specific sHsps. The first mutant will have both class II sHsps, Hsp17.6 and Hsp17.7, knocked out. Both genes will be knocked out via the CRISPR/Cas9 mechanism, or by starting with a T-DNA knockout of Hsp17.6 combined with a CRISPR/Cas9 mutation of Hsp17.7. The other mutant will eliminate the organelle-targeted sHsps for both the mitochondria and chloroplasts, identified as 26.5_MT, 25.3_CP, 23.5_C/MT and 23.6_C/MT. These mutants will be created by crossing plants carrying single or double gene knockouts already available in the lab in order to create plants that are triple knockouts for all chloroplast and all mitochondrion-targeted sHsps, knocking out out each entire class of sHsps. Understanding how these mutants behave and handle different stresses by generating null mutants of the genes encoding these proteins will provide a key to understanding how important sHsps are in these plants and in other organisms, including humans.

Introduction

Molecular Chaperones

Molecular chaperones are proteins that assist in the unfolding or refolding of other macromolecular structures. These proteins evolved as cells needed ways to protect themselves from stress conditions (Haslbeck & Vierling 2015). They interact with either completely unfolded proteins, partially folded proteins, or even nascent chains that are emerging from the ribosome, to ensure stable protein structure and function. Many chaperones are essential for viability and are also highly expressed during times of cellular stress (Haslbeck & Vierling 2015).

Small Heat Shock Proteins (sHsps)

Small heat shock proteins (sHsps) are key chaperones that play a role in stress tolerance and that are present in all three domains of life (Haslbeck & Vierling 2015). These proteins prevent the irreversible aggregation of proteins that have become misfolded and therefore nonfunctional. sHsps are proposed to serve as the first line of defense against various stressors and are considered a critical component of the cellular "protein quality control network" (Basha, O'Neill & Vierling, 2012). It is extremely important that protein homeostasis is maintained in cells, and through the interactions of sHsps with other molecular chaperones and proteases, accumulation of damaged proteins is prevented and homeostasis is sustained.

sHsps in Human Disease

There has been considerable research on the role of sHsps in human health. When the protein quality control network is either overwhelmed with damaged proteins or is somehow defective, it often leads to diseased states (Basha, O'Neill & Vierling, 2012). As molecular

chaperones, sHsps play a role in protecting protein structure and function, therefore preventing disease. When sHsps are mutated or perturbed in any way, this can contribute to cellular malfunction. For example, mutated sHsps have been discovered to be genetically linked to diseases such as myopathy and neuropathy, and their expression is associated with other neurological disorders (Basha, O'Neill & Vierling, 2012). With the number of disease states in which these chaperones play a role, there is a significant potential for therapeutic intervention. Understanding the mechanism of sHsps and their interactions with other chaperones has a wide ranging implication, including fully recognizing the roles these proteins play in cellular stress, as well as in disease processes (Basha, O'Neill & Vierling, 2012). Using plants, in particular *Arabidopsis thaliana,* we can examine these proteins and how exactly they are expressed and function under times of stress.

sHsps in Plants

The reason *Arabidopsis thaliana* was utilized as a model organism is due to the similarity of their sHsps to other eukaryotes. Land plants have sHsps that are found not only in the cytosol, but also in essentially all membrane-bound organelles. The sHsps found in various organelles act similarly to the other Hsps present throughout the cell (Haslbeck & Vierling 2015). In addition, the mechanism by which cytosolic sHsps are shuttled into the nucleus is seen under specific circumstances in almost all eukaryotic organisms (Haslbeck & Vierling 2015). Once plants evolved to live on land, they were confined to a sessile lifestyle, which may have been the reason why specific chaperones evolved to ensure the protection of proteins in all cellular compartments, making sHsps almost a universal aid in the protein misfolding process.

The *Arabidopsis thaliana* genome has a total of nineteen sHsps, which includes cytosolic

and nuclear proteins, as well as proteins that are targeted to the chloroplasts, mitochondria, peroxisomes and endoplasmic reticulum (Waters, Aevermann, & Sanders-Reed, 2008). The sHsps that are targeted to organelles only exist in plants, with varying functions compared to the cytosolic sHsps. With there being many different families of sHsps, the focus for this project will be the Class II cytosolic sHsps, chloroplast-targeted sHsps, and mitochondrion-targeted sHsps. There is evidence that these proteins are required for plants to tolerate heat stress, mostly from gain-of-function experiments in which sHsps have been over-expressed. When a plant is subjected to high temperatures, a large shift in gene expression is induced (Vierling et al., 1988). The genes encoding the sHsps are quickly transcribed to ensure that these proteins will be there to prevent the aggregation of proteins that are denatured by the heat stress (Vierling et al, 1988). Numerous studies have supported this conclusion, due to the high number of specific sHsp transcripts that are produced under abiotic and biotic stress conditions in addition to being expressed during specific stages of normal development and in organ specific patterns (Giorno, 2010). Not much is known about how a plant would handle stress if sHsps were not produced. To test for this, it is necessary to generate null mutants of the genes encoding these proteins and then determine resulting plant phenotypes under optimal or stressful growth conditions. There are different mechanisms that can be utilized to successfully knock out these genes to produce the mutants of interest.

Goals of this Project

The goal of this project is to analyze the phenotypes of *Arabidopsis thaliana* plants carrying mutations in genes that code for specific sHsps. The first mutant will have both class II sHsps, Hsp17.6 (At5g12020) and Hsp17.7 (At5g12030), knocked out. Both genes will be

knocked out either via the CRISPR/Cas9 mechanism, or starting with a T-DNA knockout of Hsp17.6 combined with CRISPR/Cas9 mutation of Hsp17.7. The other mutants will be combinations of knockouts, in which the sHsps are eliminated from either the mitochondria or chloroplasts. These genes are identified as mitochondrion-targeted 26.5_MT (At1g52560), chloroplast-targeted 25.3_CP (At4g27670), and dual mitochondrion- and chloroplast-targeted 23.5_C/MT (At5g51440) and 23.6_C/MT (At4g25200). These mutants will be created by crossing plants carrying single (26.5_MT or 25.3_CP) or double gene (23.5_C/MT, 23.6_C/MT) knockouts already available in the lab in order to create plants that are triple knockouts for all chloroplast or all mitochondrion-targeted sHsps. In the absence of these organelle sHsps, one goal will be to see how these plants will tolerate heat stress.

The mutant being generated by the CRISPR/Cas9 mechanism is a great candidate for deletion due to the close proximity of the genes encoding the class II sHsps. CRISPR/Cas9 is a unique technology that has enabled researchers to edit specific parts of the genome by either inserting, deleting, or knocking out parts of the DNA sequence (Feng et al., 2014). This system consists of two key molecules that are responsible for introducing the change in the DNA. The first molecule is an enzyme called Cas9. This enzyme cuts the double stranded DNA at a specific location of interest in the genome so that bases can be either added or deleted (Feng et al., 2014). The other molecule is a piece of RNA called the guide RNA (gRNA). Each gRNA is a pre-designed sequence that ‘guides’ Cas9 to the correct part of the genome that is being targeted (Feng et al., 2014). This is a way of ensuring that the DNA is being cut in the right place in the genome. The gRNA sequence dictates specificity due to the fact that it is designed to bind only to the target sequence and nowhere else in the genome (Feng et al., 2014). This makes it possible to create a CRISPR vector with guide RNAs to target each gene that encodes a class II sHsp. It

is, however, still difficult to obtain a double gene knockout using this method so another strategy was also initiated. A SALK T-DNA insertion line (SALK_086201.25.80) targeting one of the genes, Hsp 17.6, was ordered and genotyped to identify a homozygous mutant with this single gene knocked out. From there, a CRISPR vector with gRNAs targeting the other gene, Hsp 17.7, was created and used to transform the SALK plants carrying the mutation for Hsp 17.6 knocked out via *Agrobacterium tumefaciens*- mediated transformation.

Crossing plants is a way for researchers to combine genotypes and to assess the transmission of genes from generation to generation. Various mutations can be combined through this process, allowing for numerous outcomes. Once plants with a desired mutation have been identified, the next step is "selfing." This means allowing plants to self-pollinate followed by obtaining seeds, then growing up the next generation, and finally rechecking each plant to ensure they are the desired homozygous mutants. In order to combine different mutations, the next step involves creation of the F1 generation between the two mutant parents of interest. The plant to be used as the female parent is emasculated, i.e. the anthers are removed before maturity and the shedding of pollen. This step is to prevent self-pollination, ensuring that when the cross is made it is purely from each plant of interest. Next, pollen from the desired male parent is dusted onto the stigma of the female plant and seed maturation is allowed to proceed. These F1 seeds are then grown into plants, which can be genotyped to endure the cross occurred and then are allowed to self, producing the F2 generation of seeds. Plants grown from the F2 seeds are then screened for all of the possible mutations through genotyping by PCR.

Future Directions

The mutants of interest being produced for this project have never been made before.

They are specifically targeting key genes that encode significant plant sHsps. Once these sHsps are no longer present or being expressed under times of stress, other mechanisms of handling stress can be investigated. When confirmed homozygous mutants have been isolated, plants will first be assessed for general growth phenotypes compared to wild type, such as time of flowering, leaf shape, chlorophyll content and seed production. Heat stress assays will be performed on these mutants in order to analyze how certain temperatures affect them. These assays will consist of treating different aged seedlings for a period of time with a high, but survivable temperature followed by the regular growing temperature to allow plants to recover (Larkindale et al., 2005). If the phenotypes of these mutants allow them to survive, mass spectrometry can also be utilized to look further into what other proteins or molecules are aiding the refolding of proteins that have been denatured. This information can be particularly useful for designing new therapies for human disease with the knowledge that is gained from the mechanisms at play by the sHsps. Understanding how these mutants behave and handle different stresses will provide insight into how organisms can survive without these proteins, if they can at all. Ultimately, these mutants will provide a key understanding into how important sHsps are in these plants and in life.

Chapter 1: A review of the Small Heat Shock Protein (sHsp) literature

Introduction

Small heat shock proteins (sHsps) are ubiquitous proteins whose expression is largely determined by a variety of stressors in the environment. They exist and function constitutively in a number of cell types in multiple different organisms (Sun & MacRae 2005). sHsps are proposed to act as ATP-independent molecular chaperones by taking on the role of binding to proteins that are unfolding due to various stressors and preventing damage to cells by not allowing these denatured proteins to form aggregates (Sun & MacRae 2005). The sHsps have a diverse evolutionary history, have been shown to be linked to specific human diseases, and have interesting, unique protein dynamics. For these reasons, sHsps are of great interest to researchers in biology, biochemistry and medicine (Sun & MacRae 2005).

sHsp structure and ability to interact with other proteins.

Monomers of sHsps consist of an N-terminal domain, a conserved alpha-crystallin domain (ACD) of approximately 90 amino acid residues, and a C-terminal extension (Sun & MacRae 2005). Each one of these sHsp domains has specific proposed roles in function of the protein. sHsps undergo transitions between mono- and poly-dispersed oligomers where each of these forms have different binding affinities and rates of disassembly that has an effect on chaperone activity (Sun & MacRae 2005). The ACD is of particular importance in the stability of sHsps. Within this region, there are several beta-strands that are organized into beta-sheets which is important in the formation of the dimer state of the sHsp. The dimer is the building block of most sHsps (Sun & MacRae 2005). The N-terminal domain has a strong role in the formation of oligomers from monomers in addition to substrate binding. The C-terminal extension also plays a role in the oligomerization and in promoting solubility and chaperone activity (Sun & MacRae

2005). It is the cooperation of these three domains of the sHsp that allows it to function.

While some sHsps exist as monomers, most actually exist as oligomers. The monomeric masses of sHsps of diverse organisms range from 12 to 42 kiloDaltons (kDa), but in their native state they typically assemble into multimers of 12 to over 32 individual monomers (Basha, O'Neill & Vierling, 2012). The idea behind the mechanism of how sHsps work proposes that the dimeric subunits expose themselves to the stressed cellular proteins once the sHsp becomes activated (Figure 1). There is a shift in equilibrium to the dimeric form when the system is stressed because of the dimer's ability to bind misfolded and unfolded proteins. It is the oligomeric form that acts as a reservoir of the active dimeric units of the sHsp, aiding in the stress response (Stengel et al., 2010; Van Montfort et al., 2001b). The oligomers essentially sequester the binding sites for denaturing proteins from the outside cellular contents until these sites are needed under stress.

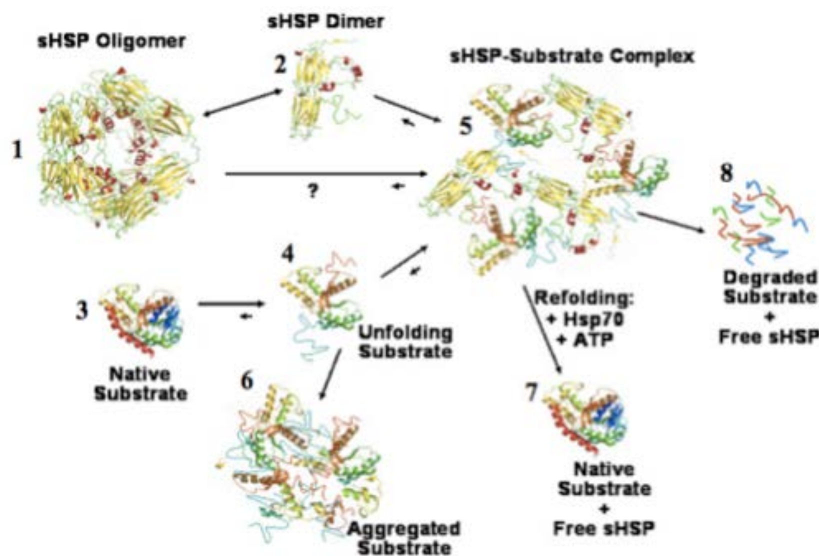


Figure 1. Role of sHSPs in the refolding or degradation of misfolded proteins. sHSPs are believed to act as oligomers in their native state. Once they are activated by any kind of stress, sHSPs dissociate into active dimer species to aid in the cellular stress response (Jaya, 2009).

Roles of sHsps in various forms of human disease.

There has been considerable research on the link between sHsps and human disease since they were first suspected to play a role in various forms of disease. Hsps in general were discovered in the 1960s accidentally in Italy during genetic research on *Drosophila* fruit flies. These flies were exposed to higher temperatures compared to the normal environmental conditions, making the activation of Hsps apparent. From this moment on, extensive research has been conducted to look at the significant roles Hsps could be playing in various homeostatic processes of living organisms. sHsps appear to protect cells from many different conditions. They are key in helping maintain homeostasis when the cells are disturbed by either heat, oxidative stress, heavy metals, or ischemic injury (Basha, O'Neill & Vierling, 2012). Currently, there is a gap in knowledge about how exactly defective sHsps play a role in these disease states. Certain sHsps are required for lens clarity in the eye, and when mutated lead to cataracts. Additionally, there is a high level of sHsps in muscle tissues, so when sHsps are mutated it is particularly detrimental because it results in both cardiac and skeletal myopathies. Another disease state that sHsps play a role in is inherited neuropathies (Basha, O'Neill & Vierling, 2012). The mutations in the sHsps that have been linked to disease are usually dominant and result in changes in the amino acid sequence. These changes occur in all three domains of the protein, the N-terminal arm, the alpha crystallin domain (ACD), and the C-terminal extension (Basha, O'Neill & Vierling, 2012). The effects of the dominant mutations are likely a result of disrupting sHsp structure. Changes in structure can alter the way in which the sHsp interacts with damaged or native substrate proteins. For example, there is a single position in the ACD that is altered in many of the disease-linked sHsps (Arg120 in human HspB5 and corresponding positions in other sHsps) (Basha, O'Neill & Vierling, 2012). It is this altering in structure caused

by a dominant mutation that is responsible for disrupting the protein's interactions with other cellular molecules. This can potentially lead to blocking essential cellular processes (Basha, O'Neill & Vierling, 2012). The therapeutic potential of manipulating sHsps has only recently been explored, and may prove useful for ameliorating disease states.

Specific sHsps being analyzed in this project- Class I, Class II, Chloroplast-targeted, Mitochondrion-targeted.

Now that it is clear how versatile sHsps are and how much of a role they could play in various forms of disease, it is important to clarify which sHsps will be investigated in this project. There are eleven different families of nuclear-encoded plant sHsps (Waters & Vierling, 1999). During heat stress in plants, these proteins are mass produced such that they make up a large portion of the total protein made in response to the stress. Plants have sHsps that localize to different parts of the cell. There are several plant sHsp families localized to the cytosol. Class I (CI) cytosolic sHsps, which have served as the main model in the mechanistic studies of how sHsps prevent the irreversible aggregation of proteins, in addition to four more classes designated Class II (CII) through Class V (CV), make up the classes of sHsp in the cytosol (Santhanagopalan et al., 2015). There are also genes that encode organelle-targeted sHsps, one plant sHsp family is localized to the chloroplasts, the mitochondria, the peroxisomes, and the endoplasmic reticulum (Waters, 2013). Out of these families of sHsps, the ones of relevance to this study are the class II cytosolic sHsps, the chloroplast-localized sHsps, and the mitochondrial-localized sHsps.

All three of these families of sHsps exhibit similar expression in *Arabidopsis thaliana*. The eFP browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) contains expression profiles of a majority of genes in the *Arabidopsis* genome. Looking at each individual sHsp gene, it is clear

that in the absence of stress, they all have the highest expression in dry seeds. In all other parts of the plant, there appears to be little to no expression. The only other area among these different families of sHsps that have similar expression levels is in flowering stage of the plant. These similarities among the sHsps provides an additional reason as to why they are being looked at collectively.

Class I and Class II Cytosolic sHsp function.

The CI and CII cytosolic sHsps are the most well characterized sHsps involved in the stress response, becoming highly abundant during times of heat stress. The CI sHsps are actually produced at a great rate in heat stressed cells, estimated to being close to 1% of the total cell protein, while the CII sHsps also accumulate, but not as much as the CI, up to about 0.25%, making both of these classes combined over 1% of total cell protein just within a few hours of the stress (Derocher et al., 1991). Through many experiments, it was discovered that both the CI and CII sHsps form dodecameric oligomers only with members of the same class. The classes do not associate with each other to form heterooligomers, but they heterooligomerize with other proteins in that class from either the same or different species. So it is evident that these classes of proteins are structurally distinct, but the way in which each functions still needs to be explored. Looking at studies in vitro, the behavior of the CI and CII appears to be distinguishable, but more evidence is needed. The CII sHsps did not dissociate into stable dimers upon heat stress, when using the same assays use to test the CI sHsps. It is suspected that the role of CI sHsps is its importance in substrate protection (Basha et al., 2010). Further tests were done utilizing polyacrylamide gel electrophoresis and mass spectrometry which revealed the CII sHsps do exchange subunits at higher temperatures, demonstrating they also can expose

hydrophobic substrate binding sites (Santhanagopalan et al., 2015).

Being one of the largest families of sHsp genes, there are six CI genes in *Arabidopsis thaliana*. The gene family for the CII sHsps is smaller, with there being only two genes encoding this class. With there being no CI *Arabidopsis* T-DNA insertion lines, most likely attributed to the small size of these intron-less genes, and the chromosomal arrangement of the CII *Arabidopsis* genes, a problem for genetic analysis of mutants exists. Limited studies have defined the function of these cytosolic sHsps, providing the need to generate the mutants of interest in order to develop more knowledge about at least one of the cytosolic classes (CII) of sHsps.

In order to show the similarity of the two Arabidopsis Class II sHsps, their amino acid sequences were aligned (Figure 2). There are many conserved areas throughout each of these sequences, showing how they are derived from similar arrangements of amino acids. This could indicate these sHsps serve redundant functions, such that deletion of both genes would be required to observe a phenotype in mutant plants.

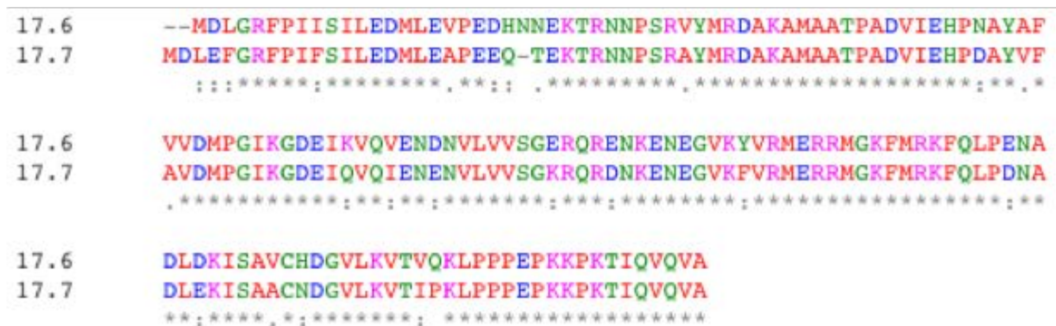


Figure 2. Alignment of the protein sequences of Hsp17.6 (At5g12020) and Hsp17.7 (At5g12030). Each of the amino acid sequences were aligned and showed conserved regions. The stars indicate an exact match, the single dot indicates an amino acid with some similar properties, and two dots indicate a different amino acid that is found as a frequent substitution with similar properties. The colors indicate amino acids that are members of the same classes, making them characteristically similar.

Chloroplast-targeted sHsp function and characteristics.

The increase of sHsps in response to high temperature stresses provides some evidence that these chaperones contribute to the development of thermotolerance in eukaryotes. But the exact mechanism by which the chloroplast-targeted sHsps do so is still being explored. The homology between the cytosolic family of sHsps and the chloroplast-targeted sHsps suggests the possibility of a common evolutionary ancestor (Vierling et al, 1988). This relationship could indicate that the chloroplast Hsps may have similar function to the cytosolic Hsps. So far, most studies on these organelle specific sHsps have analyzed HSP21, which was the first chloroplast-targeted sHsp discovered and though to have potential for playing a role in the development of thermotolerance in plants. Chloroplast-targeted sHsps are not expressed under normal temperature conditions, so they are not constitutively expressed, but they do rise to levels of detection in both the leaves and roots after heat stress (Suzuki et al., 1998). HSP21 is a nuclear-encoded protein that is targeted to the chloroplast through its amino-terminal transit peptide. In its native state, the protein is a large oligomer composed of twelve or more HSP21 subunits, often times seen as a dodecamer, which is likely to exchange subunits during heat stress (Rutsdottir et al., 2017). In addition, phosphorylation of sHsps generally disrupts oligomer formation, which has an unknown regulatory role in sHsp function of mammalian cytosolic sHsps. Phosphorylation is not seen so far in plant sHsps, except for one instance in the mitochondrion. HSP21 does not appear to be regulated by phosphorylation, but this is an interesting property that would require further research in order to gain knowledge about this protein's basic properties and what the possible functional consequences of the modification could be.

Mitochondrion-targeted sHsp function and characteristics.

Proteins have also been identified in the family of sHsps that localize to mitochondria. In addition to the chloroplast, the mitochondrion is another powerhouse in plant cells and is a significant regulator of stress responses. To understand better the regulation of mitochondrial responses, it became apparent that a variety of nuclear genes encoding mitochondrial proteins are transcribed and produced in order to handle a large range of stress conditions. These nuclear genes have been identified as alternative oxidases (AOXs), NAD(P)H dehydrogenases (NDs), and finally HSPs (Van Aken et al., 2009). All of these genes have proven to be strongly induced by many different stresses. Specifically, when a stress occurs and the effect is a change in gene expression of mitochondrial proteins, it is possible that organelle function can be directly modified. This means that it can directly lead to retrograde signaling from the organelle to the nucleus (Van Aken et al., 2009). Organelle function can also be indirectly targeted, and this would result in the stress not actually affecting the organelle at all, which leads to anterograde signaling, so from the nucleus to organelle (Van Aken et al., 2009). Looking at mainly the sHsps and their direct role in the mitochondrial stress response, two homologous sHsps known as HSP23.5 and HSP23.6 were identified. Because these two proteins are co-expressed, it is likely that they form a functional pair in the stress response and that they are actually both necessary for stabilizing mitochondrial proteins (Van Aken et al., 2009). It is the responsiveness of these mitochondrial-targeted sHsps to the large variety of stressors that indicate it is of the utmost importance to stabilize and correctly fold mitochondrial proteins exposed to unfavorable conditions (Van Aken et al., 2009). Knowing their critical role in the mitochondrial stress response makes these sHsps particularly important to study.

Aligning the protein sequences of both the chloroplast-targeted and mitochondrial-

targeted sHsps from Arabidopsis provides additional information on their relationship. This alignment is shown in Figure 3. There appear to be conserved areas of amino acids, but there are also a few spaces where the amino acids are not the same. The conserved regions indicate a possible common ancestor, but the differences show how each of these sHsps have evolved for each of the organelles. Each organelle is responsible for different, key functions which could explain how they evolved differently.

Figure 3. Alignment of the protein sequences of the chloroplast-targeted sHsp, 25.3_CP (At4g27670), and mitochondrial-targeted sHsp, 26.5_MT (At1g52560). The two amino acid sequences were aligned and showed some conserved regions as well as some differences. The stars indicate an exact similarity between amino acids, the single dot indicates a completely different kind of amino acid, and two dots indicate a different amino acid but it belongs to the same class of amino acids.The colors indicate amino acids that are members of the same classes, making them characteristically similar.

Conclusion

Understanding these key families of sHsps can help to fill the knowledge gap of how important these proteins are in life. Each of the sHsps being analyzed has a purpose for this study. Whether they have similar function to sHsps in mammalian cells or have been repeatedly demonstrated to play a key role in stress responses, the information gathered can provide insights

on the mechanisms aiding in the stress response. This information can then be applicable to identifying therapeutic approaches or designing other mutants to further analyze the function of other families of sHsps. It is the structure of these proteins that allow them to carry out their functions in cells. The ability of sHsps to bind to proteins that are actively unfolding such that no irreversible aggregates are formed is truly a novel process with a lot of potential. The aim of analyzing these three separate families of proteins and knocking out the genes encoding them can provide researchers with a lot of information about the role they previously may have served and how organisms can survive without them, if they are able to at all.

Chapter 2: Generating the double knockout of the class II sHsps

Material and Methods

Generating the CRISPR Construct

Using the protocol created by QiJun Chen of China Agricultural University, College of Biological Sciences (Xing et al., 2014), the possible CRISPR/Cas vectors were evaluated for construction of a vector to knockout the Arabidopsis Class II genes, Hsp17.6 (At5g12020) and Hsp17.7 (At5g12030). It was decided that the PHEC401 vector would be used, because it is expressed in the egg cell, making it easier to obtain homozygous mutants. In addition, pCBCDT (pCBC-DT1T2) is the vector that is used to link two target sequences. These sequences are recognized by the gRNA-Cas9 and the specificity of each of these gRNA spacer sequences is important due to the fact these allow for the gRNAs to bind the DNA in the right location.

First, the target site on both sides the gene of interest needed to be determined using the website http://www.genome.arizona.edu/crispr/CRISPRsearch.html. Next, primers were designed based off of this target sequence, shown in Table 1.

Table 1: Primers with specific features for generating the CRISPR construct

	Primer Name	Sequence
Set One	DT1-BsF_177b	5'ATATATGGTCTCGATTGATCTCTGCGGCTTGTAATGAGTT
Set One	DT1-F0_177b	5'TGATCTCTGCGGCTTGTAATGAGTTTTAGAGCTAGAAATAGC
Set Two	DT2_R0_177A	AACGGTGCTCGATAACGTCAGCTCAATCTCTTAGTCGACTCTAC
Set Two	DT2_BsR_177A	ATTATTGGTCTCGAAACGGTGCTCGATAACGTCAGCTC
Set Three	DT2_R0_176A	AACCGACGAATGCATATGCGTTACAATCTCTTAGTCGACTCTAC
Set Three	DT2_BsR_176A	ATTATTGGTCTCGAAACCGACGAATGCATATGCGTTAC

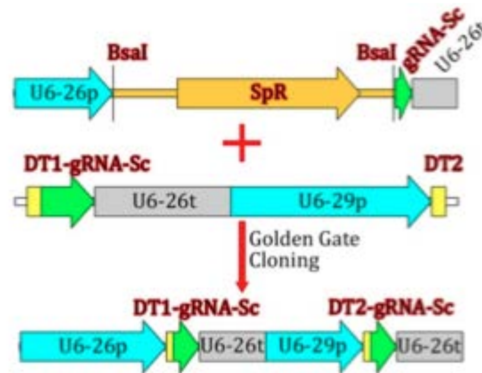
Next, PCR amplification was conducted using 100-fold diluted pCBC-DT1T2 as template in the four-primer PCR amplification. Specifically, the -BsF/-BsR primers were used at normal primer concentration (final concentration of 0.2uM) and the -F0/-R0 primers were used in a 20-fold dilution. Set one and set two (Table 1) were the primers needed to make the Hsp17.7 knockout. The PCR product was then purified and put through a restriction-ligation system setup as follows in Table 2.

Table 2. Master Mix for creating construct of interest

Component	Vol (µL)	Conditions
PCR fragment (626-bp)	2	5 hours at 37°C 5 min at 50°C 10 min at 80°C
pHEC401	2	
10x NEB T4 Buffer	1.5	
10x BSA	1.5	
BsaI (NEB)	1	
T4 Ligase (NEB) / High concentration	1	
ddH_2O	6	
Total	15	

After these products were obtained, they were sent out for sequencing to confirm that the construct contains everything it needs to, including the gRNAs and target sequence. This product was then transformed into bacteria.

The overall strategy employed Golden Gate cloning, which is imaged in Figure 4. This method compiles multiple inserts that get assembled into a vector backbone using restriction enzymes and the T4 DNA ligase.



Red: Primer Sites

[U6-26p-gRNA-U6-26t]-[U6-29p-gRNA-U6-26t]

```
AAGCTTCGACTTGCCTTCCGCACAATACATCATTTCTTCTTAGCTTTTTTTCTTCTTCTT
CGTTCATACAGTTTTTTTTTGTTTATCAGCTTACATTTTCTTGAACCGTAGCTTTCGTTT
TCTTCTTTTTAACTTTCCATTCGGAGTTTTTGTATCTTGTTTCATAGTTTGTCCCAGGAT
TAGAATGATTAGGCATCGAACCTTCAAGAATTTGATTGAATAAAACATCTTCATTCTT
AAGATATGAAGATAATCTTCAAAAGGCCCCTGGGAATCTGAAAGAAGAGAAGCAGG
CCCATTTATATGGGAAAGAACAATAGTATTTCTTATATAGGCCCATTTAAGTTGAAAA
CAATCTTCAAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCTGAGTTTATATACA
GCTAGAGTCGAAGTAGTGATTGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAA
ATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG
GTGCTTTTTTTTGCAAAATTTTCCAGATCGATTTCTTCTTCCTCTGTTCTTCGGCGTTCA
ATTTCTGGGGTTTTCTCTTCGTTTTCTGTAACTGAAACCTAAAATTTGACCTAAAAAAA
ATCTCAAATAATATGATTCAGTGGTTTTGTACTTTTCAGTTAGTTGAGTTTTGCAGTTC
CGATGAGATAAACCAATATTAATCCAAACTACTGCAGCCTGACAGACAAATGAGGAT
GCAAACAATTTTAAAGTTTATCTAACGCTAGCTGTTTTGTTTCTTCTCTCTGGTGCACC
AACGACGGCGTTTTCTCAATCATAAAGAGGCTTGTTTTACTTAAGGCCAATAATGTTG
ATGGATCGAAAGAAGAGGGCTTTTAATAAACGAGCCCGTTTAAGCTGTAAACGATGT
CAAAAACATCCCACATCGTTCAGTTGAAAATAGAAGCTCTGTTTATATATTGGTAGAG
TCGACTAAGAGATTGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAA
GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
TTTTGCAAAATTTTCCAGATCGATTTCTTCTTCCTCTGTTCTTCGGCGTTCAATTTCTGG
GGTTTTCTCTTCGTTTTCTGTAACTGAAACCTAAAATTTGACCTAAAAAAAATCTCAA
ATAATATGATTCAGTGGTTTTGTACTTTTCAGTTAGTTGAGTTTTGCAGTTCCGATGAG
ATAAACCAATAAGCTT
```

Figure 4. CRISPR Construct generated through Golden Gate Assembly. This shows how the construct was designed with the PHEC401 vector. Target sequences are denoted in yellow, primer sites are in red, guide RNAs primer sites are in green, and the promoters are denoted by light blue. This shows the general assembly of the sequences designed, both target and guide RNA, and how they were incorporated into the chosen vector.

<u>*Agrobacterium tumefaciens* mediated transformation</u>

The plasmid containing the CRISPR vector was added to 50µL of transformation-competent GV3101 *Agrobacterium* cells, made by Damian Guerra, already available in the lab in a 1.5mL tube. Next this mixture of plasmid and *Agrobacterium* was transferred into a cuvette, without transferring bubbles, and was chilled for 20 min at 20°C. The ECM 399 (BTX,

Hawthorne, NY), which is an electroporator, was utilized to permeabilize the *Agrobacterium* cell membranes and allow plasmid DNA to incorporate itself into the bacterium. The electroporator was set to 1800V and the cuvette was snapped into the right position, in the right orientation. The sample received constant pulses for about 5 msec. Immediately after the shock, 1 mL LB was added to cuvette and the mixture was pipetted up and down to mix. It was then transferred to a 1 mL microfuge tube and incubated at 28°C, 200 rpm for about 2-3 hours before 20 µl was plated on LB+50 KAN+20 Gent+ 20 Rif plates, where the concentration of each antibiotic is noted in µg/mL. Gentamycin and rifampicin are necessary for *Agrobacterium* selection and kanamycin selects for the introduced plasmid. Finally the plates were wrapped with micropore tape and incubated at 28°C for 2-3 days. The final step was to ensure that the colonies have taken up the vector by conducting colony PCR using primers given by QiJun Chen shown in Table 3 (CRISPR/ Cas9 vector protocol). The PCR is set up using a master mix containing 10X Standard Taq reaction buffer used in a 1x concentration, dNTPs used at 200uM, each primer was at a final concentration of 0.2uM, and 1.25 units of Taq polymerase was added per 50 uL of PCR cocktail. The colonies themselves served as the template DNA in the reaction.

Table 3. Primers for colony PCR & sequencing:

Primer Name	Sequence
U626-IDF	TGTCCCAGGATTAGAATGATTAGGC
U629-IDF	TTAATCCAAACTACTGCAGCCTGAC
U629-IDR	AGCCCTCTTCTTTCGATCCATCAAC

For introduction into plants, *Agrobacterium* colonies were grown in overnight culture and infiltration media was added directly to the culture (5% sucrose, 0.44mM 6-benzyladenine, 0.3%

silwet) for dipping wild type and the SALK Hsp17.6 T-DNA insertion plants that had been previously genotyped to find the homozygous mutant lines. The plants had been grown in small pots, 6x6 cm in size, filled with autoclaved soil with one plant on each corner. They had been given time to grow, while also cutting back on the number of times the plants were watered so that when they were dipped they would absorb the agrobacterium more readily. The plants were dipped in the *Agrobacterium* solution, submerging all the aerial parts of the plant, and returned to the growth chamber to allow them to grow.

Seed sterilization and plant maintenance

Seeds were harvested from the transformed plants to be screened for the gene insertion. Seeds were surface sterilized with 70% ethanol and 50% bleach for up to 10 min, and then washed up to 6 times with sterile water, discarding the water each time. Washing the seeds with ethanol and detergents is performed to kill bacteria or other organisms on the surface of the seeds. Next, the seeds are transferred to plates with plant PNS media. PNS media consists of 0.8% plant media grade agar, 0.5% sucrose, 5mM KNO_3, 2mM $MgSO_4$, 2mM $Ca(NO_3)_2 4H_2O$, 50uM FeEDTA, 2.5mM KPO_4 (pH 5.5), 70uM H_3BO_3, 14uM $MnCl_2$, 0.5uM $CuSO_4$, 1uM $ZnSO_4$, 0.2uM Na_2MoO_4, 10uM NaCl, and 0.01uM $CoCl_2$. PNS media was made in increments of 500 mL and then autoclaved for 30 mins before pouring 15-20 mL of the media into the plate, which were 10 cm in diameter. The antibiotic hygromycin was added to the media due to the fact that this resistance marker is included in the CRISPR vector, and this will demonstrate the plants that have taken up the agrobacterium and therefore the vector. The plates with seeds are placed in the cold (4°C) for 2-3 days. This allows the seeds to imbibe water and promotes more synchronous germination when removed to room temperature. The plates are then transferred to

the growth chamber set to 22°C at 41% light intensity where they will germinate and start to grow. The plants that grow have the hygromycin resistance gene, which means they are likely plants that have been transformed by the agrobacterium containing the CRISPR construct targeting Hsp 17.7.

Once they produced their first true leaves after about one week, plants were transferred to autoclaved soil available in the lab with four plants to a pot, one in each corner. The soil was moistened with water prior to planting each of the small seedlings. The growth chamber was set to growing conditions of 12 h days at 22°C at 41% light intensity, and then 12 h nights with no lights. From here, experiments will be done using their genomic DNA to test for mutations in the search for the knockouts. This procedure will be in a constant cycle as more plants grow and are discovered to be wild type, this will open up space for more seeds to be grown in the hopes one of them contain the knockout. These plants need to be monitored for their growth and receive appropriate watering and fertilizer.

Genomic DNA Extraction

In order to determine the genotype of the plants of interest, genomic DNA needs to be extracted from each plant sample. From each individual plant, 1.5-2 cm sized leaves were obtained and placed into a 1.5mL centrifuge tubes. These leaves were ground with a blue pestle (Sigma Aldrich, St. Louis, MO, USA) for about 10 sec and then 500 µL of Edwards Extraction buffer (200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) was added. The samples were reground and then vortexed to ensure homogeneity, leaving all preps at room temperature until all were completed. Next, these samples were spun down in the microcentrifuge for two min and then 300 µL of each prep was moved to a fresh tube to avoid all

of the plant matter that could contaminate the genomic DNA. A total of 310 µL of isopropanol was added and the tubes were inverted to mix, then samples were left for two min at room temperature. Finally, the tubes were centrifuged for 7 min and the supernatant was discarded, allowing the pellet time to air dry. After about an hour, the pellets were resuspended in 100 µL of 0.1X TE (25mM Tris-HCl pH8.0, 1mM EDTA), then vortexed and finally placed in the -20°C freezer for storage.

Genotyping Hsp17.6 SALK line T-DNA insertion mutants via Polymerase Chain Reaction

To genotype the plant samples, I performed a series of PCR using primers specific for the Hsp 17.6 gene and the T-DNA insertion. This reaction aided in identifying the presence of a knockout by determining whether each plant is homozygous wild type, heterozygous or homozygous mutant.

The gene for Hsp 17.6 (At5g12020.1) was further investigated using the website http://signal.salk.edu/cgi-bin/tdnaexpress to visualize all the possible T-DNA insertion lines within it that are available from the Salk Institute and other sources. Two or three T-DNA insertion lines were identified for the gene, but ultimately the point at which the T-DNA is being inserted into the gene was looked into and resulted in identifying one line with an insertion in the gene exon. The SALK line (SALK_086201.25.80) was ordered through the Salk Institute using the same website, and is noted in Figure 4.

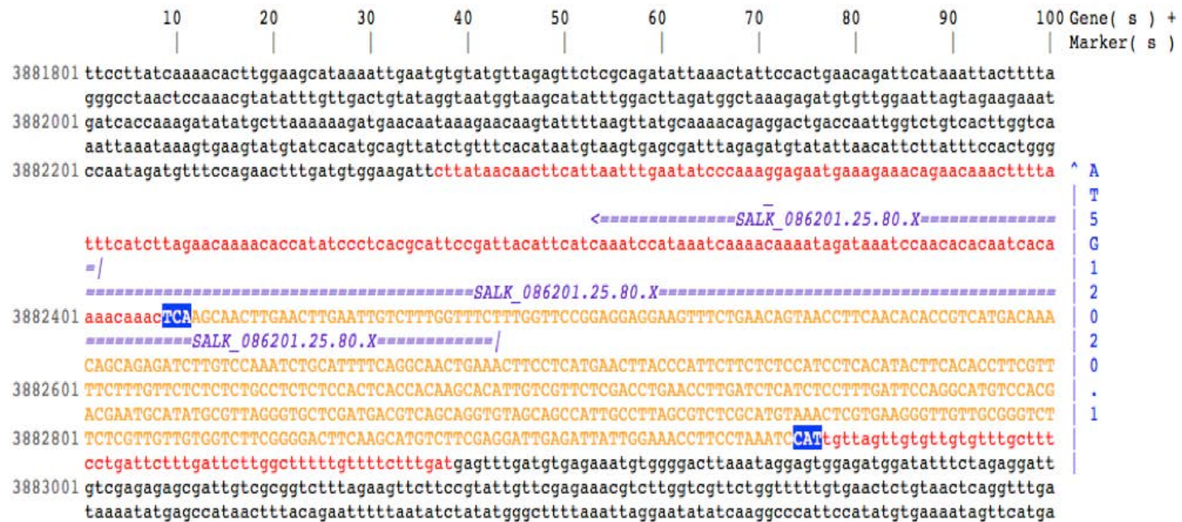


Figure 4. Position of SALK_086201.25.80 within Hsp 17.7 (AT5G12020.1) The SALK line was inserted within the exon of the gene. It is denoted in purple and indicates where exactly within the sequence of the Hsp17.7 gene it lies. Orange capital letters are the coding sequence, lower case red letters are the 5’ and 3’ non-coding regions and black lower case letters are intragenic regions. The start and stop codons are highlighted in blue. Note that the genes is shown in reverse orientation relative to translation.

These SALK line seeds were sterilized, plated, and then transferred to soil. Genomic DNA was extracted when the plants were large enough to remove leaves for genotyping reactions. Genotyping was designed to detect the presence of the T-DNA insertion within the gene of interest by a primer that binds to the T-DNA insertion (LBb1.3), whose sequence is shown in Table 3, along with a primer specific to the Hsp17.6 gene that was designed to amplify the gene in the forward or reverse direction in relation to the site where the T-DNA was inserted. Genotyping was performed with two different PCR reactions on genomic DNA isolated from the Salk lines using two different sets of primers (Table 4): 1) The forward and reverse gene primers would result in the WT band from the Hsp17.6 gene, and 2) the reverse gene primer and T-DNA primer (LBb1.3) would result in the mutant band. Using the reverse primer with the T-DNA primer instead of the forward was determined experimentally to yield the mutant band. If the

genomic DNA amplified with both sets of primers, then the plant was heterozygous for the T-DNA insertion because of the presence of both the WT and mutant bands. If amplification was obtained only in the reactions containing the reverse and T-DNA primer, the plant was designated as homozygous mutant.

Table 4. Primers for genotyping the SALK T-DNA insertion plants

SALK_086201.25.80 Primer Name	Sequence
176B_FOR	ATATATGGTCTCGATTGATCTCTGCGGCTTGTAATGAGTT
176AB_REV	TGATCTCTGCGGCTTGTAATGAGTTTTAGAGCTAGAAATAGC
LBb1.3	ATTTTGCCGATTTCGGAAC

Each PCR reaction was performed with reagents acquired from New England BioLabs in Ipswich, MA, US. A master mix was made for each individual set of reactions and varied based on the primers required with the volume adjusted to be sufficient for all samples. When calculating the contents of each reaction, the reagents were kept to consistent final concentrations. 10X Standard Taq reaction buffer was used in a 1x concentration, dNTPs were used at 200uM, each primer was at a final concentration of 0.2uM, and 1.25 units of Taq polymerase was added per 50 uL of PCR cocktail. There was less than 1ug of template DNA for each reaction. The reactions were run on a PTC-100 thermocycler (MJ Research Inc, Waltham, MA, US).

The PCR protocol was designed as follows: denaturing of the template DNA was done at 95˚C for 3 min, followed by another step of 95˚C for 30 sec, then an annealing step at 58˚C for 30 sec and an elongation step at 72˚C for 1 min and 10 sec. The cycle was run 36x before a final

step of 72˚C for a final 5 min before being stored at -20˚C. After obtaining the PCR products, they were separated on a 1% agarose and TAE (40 mM Tris, 20 mM acetate, 1 mM EDTA, PH~8.6) buffer gel for 30 min at 125V with 0.001% Gel Red (Biotium, Hayward, CA, US). Gels were then imaged with a G-box (Syngene, Frederick, MD).

To illustrate the location at which the T-DNA was inserted into Hsp 17.6 and the way Hsp17.7 was targeted using the CRISPR vector, the map of the chromosome where both of these genes are present is shown in Figure 6.

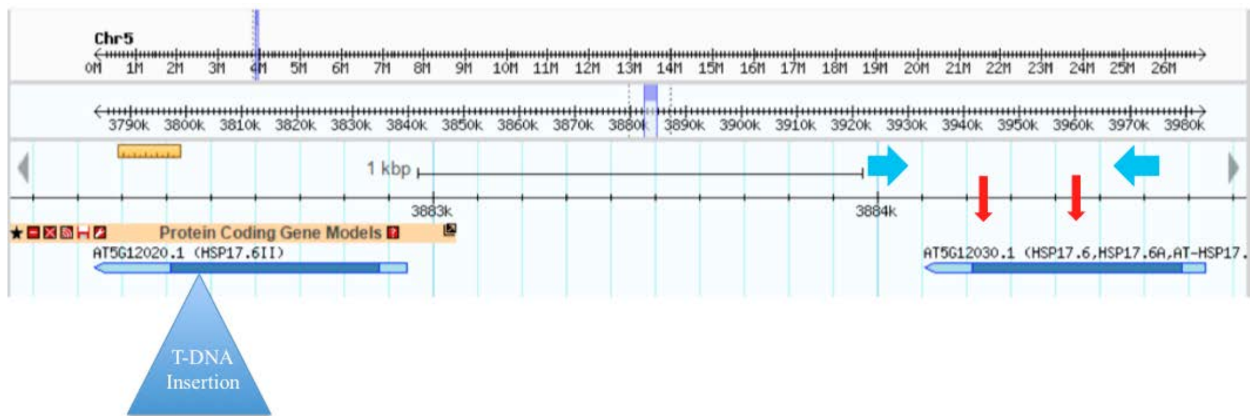


Figure 6. Map of Hsp 17.6 and Hsp 17.7 on the Arabidopsis chromosome along with the gene editing tools. The gene for Hsp 17.6 is shown on the left of the image, demonstrating the location of the SALK T-DNA insertion. The gene for Hsp 17.7 in on the right of the image, where the blue arrows indicate primer sites and the red arrows indicate cut cites for the Cas9. These two genes are within 1kb of each other on chromosome 5.

<u>Western Blots to check for protein expression</u>

As another way to check and ensure that the genes encoding the Hsp 17.6 and Hsp 17.7 are knocked out, western blots were utilized. From each plant, leaf samples were taken by cutting off a large enough part of the leaves without killing the plant. The leaf was transferred to a 1.5 mL tube and was weighed to determine how much sample buffer needs to be added. Based on each individual mass, 1X SDS Sample Buffer (2% w/v SDS, 12% v/v glycerol, 5% v/v β-mercaptoethanol, 62.5 mM Tris pH 6.8 and 0.0025% w/v Bromophenol blue) was added in a

volume that is 5 times the weight of the leaf measured in micrograms. The leaf samples were ground in the buffer and then heated at 94˚C for about 10 min to denature the proteins and then 20 μL of each sample were loaded onto a 12% SDS-PAGE gel. They were run for about 90 min at 150V to ensure full separation. Next, proteins were transferred from the gel to a nitrocellulose membrane. The membrane is essentially sandwiched between thick filter paper and placed in the transfer apparatus set to 180 mA for 1.5 hr.

After the transfer, the membrane was placed in TBS buffer containing 2.5% milk for 15 min to bind to all the sites on the membrane that do not have any transferred protein. To this buffer, primary antibodies specific for class II Arabidopsis sHsps (Hsp17.6 and Hsp17.7) that were raised in rabbits and are available at Agrisera (Vännäs, Sweden) were added in a ratio of 1:5000, so about 5 μL was added to 50 mL of TBS buffer (20 mM Tris and 150 mM NaCl). The membrane was incubated with the primary antibody for two hr at room temperature on a rocker. Next, the primary antibodies are washed off using TBST buffer (Tris-buffered saline, 0.1% Tween 20) about three times. Then, secondary antibody was added, which was raised against rabbit IgG, produced by Life Science Technologies, which is a Thermofisher company (Waltham, MA, USA). It was added in the same ratio as the primary and given another two hr of incubation. The membrane was washed again with TBST buffer to clear the membrane of antibody. It was imaged using enhanced chemiluminescence (ECL), where substrate is added to the membrane and it binds to reveal the presence of proteins. The specific kit utilized was the Pierce ECL Western Blotting Substrate created by Thermo Scientific (Agawam, MA). There are two reagents that need to be added in equal ratios, one being the peroxide solution and the other being the luminol enhancer solution. It needs to be enough to cover the whole membrane, so usually 300 μL of each reagent is added. The membrane needs to be imaged immediately using

the G-box (Syngene, Frederick, MD) under an exposure time of at least 8 min. This method can be used to confirm that the mutants do not express the specific sHsps being studied, ensuring that they are in fact knockout mutants.

Results

Identifying Hsp17.6 homozygous T-DNA insertion mutants

To create a double knock out of the Arabidopsis class II sHsp genes, I first identified homozygous T-DNA insertion mutants of one of the class II genes, Hsp17.6. These plants were then to be used of CRISPR/Cas9 knockout of the second class II gene, Hsp17.7. There was no T-DNA insertion available for Hsp 17.7 and since the class II sHsps are within 1.5 kb of each other on the chromosome, this would make it very difficult to isolate a double mutant even if there was a T-DNA insertion mutant available. In order identify the Hsp17.6 homozygous knockout plants, each plant sample was genotyped by PCR using primers 176B_FOR and 176AB_REV (Table 4) to test for the presence of either the WT gene, mutant gene, or both, ultimately determining the plant's genotype as homozygous, heterozygous, or WT.

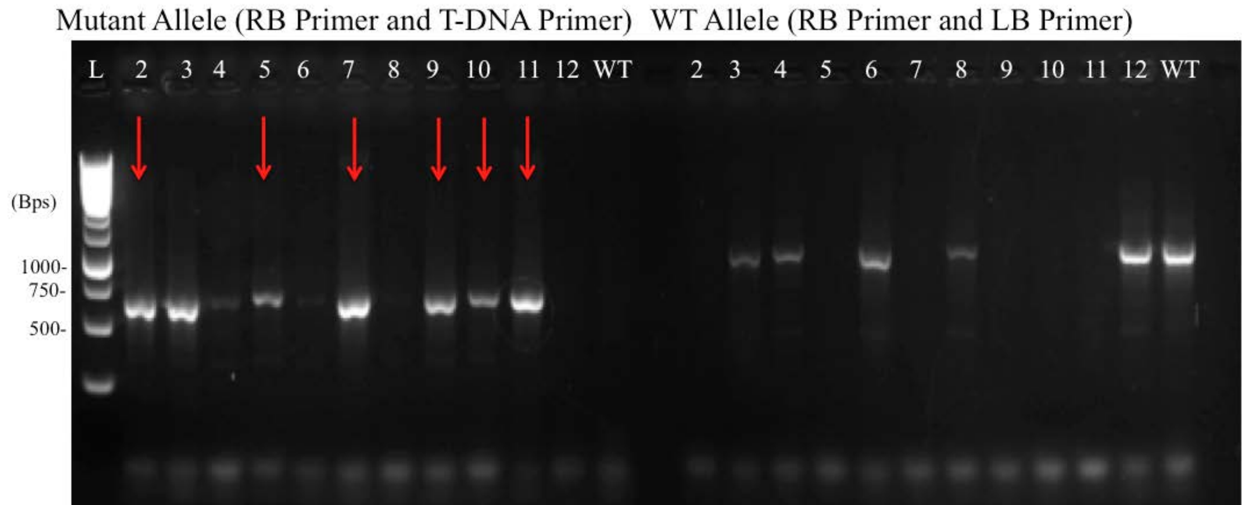


Figure 7. Confirmation of single T-DNA insertion knockout in Hsp17.6 by genotyping using PCR. Each plant sample was analyzed in two separate PCR reactions to test for the WT or mutant allele of Hsp17.6. The [RB Primer and T-DNA Primer] was used to amplify the mutant allele and the [RB Primer and LB Primer] to amplify the wild type allele. Samples were loaded in the same order for both sets of PCR. Red arrows indicate those samples that appear to be homozygous mutant. Analysis was completed by going through each lane and determining the presence of each band to determine the genotype of the plant.

Presence of the PCR product for the mutant allele indicates that the T-DNA insertion was present in the Hsp17.6 gene (Figure 7). From here, six total plants were identified to be homozygous and seeds from each were collected. Out of these six, two were selected to grow up the next generation and seeds were collected. These plants were named based on their generation, the plant they emerged from, and finally the order at which the seeds were planted. An example for how these plants would be named is F2_6_1, for the F2 generation, it came from SALK Plant #6, and it is the first seed planted.

Five pots containing 5 seeds, one in each corner and one in the center, each from the Hsp17.6 homozygous mutant were planted and transformed with the CRISPR vector targeting the other class II sHsp, Hsp17.7. We utilized seeds from two of the confirmed homozygous SALK T-DNA insertion lines, identified as SALK 6 and SALK 9 based on the seeds that were planted initially, and transformed both, which are shown in Figure 8. There were a total of 50

plants that were transformed. These plants look healthy and are growing normally, producing a lot of siliques containing 15-20 seeds each.

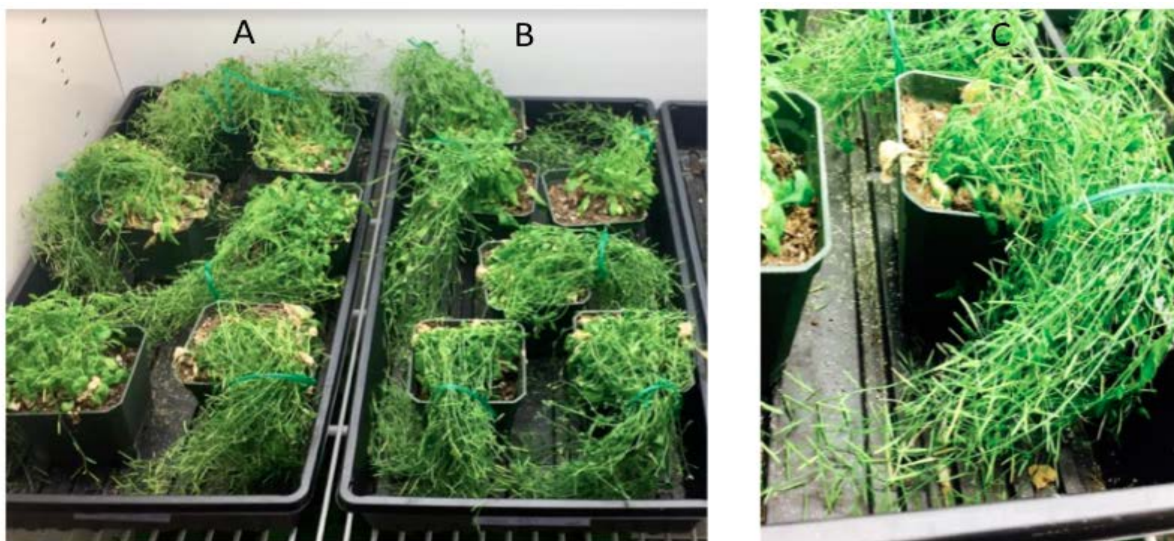


Figure 8. WT and the T-DNA SALK line insertion of Hsp 17.6 transformed for deletion of class II sHSPs. Six total SALK line mutants were identified, but only two lines were used for transformation. (A) This tray contains five pots that each has five plants from the seeds of one of the homozygous SALK line mutants (SALK 6). (B) This tray has a similar setup except these were seeds from a sibling homozygous SALK line (SALK 9). (B) A close up image of the siliques on the transformed plants.

Once the siliques were ready on these plants, seeds were collected. The correct plants will be identified by growing the transformed seeds on plates with hygromycin, which is the antibiotic the constructs are resistant to due to the makeup of the CRISPR design.

Discussion

When reviewing other literature on CRISPR success rates in generating knockouts in *Arabidopsis thaliana,* it ultimately lead to analyzing the nucleosome structure surrounding the genes of interest (Zhang, 2016). The nucleosome is a structural unit of a eukaryotic chromosome, consisting of a length of DNA wrapped around histones within this structure. Nucleosomes package DNA so that it is coiled tightly within the cell nucleus. Chromosomal DNA can exist in

two forms that are indicative of the level of activity of the cell. One form is heterochromatin, which is tightly packed DNA that accumulates in cells or in genomic regions that are less active or inactive. The other form is euchromatin, which is a more open form of chromatin that contains genes undergoing consistent, frequent transcription. When trying to mutate DNA using CRISPR/Cas9, the extent of chromosome compaction and nucleosome occupancy could limit the effectiveness of the method.

The genes that encode Hsp17.6 and Hsp17.7 are in very close proximity. Examining the positions of Arabidopsis genes for which other knockouts have been generated with CRISPR revealed that most of the genes had high expression throughout the plant in addition to low nucleosome occupancy, making it less compact and therefore the DNA more easily accessible (Hyun, 2015). This shows that the genes are consistently transcribed so therefore probably in a region of euchromatin (Hyun, 2015). However, this was not seen with the genes of interest encoding the class II sHsps. These two genes are in an area of the genome where there is a high nucleosome occupancy and they have little to no expression within most parts of the plant. This information was obtained through the Plant DNase I hypersensitive Sites (DHSs) Database (http://plantdhs.org). As the use of CRISPR in plants is still new, it is possible that this genomic configuration will limit the effectiveness of the CRISPR mutagenesis in these experiments.

Conclusions

Once these plants are screened and genotyped, double knockouts of the class II sHsps can be identified. Noting their growth and stress tolerance without the Class II sHsps will provide some key insights into the importance of this class of sHsps. To look at the function of these mutants, heat stress assays will be conducted. The Vierling lab group has designed protocols to

provide quantitative data on the ability of seedlings to tolerate high temperature stress. One assay examines the ability of dark grown seedling hypocotyls to elongate after an acclimation treatment followed by a period of 45°C heat stress. Sterile seeds are put on plates in lines, evenly spaced to make sure each seedling has space to grow, and plates are wrapped in aluminum foil to ensure total darkness, because hypocotyls elongate more in the dark. After germination, a heat acclimation treatment is performed for which the foil is removed and the plate is put in an incubator set to 38 °C for 1.5 hr and then the plates are allowed to recover for 2 hr by placing them in an incubator set to 22 °C in the dark. The next step is a severe, acute heat stress; the plates are put in an incubator set to 45 °C for a specific amount of time, ranging from 1 to 3 hr. After this stress the hypocotyl length is noted by marking each seedling and the plates are rewrapped in foil for growth at 22 °C for an additional 2.5 days. Growth of these plants is then analyzed is by measuring how much the hypocotyls was have grown after the 45 °C heat stress compared to untreated, or 38 °C treated seedlings. Susceptibility to heat stress can also be tested in assays with light grown seedlings, with seeds, and with more mature plants. Biochemical phenotypes can be further assayed, potentially even looking to see what other proteins are involved in the stress response.

Chapter 3: Generating triple mutants of all mitochondrial- or all chloroplast-targeted sHsps.

Materials and Methods

Crossing of plants containing mutations of interest (conducted by Minsoo Kim)

A series of crosses were done previously to yield mutants that will be utilized in generating the triple mutants of either all mitochondrial-targeted sHsps or all chloroplast-targeted sHsps. By crossing plants carrying single gene knockouts in 26.5_MT (At1g52560) or 25.3_CP (At4g27670) with double gene knockouts of 23.5_C/MT (At5g51440) and 23.6_C/MT (At4g25200), it should be possible to create the plants of interest. Minsoo Kim generated the single knockouts, while Olivier Van Aken (Van Aken et al., 2009) generated the double knockout and provided the seeds for these specific knockout plants. Each of these previous created knockouts were generated from both SALK and SAIL line insertions within each gene of interest. For 23.5_C/MT, the SALK line named SALK_118536 was inserted into the gene. For 23.6_C/MT, it was a SAIL insertion line identified as SAIL_373_B09. For 26.5_MT, the SAIL insertion line was called SAIL_423_G06/CS874042. Finally for the 25.3_CP gene, it was a point mutation generated and this will be genotyped differently from the other genes using the Derived Cleaved Amplified Polymorphic Sequences (dCAPS) assay that allows researchers to identify specific mutations within the specific gene of interest.

Currently there are no mutants available in which all three of these genes are knocked out. Using a Punnett square, the probability of achieving a triple mutant can be determined. By selfing the F1 generation produced from the cross, a total of eight different gamete types can result. Putting these possibilities into a Punnett square reveals that theoretically 1/64 progeny will be a homozygous triple mutant. This ratio is expected provided the genes are not linked, which in the case for the mitochondrial sHsps. This means that the F1 seeds needs to be grown

and allowed to self, after which F2 generation plants will need to be genotyped for each of the mutant alleles. If a triple mutant is not found in the F2 plants tested, F2 plants that are homozygous in two of the three sHsps can be grown to obtain the F3 generation, in which segregation should result in ¼ of the plants being triple mutants.

Diagrammed below is the first cross conducted followed by the genotype of the F1 generation. From this F1 generation, the resulting Punnett square in Table 5 demonstrates the genotypes that could result in the F2. In red is the genotype that corresponds to the triple mutant.

1st Cross: $\frac{++25.3}{++25.3}$ X $\frac{23.5}{23.5}$ $\frac{23.6}{23.6}$ $\frac{+}{+}$

(25.3 = Chloroplast-targeted sHsp) (23.5/23.6= Mitochondria and Chloroplast-targeted sHsp)

F1: $\frac{23.5}{+}$ $\frac{23.6}{+}$ $\frac{25.3}{+}$

Table 5. Expected Result of Selfing the F1 to obtain the mitochondrial triple sHSP mutant

	+/+/+	+/+/23.5	+/+/23.6	+/+/25.3	+/23.5/23.6	+/23.5/25.3	+/23.6/25.3	23.5/23.6/25.3
+/+/+	+/+/+ +/+/+	+/+/23.5 +/+/+	+/+/23.6 +/+/+	+/+/25.3 +/+/+	+/23.5/23.6 +/+/+	+/23.5/25.3 +/+/+	+/23.6/25.3 +/+/+	23.5/23.6/25.3 +/+/+
+/+/23.5	+/+/+ +/+/23.5	+/+/23.5 +/+/23.5	+/+/23.6 +/+/23.5	+/+/25.3 +/+/23.5	+/23.5/23.6 +/+/23.5	+/23.5/25.3 +/+/23.5	+/23.6/25.3 +/+/23.5	23.5/23.6/25.3 +/+/23.5
+/+/23.6	+/+/+ +/+/23.6	+/+/23.5 +/+/23.6	+/+/23.6 +/+/23.6	+/+/25.3 +/+/23.6	+/23.5/23.6 +/+/23.6	+/23.5/25.3 +/+/23.6	+/23.6/25.3 +/+/23.6	23.5/23.6/25.3 +/+/23.6
+/+/25.3	+/+/+ +/+/25.3	+/+/23.5 +/+/25.3	+/+/23.6 +/+/25.3	+/+/25.3 +/+/25.3	+/23.5/23.6 +/+/25.3	+/23.5/25.3 +/+/25.3	+/23.6/25.3 +/+/25.3	23.5/23.6/25.3 +/+/25.3
+/23.5/23.6	+/+/+ +/23.5/23.6	+/+/23.5 +/23.5/23.6	+/+/23.6 +/23.5/23.6	+/+/25.3 +/23.5/23.6	+/23.5/23.6 +/23.5/23.6	+/23.5/25.3 +/23.5/23.6	+/23.6/25.3 +/23.5/23.6	23.5/23.6/25.3 +/23.5/23.6
+/23.5/25.3	+/+/+ +/23.5/25.3	+/+/23.5 +/23.5/25.3	+/+/23.6 +/23.5/25.3	+/+/25.3 +/23.5/25.3	+/23.5/23.6 +/23.5/25.3	+/23.5/25.3 +/23.5/25.3	+/23.6/25.3 +/23.5/25.3	23.5/23.6/25.3 +/23.5/25.3
+/23.6/25.3	+/+/+ +/23.6/25.3	+/+/23.5 +/23.6/25.3	+/+/23.6 +/23.6/25.3	+/+/25.3 +/23.6/25.3	+/23.5/23.6 +/23.6/25.3	+/23.5/25.3 +/23.6/25.3	+/23.6/25.3 +/23.6/25.3	23.5/23.6/25.3 +/23.6/25.3

23.5/23.6/ 25.3	+/+/+ _ 23.5/23.6/ 25.3	+/+/23.5 23.5/23.6/ 25.3	+/+/23.6 23.5/23.6/ 25.3	+/+/25.3 23.5/23.6/ 25.3	+/23.5/23.6 23.5/23.6/ 25.3	+/23.5/25.3 23.5/23.6/ 25.3	+/23.6/25.3 23.5/23.6/ 25.3	**23.5/23.6/25.3 23.5/23.6/25.3**

Genotyping by Polymerase Chain Reaction

While genotyping it was important to carefully keep track of each plant tested and what primers were necessary for detecting the sHsp alleles of interest. For the cross designed to knock out the mitochondrial-targeted sHsps, the primers include primer set one, two, and three, corresponding to the three different genes as listed in Table 4. For the cross designed to knock out the chloroplast-targeted sHsps, the primers also include primer set one, two, as well as a fourth primer (set four) (Table 6).

Table 6. Primers for genotyping

Insertion Line		Gene	Primer Name	Sequence
SALK_118536	Set One	AT5G51440	23.5-F1	CTTCGCATCGAACTTCTCATC
	Set One	AT5G51440	23.5-R1	CCTACTCGTAAACCTCCGTCC
	Set One	AT5G51440	LBb1.3	ATTTTGCCGATTTCGGAAC
SAIL_373_B09	Set Two	AT4G25200	23.6-F1	CGAGTCTTCTTGGTCTTTTCG
	Set Two	AT4G25200	23.6-R1	ATCTCCGATTACCGCTCTCTC
	Set Two	AT4G25200	SAIL-LB3	TAGCATCTGAATTTCATAACCAATCT CGATACAC
SAIL_423_G06/CS8 74042	Set Three	AT1G52560	26.5m-1	TCTAGCTCGTCTGGCTTTGAG
	Set Three	AT1G52560	26.5m-2	AAGAACACAAAAACGACACCG
	Set Three	AT1G52560	SAIL-LB3	TAGCATCTGAATTTCATAACCAATCT CGATACAC
Point Mutation	Set Four	AT4G27670	25.3p-3	AAACAATGTTCTGTTTTAATCTAACC ACC
	Set Four	AT4G27670	25.3p-4	AGAGACCAGGCATGTCGAAA

	Set Four	AT4G27670	dCAPS	n/a

Genomic DNA was extracted from each of the plants for each cross as described in Chapter 2. These genotyping reactions were similarly designed with respect to the PCR protocol created for genotyping the Hsp17.7 gene in the double knockout described in Chapter 2. The goal of these reactions is to detect the presence of the insertion within each individual gene.

Each PCR reaction was performed with the same reagents as stated in the previous chapter. When calculating the contents of each reaction, the reagents were kept to consistent final concentrations that aided in the calculations, but these were slightly different from prior genotyping reactions due to the different polymerase utilized. 5X High-Fidelity reaction buffer from New England Biolabs (Ipswich, MA) was used in a 1x concentration, dNTPs were used at 200µM, each primer was at a concentration of 0.2µM, and 1.0 unit of Phusion High-Fidelity DNA polymerase was added per 50 µL of PCR reaction. Still, less than 1 µg of genomic template DNA was used in each reaction. In these reactions, most were set up with all three primers together in the master mix due to the fact that they have the same annealing temperature, so all primers should effectively bind the DNA, producing the bands of interest to reveal homozygous, heterozygous, or WT genotypes.

The PCR protocol differed from the one utilized in the genotyping of the SALK line plants due to the fact the enzyme used here was Phusion DNA polymerase, which can withstand higher temperatures. This protocol was designed as follows: denaturing of the template strand was done at 98˚C for 3 min, followed by another step of 98˚C for 30 sec, then an annealing step at 58˚C for 30 sec and an elongation step at 72˚C for 1 min and 10 seconds. The cycle was run 36x before a final step of 72˚C for 5 min before being stored at -20˚C. The PCR products were

separated on a 1% agarose and TAE gel for 30 min at 125V with 0.001% GelRed (Biotium, Hayward, CA, US). Gels were then imaged with a G-box (Syngene, Frederick, MD).

Results

In total, 128 plants from the F2 of each cross were planted, with four plants per 6 x 6cm pot and given ample time to grow, making a total of 256 plants to be analyzed in three separate, individual PCR reactions. Amplification of each allele for each gene produces an expected fragment size that aids in identifying the genotypes (Table 7).

Table 7. Expected Lengths of Amplification for each gene of interest

Gene	Band	Expected Size
23.5_C/M	WT	1134 bp
23.5_C/M	Mutant	700 bp
23.6_C/M	WT	1067 bp
23.6_C/M	Mutant	520 bp
26.5_M	WT	1008 bp
26.5_M	Mutant	450 bp

Finally, for the 25.3C gene (encoding chloroplast Hsp21), needs to be genotyped differently because it is a point mutation. The Derived Cleaved Amplified Polymorphic Sequences (dCAPS) assay was utilized, in which the gene region is first amplified with primers designed to introduce a restriction site into either the wild type or mutant DNA. After amplification the DNA is then subjected to digestion by the restriction enzyme. The expected digestion product size for the WT allele is 222 bp while the mutant allele should yield bands of 190 bp and 32 bp. This assay will be explained while discussing the results of the cross to knockout of all chloroplast-targeted sHsps.

Knockout of all mitochondrial-targeted sHsps: 26.5M x 23.5C/M,23.6C/M

Out of the 128 plant samples available, 81 plants were all analyzed with each PCR, logging each time a reaction was run and the genotype of each gene for each plant sample.

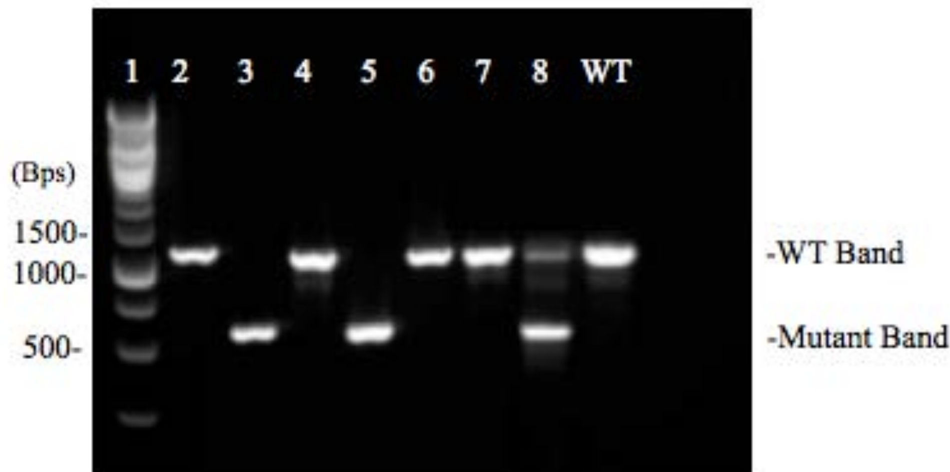


Figure 9. Example of genotyping reaction for the 23.5_C/MT (At5g51440) gene. Adding all three primers (one, two and three – Table 6) in this reaction results in bands that either are identified at WT or mutant. The mutant band lies just over the 500 bp mark on the ladder, while the WT band lies just over the 1000 bp mark on the 1 kb ladder from New England Biolabs (Ipswich, MA). If the sample reveals one band in the WT position, its genotype is identified as WT (lane 2). If the sample reveals one band in the mutant position, its genotype is identified as homozygous mutant (lane 3). If the sample reveals bands in both of positions, its genotype is identified as heterozygous (lane 8).

Out of the 128 plants available, no triple homozygous mutant lines were identified, but plants that were homozygous for two of the genes and heterozygous for the third gene were saved and allowed to self to further screen for the mutants of interest in the F3 generation. In the F3 generation, because the plants will be segregating for only one of the three genes (the one that was heterozygous) the chances of obtaining the triple mutant from these plants is 1 out of 4, based on the expected ratio 1:2:1 (WT:HET:MUT) according to the father of genetics, Gregor

Mendel. In fact, the ratio that was observed in the generation of plants genotyped was 1:2.25:1.25, which is pretty close to what we expected to see. This is significant because it showed how the genes assorted themselves independently.

In total, there were 7 plants with these characteristics, detailed in Table 8.

Table 8. Genotypes of F2 plants for the three mitochondrial-localized sHsps

Plant ID	23.5C/M	23.6C/M	26.5M
F2_M_26	homozygous	homozygous	heterozygous
*F2_M_34	heterozygous	homozygous	homozygous
F2_M_37	homozygous	homozygous	heterozygous
F2_M_71	homozygous	homozygous	heterozygous
*F2_M_97	homozygous	homozygous	heterozygous
*F2_M_99	homozygous	homozygous	heterozygous
*F2_M_124	homozygous	heterozygous	homozygous

The asterisks in Table 5 indicate those plants for which that the F3 generation was planted and screened. A total of 32 seeds from each of these plants were plated and then transferred to soil, followed by the genotyping reactions.

From this set of plants so far, 9 plants have been identified as homozygous triple mutants for all mitochondrial-targeted sHsps. This was validated through re-extraction of genomic DNA and re-running these samples through all three PCR genotyping reactions (Figure 10).

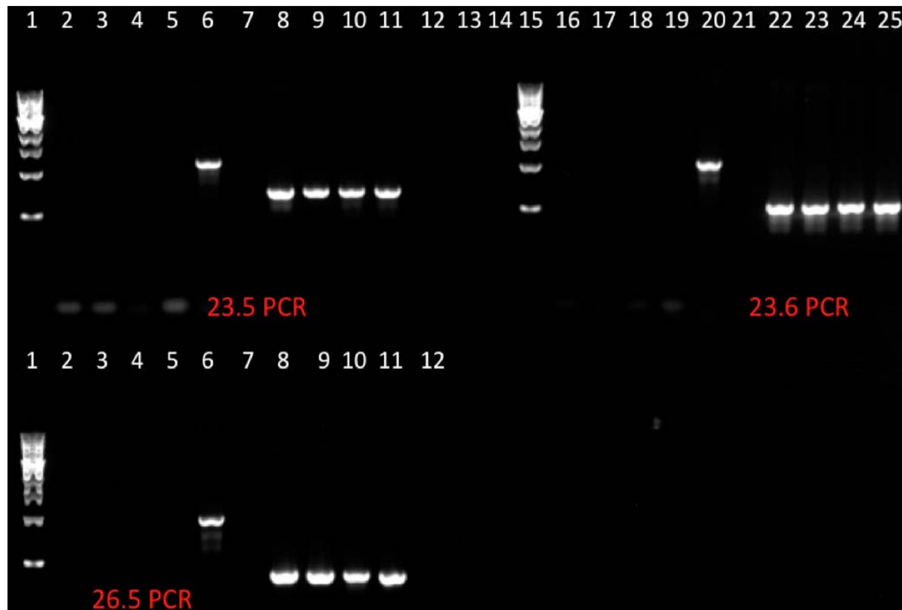


Figure 10. Confirmation of the first four homozygous triple mutants for mitochondrial-targeted sHsps The same samples were run through each PCR demonstrated on the gel. Samples were loaded in the same order for each set of PCR. The PCR identifying the WT band was loaded first, shown in lanes 2-5 (top and bottom of gel) and lanes 16-19, followed by the PCR identifying the mutant band, shown in lanes 8-11(top and bottom of gel) and lanes 22-25.

As seen in Figure 10, the WT sample revealed the band of interest while none of the other samples appear to have the band. All the samples reveal only the mutant band, therefore confirming the initial conclusion that these 4 plants are triple mutants.

Currently, these plants have been isolated into their own pots and are producing seeds. These plants are named depending on the generation it is currently in, the parent plant ID, and then the order at which the seeds were planted plus a “C” or an “M” to indicate which cross this plant is from. For example, there exists F3_34_1M, which means this plant is from the F3 generation, comes from parent plant F2_34_M, it was the first seed planted, and it is from the mitochondrial cross. After harvest, seeds will be planted, the genotypes will be rechecked and plants utilized to characterize the growth and heat stress phenotypes of these plants.

Knockout of all chloroplast-targeted sHsps: 25.3C x 23.5C/M, 23.6C/M

All 128 samples were analyzed by PCR for the 23.5C/M and 23.6 C/M allele to identify which of these samples were homozygous in at least these two genes. The third gene for the chloroplast-targeted sHsps requires a step in addition to the PCR reaction. This is because the mutation within the 25.3C gene is a point mutation, so a different method needed to be utilized. The dCAPS assay is usually used for the detection of Single Nucleotide Polymorphisms (SNPs), which in this case is what is occurring within this third gene. This technique introduces restriction enzyme recognition sites by using primers that contains one or more mismatches in its sequence (NCBI, 2017). The genomic DNA is still amplified by PCR with Set Four forward and reverse primers listed in Table 6, but after this, the PCR product is subjected to digestion by BstXI. If the sample is a mutant, the DNA will be cut and produce fragments of 190 bp and 32 bp. This restriction enzyme digestion determines the presence or absence of the mutation of interest.

After finishing the genotyping on this set of 128, there were no true triple homozygous mutants, but the same situation occurred as with the mitochondrial cross; plants that were homozygous mutant in two of the three genes and heterozygous for the third gene were identified. There were 12 plants with these characteristics, shown in Table 9.

Table 9. Final plants from initial screening for triple mutant for mitochondrial cross

Plant ID	23.5C/M	23.6C/M	25.3C
*F2_C_139	homozygous	heterozygous	homozygous
*F2_C_141	homozygous	heterozygous	heterozygous
F2_C_149	homozygous	homozygous	heterozygous
*F2_C_153	homozygous	heterozygous	homozygous
F2_C_157	heterozygous	heterozygous	homozygous

*F2_C_175	homozygous	heterozygous	heterozygous
F2_C_181	homozygous	homozygous	WT
F2_C_186	homozygous	heterozygous	heterozygous
F2_C_192	heterozygous	heterozygous	homozygous
F2_C_201	homozygous	heterozygous	heterozygous
F2_C_208	homozygous	heterozygous	heterozygous
F2_C_225	homozygous	heterozygous	heterozygous

The starred plants in Table 9 indicate those that the F3 generation was planted and screened. As for the mitochondrial mutants, 32 seeds from each of these plants were plated and then transferred to soil. It took a long time for these plants to start growing after initial planting, so there were some delays in completing the genotyping reactions.

Currently, the plants have finally grown enough to start extracting genomic DNA. Genotyping reactions have begun on those plants that were large enough. There exists some double mutants within 23.5 C/M and 23.6 C/M, but these still need to be checked for the third gene of 25.3C to see if any of these are in fact triple mutants.

Discussion

When determining the probability of obtaining a triple mutant from selfing plants that are homozygous in two genes and heterozygous in the third gene of interest, it is important to note the location of these genes on the chromosome in order to see how and the predicted frequency with which these traits will be recovered in the next generation.

For the mitochondrial cross, all three genes (23.5_C/M, 23.6_C/M, and 26.5M) exist on separate chromosomes within the *Arabidopsis thaliana* genome. Being dispersed between chromosome 1, 4, and 5, this demonstrates the law of independent assortment which states that

individual factors assort independently from the others, allowing for an equal opportunity for occurring together. This helps explain why after the initial screening of plants from this cross, the odds of obtaining a triple mutant went from a 1/64 chance to a 1/4 chance, thus improving the possibility of obtaining actual mutants. This final ratio of 1:2:1 (WT:Het:Mut) was actually pretty closely determined experimentally amongst the current generation of F3 plants, giving 19 total plants with all of the mitochondrial-targeted sHsps knocked out.

For the chloroplast cross, things are a little more complicated. The 23.5C/M gene exists on a separate chromosome, while 23.6C/M and 25.3C are on the same chromosome. This could be problematic due to the fact the genes on the same chromosome do not segregate independently. When genes are on the same chromosome recombination must occur between the genes in order for segregation to occur. In order to determine the frequency and rate at which recombination will occur for the specific genes of interest, a value called a centimorgan (cM), or map unit, can be calculated. It measures genetic linkages and it defined as the distance between chromosome positions for which the expected average number of chromosomal crossovers in a single generation is 0.01. With both of these genes being on *Arabidopsis* chromosome 4, this value can be calculated starting with the entire size of the chromosome, which is 19 million basepairs (bp). It was determined that chromosome 4 is 75.9 cM, which means that ~250,000 bp is equal to 1 cM (Singer et al., 2006). Looking at the location of the genes and the distance between them revealed that 23.6C/M and 25.3C are separated by a distance of ~4 cM. This means that there is a 4 out of 100 chance of obtaining a mutant that is homozygous in both of these genes. For the third gene to make the triple mutant (23.5C/M), it will segregate independently from these other two genes. Because of the fact that two out of the three genes are linked, more progeny will need to be screened to actually obtain a triple mutant from this cross.

Conclusion

Obtaining the triple mutants of the mitochondrial-targeted sHsps was a great success. It will be interesting to see how these plants survive without their sHsps targeted to the mitochondria. The next step will be to obtain seeds from them and start various assays to analyze the plant phenotype. Currently, there appears to be no phenotypic difference between the mutants and the WT plants, based on the physical appearance of the F3 generation. There is a normal growth pattern amongst the mutants, same green color as the wild type plants and they have approximately the same leaf size.

In Chapter 2, the conclusion described one of the heat stress assays that will aid in identifying the possible phenotypes of the mutants by stressing seedlings and analyzing the growth of the hypocotyl. There are other assays that also would serve a purpose in characterizing the mutants. Another assay includes looking at 7-10 day old seedlings, positioning these seeds for each mutant in different sectors of the plate. These plants are given some time to germinate and grow, but then are heat stressed in a similar manner as the young seedlings described earlier. After the heat stress, the plates are given time to recover for about 5-8 days. After this time, the phenotypes of the mutants can be scored as seedling survival. If they are able to grow normally, then the next step would be to identify other key factors that allow the mutant plants to survive the stress. It is possible that new interactors can be targeted as potential means for controlling the stress response. By knocking out these classes of sHsps, a lot can be learned about the mechanisms at which these sHsps normally function and ultimately from there it can be seen if this specific classes of sHsps are important in the progression of protein-misfolding human diseases.

References

Basha E, Jones C, Wysocki V, Vierling E (2010) Mechanistic differences between two conserved classes of small heat shock proteins found in the plant cytosol. J Biol Chem 285:11489–11497

Basha, E., O'Neill, H., & Vierling, E. (2012). Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *TIBS*, *37*(3), 106-117.

Derocher AE, Helm KW, Lauzon LM, Vierling E (1991) Expression of a conserved family of cytoplasmic low molecular weight heat shock proteins during heat stress and recovery. Plant Physiol 96:1038–1047

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proceedings of the National Academy of Sciences*, *111*(12), 4632-4637.

Giorno F, Wolters-Arts M, Grillo S, Scharf KD, Vriezen WH, Mariani C (2010) Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anthers. J Exp Bot 61:453–462

Haslbeck, M., & Vierling, E. (2015). A first line of stress defense: small heat shock proteins and their function in protein homeostasis. *Journal of molecular biology*, *427*(7), 1537-1548.

Hyun, Y., Kim, J., Cho, S. W., Choi, Y., Kim, J. S., & Coupland, G. (2015). Site-directed mutagenesis in Arabidopsis thaliana using dividing tissue-targeted RGEN of the CRISPR/Cas system to generate heritable null alleles.*Planta*, *241*(1), 271-284.

Jaya, Nomalie (2009). Substrate binding site flexibility of small heat shock proteins and factors contributing to efficient chaperone activity. PhD dissertation advised by Dr. Elizabeth Vierling

Larkindale, J., Hall, J. D., Knight, M. R., & Vierling, E. (2005). Heat stress phenotypes of Arabidopsis mutants implicate multiple signaling pathways in the acquisition of thermotolerance. *Plant Physiology*, *138*(2), 882-897.

McLoughlin, F., Basha, E., Fowler, M. E., Kim, M., Bordowitz, J., Katiyar-Agarwal, S., & Vierling, E. (2016). Class I and II Small Heat Shock Proteins Together with HSP101 Protect Protein Translation Factors during Heat Stress. *Plant Physiology*, *172*(2), 1221-1236.

NCBI. Derived Cleaved Amplified Polymorphic Sequences (dCAPS). Retrieved April 25, 2017, from https://www.ncbi.nlm.nih.gov/genome/probe/doc/TechDCAPS.shtml

Osteryoung, K.W. and Vierling, E. (1994). Dynamics of small heat shock protein distribution within the chloroplasts of higher plants. J. Biol. Chem.269, 28676-28682.

Rutsdottir, G., Härmark, J., Weide, Y., Hebert, H., Rasmussen, M. I., et al. (2017). Structural Model of Dodecameric Heat-shock Protein Hsp21-Flexible N-terminal Arms Interact with Client Proteins while C-terminal Tails Maintain the Dodecamer and Chaperone Activity. *Journal of Biological Chemistry*, jbc-M116.

Santhanagopalan, I., Basha, E., Ballard, K., Bopp, N., & Vierling, E. (2015). Model Chaperones: Small Heat Shock Proteins from Plants (R. Tanguay, & L. Hightower, Eds.). In *The Big Book on Small Heat Shock Proteins*. Springer International Publishing Switzerland (pp. 2-3)

Siddique M, Gernhard S, von Koskull-Doring P, Vierling E, Scharf KD (2008) The plant sHSP superfamily: five new members in Arabidopsis thaliana with unexpected properties. Cell Stress Chaperones 13:183–197

Singer, T., Fan, Y., Chang, H. S., Zhu, T., Hazen, S. P., & Briggs, S. P. (2006). A high-resolution map of Arabidopsis recombinant inbred lines by whole-genome exon array hybridization. *PLoS Genet*, *2*(9), e144.

Stengel F, Baldwin AJ, Painter AJ, Jaya N, Basha E, Kay LE, Vierling E, Robinson CV, Benesch JL (2010) Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. Proc Natl Acad Sci U S A 107:2007–2012

Sun, Y., & MacRae, T. H. (2005). Small heat shock proteins: molecular structure and chaperone function. *Cellular and Molecular Life Sciences CMLS*, *62*(21), 2460-2476.

Suzuki, T.C., D.C. Krawitz, E. Vierling. (1998). The chloroplast small heat shock protein oligomer is not phosphorylated and does not dissociate during heat stress *in vivo*. Plant Physiol. 116:1151-1161.

Van Aken, O., Zhang, B., Carrie, C., Uggalla, V., Paynter, E., Giraud, E., & Whelan, J. (2009). Defining the mitochondrial stress response in Arabidopsis thaliana. *Molecular plant*, *2*(6), 1310-1324.

Van Montfort RL, Basha E, Friedrich KL, Slingsby C, Vierling E (2001b) Crystal structure and assembly of a eukaryotic small heat shock protein. Nat Struct Biol 8:1025–1030

Vierling, E., Nagao, R. T., DeRocher, A. E., & Harris, L. M. (1988). A heat shock protein localized to chloroplasts is a member of a eukaryotic superfamily of heat shock proteins. *The EMBO journal*, *7*(3), 575.

Waters ER (2013) The evolution, function, structure, and expression of the plant sHSPs. J Exp Bot 64:391–403

Waters, E. R., Aevermann, B. D., & Sanders-Reed, Z. (2008). Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. *Cell Stress and Chaperones*, *13*(2), 127-142.

Xing, H. L., Dong, L., Wang, Z. P., Zhang, H. Y., Han, Chen, Q. J., et al. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC plant biology*, *14*(1), 327.

Zhang, Tao T. Plant DNase I hypersensitive Sites (DHSs) Database. Retrieved April 25, 2017, from http://plantdhs.org/