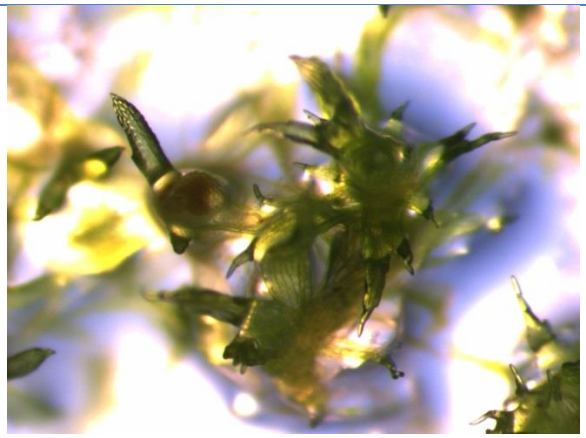
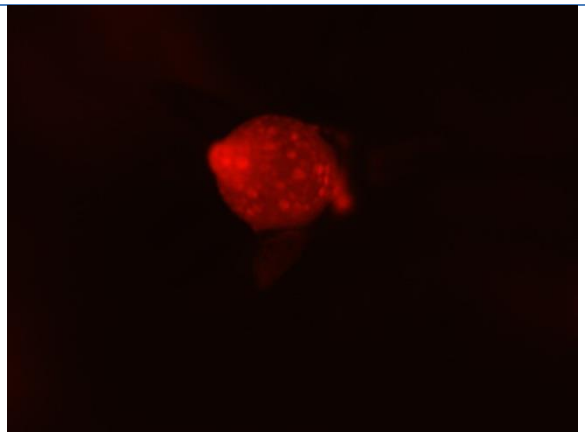
Isolation of Temperature-sensitive Mutants in the Moss *Physcomitrella patens* and Mapping of their Causal Mutation by Genome Sequencing of Pooled Segregants





A Major Qualifying Project Report
Submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
In partial fulfillment of the requirements for the
Degree of Bachelor of Science
On May 1st, 2014
By:

Xinxin Ding

Submitted to:
WPI Advisors: Patrick J. Flaherty, Luis Vidali.

Abstract

Plants are important as sources of oxygen, food, medicine and as a potential source for clean energy. Thus, plants can play a critical role in encountering the global challenge of food and energy shortage. Although encouraging progress has been made in understanding the cellular and molecular basis of plant growth, many key processes and mechanisms still remain a puzzle. In order to use plants to achieve a sustainable future, it is critical to understand how genes direct cell growth and cell division. Nonetheless, this has been difficult to study in plants because disruption of genes vital for growth often leads to embryonic lethality. An effective solution often used in other systems is to isolate conditional mutations. This project used the moss *Physcomitrella patens* as the model organism and focused on isolating temperature-sensitive (TS) mutants with impaired growth, characterizing their morphology under the permissive and restrictive conditions, as well as trying to identify the genetic mutation that causes the TS phenotype. UV mutagenesis was used to generate the mutants and growth assays were conducted to obtain images of individual mutant plants for statistical analysis of morphological parameters. In an attempt to identify the affected genes, genome sequencing of one TS mutant (LV768) and pooled-sequencing of F1 recombinants of another TS mutant (LV767) were conducted. **I isolated eight mutants and characterized their morphological and growth features in detail, of which six mutants showed a significant reduction in growth rate and decrease in polarity under the restrictive condition while the other two showed decrease in polarity at all temperatures. I also designed an experiment to map the TS causal mutation of LV767 by genome sequencing of 24 pooled F1 segregants.**

Acknowledgements

I would like to thank the following individuals and organization for their support and contributions to the success of our project:

- **Jeffrey Philippe Bibeau**, a graduate student of Vidali's Lab, for writing the ImageJ macro for plant area measurement and solidity calculation.
- The Microscopy Manager **Victoria M. Huntress** for her kindly assistance in the microscopy of growths assays.
- Advisors **Dr. Patrick J. Flaherty** for continued support with the computational analysis in the experiment design of genome sequencing by pooled segregants, and **Dr. Luis Vidali** for his continued support and help with all the experiments and assays in this study.

Table of Contents

Contents

Table of Tables

Table of Figures

Introduction

Plants are an important part of ecosystems and a critical source of food, renewable energy and medicine. Serious challenges have been posed by the rapid increase in world population, limited fossil fuel reserves and clean water resources, aggravating pollution and changing climate. In order to meet the global demand of food and energy and alleviate pollution, the world needs to increase the yields of food and energy crops in a sustainable way. The way of increasing crop yields need to contain the use of water for irrigation and conversion of natural habitats to arable lands as well as reduce the use of agrochemicals and fossil fuels. Modern plant biology provides a way to achieve this goal through genomic marker-assisted crop breeding and transformation of foreign genes into crops, which can accelerate the process of traditional crop breeding and introduce desired traits to crops such as increase in resistance to biotic and abiotic stresses. Identification of genes that are critical in plant cell growth, division and response to various biotic and abiotic stresses, therefore, is one key to the improvement of crops.

Forward and reverse genetics are the two main strategies used to study gene function. In forward genetic studies, researchers generate mutants, select those with desired phenotypes and ultimately map causal mutation to the genome. In reverse genetic studies, researchers in contrast have a known gene and try to deduce the function by mutating or silencing the gene. It is hard to study the function of essential genes since silencing of those genes is likely to be lethal. One effective approach to study these genes is the isolation of temperature-sensitive (TS) mutants where TS mutations affect cell function differently at different temperatures. *Physcomitrella patens* is a good model system to study function of plant genes since it is easy to keep and grow in lab, has a simple structure and a dominant haploid phase as well as some powerful genomic and genetic tools. In addition, *P. patens* has a much simpler structure and growth pattern than those of flowering plants and many genes important for the development of flowering plants are conserved in *P. patens*.

The goal of this study is to identify genes that are critical for plant and cell growth in *P. patens*. In order to achieve this goal, I isolated TS mutants that had impaired growth at restrictive temperature using *P. patens*, characterized their morphology, and tried to map the causal mutation by pooled-sequencing of segregants. Thus, the objectives of this study were to:

1. Generate random mutants by ultraviolet (UV) mutagenesis and select TS mutants by comparing their growth at the permissive (25ºC) and restrictive (32ºC) temperature,
2. Evaluate the inhibited growth of mutant plants in terms of growth rate and morphology by microscopy and analysis of morphological parameters such as area and solidity,
3. Generate F1 segregants of the isolated TS mutants by outcrossing them to a fluorescently labelled polymorphic strain of *P. patens*, and
4. Design an experiment to map the causal mutation of one TS mutant by pooled-sequencing of its F1 segregants that still retained the TS mutations.

In order to study the function of critical genes without any bias, I chose to use the strategy of forward genetics by isolating temperature-sensitive (TS) mutants, which have impaired growth patterns at restrictive temperature (32ºC) and mapping the causal mutation by genomic sequencing. First, random mutants were generated by ultraviolet (UV) mutagenesis and TS mutants were selected by comparing the growth at the permissive (25ºC) and restrictive (32ºC) temperature. Second, the inhibited growth of TS mutant plants in terms of growth rate and morphology was evaluated by fluorescence microscopy and analysis of morphological parameters such as area and solidity. Third, the F1 segregants of the isolated TS mutants were generated by outcrossing them to a fluorescently labelled polymorphic strain of *P. patens*. Lastly, the reasonable size of mapping population was estimated using a Monte Carlo simulation based on crossover landscape of F2 outcrossed *Arabidopsis*.

A total of eight mutants were selected after UV mutagenesis and screening. According to the morphological characterization of the eight mutants at 20ºC, 25 ºC and 32 ºC, six of the eight mutants displayed defects in cell growth and division at 32ºC while had no significant difference from the wild type control at lower temperatures. The control of the six mutants showed only small changes in size at 32ºC on day4. In order to map the causal mutation of the TS phenotype of the isolated LV767, outcrossing was conducted with a fluorescently labeled polymorphic strain Vx::mCherry and 24 F1 segregants with the TS phenotype at 32ºC were selected as the mapping population. According to the Monte Carlo simulation, 24 F1 segregants should narrow down the range of the causal mutation to one chromosome within 3 million base pairs. Based on existent and preliminary studies, the sequencing was designed to be done pair-ended (90nt per read) with10X coverage. The most important future study will be mapping the causal mutation using the genome sequencing reads of pool segregants.

Background

1.1 Why Do We Need to Study Plants?

The following sections explain why it is important for us to study plants. The topics include what plants are, what some critical function plants have, and what some serious challenges the world is facing that are also closely related to plants.

1.1.1 What Are Plants?

The definition and classification of plants has been changing while the knowledge of plant structure and function improves (Batra, 2009). Some organisms that once were categorized as plants, such as several algae and fungi, were removed from the Kingdom Plantae (Batra, 2009; Gaston and Spicer, 2004). According to the classification of Margulis and Schwartz, there are four Kingdoms of organisms including Bacteria, Protoctista, Animalia and Plantae (Gaston and Spicer, 2004). There are about 350,000 plant species, and up until 2004, 287,655 of those had been identified as flowering plants, bryophytes, ferns or green algae (Abedon, 1997; Gaston and Spicer, 2004; Raven et al., 2005). The classification of plants is based on their structures and evolutionary relationships (Abedon, 1997; Gaston and Spicer, 2004; Prigge and Bezanilla, 2010). Similar to animals, it is suggested that plants were evolved from living in water (algae), to wetland (bryophytes), and ultimately to dry land (vascular plants) (Abedon, 1997; Gaston and Spicer, 2004; Prigge and Bezanilla, 2010). All plants have cell walls made of cellulose and most of them (except for about 300 parasites) are autotrophs which obtain energy through photosynthesis (Abedon, 1997; Batra, 2009).

1.1.2 Why Are Plants Important?

Plant vegetation is a critical part of the biosphere which is defined as the space where life exists on the Earth (Batra, 2009; Robbins, 1944). Plants play an important role not only in the ecosystem by fixing carbon, assisting water and nitrogen cycles, preventing soil erosion, providing shelters for animals, etc., but also in the development of human society by providing products such as food, medicine, fibers, and clean energy and helping scientific research to understand mechanisms of life.

Some Important Ecological Function

Plants are indispensable components of ecosystems and exist in most habitats of the Earth (Batra, 2009; Hussain, 2009). Through photosynthesis, plants utilize solar energy, covert carbon dioxide into carbohydrates and produce oxygen (Hussain, 2009; Robbins, 1944). Many organisms that are not able to use sunlight or other abiotic energy or live in anaerobic condition must rely on plants as their source of energy and oxygen (Raven et al., 2005; Robbins, 1944). Therefore, plants play an important role in the carbon cycle of the ecosystem.

In addition to carbon cycle, plants also participate in nitrogen cycle and water cycle and inhibit soil erosion (2010; Batra, 2009; Gyssels and Poesen, 2003; Hussain, 2009). Some plants can fix nitrogen using bacteria which have coevolved with the plants (Batra, 2009). In the water cycle, different types and different parts of plants have different function. Tree canopies, for instance, transfer water absorbed from soil into the atmosphere through evaporation from leaves and stems in a process called "transpiration" (2010; Hussain, 2009). The root system of plants, living and dead, are underground pathways which help water flow from ground to lower soil levels (Gyssels and Poesen, 2003). While assisting the water cycle, plants also protect the ground soil from erosion. Tree canopies can prevent strong wind blowing surface soil away, whereas root systems secrete binding agents which promote the production of other binding agents from the microorganisms (Gyssels and Poesen, 2003; Hussain, 2009).

Contribution to Human Society

Plant products have been facilitating and promoting the existence and development of human society and civilization. Three examples of the most important products are food, fuel and medicine.

Plants are one of the major food sources of human. Approximately 12,500 flowering plants are edible while 200 species have been domesticated (Gaston and Spicer, 2004). According to the statistics of the Food and Agriculture Organization of the United Nations (FAO), grains and stable crops are the major sources of food globally although their proportions in total food consumption have been decreasing along with income growth (Food and Agriculture Organization of the United Nations). The estimated global progress in food consumption is shown in **Figure 1**.

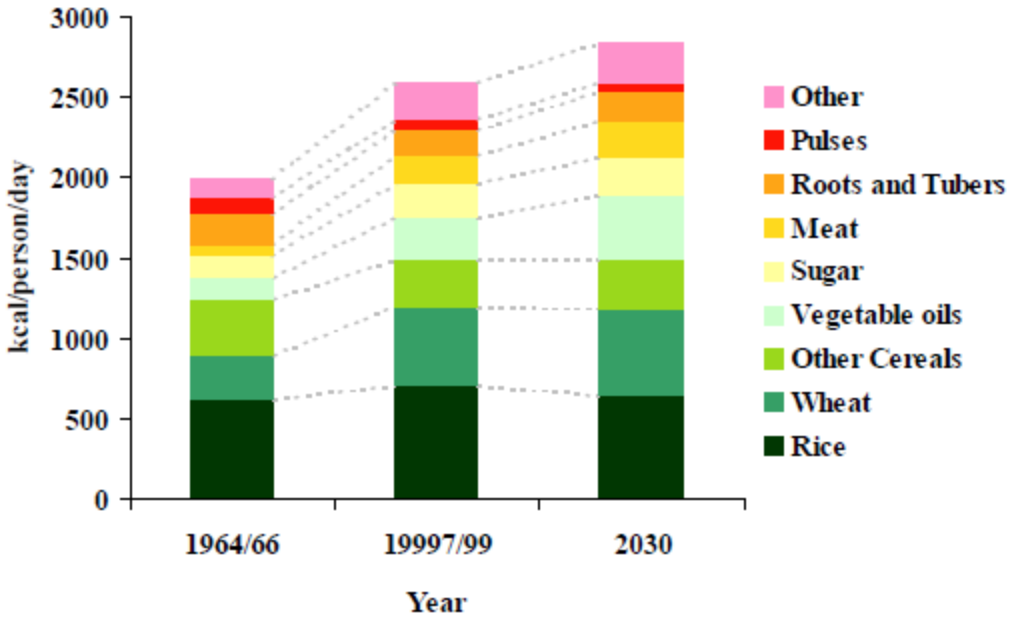


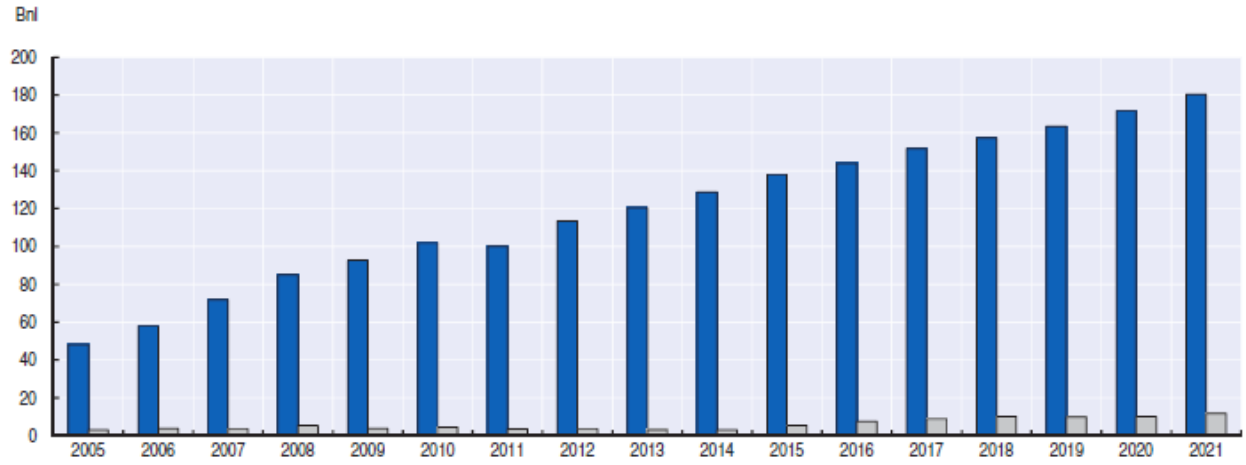
Figure 1: Global food consumption comparisons. The estimated global progress in food consumption (kcal/person/day) in 1964-66, 1997-99 and 2030 (Food and Agriculture Organization of the United Nations).

Additionally, plants are also one of the major sources of energy and potential sources of renewable energy. Fossil fuels were formed from prehistoric plants and animals (U.S. Department of Energy, 2014). Among all the fossil fuels, coal was formed from the remains of plants which lived 300-400 million years ago (U.S. Department of Energy, 2014). According to the statistics from the World Coal Association, nearly 30% of global energy needs are/were fulfilled by coal burning (World Coal Association, 2014). Nonetheless, fossil fuels are not renewable energy and their burning produces enormous pollutants and greenhouse gases.

People have been exploring replacements of fossil fuels that are renewable and cause less pollution. Biofuels generated from plants are one of the candidates (Timilsina and Shrestha, 2010). Ethanol and biodiesel are two major biofuels used for transportation (Chang, 2007; Nations et al., 2012; Yuan et al., 2008). Ethanol can be generated by fermentation of simple sugars, starches or other processed biopolymers from plants (Chang, 2007) while biodiesels is produced by transesterification of lipids from plants (Timilsina and Shrestha, 2010). Currently, agricultural feedstock such as sugarcane, maize, and oilseeds are the major raw materials for production of bio-ethanol (Chang, 2007; Food and Agriculture Organization of the United Nations). Although the shares of biofuel in transport fuel use of most biofuel producing countries are currently less than 5%, the production of ethanol and biodiesel from plants are growing

steadily as shown in **Figure 2** (Food and Agriculture Organization of the United Nations et al., 2012).

A.



B.

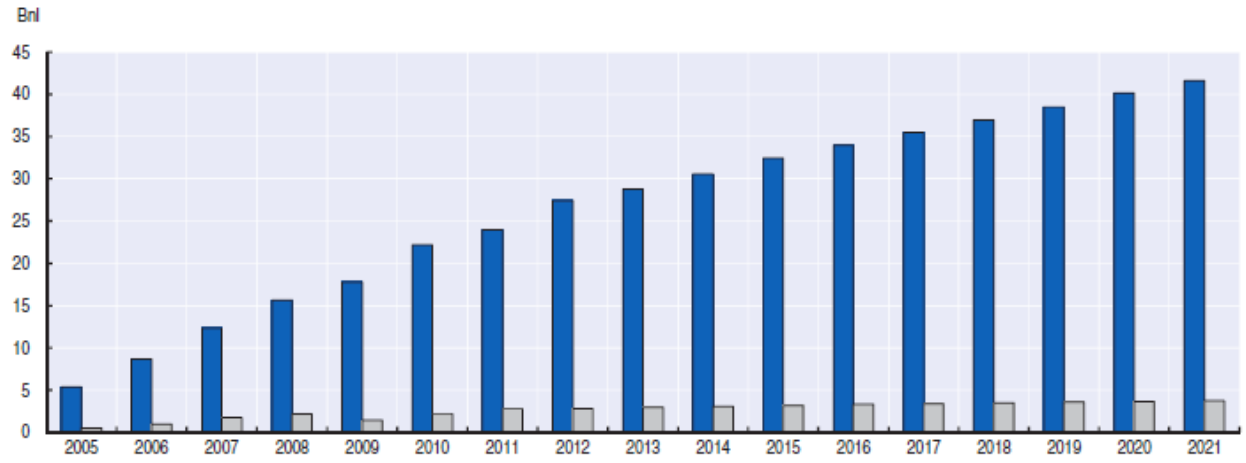


Figure 2: Estimated world trade of ethanol and biodesel (2005-2021). Estimated ethanol (Panel A) and biodiesel (Panel B) production and trade from 2005 to 2021 (Food and Agriculture Organization of the United Nations et al., 2012). The blue bars represent production and gray bars represent trade.

Some plants have been used as herbal medicine for hundreds and thousands of years. Before the chemically synthesized drugs were developed, herbal drugs were essential in medication. Records of herbal medicine were found all over the world such as in China, Greece, India, Rome and the Middle East (Fowler, 2006). Chinese people started to use and document herbal medicine about 2,000 years ago (Benzie and Wachtel-Galor, 2011; Fowler, 2006). Even today, it is estimated that about 90% and 70% of the population in Africa and India respectively,

depend on herbal medicines for their primary health care, while more than 90% of the general hospitals in China still uses herbs in their traditional medicine units (Benzie and Wachtel-Galor, 2011).

Not only have herbs been used directly as a medicine, they also have significant contributions to pharmaceutical discovery and drug development. According to the statistics from WHO in 2001, approximately 11% of drugs were derived from plants (Fowler, 2006). Some examples of major plant-derived drugs are displayed in **Table 1**. The reason why plants are so useful as drugs is that they have developed systems to contend with disorder of their internal systems and external pressure in evolution while causing least harm to themselves (Fowler, 2006; Gurib-Fakim, 2011). Therefore, herbal medicines usually have less adverse effects and sometimes even more efficient in treating diseases than synthetic drugs (Benzie and Wachtel-Galor, 2011; Fowler, 2006; Garg and Adams, 2012).

Table 1: Some major drugs derived from plants (Fowler, 2006, p1799).

Drug	Source	Pharmacological action	Chemical family
Artemisinin	*Artemisia annua*	Antimalarial	Sesquiterpene, lactone
Codeine, morphine	*Papaver somniferum*	Analgesics	Opiate alkaloids
Cocaine	*Erythroxylum coca*	Local anaesthetic	Cocaine alkaloid
Digoxin	*Digitalis pupurea*	Cardiotoinc	Steroidal glycoside
Galanthamine	*Leucojum aestivum*	Cholinesterase inhibitor	Isoquinoline alkaloid
Quinine	*Cinchona ledgeriana*	Antimalarial	Quinoline alkaloid
Taxol	*Taxus brevifolia*	Antineoplastic	Diterpene
Vincrystine, vinblastine	*Catharanthus roseus*	Antineoplastic	Bis-indole alkaloids

Some Critical Challenges Faced by Human Society

With the increase in world population, the demand for food and energy has been increasing. Although the production of energy and food has been increasing rapidly in the past half century (Fuglie and Nin-Pratt, 2012; U.S. Energy Information Administration, 2013), it is questionable whether such an increase is sustainable or not. The growth in crop yields can mainly be attributed to the expansion of farming land, the increased use of agrochemicals such as fertilizer, herbicides and pesticides, and the increase in irrigated area (Food and Agriculture Organization of the United Nations, 2003; Kiers et al., 2008). Although the amount of renewable and nuclear energy generated has been increasing in the past two decades, most energy use still relies on fossil fuels including petroleum, coal and natural gas (U.S. Energy Information Administration, 2013). As a result, there are some serious side-effects and potential problems that need to be contended with.

It is estimated by the U.S. Census Bureau that the world population is currently close to seven billion and may reach nine billion before 2050 as shown in **Figure 3**. In order to fulfill the demand of food and energy, the world must keep increasing the production of crops and fuels. Nonetheless, the trade-offs and limitations of the aforementioned approaches to achieve this goal must be aware and dealt with first. In terms of crop production, modern agriculture has several major limitations and negative impacts on the environment. For instance, arable lands and clean water are limited and the conversion of forests to farming land causes land degradation (Food and Agriculture Organization of the United Nations, 1997, 2003; Pimentel et al., 2005). In addition, the use of agrochemicals have caused increased emission of greenhouse gases such as methane and nitrous oxide and also have caused nitrate and herbicide leaching which causes pollution of water sources (Food and Agriculture Organization of the United Nations, 1997, 2003; Pimentel et al., 2005). As for energy production, fossil-fuel reserves are limited and overexploitation will likely lead to higher greenhouse gases emission and dramatic change in global climate (Food and Agriculture Organization of the United Nations; Zecca and Chiari, 2010). Therefore, it seems that the world should increase its crop and energy production in a more sustainable way which relies less on fossil fuels and causes less environmental damage.

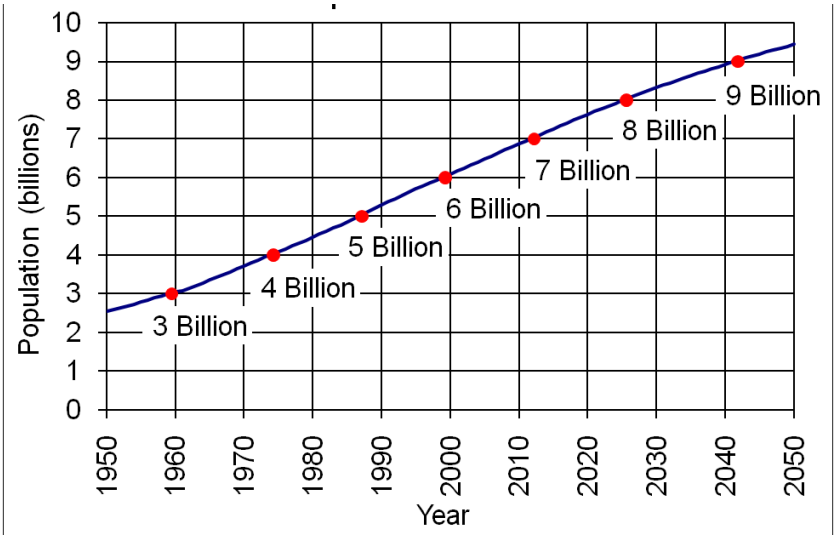


Figure 3: The estimated world population from 1950 to 2050 (U.S. Census Bureau, 2011).

1.2 How Can We Sustainably Increase Crop Yields and Energy Generation?

After learning why plants are important and how they are closely related to some serious challenges and problems the world is facing, it is critical to explore and seek ways to encounter the challenges and solve the problems. The following sections address how we can encounter the

aforementioned challenges with plants and plant biology and how can we study critical gene function in plants.

1.2.1 What Can We Do With Plants to Encounter Challenges Faced by Human Society?

As mentioned in the previous section, the challenges to feed and power the world in a sustainable way are closely related to how to increase crop yields without over-expansion of arable lands and massive use of agrochemicals and water (Food and Agriculture Organization of the United Nations). In other words, agriculture needs to be transformed from adapting the environment to crop growth into adapting crops to environmental changes (Job, 2002; Mannion, 1995). Therefore, modifications need to be done to crops so that they can have features such as resistance to pests and diseases, tolerance to different temperatures, drought, salinity and herbicides, and increased nutritional quality (Food and Agriculture Organization of the United Nations; Job, 2002; Sharma et al., 2002). The achievement of such modifications relies on biotechnology, such as plant breeding and genetic engineering, to transfer genes that control specific traits into the target plant.

Traditional plant breeding relies on the principles of heredity discovered in the nineteenth century and has been working effectively so far in producing crops with better quality (Mannion, 1995; Sharma et al., 2002). However, two limitations of traditional plant breeding are that it can only be done with relatively close-related species and that the process takes a long time (usually seven to ten years) and requires a lot of progenies for selection (Sharma et al., 2002). Modern genetic engineering and genomics have made encouraging progress in overcoming these two limitations. Transformation and site-specific recombination make it possible to transfer genes of one organism to another totally unrelated organism (Job, 2002; Mannion, 1995). Advances in genomics and transcriptomics of plants, especially the sequencing of complete genome and the identification of genes, have helped to enhance understanding of plant genome and transcriptome function, which in turn helps identification of more genes and gene pathways (Varshney et al., 2005). In modern agriculture, some genes are used as functional markers in marker-assisted crop selection where molecular markers are used for indirect selection of desired traits at the seedling stage (Sharma et al., 2002; Varshney et al., 2005). The combination of transformation and marker-assisted crop selection can speed up the crop breeding which usually takes three to six years (Sharma et al., 2002; Varshney et al., 2005).

Identification of genes that are critical in plant cell growth, division and response to various biotic and abiotic stresses, therefore, is one key to the improvement of crops. Although some genes that play an important role in plant metabolism, growth, photosynthesis, resistance to pests and diseases, and tolerance to abiotic stresses have been identified and modified in crops to increase their yields, some mechanisms such as plant response and tolerance to abiotic stresses are not yet well-understood (Job, 2002; Mannion, 1995; Sharma et al., 2002; Varshney et al., 2005). For example, plants' heat stress response, which is essential in plants' heat tolerance, is well known for its complexity (Kotak et al., 2007; Mittler et al., 2012; von Koskull-Döring et al., 2007). Although many heat stress sensors and signaling molecules and pathways have been identified in plants, such as calcium channels, various kinases and transcription factors, lipid signaling molecules, unfolded protein response pathway, reactive oxygen species, heat shock proteins, etc., and the signaling processes and mechanisms are not well-understood (Kotak et al., 2007; Mittler et al., 2012; von Koskull-Döring et al., 2007). In order to improve plants adaptation to a changing environment, it is crucial to study and understand these process and mechanisms in plants, especially how genes control them.

1.2.2 How Can We Study Genes That Are Critical for Plant Growth?

The following sections summarized some major ways to study gene function. The emphases are in major ways of studying gene function which include forward and reverse genetics and one specific strategy of forward genetics – the isolation of TS mutants, some important features of *P. patens*, which is the model system used in this study, and some background on genomic mapping of mutated genes.

Forward Genetics, Reverse Genetics and TS Mutants

There are two major ways of studying gene function: first, mutants are generated where those with desired phenotypes are selected and the causal mutation is then mapped to the genome; second, a gene is already known and the function of the gene is then deduced by mutating or silencing the gene (Alonso and Ecker, 2006; Lackie, 2010; Martin and Hine, 2008). The first and second approaches are named forward and reverse genetics respectively. There are advantages and disadvantages in each of the two approaches. Reverse genetics is good at revealing overlapping and specific functions of related gene families, since multiple genes can be knocked out at the same time, but this approach has limitations in detecting new genes because it requires genes with known sequences (Alonso and Ecker, 2006). Forward genetics can produce specific

mutant phenotypes with high reproducibility and gene discovery in this process is unbiased, but it is usually hard to analyze genes with high redundancy and the mapping of causal mutation can be laborious and time-consuming (Alonso and Ecker, 2006). Fortunately, progress in high-throughput sequencing, genome alignment, and genomic mutation detection and mapping algorithms has reduced the amount of work required for causal mutation mapping in forward genetics (Darby and Hall, 2008).

Although loss-of-function phenotypes produced by gene silencing or gene deletion provide valuable information on gene function, this strategy is not suitable to study essential genes which are required for viability and pleiotropic genes which function at multiple places or times in the life cycle (Tan et al., 2009). One effective approach to study these genes is the isolation of temperature-sensitive (TS) mutants (Seiler and Plamann, 2003; Suzuki, 1970; Tan et al., 2009). In most cases, TS mutations affect cell function differently at different temperatures where it is called restrictive temperature if the mutants are less active than the wild type and called permissive temperature if the mutants' phenotype and functionality are similar to the wild type (Bajaj et al., 2008; Lockwood et al., 2011; Tan et al., 2009).

Since TS mutations provide a way to study gene function *in vivo*, TS mutants have been widely used in the study of gene function in various organisms including viruses, bacteria, fungi, *Drosophila*, *Caenorhabditis elegans*, plants, and mammalian cells (Lockwood et al., 2011). The classical method of generating TS mutants involves random mutagenesis and genetic screening which could be laborious in multicellular organisms (Lockwood et al., 2011). As a result, some studies have tried to design TS mutants using the sequence and tertiary structure of proteins and create the mutants via site-directed mutagenesis (Bajaj et al., 2008; Vidali et al., 2009). It was found in these studies that although TS mutations have no simple patterns, they can be located on ligand-binding and buried sites of proteins. Another study investigated the sequence and structural features of TS mutations using machine learning and has a similar conclusion to previous studies that TS mutations are like to occur at rigid or buried sites in conserved domains (Lockwood et al., 2011). TS mutations have been found in various genes and cellular processes. For example, in *Neurospora crassa*, a filamentous fungus, TS mutations have been isolated that affect mycelial growth and morphology (Schmit and Brody, 1982), circadian clock (Hunt et al., 2012), protein synthesis (Loo, 1975), riboflavin synthesis (Mitchell and Houlahan, 1946) and hyphal growth (Seiler and Plamann, 2003). TS mutants have not been widely used in the study of

plant genes, but some studies, such as CDKG1 protein kinase's function in synapsis and male meiosis (Zheng et al., 2014) and critical genes affecting organogenesis in *Arabidopsis* (Sugiyama, 2014), showed the great potential to study plant genes with TS mutants.

The Moss *Physcomitrella patens*

P. patens is in the family *Funariaceae*, which are a group of "short-lived, minute to medium-sized, light to yellow-green and annual to biennial plants that grow gregarious to open tufts", and order *Funariales* as a member of the class *Bryopsida* (Lang et al., 2008). Mosses (Bryophytaea) are one of the oldest plant groups and are very likely to be a representative during plants' water-to-land transition (Lang et al., 2008; Prigge and Bezanilla, 2010; Schaefer and Zrÿd, 2001). In support of this argument, it has been found that many critical genes in the development of flowering plants are conserved in *P. patens*, but the function of some genes has been changed in flowering plants (Prigge and Bezanilla, 2010). Proposed phylogenetic relationships among green plants are shown in **Figure 4**. As a result, *P. patens* is a good model for studying the function and structure of related genes by comparing to genes of other plants.

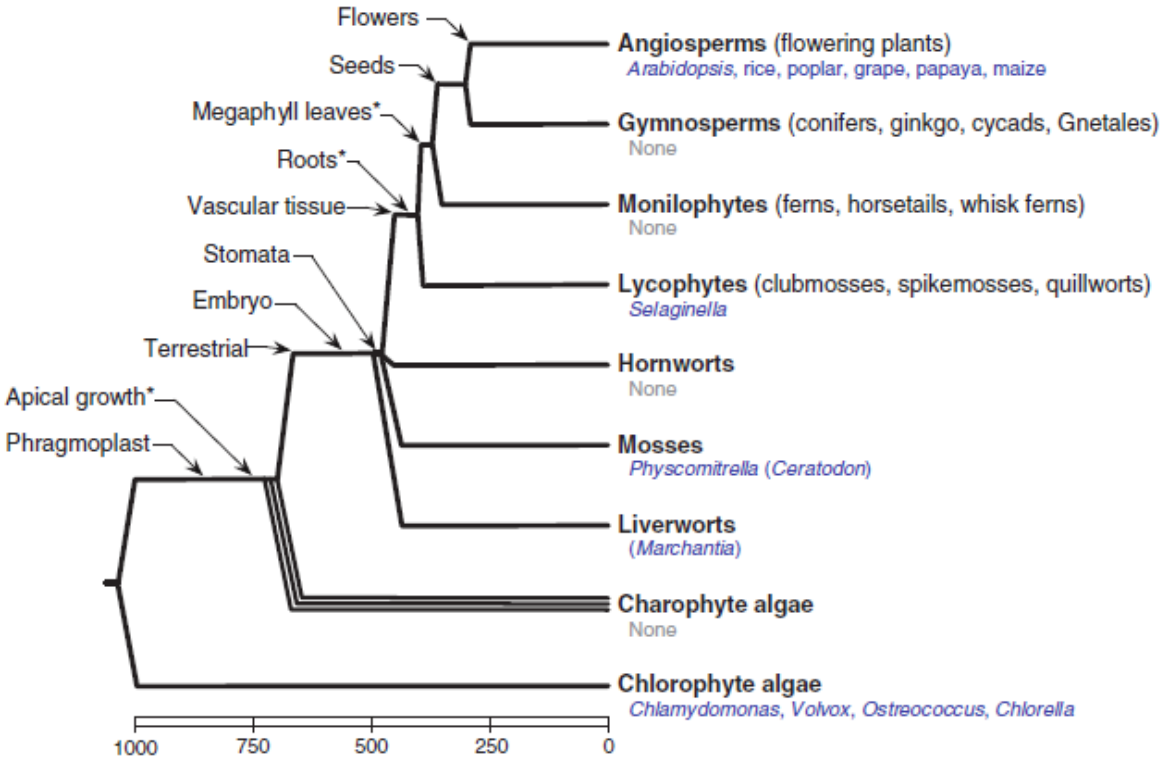


Figure 4: Evolutionary relationships of major types of plants. This dendrogram displays the proposed relationships between main groups of green plants. The time scale at the bottom indicates the approximate time of lineage divergence (million years). Asterisks indicate the specific trait may have evolved independently in multiple lineages after the point of divergence. Blue texts are plants with complete genome sequences. (Prigge & Bezanilla, 2010, p3536).

The life cycle of mosses, similar to those of ferns and flowering plants, is consisted of a haploid phase and a diploid phase (Cove, 2005). In contrast to ferns and flowering plants, the gamete-producing haploid state (gametophyte) is the dominant phase of mosses whereas the spore-producing diploid state (sporophyte) only exists for a short time (Cove, 2005; Frank et al., 2005). Different stages of the life cycle of *P. pates* have been well studied and the following description gives a brief summary (Cove, 2005; Frank et al., 2005; Prigge and Bezanilla, 2010; Schaefer and Zrÿd, 2001). The gametophyte phase begins with the germination of a haploid spore where a germ tube grows out. The germ tube extends by serial divisions of its apical cell and forms protonema filaments which can be classified into chloronema and caulonema cells. Chloronema cells have a great number of chloroplasts and grow relatively slow (2-5 μm/h and divide every 22-26 h) whereas caulonema cells have fewer chloroplasts and extend much faster (25-40 μm/h and divide every 6-8 h). Subapical cells (chloronema and caulonema) also undergo division to produce branches. A few caulonema branches will develop into gametophores which are leafy shoots and will later develop into gametangia. *P. patens* is monoecious, which means its antheridia (male sex organs) and archegonia (female sex organs) are produced on one plant. Then spermatozoids (male gametes) are generated by the antheridia and swim on a water surface using flagella to reach the archegonium where the egg cell is generated. After fertilization, the zygote will develop into a sporophyte from which about 4,000 haploid spores can be generated. A diagram of *P. patens'* life cycle is shown in Figure 5**.**

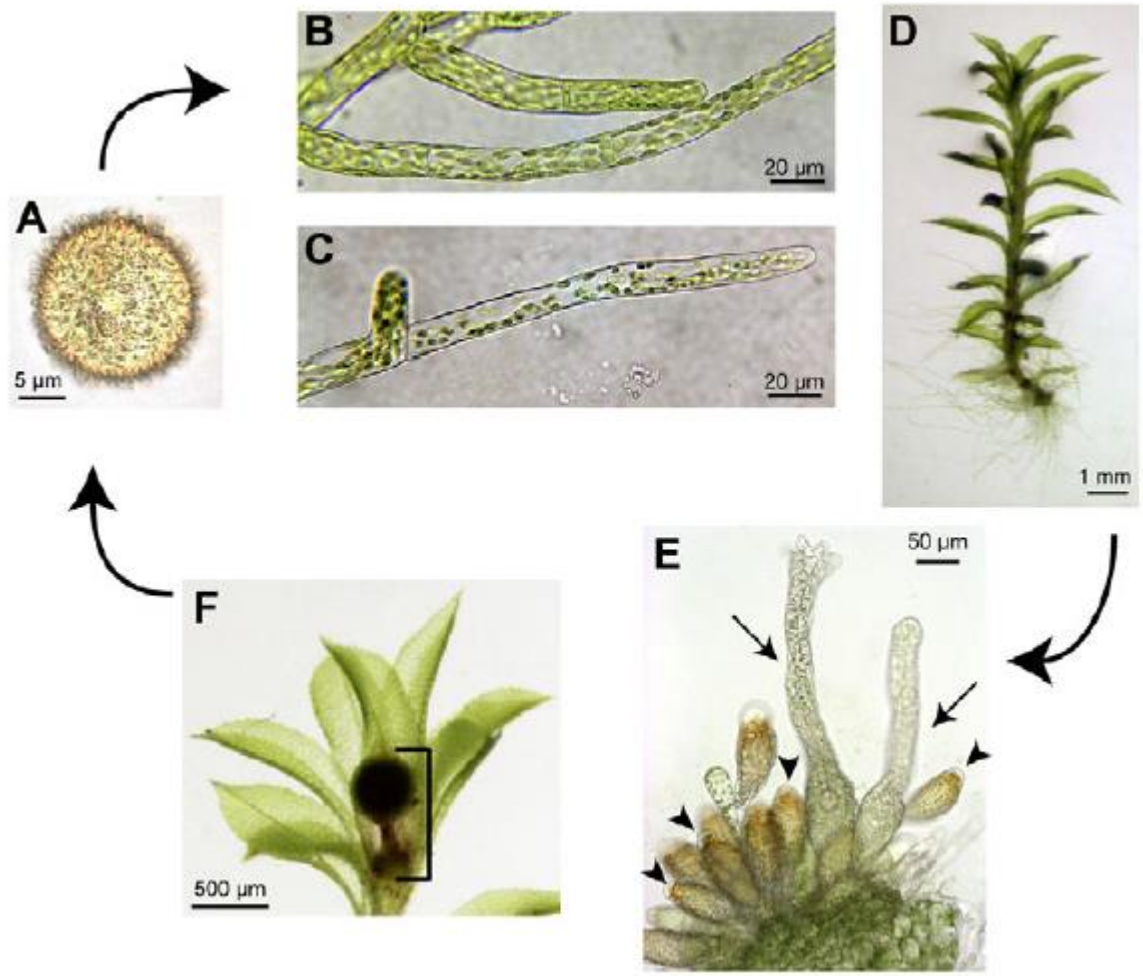


Figure 5: The life cycle of *P. patens*. A. a haploid spore. B. chloronema cells which can differentiate into C. caulonema cells. D. a gametophore. E. at the apex of gametophores, antheridia (arrowheads) and archegonia (arrows) form. F. a sporophyte. (Prigge & Bezanilla, 2010, p3537).

P. patens is very easy to keep and grow in the lab and several powerful genetic and genomic tools have been developed in *P. patens* which makes it an ideal model system to study gene function and cellular processes. *P. patens* can be expanded in lab by sexual reproduction or tissue culture and can be grown on solid or liquid medium (Cove, 2005). Mutagenesis with irradiation and chemicals can be done using spores or protonema-digested protoplasts (Cove, 2005). Transformation and RNA interference (RNAi) can also be achieved in *P. patens*. Transformation is usually achieved through polyethylene-glycol (PEG)-mediated DNA uptake by protoplasts and homologous recombination into specific genomic sites (Cove, 2005; Frank et al., 2005). Transformation can be used for targeted gene knock-in and knockout (Cove, 2005; Frank et al., 2005). In some cases of gene inactivation, knockout of a gene using transformation

may not give rise to phenotypic changes due to redundancy of gene function (Cove, 2005). RNAi may then be a solution to the problem since it relies on the generation of short double-stranded RNA to silence the expressed mRNAs with similar sequences, and thus RNAi may be able to silence mRNAs expressed by families of genes with similar sequences (Cove, 2005). In addition, a draft genome sequence of *P. patens* was completed, assembling into about 480 Mbp of scaffold sequence (Rensing et al., 2008). Expressed sequence tag (EST) covers over 98% of the whole genome and 35,938 gene models are included in the genome sequence (Rensing et al., 2008).

With all the aforementioned features and tools, *P. patens* is a useful model system for both forward and reverse genetic studies. The dominant haploid state and complete genome sequence provide a great advantage in mutant screening and causal mutation mapping of forward genetic studies while transformation and RNAi are great tools in reverse genetic studies.

Genomic Mapping of Mutated Genes

In forward genetics, mapping and cloning of the mutated gene is a key step to study gene function and how the mutation affects the phenotype after the isolation of mutants. The conventional strategies of mutation mapping involve genome-wide mapping, which usually is conducted through recombination analysis using known genetic markers, and candidate gene sequencing (Galvão et al., 2012; Lodish et al., 2000; Schneeberger et al., 2009). Such a process is usually tedious, time-consuming and as a result not very cost-effective (Hobert, 2010). With the rapid progress in whole genome sequencing (WGS) and subsequent data analysis algorithms, many studies have successfully identified the causal mutation in their model organisms such as *C. elegans* and *Arabidopsis* which already have an assembled genome (Austin et al., 2011; Doitsidou et al., 2010; Galvão et al., 2012; Sarin et al., 2008; Schneeberger et al., 2009; Zuryn et al., 2010).

In the aforementioned studies using WGS for mutation detection, two main strategies were used which include direct mapping and mapping with pooled segregants. Direct mapping can be conducted by sequencing individual backcrossed mutants which maintain the phenotype of interest (James et al., 2013; Zuryn et al., 2010). The purpose of backcrossing is to eliminate enormous background mutations that are introduced during mutagenesis (James et al., 2013; Schneeberger and Weigel, 2011). The advantage of this approach is that it does not require a large population of segregants which are usually needed for mapping using pooled segregants. Nonetheless, a few (4-6) rounds of backcrossing and signature mutation types of the mutagenesis

are required in order to detect the causal mutation (Zuryn et al., 2010). It is also recommended to independently sequence at least two distinct mutants from the same mutagenesis in order to subtract common variants between the two mutant lines which are likely to be natural variants accumulated in the wild type (Zuryn et al., 2010). By detecting the region with relatively dense background mutations (signature mutations of the specific mutagenesis methods used), the approximate location of the causal mutation can then be determined (Zuryn et al., 2010). Mapping by pooled-sequencing of segregants, instead, requires a relatively large population of segregants which maintain the phenotype of interest and are commonly generated by outcrossing mutants to a polymorphic strain, preferably with known single nucleotide polymorphism (SNP) markers (Austin et al., 2011; James et al., 2013; Sarin et al., 2008; Schneeberger et al., 2009). With enough sequencing depth, the pooled-sequencing of a population of F2 progeny (diploid organisms) or F1 progeny (haploid organisms) with sufficient individuals can be used to locate the causal mutation to a relatively small region of the genome (Austin et al., 2011; James et al., 2013; Sarin et al., 2008; Schneeberger et al., 2009). By observing the recombination of genetic markers such as SNPs from the polymorphic line with the mutant genome, the genomic region containing the causal mutation will have decreasing rates of the genetic marker recombination on both ends (Austin et al., 2011; Doitsidou et al., 2010; Schneeberger et al., 2009).

In the experimental design of causal mutation mapping using WGS, two critical parameters are the size of the mapping population and the amount of sequencing data (Austin et al., 2011; Doitsidou et al., 2010; James et al., 2013; Zuryn et al., 2010). In a study where direct mapping of causal mutations in *C. elegans* was conducted, it is suggested that a paired-ended sequencing of 57nt/read and an average depth of 13x should be sufficient. In studies using pooled-sequencing of segregants, the pool size and average sequencing depth varies from 20 to 500 and from 6x to 41x respectively while the detected range of the causal mutation does not change linearly with the two parameters (Austin et al., 2011; Doitsidou et al., 2010; James et al., 2013; Schneeberger et al., 2009). According to experimental studies and *in silico* simulations, it is suggested that the size of the mapping population and amount of sequencing data needed to detect the causal mutation should be affected by recombination rates during crossover and the location of the causal mutation (Austin et al., 2011; Doitsidou et al., 2010; James et al., 2013; Schneeberger et al., 2009). As a result, there is no uniform guideline for experimental design,

and the two key parameters need to be adjusted based on the specific model organisms used and the trade-off between mapping accuracy and costs.

To detect the region that contains the causal mutation, the analysis of sequencing data can be achieved using existent software such SHOREmap and CloudMap (Minevich et al., 2012; Schneeberger et al., 2009). SHOREmap is an extension of the sequencing read analysis pipeline SHORE and it detects the region containing causal mutation by calculating and plotting the relative allele frequencies of mapping parents (Schneeberger et al., 2009). CloudMap is a cloud-based, genome analysis pipeline running on Galaxy platform (Minevich et al., 2012). In theory, both SHOREmap and CloudMap can be used for causal mutation mapping in any model organism with an assembled genome but a potential advantage CloudMap has is that it is browser-based and therefore requires less technical expertise (Minevich et al., 2012).

Materials and Methods

The following sections explain the materials and methods used to achieve the stated objectives. Details about moss proliferation and maintenance, protoplasting, UV mutagenesis, TS mutant screening, morphological characterization of TS mutants through growth assays, crossing and genomic analysis are described here.

2.1 Moss Proliferation and Maintenance

All plants used in this study (except during crossing) were proliferated on solid PpNH4 medium agar plates at the designated temperature (20ºC, 25ºC or 32º C), depending on purpose of use, under a cycle of 16 h light (90 µmol m-2sec-1) and 8 h dark. Components of required stock solutions and the components of liquid PpNH4 medium are listed in Table 2**.** To prepare solid $PpNH_4$ medium, plant agar, 7g/L, was added to the liquid medium before autoclave. The plants were maintained and passed on fresh medium every week. Plant tissue was ground with a homogenizer (Power Gen 125, Fisher Scientific) and transferred onto solid PpNH4 medium plates overlaid with a piece of cellophane.

Table 2: Stock solutions for media and ingredients of 1L PpNH4 medium.

Stock Solutions for Media	Stock Solutions	Final Volume (mL)	Weight (g)
	$MgSO_4 \cdot 7H_2O$ (500x)	500	61.6
	KH_2PO_4 (500x)	500	62.6
	$Ca(NO_3)_2 \cdot 4H_2O$ (500x)	500	200.7
	$CaCl_2$ (1M)	500	29.4

Components of 1L PpNH4 Medium	Ingredients	Amount
	$MgSO_4 \cdot 7H_2O$ (500x)	2mL
	KH_2PO_4 (500x)	2mL
	$Ca(NO_3)_2 \cdot 4H_2O$ (500x)	2mL
	Di-ammonium tartrate	0.5g
	$FeSO_4 \cdot 7H_2O$	12.5mg
	Micro Elements (1000x)	1mL
	H_2O	Up to 1L

2.2 Moss Protoplasting

One-week old moss, which was proliferated according to description in 3.1, from 1-2 plates was harvested and incubated for 1 hour with a digestive solution [0.5% (w/v) driselase (laminarinase, xylanase and cellulase)/8% (w/v) mannitol] in order to get rid of the cell wall. The protoplasts were sieved through 70µm mesh to remove large debris and centrifuged at 250x g for 5 min. The pellet of protoplasts was re-suspended in 10 ml of 8% mannitol and washed two more times.

2.3 UV Mutagenesis and Mutant Selection

Wildtype *P. patens* Gransden were protoplasted according to procedure described in 2.2. The protoplasts were re-suspended in 1-2mL liquid $PpNH_4$, which also contains 8% mannitol and 10mM $CaCl_2$, after the final centrifugation. The cells from the protoplast suspension of each strain were counted using a hemocytometer. For each strain, ten PRMB plates were labeled with the name of each strain and overlaid with a piece of cellophane, and 500,000 re-suspended protoplasts of the corresponding strain were distributed onto each plate and spread evenly. The ingredients of PRMB agar medium were different from those of $PpNH_4$ agar medium only in the addition of 60g/L mannitol, the amount of plant agar, which was 10g/L instead of 7g/L and the addition of 10mL/L 1M $CaCl_2$ before pouring the medium into 90mm petri dishes. The plates were uncovered and irradiated with 1,100J/m^2 UV light (about 90% killing rate) and were cultivated at 25ºC under a cycle of 16 h light (90 µmol m-2sec-1) and 8 h dark for 4 days. After 4 days, the plates were cultivated at 32ºC with the same light condition as previously for one week. The mutant plants were then re-suspended with 12mL sterile $PpNH_4$ medium and selected by sieving the plants through a 400µm mesh. The selected mutants were re-plated at 32ºC for another week. The plants showing a mutant phenotype were identified, cultured at 25ºC on $PpNH_4$ agar plates, and tested for temperature-sensitivity by proliferating each line on two $PpNH_4$ agar plates and cultivating at 25ºC and 32ºC respectively.

2.4 Growth Assays

The control and TS mutants were protoplasted according procedure described in 3.2. The protoplasts were re-suspended in 1-2mL 8% mannitol. The cells from the protoplast suspension of each strain were counted using a hemocytometer. Each moss line was re-suspended in 2mL

melted PRM-T agar kept at no higher than 47 ºC, 5K, 25K and 50K cells/mL and, distributed onto PRMB plates (90mm petri dishes) overlaid with a piece of cellophane, and quickly spread the agar evenly. The protoplasts were regenerated at 25ºC under a cycle of 16 h light (90 µmol m-2sec-1) and 8 h dark for 4 days. After 4 days, the cellophanes were each cut into three equal pieces where each one was transferred to a PpNH4 plate (Day 0). The three cellophane pieces were cultivated at 20ºC, 25ºC and 32ºC respectively under a cycle of 16 h light (90 µmol m-2sec-1) and 8 h dark.

Microscopy was performed on Days 2-4 using the following concentrations per day: Day 2: 50K. Day 3: 25K. Day 4: 5K. The different concentrations were used to ensure sufficient numbers of separated moss plants can be captured in micrographs while the plant sizes increase. The cell walls of moss were fluorescently stained with10 µg/mL calcofluor diluted in water. Squares about the size of a coverslip were cut from the cellophane with PRM-T agar poured on top. Then, 30 µL of the calcofluor was added to the slide and the sample was inverted onto the slide. The cellophane was removed, another 20 µL calcoflour was added and a coverslip was placed with caution to avoid bubbles. The slide was sealed with melted VALAP and kept away from light to prevent bleaching.

Imaging was performed at 10X objective magnification (no optovar) using a Zeiss Axiovert 200M microscope, which is based on an inverted microscope base equipped with a motorized platform and Coolsnap fx CDD camera. Zeiss Axiovision software was used to create an overlapping grid pattern of 200-300 pictures. Parameters such as plant area and solidity (plant area/convex hull area) were measured using existing macros of ImageJ (**Appendix A**).

Log(area) was computed for all plant areas in order to normalize the data. In order to determine if there was a significant difference in plant area and solidity when the moss mutants and controls were grown at 20°C, 25°C and 32°C, ANOVA-Tukey test and non-parametric Kruskal-Wallis test were conducted using the GraphPad to reject equivalence of means and the statement that the samples come from populations with the same data distribution. The adjusted P values of comparing log(area) and solidity respectively at 20°C vs. 25°C , 20°C vs. 32°C, and 25°C vs. 32°C on both day3 and day4 of each mutant line and control were obtained from the tests. It was assumed in this study that the difference was not statistically significant if the adjusted P values were smaller than 0.05. In addition, four box plots were created using

MATLAB to summarize the data distribution of log(area) and solidity of all mutants and control at 20°C, 25°C and 32°C on day3 and day4.

2.5 Moss Crossing

TS mutants and fluorescently labeled *P. patens* Villersexel (Vx::mCherry) were proliferated according to 3.1 and were harvested for crossing at one week old. Special solid medium (BCD medium with low nitrogen) was prepared to help sporophyte development (Perroud *et al.*, 2011). The ingredients are listed in Table 3**.** Before the addition of plant agar, the pH of the medium was adjusted to 6.5 with 4M KOH. Before the medium was poured into deep petri dishes, 1mL/L 1M $CaCl_2$ was added to the medium. About 90mL melted medium was poured into each plate.

Table 3: Recipe for BCD medium with low nitrogen supply (1L stock).

Ingredients	Amount
$MgSO_4 \cdot 7H_2O$	250mg
KNO_3	0.4g or 4ml 1M solution
KH_2PO_4	250mg
$FeSO_4 \cdot 7H_2O$	12.5mg
Micro Elements (1000x)	1ml
Plant agar	8g

Plant tissue of all the mutants and the polymorphic strain was ground with a homogenizer (Power Gen 125, Fisher Scientific). The ground tissue of each TS mutant was mixed with same amount of ground tissue of the fluorescently labeled wild type strain. Mixed moss tissue was grown under the same light condition as moss proliferation for 3 weeks. After 3 weeks, the plates were transferred to be cultivated at 15ºC under a cycle of 8 h light (90 µmol m-2sec-1) and 16 h dark. After 2 weeks, sterile distilled water was added to each plate to just submerge the moss and the water was removed after one day. The same procedure was repeated after 3 weeks' of culture at 15ºC. Sporophytes were picked when capsules turned brown. To identify crossed sporophytes, the moss was observed using a dissecting microscope under green light with a red filter. The sporophytes of plants where only capsules were fluorescent were picked.

One or more sporophytes were harvested in a sterile 1.5-mL microcentrifuge tube. The following sterilization was conducted according to a published protocol in Cold Spring Harbor Protocols (Cove *et al.*, 2009). In order to sterilize the sporophytes, 1 mL of 70% ethanol was added to the tubes and the tubes were incubated for 4 min at room temperature. The ethanol was then removed. The sporophytes were gently rinsed three times with 1 mL of sterile distilled water at room temperature. Another 1 mL of sterile distilled water was added to all tubes which were then placed for 7 days in the dark at 4°C. To germinate the spores, the sporophyte capsules were crushed and mixed to produce a spore suspension. About 400µL of the spore mixture was then added to germination agar plates in 60mm petri dishes. The germination medium was made by adding 10mL/L 1M $CaCl_2$ to melted $PpNH_4$ agar medium before pouring plates. After the plants grew big enough, they were harvested on to $PpNH_4$ agar plates.

To screen for the F1 segregants that retained the TS phenotype, each segregant and a control were proliferated on two $PpNH_4$ agar plates and cultivated at 25ºC and 32ºC respectively for one week. Imaging was performed on a stereo microscope under white light at a magnification of 64X.

2.6 Genomic Analysis

For the mapping of the TS phenotype causal mutation, genome sequencing of pooled segregants was chosen as the mapping strategy due to its advantages discussed in Section 1.2. Outcrossing was chosen to generate the segregants because the strain, Gransden, used to create TS mutants in this study has a relatively low self-sterilization rate compared to its outcrossing rate to the strain Villersexel. The crossing was conducted as described in Section 2.7. As mentioned in section 1.2.2, three important components of the experimental design of mapping by pooled-sequencing are the size of the mapping population, sequencing depth, and sequencing strategy (single-ended or paired-ended). The sequencing depth and sequencing strategy were decided to be 10X coverage and paired-ended with 90nt read length based on previous TS and whole genome sequencing studies (Austin et al., 2011; Doitsidou et al., 2010; James et al., 2013; Zuryn et al., 2010) and the direct sequencing result of LV768 (data not shown). Nonetheless, the size of the mapping population could not be decided by simple calculation since a number of parameters affect the pattern of crossovers during the outcrossing. In order to estimate the

relationship between the size of mapping population and the detected range of the causal mutation, a Monte Carlo simulation was conducted using MATLAB.

The simulation was written to find the distribution of the detected range of the causal mutation when different sizes of mapping population (F1 segregants) were used. The detected range of the causal mutation was defined as the chromosomal range where no crossover occurred. The following are some assumptions made for the simulation:

i. There is one causal mutation which is located randomly on one nucleotide;
ii. There is either 0, 1, or 2 crossover(s) per chromosome;
iii. The chance of 0, 1, or 2 crossover(s) is correlated with the chromosome length;
iv. If there is only 1 crossover on a chromosome, the crossover point is randomly chosen from all nucleotides of that chromosome (i.e. no influence from centromere or telomere) ;
v. If there is 2 crossovers on a chromosome, the crossover point of the first crossover is random, and the distance between the 2^{nd} crossover and the 1^{st} crossover is generated according to a gamma distribution whose shape and scale parameters are correlated with the chromosome length;
vi. During meiosis, the segregation of chromosomes is random.

Assumptions i and iv were made because the locations of the causal mutation, the centromeres and the telomeres were unknown. Assumptions ii, iii and v were based on a study of recombination landscape in *Arabidopsis thaliana* F2 populations (Salome et al., 2012). The chromosome lengths of *P. patens* were obtained from the V3 genome assembly (Joint Genome Institute and Center for Integrative Genomics, 2014). The chances of 0, 1 and 2 crossovers and the shape and scale parameters of gamma distributions were calculated accordingly (Appendix B). The algorithm of the simulation was formulated to determine the crossover landscape of a defined number of haploid spores (the mapping population) of *P. patens* and summarize the chromosomal regions that no crossover ever occurs in any haploid spores (possible region containing causal mutation). The simulation includes some important steps:

i. Randomly select the causal mutation for a population of a defined number of mutants.
ii. For each chromosome of one haploid spore, determine the number of crossovers according to its chances of having 0, 1 and 2 crossover(s).
iii. Determine crossover location for chromosomes with 1 or 2 crossover(s).

iv. Determine how chromosomes are segregated into one haploid spore assuming random segregation and selection for causal mutation.

v. Update the range of causal mutation with known crossover locations.

vi. Repeat ii-v for the defined number of mutants in a mapping population.

vii. Count numbers of chromosomes and base pairs within the detected range of causal mutation.

viii. Repeat i, vi and vii for 100 times.

ix. Generate histograms for counted numbers of chromosomes and base pairs.

The simulation was run for population size of 10, 20, 24, 50, 70 and 90 in order to find the lower limit of a reasonable size and how the detected range of the causal mutation decrease while increasing the size of mapping population.

Results

The goal of this study was to find and study genes that are critical in plant cell growth and division. In order to achieve this goal, the objectives of this study were to isolate TS mutants in *P. patens,* characterize the morphology of the isolated mutants and design a strategy to map the causal mutation by genome sequencing of pooled segregants. Eight mutants were isolated using two different isolates of the wild type strain of *P. patens* (Gransden). Six of the eight mutants displayed significant decrease in plant areas and increase in plant solidities at 32°C compared to the control at 32°C and themselves at lower temperatures according to the results of the growth assays and statistical test results. Six of the eight mutants achieved outcrossing with a fluorescently labeled polymorphic strain Vx::mCherry, and 24 F1 segregants of mutant LV767, which displayed the TS phenotype, were selected as the mapping population for genome sequencing of pooled segregants. It was estimated using a Monte Carlo simulation that the mapping population of size 24 should narrow down the causal mutation to one chromosome within 300 million base pairs.

3.1 Isolation and Screening of TS Mutants

Eight mutants displaying growth defect at 32°C were selected from candidates generated by mutagenesis of two individual strains of *P. patens* Gransden. In order to obtain TS mutants which have growth defects at 32°C, I treated wild type moss protoplasts with 1,100 J/m2 UV light and selected those with sizes smaller than 400µm after four days of regeneration at 25°C and one week of culture at 32°C. Two different strains of *P. patens* Gransden (LV777 and LV689) were used as the wild type in mutagenesis. To test temperature sensitivity, each mutant and its corresponding wild type control were proliferated on two $PpNH_4$ agar plates and cultivated at 25ºC and 32ºC respectively. A total of eight mutants (LV767, LV768, LV769, LV770, LV771, LV772, LV774 and LV776), which showed decrease and abnormality in growth at 32°C compared to themselves at 25°C and 20°C and the corresponding wild type at 32°C, were selected. Among the mutants, LV767, LV768, LV769, LV770, LV771 and LV772 were generated by mutagenizing LV777 while LV774 and LV776 were generated by mutagenizing LV689.

3.2 Morphological Characterization of TS Mutants

To characterize the phenotypes of the eight selected TS mutants growing at 20°C, 25°C and 32°C, growth assays were conducted on all mutants and their corresponding controls. The protoplasts of the eight mutants and two controls were regenerated for four days at 25°C and were transferred to the $PpNH_4$ growth medium plates growing at 20°C, 25°C and 32°C for another four days. Micrographs were taken under the microscope on day3 and day4 after the protoplasts were transferred to $PpNH_4$ plates to record their growth and morphology at the three different temperatures.

Mutants generated from LV777 (LV777 mutants) and from LV689 (LV689 mutants) displayed different TS phenotypes. LV777 (C1) and LV777 mutants showed similar growth at 20°C and 25°C while LV777 mutants showed significant decrease in cell division and elongation at 32°C (**Figure 6**). In contrast, LV777 did not display visible changes at different temperatures (**Figure 6**). Both LV689 (C2) and LV689 mutants displayed different TS phenotypes when compared to LV777 and LV777 mutants. Unlike LV777, LV689 showed decrease in cell growth at 32°C and seemed to grow fastest at 25°C (Figure 6). LV689 mutants and LV689 showed similar growth with respect to size, while LV689 mutants had severely curled cell filaments at 20°C, 25°C and 32°C (**Figure 6**).

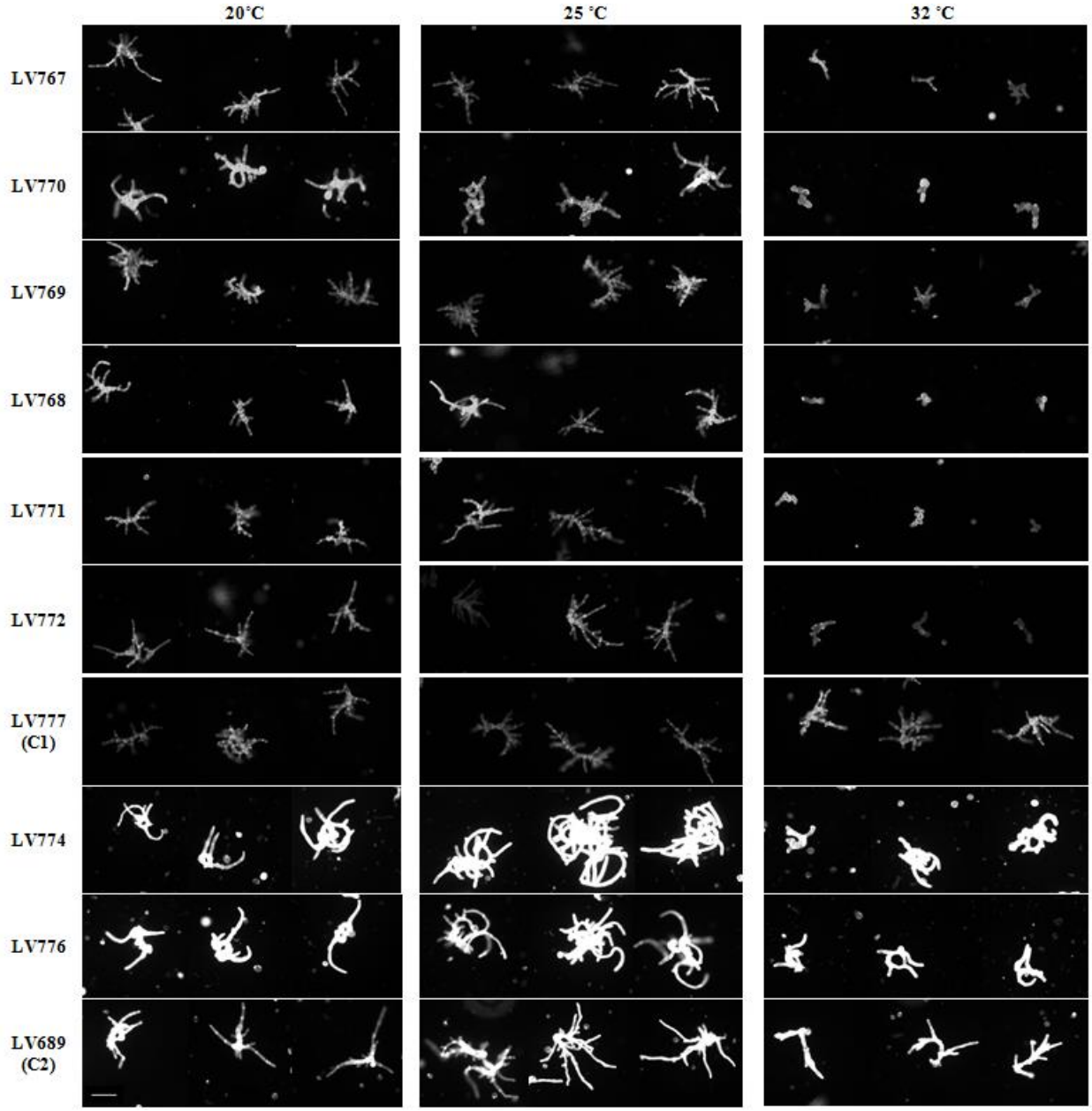


Figure 6: Representative images of plants that were cultured at 20°C, 25°C, and 32°C. The plants were photographed on day3 after passing to growth medium. C=control. White bar=200µm.

In order to quantitatively characterize the morphology and measure the difference of mutant and control populations, plant areas were measured using the micrographs taken in the growth assays. Meanwhile, solidity (plant area/convex hull area) was also calculated. Plant areas were used to assess growth rate whereas plant solidities were used to assess the extent of polarization and branching of cell filaments. Theoretically, the larger the plant area, the faster the plant grows and the smaller the solidity, the more the plant extends out its filaments. If the plants grow slower, then smaller plant areas and larger solidities would be expected whereas if the

growth rate does not change but the cell filaments curl up instead of extending straight, then only larger solidities would be expected. The temperature sensitivity of the mutants and controls observed in the micrographs (**Figure 6**) were confirmed by the ANOVA-Tukey test and the non-parametric Kruskal-Wallis test. In addition, more insights were obtained after the data distributions of the mutants and controls were computed using box plots (**Figure 7**).

Based on the statistical results and sample distributions of comparing the log(area) and solidity at 20°C, 25°C and 32°C, LV777 mutants can be categorized into three groups. The first group includes LV767, LV768, LV771 and LV772 which had log(area) and solidity similar to those of the LV777 control at 20°C and 25°C while displayed significant decrease in log(area) and increase in solidity at 32°C compared to at 20°C and 25°C. The second group includes LV769 which had log(area) and solidity similar to those of the LV777 control at 20°C and 25°C but significantly smaller change in log(area) and solidity at 32°C when compared to the changes of mutants in the first group. The third group includes LV770 which had log(area) similar to that of the LV777 control yet a significantly larger solidity at 20°C and 25°C while it displayed significant decrease in log(area) and increase in solidity at 32°C compared to at 20°C and 25°C.

The data distributions displayed in the Whiskers box plots reveal all LV777 mutants have growth defects at 32°C compared to the LV777 control. In addition, some variance was also revealed in TS phenotypes of LV770 and LV769 compared to the other LV777 mutants. The sample distributions of log(area) of all LV777 mutants and LV777 control were very similar at 20°C and 25°C (**Figure 7A and 7B**) whereas the mutants (except for LV769) showed decrease in log(area) and LV777 control showed no significant change at 32°C. The differences in log(area) of LV777 grown at 20°C vs. 25°C and at 20°C vs. 32°C were all much smaller than the differences in log(area) of LV777 mutants grown at 20°C vs. 32°C and at 25°C vs. 32°C (**Figure 7A and 7B**). Mutant LV769 showed a smaller decrease in log(area) at 32°C compared to other LV777 mutants (**Figure 7A and 7B**). Similarly, the sample distributions of solidity of all LV777 mutants (except for LV770) and LV777 control were very similar at 20°C and 25°C (**Figure 7C and 7D**) whereas all the mutants showed increase in solidity and LV777 control showed no significant change at 32°C. Although the solidities of LV770 at 20°C and 25°C was smaller than that of LV770 at 32°C, they were still larger than those of the other LV777 mutants at 20°C and 25°C (**Figure 7C and 7D**). Similar to the change of log(area), LV769 showed a smaller increase in solidity at 32°C compared to other LV777 mutants (**Figure 7C and 7D**).

The differences in phenotypes at 20°C, 25°C and 32°C observed in box plots were confirmed by the statistical test results. The control LV777 had no statistically significant difference when its log(area) and solidity on day3 and its solidity on day4 from plants grown at 20°C, 25°C and 32°C were compared (**Table 4 and Table 5**). LV777 showed statistically significant difference only in log(area) when compared at 20°C vs. 25°C and at 20°C vs. 32°C on day4 (**Table 4**). All LV777 mutants, except for LV769, showed no statistically significant difference when the log(area) and solidity of plants grown at 20°C were compared to those of plants grown at 25°C, but had statistically significant difference when the log(area) and solidity of plants grown at 32°C were compared to those of plants grown at 25°C or at 20°C (**Table 4 and Table 5**). The only discrepancy was the non-parametric test result which indicated no statistical significant difference when comparing the solidity of LV770 at 20°C vs. at 32°C on day3 (**Table 5**). When the log(area) of LV769 grown at 32°C on day4 was compared to that of LV769 grown at 20°C and 25°C, only one ANOVA test result indicated a statistical significant difference (**Table 4**).

Combining with the observations obtained from the micrographs, the LV689 mutants and control all showed decrease in log(area) and increase in solidity at 32°C compared to at 20°C and 25°C, and the LV689 mutants also showed increases in solidity at all three temperatures when compared to the LV689 control (**Figure 7**). There was an increase in the log(area) of LV689 mutants and control at 25°C compared to at 32°C on day3 and day4, and there was a smaller increase in the log(area) of LV689 mutants and control at 25°C compared to 20°C (**Figure 7A and 7B**). The sample distributions of the log(area) of the LV689 mutants were similar to those of the LV689 control at all temperatures on day3 and day4 except for at 25°C on day4 (**Figure 7A and 7B**). In terms of solidity, the sample distributions of the LV689 control and mutants were similar at 20°C and 25°C but there was an increase in solidity at 32°C on day3 and day4 (**Figure 7C and 7D**). When the sample distributions of solidity of the LV689 mutants and control were compared at each temperature, there was an increase in the solidity of LV689 mutants compared to that of the LV689 control on day3 and day4 (**Figure 7C and 7D**).

In **Table 4 and 5**, there are a few discrepancies between the results of ANOVA and non-parametric tests concerning the LV689 control and mutants. Based on the consistent ANOVA and non-parametric test results, both LV689 mutants and the control displayed statistically significant changes when their log(area) and solidity of plants grown at 20°C, 25°C and 32°C

were compared. The LV689 control had significant differences when the log(area) and solidity of the plants grown at 32°C were compared to that of the plants grown at 25°C on day3 and to that of the plants grown at 20°C on day4 (**Table 4 and Table 5**). LV774 showed statistically significant differences when the log(area) of the plants grown at 25°C was compared to that of the plants grown at 32°C on day3 and day4 (**Table 4**), and also when the solidity of the plants grown at 32°C was compared to that of the plants grown at 20°C and 25°C on day3 and that of plants grown at 25°C on day4 (**Table 5**). LV776 showed statistically significant differences when the log(area) and solidity of the plants grown at 32°C were compared to those of the plants grown at 20°C and 25°C on day3 and day4 (**Table 4 and Table 5**).

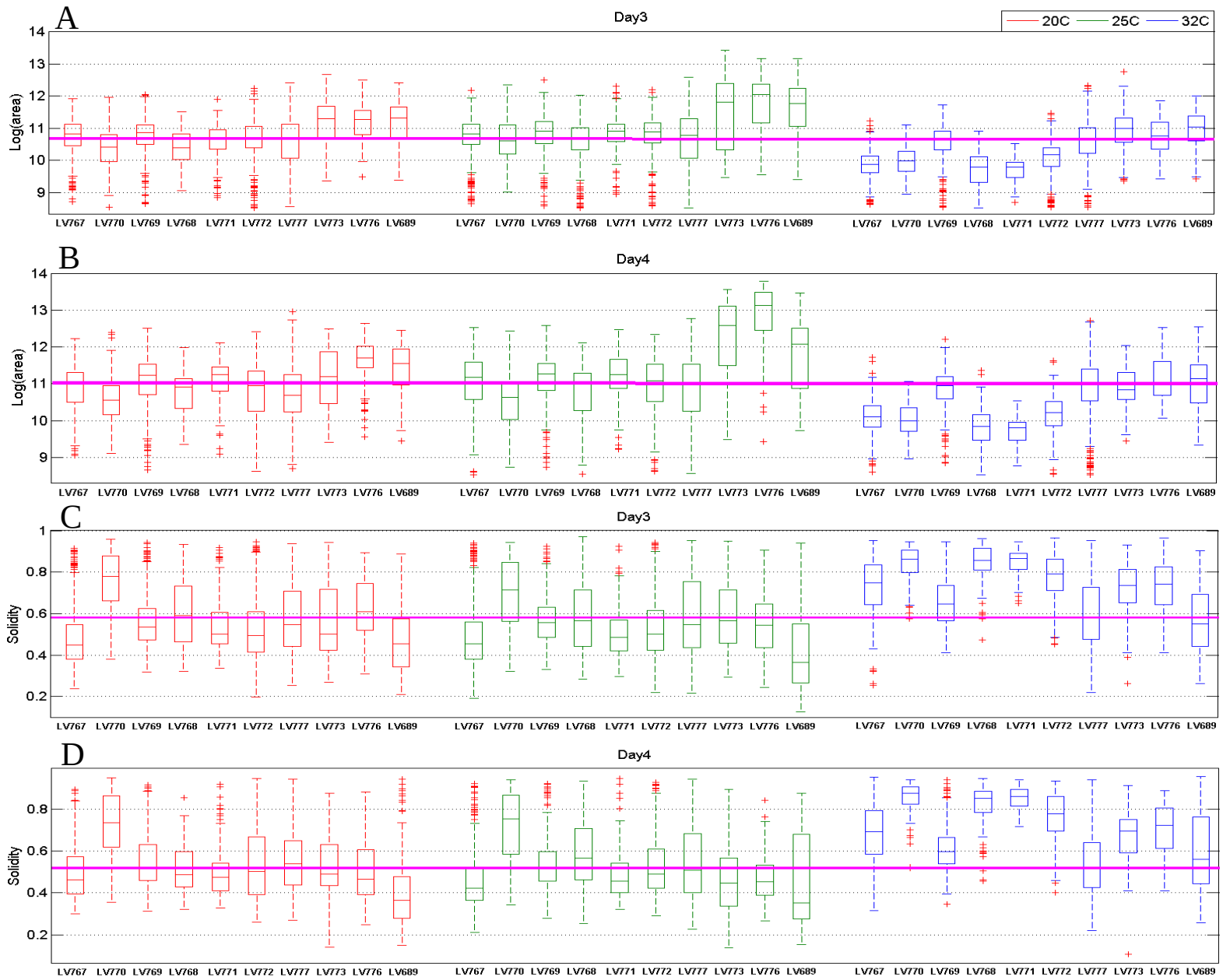


Figure 7: Quantification of the phenotypes of temperature sensitive mutants and the controls. Quantification of log(area) and solidities of all mutants and controls at 20°C, 25°C, and 32°C on day3 and day4 after passing to growth medium displayed as Whiskers box plots. C=control. Red+ represents any outlier. The magenta line indicates the mean of C1 at 32°C. A. plant log(area) on day3; B. plant log(area) on day4; C. plant solidities on day3; D. plant solidities on day4.

Table 4: Statistical test results of comparing log(area) of all isolated mutants and controls. Adjusted P values are shown for rejecting equivalence of means of log(area) and the statement that the samples come from populations with the same data distribution using ANOVA-Tukey and non-parametric Kruskal-Wallis tests. C=control. d=day. Values in bold text indicate that the difference is not statistically significant at the 0.05 level. For lines listed in the tables, between 30-562 plants were analyzed per condition.

	Log(area)											
	Compare to the mutant itself (temperature1 vs. temperature2, day number)											
	20°C/25°C d3		20°C/32°C d3		25°C/32°C d3		20°C/25°C d4		20°C/32°C d4		25°C/32°C d4	
	ANOVA	Non-par	ANOVA	Non-par	ANOVA	Non-par	ANOVA	Non-par	ANOVA	Non-par	ANOVA	Non-par
LV767	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001	**0.9850**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001
LV770	**0.2102**	> **0.9999**	<0.0001	0.0031	<0.0001	<0.0001	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001
LV769	> **0.9999**	> **0.9999**	0.0034	0.0148	0.0001	0.0003	> **0.9999**	> **0.9999**	**0.3023**	> **0.9999**	0.0075	**0.1663**
LV768	**0.9740**	> **0.9999**	<0.0001	0.0013	<0.0001	<0.0001	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001
LV771	**0.2394**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001
LV772	**0.6214**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001	**0.9884**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001
LV777 (C1)	**0.3554**	> **0.9999**	> **0.9999**	> **0.9999**	**0.9994**	> **0.9999**	0.0293	**0.3469**	<0.0001	<0.0001	> **0.9999**	> **0.9999**
LV774	**0.4602**	> **0.9999**	**0.2707**	**0.1033**	< 0.0001	0.0006	0.0004	**0.7051**	**0.9897**	> **0.9999**	< 0.0001	< 0.0001
LV776	0.0401	**0.3303**	0.0114	0.0116	< 0.0001	< 0.0001	< 0.0001	**0.3516**	0.0025	0.0005	< 0.0001	<0.0001
LV689 (C2)	**0.0505**	**0.3625**	**0.9552**	> **0.9999**	< 0.0001	< 0.0001	**0.9977**	> **0.9999**	0.0077	0.0076	0.0039	**0.3288**

Table 5: Statistical test results of comparing solidities of all isolated mutants and controls. Adjusted P values are shown for rejecting equivalence of means of solidities and the statement that the samples come from populations with the same data distribution using ANOVA-Tukey and non-parametric Kruskal-Wallis tests. C=control. d=day. Values in bold text indicate that the difference is not statistically significant at the 0.05 level. For lines listed in the tables, between 30-562 plants were analyzed per condition.

	Solidity											
	Compare to the mutant itself (temperature1 vs. temperature2, day number)											
	20°C/25°C d3		20°C/32°C d3		25°C/32°C d3		20°C/25°C d4		20°C/32°C d4		25°C/32°C d4	
	ANOVA	Non-par	ANOVA	Non-par	ANOVA	Non-par	ANOVA	Non-par	ANOVA	Non-par	ANOVA	Non-par
LV767	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001	**0.9366**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001
LV770	**0.0617**	> **0.9999**	0.0059	**0.5884**	<0.0001	< 0.0001	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	0.0002
LV769	> **0.9999**	> **0.9999**	< 0.0001	< 0.0001	<0.0001	<0.0001	> **0.9999**	> **0.9999**	0.0034	0.0139	< 0.0001	0.0007
LV768	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001	**0.8849**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001
LV771	**0.9998**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001
LV772	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001	> **0.9999**	> **0.9999**	<0.0001	<0.0001	<0.0001	<0.0001
LV777 (C1)	> **0.9999**	> **0.9999**	**0.8387**	> **0.9999**	> **0.9999**	> **0.9999**	> **0.9999**	> **0.9999**	**0.9998**	> **0.9999**	> **0.9999**	> **0.9999**
LV774	**0.9998**	> **0.9999**	< 0.0001	< 0.0001	< 0.0001	<0.0001	> **0.9999**	> **0.9999**	0.0107	**0.1184**	< 0.0001	0.0053
LV776	**0.8823**	> **0.9999**	0.0024	0.0204	< 0.0001	<0.0001	> **0.9999**	> **0.9999**	< 0.0001	<0.0001	< 0.0001	<0.0001
LV689 (C2)	**0.5469**	> **0.9999**	0.0180	**0.9326**	< 0.0001	0.0044	> **0.9999**	> **0.9999**	< 0.0001	< 0.0001	0.0015	**0.7900**

3.3 Mapping of the TS Causal Mutation

3.3.1 Outcrossing of the TS Mutants to Vx::mCherry

In order to find the causal mutation of the TS mutants among all the background mutations introduced by UV mutagenesis, the approach of genome sequencing of pooled segregants was used. Since the wild type strain Gransden used to generate TS mutants in this study had a relatively low self-fertilization rate compared to its outcrossing rate to the strain Villersexel, outcrossing the mutant to a fluorescently labeled polymorphic strain Vx::mcherry (Perround *et al.*, 2011) was conducted to generate the mapping population (F1 segregants). By selecting F1 segregants that displayed the TS phenotype at 32°C, the causal mutation should be conserved in all selected segregants whereas other parts of the genome undergo random chromosomal crossover and recombination. Therefore, it will be expected that the chromosomal recombination rates get lower on chromosomal sections that are closer to the causal mutation. The approximate location of the causal mutation can be detected by computing the recombination frequencies at genome marker locations of the polymorphic strain after sequencing the genomes of pooled segregants.

The outcrossing and selection strategies used here were large based on a crossing study done by Perround *et al.* (2011) and outcrossing was achieved for six out of the eight mutants. In that study, a transgenic polymorphic strain Vx::mCherry was created whose fluorescence was excited by green light and viewed with a red filter (**Figure 8d**). In the study of Perround *et al.* (2011), Gransden had only been found as the female parents in the outcrossing with Villersexel. Therefore, crossed sporophyte containing the mCherry fluorescent protein would be generated on a non-fluorescent TS mutant gametophyte, whereas a self-fertilized Vx::mCherry would have both fluorescent sporophyte and gametophyte. Crossings were set up for all eight mutants with Vx::mCherry. After the sporophytes and gametophytes were observed by fluorescent stereomicroscopy under green light with a red filter, outcrossed sporophytes were found in LV767, LV768, LV769, LV771, LV772 and LV774, and self-fertilized sporophyte were found in Vx::mCherry (**Figure 8a-d**). No outcrossed sporophytes of LV770 and LV776 have been found yet. In addition, the rate of self-fertilization of Vx::mCherry was much higher than the rate of outcrossing for all eight mutants.

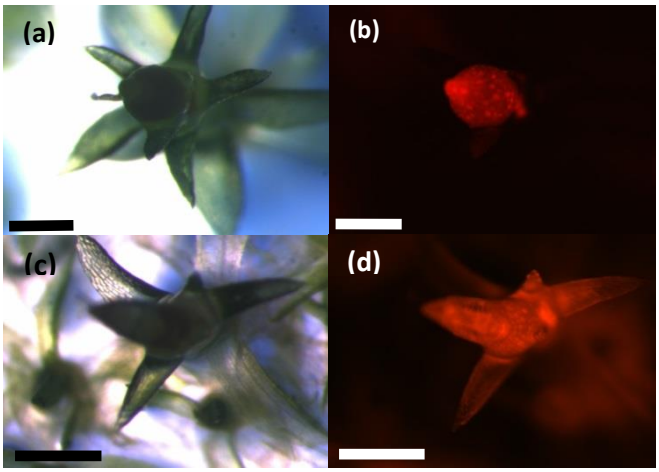


Figure 8: Gametophores and sporophytes of outcrossed mutants and self-fertilized Vx::mCherry. (a) Crossed LV767 under bright field light; (b) crossed LV767 under fluorescence excitation with green light and emission detected with a red filter; (c) selfed Vx::mcherry under bright light; (d) selfed Vx::mcherry under fluorescence excitation with green light and emission detected with a red filter. Bars are approximately 300µm.

Among all eight mutants, the F1 segregants of LV767 were screened by propagating each segregant on the growth medium at 25°C and 32°C and a total of 24 segregants which displayed the TS phenotype similar to that of LV767 at 32°C were selected (**Figure 9A-C**). The LV777 control was also propagated with all LV767 segregants as the TS-negative control which showed no significant change in phenotype when grown at 25°C and at 32°C (**Figure 9D**). The reasons of selecting LV767 for genome sequencing of pooled segregants were that it displayed a strong TS phenotype at 32°C according to the morphological characterization and it had relatively high crossing and spore germination rates which in turn generated most F1 segregants among all mutants. After screening about 120 F1 segregants of LV767, 24 segregants with strong TS phenotypes at 32°C were selected. Interestingly, different segregants sometimes showed different degree of growth defect at 32°C. While LV767 and its F1 segregants had similar phenotypes at 25°C, the segregant2 (**Figure 9C**) showed a more severe growth defect at 32°C where it had an even slower growth rate than LV767 and segregant1 did at 32°C (**Figure 9B**).

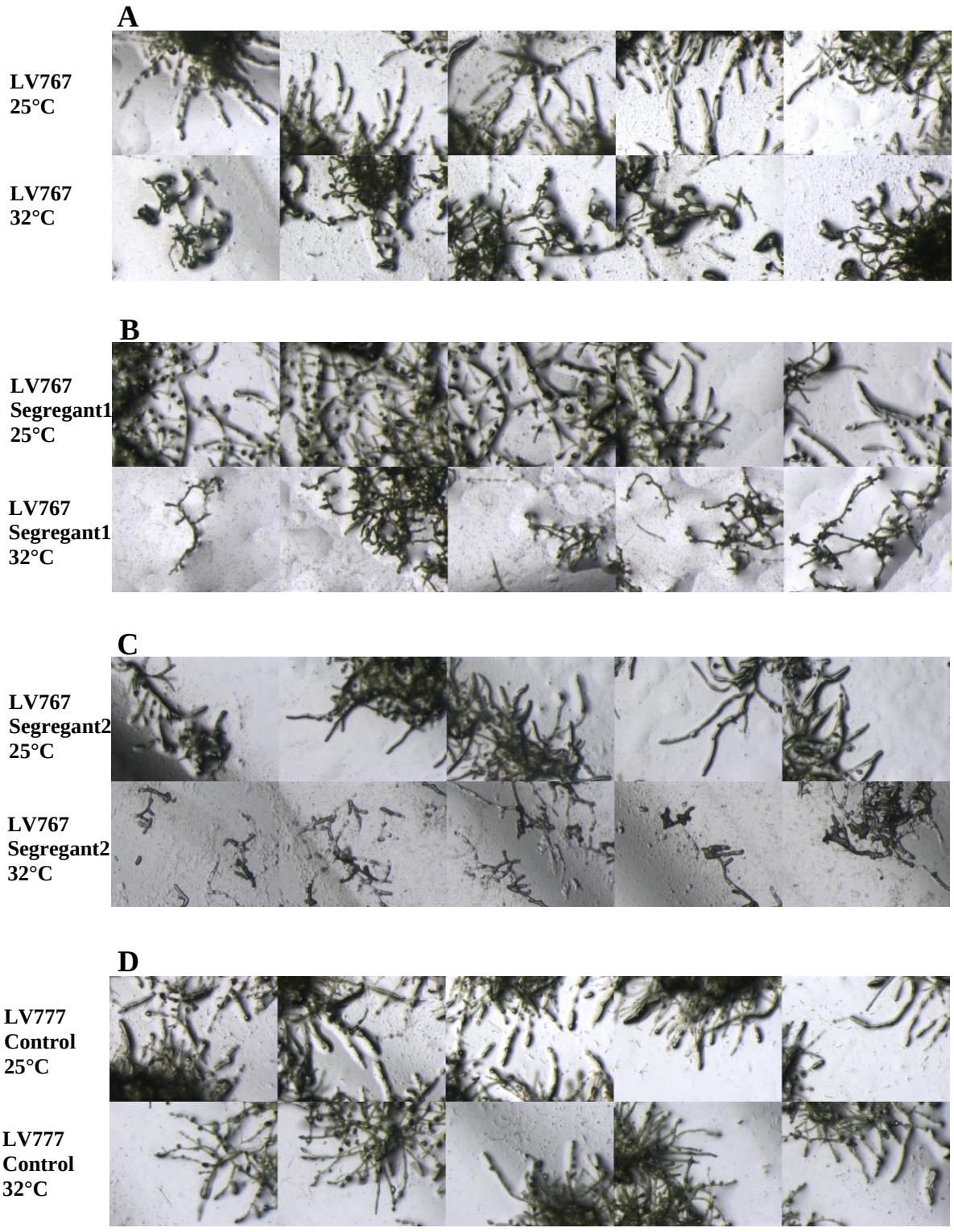


Figure 9: Representative images of LV767 segregants, LV767 and LV777 control cultured at 25°C and 32°C. The plants were photographed on day7 after propagation on growth medium at 64x magnification. A. LV767 at 25°C and at 32°C; B. an F1 segregant of LV767 at 25°C and at 32°C; C. a different F1 segregant of LV767 at 25°C and at 32°C; C. LV777 control at 25°C and at 32°C

3.3.2 Genome Sequencing of Pooled LV767 F1 Segregants

A Monte Carlo simulation was done using MATLAB to explore the relationship between the size of the mapping population (pooled segregants) and the size of detected range of the causal mutation. In theory, the more segregants included in sequencing, the smaller the probable location of the causal mutation can be narrowed down to. Nonetheless, it was hard to establish a direct relationship between the size of mapping population and the size of detected chromosomal range of the causal mutation by calculation. The crossover landscape is random on each chromosome and depends on a number of factors such as the chromosome length, crossover frequency, crossover interference, etc. Thus, a Monte Carlo simulation was done based on a few assumptions (**Section 2.6**) made according to a study on the crossover landscape of F2 *Arabidopsis* (Salome et al., 2012). The simulation simplified the crossover process by assuming there were either 0, 1 or 2 crossovers per chromosome and the number of crossovers per chromosome was linearly correlated with the chromosome length. The average number of crossovers per chromosome in this simulation was one crossover per chromosome. In addition, it was also assumed that the shape and scale parameters of the gamma distribution of crossover interference distances between two adjacent crossover points were linearly correlated with the chromosome length.

With the aforementioned assumptions, the relationship between the size of mapping population and the size of the detected range of the causal mutation was simulated by generating 100 independent crossover results with a given population size. A mapping population of sizes 10, 20, 24, 30, 40, 50, 70 and 90 were tested. The range of the causal mutation was represented by number of chromosomes and number of base pairs (**Figure 10**). As expected, the larger the population size, the smaller the range the causal mutation can be located in (**Figure 10**). With a mapping population of size 24 which was the number of F1 LV767 segregant obtained, the causal mutation should be narrowed down to one chromosome within 3 million base pairs out of the approximately 450 million base-pair genome consisting of 27 chromosomes with a good sequencing quality and an enough sequencing depth (Joint Genome Institute and Center for Integrative Genomics, 2014). Based on previous mapping studies (Austin et al., 2011; Doitsidou et al., 2010; James et al., 2013; Zuryn et al., 2010) and the direct sequencing result of LV768 (data not shown), the sequencing depth and sequencing strategy were decided to be 10X coverage and paired-ended with 90nt read length after balancing the cost and accuracy.

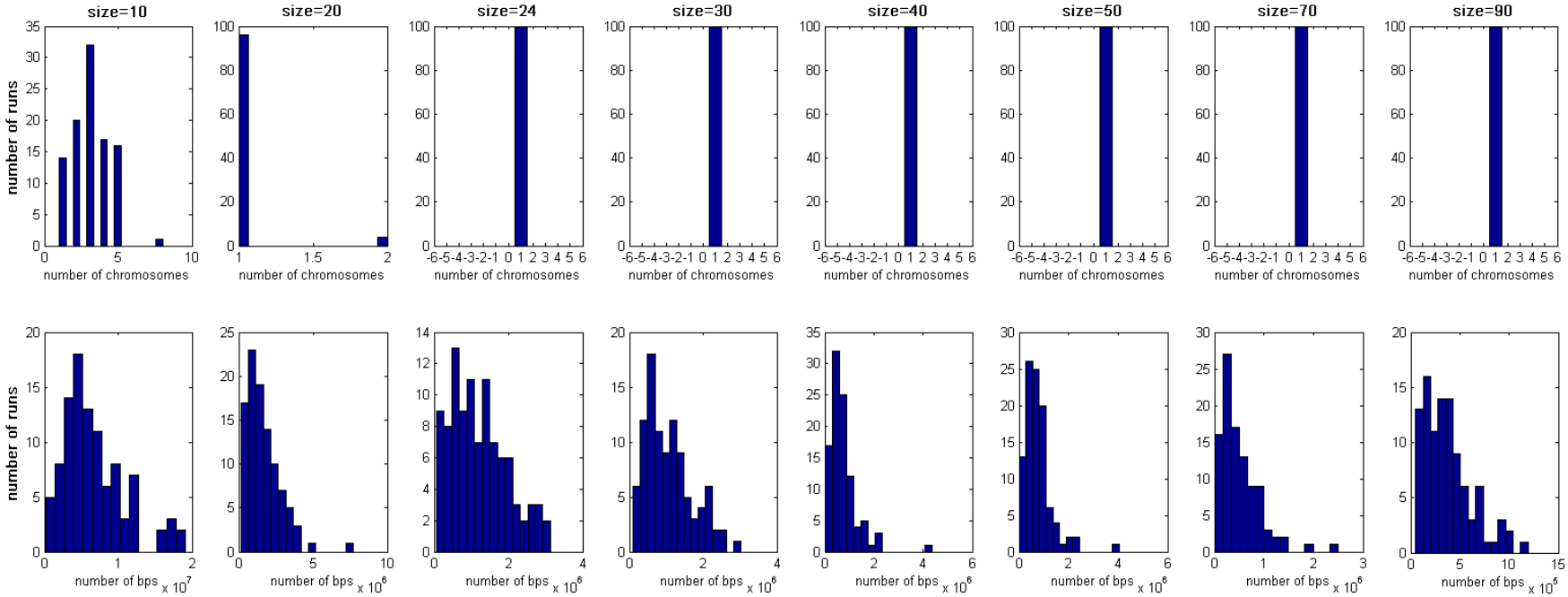


Figure 10: The results of the Monte Carlo simulation. Histograms of detected ranges (#chromosomes and bps) of the causal mutation with 10, 20, 24, 30, 50, 70, and 90 pooled mutants. With each set of pooled mutants, the results were obtained by simulating the crossover independently for 100 times using a Monte Carlo simulation.

Discussion

The isolation of conditional mutants is a strategy of forward genetics to study essential genes. Using this strategy, the goal of this study was to isolate and characterize TS mutants of *P. patens* that have defects in cell growth and division at high temperature (32°C) and to design a strategy to map the causal mutation by genome sequencing of pooled segregants. Eight mutants were selected by propagating the candidates, following UV mutagenesis, at 25°C and 32°C. The morphology of the eight mutants and their corresponding controls at 20°C, 25°C and 32°C were characterized by growth assays, and statistical tests were conducted to compare the difference of plant areas and solidities between the three different temperatures. All mutants and one control showed defects in cell growth and division at 32°C. After the isolation of mutants with desired TS phenotypes, outcrossing of mutants to a polymorphic strain Villersexel was conducted, and 24 F1 segregants of mutant LV767, which retained the TS phenotype, were selected to be the mapping population. The relationship between the size of the mapping population and the size of the detected range of causal mutation was established by a Monte Carlo simulation.

4.1 Isolation and Screening of TS Mutants

There were two levels of screening in the selection of TS mutants after UV mutagenesis; the first selection involves those having smaller plant sizes at 32°C compared to that of the wild type LV777 and the second selection involves those also having a growth rate similar to that of the wildtype at 25°C. While only the wild type strain LV777 was characterized in the preliminary studies to have no significant changes in phenotype when grown at 20°C, 25°C and 32°C, it was assumed that the other wild type strain, LV689 of the same subspecies as LV777, would also have no significant changes at the three temperatures. Both LV777 and LV689 were used in UV mutagenesis, and in the end six LV777 mutants and two LV689 mutants were selected after the two levels of screening.

Some issues were found with the screening method and the assumption of the phenotypes of the two wild type lines after the morphological characterization of the eight mutants. First, because the second level of the screening was qualitative, it was easy for miss-scoring to happen. Second, the assumption on the consistency of the phenotypes of the two wild type strains was proved to be wrong by the growth assays and statistical tests. Taking micrographs of moss

mutants grown at different temperatures under a microscope and then comparing the growth should be more objective than simply looking at the moss by eye and thus may reduce the frequency of miss-scoring. In addition, it is critical to characterize the morphology of all wild type strains before they are used in mutagenesis and as controls in morphological characterization of mutants.

4.2 Morphological Characterization of TS Mutants

Both qualitative and quantitative analyses were conducted to determine if a mutant had the desired growth defect at 32°C. The advantage of the qualitative analysis was that detailed changes and morphological effects at cellular levels, that were difficult to quantify, can be easily captured by eye. For example, the increase in solidities can be due to decrease in plant sizes (decrease in cell filament extension) or the curling of cell filaments. These two different causes would be hard to identify and differentiate based only on quantitative measurements and statistics. On the other hand, the advantages of the quantitative analyses were that some important phenotypes, such as plant areas and solidities, can be compared within and between different populations by computing descriptive statistics such as those displayed in the Whiskers box plots (**Figure 7**) and by conducting statistical tests such as ANOVA and non-parametric tests (**Table 4 and 5**). These population-wise comparisons would be hard to conduct and inaccurate simply through observations.

Most LV777 mutants (except for LV769) showed significant decrease in log(area) and increase in solidity at 32°C compared to at 20°C and 25°C on day3 and day4 whereas the LV777 control showed significant but much smaller variance in plant areas compared to LV777 mutants at different temperatures on day4 (**Table 4, Table 5 and Figure 7B**). Combined with the observations obtained from the micrographs (**Figure 6**), the decrease in log(area) and increase in solidity of LV777 mutants at 32°C is likely to be due to decrease in the rate of cell growth and division which led to smaller plants and less filament extension.

Based on their morphological changes at different temperatures, LV777 mutants were categorized into three groups. In the first group, except for LV767, LV768, LV771 and LV772 had very similar distributions of log(area) and solidities at all temperatures (**Figure 7**) which may be because their causal mutations were in genes that have closely related functions. Therefore, two explanations of LV767's different TS phenotypes were that its causal mutation

was located in the gene(s) having a different function, or its causal mutation caused a less disruptive structural change of the function molecule produced by the gene(s). Similarly, in the second group the weaker growth defect of LV769 at 32°C compared to other LV777 mutants (**Table 4 and Figure 7**) can be because its causal mutation was located in the gene(s) that are less critical in cell growth and division, or its causal mutation caused a less disruptive structural change of the function molecule produced by the gene(s). In the third group, the relatively larger solidity of LV770 at 20°C and 25°C was likely to be due to the curling of its cell filaments as displayed in **Figure 6**. This curling phenotype at 20°C and 25°C may be because the causal mutation already caused a phenotype at lower temperatures, or there were multiple mutations that one was responsible for the curling at 20°C and 25°C and the other was responsible for the growth defect at 32°C.

Although the LV689 mutants showed significant decrease in log(area) and increase in solidity at 32°C compared to at 20°C and/or 25°C, it cannot be concluded that they are TS mutants because the LV689 control also showed similar TS phenotypes at 32°C compared to at the other two temperatures (**Table 4, Table 5, Figure 6 and Figure 7**). Although LV689 mutants seemed to show a larger extent of increase in solidity at 32°C compared to the LV689 control, due to the relatively large variations and small sample sizes, more data needs to be collected in order to characterize the TS phenotypes more accurately (**Figure 7C and D**). The TS phenotypes at 32°C are likely to be due to decrease in the rate of cell growth and division and consequently decrease in filament extension (**Figure 6**). In addition, LV774 and LV776 also showed increases in solidity at all three temperatures when compared to the LV689 control which are likely to be due to the curling of it cell filaments (**Figure 6**). In Table 4 and 5, there are a few discrepancies between the results of ANOVA and non-parametric tests which are likely to be caused by their relatively small sample sizes. More growth assays are needed to obtain consistent ANOVA and non-parametric test results.

Although it is very likely that the causal mutations of the isolated TS mutants are in genes that are critical in cell growth and/or division, it is hard to infer the identity or even the type of genes where the causal mutations are located simply based on the phenotype. The reasons are as follows. First, there is not similar study that has been done in *P. patens*. Second, in *Neurospora crassa*, which is a filamentous fungus growing in a similar pattern as *P. patens*, many genes have

been discovered to be important for cell growth and filament extension where mutations or gene knockout can lead to mutants that conditionally have growth defects (Seiler and Plamann, 2003). For example, mutations in *cdc-24*, *bem-1* and *mcb* caused different types of growth defects at 39°C compared to at 25°C. At 39°C, the aforementioned mutants had cell filaments grown as chains of spheres, irregular chains of spheres, and thick, separated and swollen hyphae respectively. Although some of the *N. crassa* mutants displayed some similar phenotypes as the TS mutants in this study, there were still some significant differences, part of which was likely due to some essential differences between *P. patens* and *N. crassa* (Seiler and Plamann, 2003).

4.3 Mapping of the TS Causal Mutation

4.3.1 Outcrossing of the TS Mutants to Vx::mCherry

Using the transgenic polymorphic strain as the male parent, outcrossing has been achieved for six (LV767, LV768, LV769, LV771, LV772 and LV774) out of the eight isolated mutants. The reasons why outcrossed or self-fertilized sporophytes have not be obtained for the rest two mutants (LV770 and LV776) could be as follows. First, there may not be enough mutant gametophytes since in some plates it was even hard to find any mutant gametophyte. This could be due to insufficient mutant tissue plated onto the growth medium and/or Vx::mCherry out-competed the mutants during growth. Second, the causal mutation of TS phenotypes at 32°C affected the fertility of the mutants in normal condition. This was not a surprise since many genes have multiple functions in growth and reproduction of organisms (Kotak *et al.*, 2007). This was especially likely to be the case for LV770 since some degree of growth defects have been found at 20°C and 25°C. If the mutants were sterile, the mutation mapping would then not be able to be achieved through genome sequencing of pool segregants.

4.3.2 Genome Sequencing of Pooled LV767 F1 Segregants

Due to the lack of information on crossover landscape of *P. patens*, the data from F2 segregants of *Arabidopsis* were used to simulate chromosomal crossovers in single crosses in *P. patens*. The reasons of using data from *Arabidopsis* were that first, it is a plant and second, the chromosome lengths of *Arabidopsis* chromosomes and most *P. patens* chromosomes are comparable. *Arabidopsis* has five chromosomes whose lengths range from about 20 to 30 million base pairs, whereas 24 out 27 chromosomes of *P. patens* range from about 15 to 30 million base

pairs (Joint Genome Institute and Center for Integrative Genomics, 2014; Salome et al., 2012). To prevent overestimation of number of crossovers in the simulation, the average number of crossovers per chromosome was set to be one which was consistent with the theoretical expectation (Kamisugi *et al.*, 2008). The uncertainty and inaccuracy of this simulation would be expected to be mainly from the simplification of chromosomal crossovers and the estimated crossover interference since there was no data support from *P. patens*.

As for the simulation results, its accuracy could not be evaluated since no genome sequencing of pooled segregants study has been done in *P. patens* to date. Nonetheless, some general features of the simulation results in this study were comparable to the simulation results of outcrossing of another study done by James *et al.* in *Arabidopsis* (2013). Regardless of the sequencing depth and method, the detected range of the causal mutation was smaller when more segregants were used. In addition, the variance in the size of the detected range of the causal mutation was also smaller when more segregants were used. A sequencing depth of 10X should cover most of the *P. patens* genome, and according to a whole genome sequencing study of C. elegans, 13X coverage were more than enough to detect the causal mutation (*Zuryn et al.*, 2010).

4.4 Future Studies

First, as mentioned in Section 5.2, more growth assays need to be done in order to obtain a more accurate morphological characterization of LV774, LV777 and LV689. Second, the causal mutation needs to be mapped using the genome sequencing reads of the 24 pooled F1 segregants of LV767. The mapping can be done using published software such as SHOREmap or by direct computing of the crossover frequencies of the polymorphic markers across all 27 chromosomes of *P. patens*.

References

Abedon, S. (1997). *Supplemental Lecture*. Retrieved Mar 20, 2014, from http://www.mansfield.ohio-state.edu/~sabedon/biol3060.htm

Alonso, J., & Ecker, J. (2006). Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. *Nature Reviews, 7*(7), 524-536.

Austin, R. S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N. J., Bonetta, D., et al. (2011). Next-generation mapping of *Arabidopsis genes*. *The Plant Journal, 67*(4), 715–725.

Bajaj, K., Dewan, P. C., Chakrabarti, P., Goswami, D., Barua, B., Baliga, C., et al. (2008). Structural correlates of the temperature sensitive phenotype derived from saturation mutagenesis studies of CcdB.*Biochemistry, 47*(49), 12964-12973.

Batra, V. (2009). *Plant Ecology*. Jaipur, IND: Global Media.

Benzie, I. F. F., & Wachtel-Galor, S. (2011). *Herbal medicine: biomolecular and clinical aspects* (second ed.). Boca Raton, FL: CRC Press.

Chang, M. C. Y. (2007). Harnessing energy from plant biomass. *Model systems/Biopolymers, 11*(6), 677-684.

Cove, D. J. (2005). The Moss *Physcomitrella patens*. *Annual Review of Genetics, 39*, 339-358.

Cove, D. J., Perroud, P.-F., Charron, A. J., McDaniel, S. F., Khandelwal, A., & Quatrano, R. S. (2009). The Moss *Physcomitrella patens*: A Novel Model System for Plant Development and Genomic Studies. *Cold Spring Harbor Protocols, 2009*(2), pdb.emo115.

Darby, A., & Hall, N. (2008). Fast forward genetics. *Nature Biotechnology, 26*(11), 1248-1249.

Doitsidou, M., Poole, R. J., Sarin, S., Bigelow, H., & Hobert, O. (2010). *C. elegans* mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. *PloS one, 5*(11), e15435.

Food and Agriculture Organization of the United Nations. *How to Feed the World in 2050*: Food and Agriculture Organization of the United Nations. Retrieved Mar 20, 2014, from http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

Food and Agriculture Organization of the United Nations. (1997). Pressures on Agricultural Resources and the Environment. In *Long-term scenarios of livestock-crop-land use interactions in developing countries* (sixth ed.): Food and Agriculture Organization of the United Nations.

Food and Agriculture Organization of the United Nations. (2003). *Agriculture and the environment: changing pressures, solutions and trade-offs*. Retrieved Mar 25, 2014, fromhttp://www.fao.org/DOCREP/005/Y4252E/Y4252E00.HTM

Food and Agriculture Organization of the United Nations, PaperC, ebrary Inc, & Organisation for Economic Co-operation and Development. (2012). Biofuels. In *OECD-FAO agricultural outlook, 2012-2021* (pp. 88-117): OECD and FAO.

Fowler, M. W. (2006). Plants, medicines and man. *Journal of the Science of Food and Agriculture, 86*(12), 1797-1804.

Frank, W., Decker, E. L., & Reski, R. (2005). Molecular Tools to Study *Physcomitrella patens*. *Plant Biology, 7*(3), 220-227.

Fuglie, K., & Nin-Pratt, A. (2012). *A Changing Global Harvest*: International Food Policy Research Institute. Retrieved Mar 20, 2014, from http://www.ifpri.org/gfpr/2012/agricultural-productivity

Galvão, V. C., Nordström, K. J. V., Lanz, C., Sulz, P., Mathieu, J., Posé, D., et al. (2012). Synteny-based mapping-by-sequencing enabled by targeted enrichment. *The Plant Journal, 71*(3), 517–526.

Garg, G., & Adams, J. (2012). Treatment of neuropathic pain with plant medicines. *Chinese Journal of Integrative Medicine, 18*(8), 565-570.

Gaston, K., & Spicer, J. (2004). *Biodiversity* (Second ed.). United Kingdom: Blackwell Publishing company.

Gurib-Fakim, A. (2011). Traditional Roles and Future Prospects for Medicinal Plants in Health Care.*Development Review, 13*(7), 77-83.

Gyssels, G., & Poesen, J. (2003). The importance of plant root characteristics in controlling concentrated flow erosion rates. *Earth Surface Processes and Landforms, 28*(4), 371-384.

Hobert, O. (2010). The Impact of Whole Genome Sequencing on Model System Genetics: Get Ready for the Ride. *Genetics, 184*(2), 317-319.

Hunt, S., Elvin, M., & Heintzen, C. (2012). Temperature-sensitive and circadian oscillators of *Neurospora crassa* share components. *Genetics, 191*(1), 119-131.

Hussain, I. (2009). *Textbook of Plant Ecology*. Jaipur, IND: Global Media.

Ikehata, H., & Ono, T. (2011). The Mechanisms of UV Mutagenesis. *Journal of radiation research, 52*(2), 115-125.

James, G. V., Patel, V., Nordström, K. J., Klasen, J. R., Salomé, P. A., Weigel, D., et al. (2013). User guide for mapping-by-sequencing in *Arabidopsis*. *Genome Biology, 14*(6), R61.

Job, D. (2002). Plant biotechnology in agriculture *Biochimie, 84*(11), 1105-1110.

Joint Genome Institute, & Center for Integrative Genomics. (2014). *Physcomitrella patens v3.0 Data Usage Policy*. Retrieved Apr. 26, 2014, from http://www.phytozome.com/dataUsagePolicy.php?org=Org_Ppatens_v3.0

Kamisugi, Y., Von Stackelberg, M., Lang, D., Care, M., Reski, R., Rensing, S. A., et al. (2008). A sequence-anchored genetic linkage map for the moss, *Physcomitrella patens*. *The Plant Journal, 56*(5), 855–866.

Kiers, E. T., Leakey, R. R. B., Izac, A.-M., Heinemann, J. A., Rosenthal, E., Nathan, D., et al. (2008). Agriculture at a Crossroads. *Science, 320*(5874), 320-321.

Kotak, S., Larkindale, J., Lee, U., von Koskull-Döring, P., Vierling, E., & Scharf, K.-D. (2007). Complexity of the heat stress response in plants. *Current Opinion in Plant Biology, 10*(3), 310-316.

Lackie, J. (2010). reverse genetics, *A Dictionary of Biomedicine*: Oxford University Press.

Lang, D., Zimmer, A. D., Rensing, S. A., & Reski, R. (2008). Exploring plant biodiversity: the *Physcomitrella* genome and beyond. *Trends in Plant Science, 13*(10), 542-549.

Lockwood, S., Krishnamoorthy, B., & Ye, P. (2011). Neighborhood properties are important determinants of temperature sensitive mutations. *PloS one 6*(12), e28507.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). Genetic Mapping of Mutations. In *Molecular Cell Biology* (Fourth ed.). New York: W. H. Freeman.

Loo, M. (1975). *Neurospora crassa* temperature-sensitive mutant apparently defective in protein synthesis.*Journal of Bacteriology, 121*(1), 286-295.

Mannion, A. M. (1995). Agriculture, environment and biotechnology. *Agriculture, Ecosystems & Environment, 53*(1), 31-45.

Martin, E., & Hine, R. (2008). forward genetics, *A Dictionary of Biology* (sixth ed.): Oxford University Press.

Minevich, G., Park, D. S., Blankenberg, D., Poole, R. J., & Hobert, O. (2012). CloudMap: A Cloud-Based Pipeline for Analysis of Mutant Genome Sequences. *Genetics, 192*(4), 1249-1269.

Mitchell, H. K., & Houlahan, M. B. (1946). *Neurospora*. IV. A Temperature-Sensitive Riboflavinless Mutant.*American Journal of Botany, 33*(1), 31-35.

Mittler, R., Finka, A., & Goloubinoff, P. (2012). How do plants feel the heat? *Trends in Biochemical Sciences, 37*(3), 118-125.

Nordstrom, K. J. V., Albani, M. C., James, G. V., Gutjahr, C., Hartwig, B., Turck, F., et al. (2013). Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. *Nature biotechnology, 31*(4), 325-330.

Perroud, P.-F., Cove, D. J., Quatrano, R. S., & McDaniel, S. F. (2011). An experimental method to facilitate the identification of hybrid sporophytes in the moss *Physcomitrella patens* using fluorescent tagged lines.*New Phytol, 191*(1), 301–306.

Pimentel, D., Hepperly, P., Hanson, J., Douds, D., & Seidel, R. (2005). Environmental, Energetic, and Economic Comparisons of Organic and Conventional Farming Systems. *Bioscience, 55*(7), 573-582.

Prigge, M., & Bezanilla, M. (2010). Evolutionary crossroads in developmental biology:*Physcomitrella patens*, *Development* (Vol. 137, pp. 3535 - 3543). Cambridge, England: Company of Biologists Ltd.

Raven, P., Evert, R., & Eichhorn, S. (2005). Botany: An Introduction. In *Biology of Plants* (Seventh ed.). New York: W. H. Freeman and Company.

Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The *Physcomitrella* Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. *Science, 319*(5859), 64-69.

Robbins, W. J. (1944). The Importance of Plants. *Science, 100*(2603), 440-443.

Salome, P. A., Bomblies, K., Fitz, J., Laitinen, R. A. E., Warthmann, N., Yant, L., et al. (2012). The recombination landscape in *Arabidopsis thaliana* F2 populations. *Heredity, 108*(4), 447–455.

Sarin, S., Prabhu, S., O'Meara, M., Pe'er, I., & Hobert, O. (2008). *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. *Nature Methods, 5*(10), 865 - 867.

Schaefer, D. G., & Zrÿd, J. P. (2001). The moss *Physcomitrella patens*, now and then. *Plant physiology, 127*(4), 1430-1438.

Schmit, J., & Brody, S. (1982). Temperature-sensitive Mutant of *Neurospora crassa* that Affects Mycelial Growth and Morphology. *Journal of General Microbiology, 128*(5), 1147-1158.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., et al. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods, 6*(8), 550 - 551.

Schneeberger, K., & Weigel, D. (2011). Fast-forward genetics enabled by new sequencing technologies.*Trends in Plant Science, 16*(5), 282-288.

Seiler, S., & Plamann, M. (2003). The genetic basis of cellular morphogenesis in the filamentous fungus *Neurospora crassa*. *Molecular biology of the cel, 14*(11), 4352-4364.

Sharma, H. C., Crouch, J. H., Sharma, K. K., Seetharama, N., & Hash, C. T. (2002). Applications of biotechnology for crop improvement: prospects and constraints. *Plant Science, 163*(3), 381-395.

Sugiyama, M. (2014). Molecular genetic analysis of organogenesis in vitro with temperature-sensitive mutants. *Plant Biotechnol Rep, 8*(1), 29-35.

Suzuki, D. T. (1970). Temperature-Sensitive Mutations in *Drosophila melanogaster*. *Science, 170*(3959), 695-706.

Tan, G., Chen, M., Foote, C., & Tan, C. (2009). Temperature-Sensitive Mutations Made Easy: Generating Conditional Mutations by Using Temperature Sensitive Inteins That Function Within Different Temperature Ranges. *Genetics, 183*(1), 13-22.

Timilsina, G. R., & Shrestha, A. (2010). *Biofuels: markets, targets and impacts*. Washington, DC: The World Bank.

Transpiration: transfer of water from plants to the atmosphere. (2010). Retrieved Mar 22, 2014, fromhttp://ww2010.atmos.uiuc.edu/(Gh)/guides/mtr/hyd/trsp.rxml

U.S. Census Bureau. (2011). *World Population: 1950-2050*. Retrieved Mar 25, 2014, fromhttp://www.census.gov/population/international/data/idb/worldpopgraph.php

U.S. Department of Energy. *How Fossil Fuels were Formed*. Retrieved Mar 24, 2014, fromhttp://www.fossil.energy.gov/education/energylessons/coal/gen_howformed.html

U.S. Energy Information Administration. (2013). *International Energy Outlook 2013*. Retrieved Mar 25, 2014, from http://www.eia.gov/forecasts/ieo/world.cfm

Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genomics-assisted breeding for crop improvement.*Trends in Plant Science, 10*(12), 621-630.

Vidali, L., Augustine, R. C., Fay, S. N., Franco, P., Pattavina, K. A., & Bezanilla, M. (2009). Rapid Screening for Temperature-Sensitive Alleles in Plants. *Plant Physiology, 151*(2), 506-514.

von Koskull-Döring, P., Scharf, K.-D., & Nover, L. (2007). The diversity of plant heat stress transcription factors. *Trends in Plant Science, 12*(10), 452-457.

World Coal Association. *Coal Statistics*. Retrieved Mar 24, 2014, fromhttp://www.worldcoal.org/resources/coal-statistics/

Yuan, J. S., Tiller, K. H., Al-Ahmad, H., Stewart, N. R., & Stewart Jr, C. N. (2008). Plants to power: bioenergy to fuel the future. *Trends in Plant Science, 13*(8), 421-429.

Zecca, A., & Chiari, L. (2010). Fossil-fuel constraints on global warming. *Energy Policy, 38*(1), 1-3.

Zheng, T., Nibau, C., Phillips, D. W., Jenkins, G., Armstrong, S. J., & Doonan, J. H. (2014). CDKG1 protein kinase is essential for synapsis and male meiosis at high ambient temperature in *Arabidopsis thaliana*.*Proceedings of the National Academy of Sciences, 111*(6), 2182–2187.

Zuryn, S., Le Gras, S., Jamet, K., & Jarriault, S. (2010). A Strategy for Direct Mapping and Identification of Mutations by Whole-Genome Sequencing. *Genetics, 186*(1), 427-430.

Appendices

Appendix A. ImageJ Macro for Measuring Plant Areas and Calculating Solidities

```
//path="C:\\Test\\";
setSlice(1);
run("Set Scale...", "distance=1 known=2.084 pixel=1 unit=um global");

c=0;
PreResLabel=newArray(10000);
PreResArea=newArray(10000);
PreResPer=newArray(10000);
PreResCHA=newArray(10000);
PreResPE=newArray(10000);
PreResSol=newArray(10000);
PreResCo=newArray(10000);
PreResCi=newArray(10000);

var test = getTitle();
var drawing="Drawing of ";
var total=drawing+test;
print(total);

//Used to creat a new image stack where we can copy and paste the outlined image we did the stats on
getDimensions(w,h,channels,slices,frames);
newImage("Outline","8-bit black",w,h,slices);

//here is where you can enhance the contrast
run("Enhance Contrast", "saturated=0.4 normalize process_all");

selectWindow(test);

setBatchMode(true);

for(k=1; k<=nSlices; k++) {

setSlice(k);

//Make the threshold automatic on each individual frame
```

```
setAutoThreshold("Otsu dark");

   run("Set Measurements...", "area perimeter circularity limit display decimal=4 ");
          //Change size to limit to infinity, changed cricularity upper bound to .8
   run("Analyze Particles...", "size=12000-1000000 circularity=0.00-0.80 show=Outlines display exclude
clear record slice");

 //Copying newly made outline image to a new image
 selectWindow(total);
   run("Copy");
selectWindow("Outline");
   setSlice(k);
   run("Paste");
 selectWindow(total);
  close;
   selectWindow(test);

          n = nResults;
   area1 = newArray(n);
   length1 = newArray(n);
   area2 = newArray(n);
   length2 = newArray(n);
   circularity1 = newArray(n);
   label1 = newArray(n);

   xstart = newArray(n);
   ystart = newArray(n);
   totalArea=0;

   for (i=0; i<n; i++) {
    label1[i] = getResultLabel(i);
    area1[i] = getResult('Area', i);
    length1[i] = getResult('Perim.', i);
    circularity1[i] = getResult('Circ.', i);

    xstart[i] = getResult('XStart', i);
    ystart[i] = getResult('YStart', i);

    totalArea=totalArea+area1[i];
   }

   run("Clear Results");
```

```
   for (i=0; i<n; i++) {
     doWand(xstart[i], ystart[i]);
     run("Convex Hull");
      run("Set Measurements...", "area perimeter");
     run("Measure");
     area2[i] = getResult('Area', i);
     length2[i] = getResult('Perim.', i);
   }
   run("Select None");

   for (i=0; i<n; i++) {
   PreResLabel[c+i]= label1[i];
PreResArea[c+i]=area1[i];
PreResPer[c+i]= length1[i];
PreResCHA[c+i]=area2[i];
PreResPE[c+i]=length2[i];
PreResSol[c+i]=area1[i]/area2[i];
PreResCo[c+i]=length2[i]/length1[i];
PreResCi[c+i]=circularity1[i];
  }
c=c+n;
}

setBatchMode(false);

   for (i=0; i<c; i++) {
     setResult("Label", i, PreResLabel[i]);
     setResult("mutant", i, "gran");
     setResult("temperature", i, 32);
     setResult("day_num", i, 4);
     setResult("Area", i, PreResArea[i]);
     setResult("Perim.", i, PreResPer[i]);
     setResult("CH Area", i, PreResCHA[i]);
     setResult("CH Perim.", i, PreResPE[i]);
     setResult("Solidarity", i, PreResSol[i]);
     setResult("Convexity", i, PreResCo[i]);
     setResult("Circularity",i, PreResCi[i]);

    updateResults();

}
```

Appendix B. MATLAB Code of the Monte Carlo Simulation

The code listed here generates two column vectors that contain the numbers of chromosomes and base pairs in the detected range of causal mutation with a given size of the mapping population. Each column vector contains a defined number of results obtained from a defined number of independent runs of crossover events with the given size of the mapping population. There are a total of four functions and a .mat file included in this simulation whose roles and components are explained in the comments.

- Function 1:

```
%Input: the number of times by which the function "simulation" will be
%repeated; an empty cell that will be out put in the format depicted below;
%the size of mapping population that will be tested;
%an array that contains the ranges of 27 chromosomes; an array that
%contains the length of chromosomes; an array that contains the shape and
%scale parameters of the gamma distribution that defines the crossover
%interference distance between two crossover points.
%Note: the last four inputs are provided in "simulationData.mat".

%data_tab:
%#mutant    10      20      30      40      50
%[Mc_chr] [rch_num]
%[Mc]     [rMc_bp]
%         [range_chr]
%         [Mc_range]
function data_tab = repeated_simulation(repeated_num,
data_tab,mutant_num,chr_ranges,XO_chances,chr_length,gamma_par)
for i=1:repeated_num
    [Mc_range range_chr Mc_chr Mc rch_num rMc_bp]=
simulation(mutant_num,chr_ranges,XO_chances,chr_length,gamma_par);
    %data_tab{2,1}(i,mutant_num/10)=Mc_chr;
    %data_tab{3,1}(i,mutant_num/10)=Mc;
    data_tab{2,1}(i,1)=Mc_chr;
    data_tab{3,1}(i,1)=Mc;

    %c=mutant_num/10+1;
    c=2;
    data_tab{2,c}(i,1)=rch_num;
    data_tab{3,c}(i,1)=rMc_bp;
    %data_tab{4,c}(i,1)=range_chr;
    %data_tab{5,c}(i,1)=Mc_range;
end
end
```

- Function 2:

```
%This function simulate the process and result of XOs with 0, 1 or 2 XO(s)
%occuring per chromosome.

function [Mc_range,range_chr,Mc_chr,Mc,rch_num,rMc_bp]=
simulation(mutant_num,chr_ranges,XO_chances,chr_length,gamma_par)
```

```
%% Causal mutation definition
%randomly choose the location of causal mutation Mc. This only runs
%once for every pool of crossed mutants.
%select the chromosome: chance is positively correlated with chromosome
%length
r1=rand(1);
for i=1:27
    if r1>=chr_ranges(i,1) && r1<chr_ranges(i,2)
        Mc_chr=i;
    end
end

%select a nucleotide location: since the location of centromere is unknown,
%each nucleotide has the same chance to be Mc.
lc=makedist('Uniform','lower',1,'upper',chr_length(Mc_chr,1));
Mc=random(lc);


%% Generate defined number of crossed mutants (XO locations for 27 chr)

%Generate a table Mc_range to record the range of causal mutation. This
%table is updated with the XOs of every mutant.
Mc_range=cell(27,1);
for i=1:27
    Mc_range{i,1}(1,1)=1;
    Mc_range{i,1}(2,1)=chr_length(i,1);
end

range_chr=1:1:27;


%For each mutant, XO will be generated for each chromosome and then the
%Mc_range will be updated according to the XOs.
for n=1:mutant_num

%determine the number of XOs per chromosome (0,1,or 2). The XO_chances were
%estimated using data from Arabidopsis.
XO_num=[];
r2=rand(27,1);
for i=1:27
    if (r2(i,1)>0 && r2(i,1)<=XO_chances(i,1))
        XO_num(i,1)=0;
    else
        if (r2(i,1)>XO_chances(i,1) && r2(i,1)<=XO_chances(i,2))
            XO_num(i,1)=1;
        else
            XO_num(i,1)=2;
        end
    end
end


% XO location is randomly selected for 1 XO and the 1st of 2 XOs; the
interference distance
```

```
% is determined by a gamma distribution whose shape and scale were estimated
using
% Arabidopsis data.
XO_loc=[];
for i=1:27
    if XO_num(i,1)==0
        XO_loc(i,1)=0;
    else
        if XO_num(i,1)==1
            xolc=makedist('Uniform','lower',2,'upper',chr_length(i,1)-1);
            XO_loc(i,1)=random(xolc);
        else
            xolc=makedist('Uniform','lower',2,'upper',chr_length(i,1)-1);
            XO_loc(i,1)=random(xolc);
            xolc2=makedist('Gamma','a',gamma_par(i,2),'b',gamma_par(i,1));
            dist=random(xolc2)+2;
            %modification needed: revise the location of the 2nd XO to be
            %more random: if XO_loc(i,1)+dist<chr_length(i,1) &&
            %XO_loc(i,1)-dist>=0, then XO_loc(i,2) has equal chances to be
            %before or after XO_loc(i,1). AND RERUN THE SIMULATION.
            if ((XO_loc(i,1)+dist)<chr_length(i,1) && (XO_loc(i,1)-dist)>=2)
                rc1=rand(1);
                if rc1<=0.5
                    XO_loc(i,2)=XO_loc(i,1)+dist;
                else
                    XO_loc(i,2)=XO_loc(i,1);
                    XO_loc(i,1)=XO_loc(i,2)-dist;
                end
            else
                if ((XO_loc(i,1)+dist)>=chr_length(i,1) && (XO_loc(i,1)-
dist)>=2)
                XO_loc(i,2)=XO_loc(i,1);
                XO_loc(i,1)=XO_loc(i,2)-dist;
                else
                    if ((XO_loc(i,1)+dist)<chr_length(i,1) && (XO_loc(i,1)-
dist)<2)
                        XO_loc(i,2)=XO_loc(i,1)+dist;
                    else
                    rc2=rand(1);
                    if rc2<=0.5
                        XO_loc(i,2)=chr_length(i,1)-1;
                    else
                        XO_loc(i,2)=XO_loc(i,1);
                        XO_loc(i,1)=2;
                    end
                    end
                end
             end
           end
      end
end

%determine the segregation type -- which part is uncrossed (what to
%keep in Mc_range).
r3=rand(27,1);
range_upd=cell(27,1);
for i=1:27
```

```
    if i~=Mc_chr
        if XO_num(i,1)==1
            if r3(i,1)<=0.5
               range_upd{i,1}(1,1)=1;
               range_upd{i,1}(2,1)=XO_loc(i,1);
            else
                range_upd{i,1}(1,1)=XO_loc(i,1);
                range_upd{i,1}(2,1)=chr_length(i,1);
            end
        end

        if XO_num(i,1)==2
            if r3(i,1)<=0.5
                range_upd{i,1}(1,1)=1;
                range_upd{i,1}(2,1)=XO_loc(i,1);
                range_upd{i,1}(1,2)=XO_loc(i,2);
                range_upd{i,1}(2,2)=chr_length(i,1);
            else
                range_upd{i,1}(1,1)=XO_loc(i,1);
                range_upd{i,1}(2,1)=XO_loc(i,2);
            end
        end
    end

    if i==Mc_chr
        if XO_num(i,1)==1
            if Mc < XO_loc(i,1)
               range_upd{i,1}(1,1)=1;
               range_upd{i,1}(2,1)=XO_loc(i,1);
            else
               range_upd{i,1}(1,1)=XO_loc(i,1);
               range_upd{i,1}(2,1)=chr_length(i,1);
            end
        end


        if XO_num(i,1)==2
            if (Mc > XO_loc(i,1) && Mc < XO_loc(i,2))
               range_upd{i,1}(1,1)=XO_loc(i,1);
               range_upd{i,1}(2,1)=XO_loc(i,2);
            else
                   range_upd{i,1}(1,1)=1;
                   range_upd{i,1}(2,1)=XO_loc(i,1);
                   range_upd{i,1}(1,2)=XO_loc(i,2);
                   range_upd{i,1}(2,2)=chr_length(i,1);
            end
        end
    end
end

%update Mc_range with known XOs -- eliminate impossible (crossed) ranges
s_chr=size(range_chr);
range_chr_dup=range_chr;

for m=1:s_chr(2)
    i=range_chr_dup(1,m);
```

```
    %fprintf('check chr%d\t',i);

    if XO_num(i,1)~=0
        si=size(Mc_range{i,1});
        s=si(2);

        %if Mc_range{i,1}(2,1)==0
            %range_chr=range_chr(range_chr~=i);
            %fprintf('\n%d\n',i);
        %else
            % eliminate the former part/keep the latter part of a single XO --
- update a start point of Mc_range
            if (XO_num(i,1)==1 && range_upd{i,1}(2,1)==chr_length(i,1))
                %fprintf('keep latter %d\n',i);
                %range_upd{i,1}
               if ~(Mc_range{i,1}(1,1) >= range_upd{i,1}(1,1))
                Mc_range=s_function1(i,Mc_range,range_upd{i,1}(1,1));
               end
            end

            % eliminate the latter part/keep the former part of a single XO --
- update an end point of Mc_range
            if (XO_num(i,1)==1 && range_upd{i,1}(1,1)==1)
                %fprintf('keep former %d\n',i);
                %range_upd{i,1}
                if ~(range_upd{i,1}(2,1)>=Mc_range{i,1}(2,s))
                    Mc_range=s_function2(i,Mc_range,range_upd{i,1}(2,1));
                end
            end

            % eliminate the inside part/keep the 2 ourside parts of a double
XO -- update a start and an end point of Mc_range
            if (XO_num(i,1)==2 && range_upd{i,1}(1,1)==1 &&
range_upd{i,1}(2,2)==chr_length(i,1))
                %fprintf('keep 2 outsides %d\n',i);
                %range_upd{i,1}
                if ~(range_upd{i,1}(2,1)>=Mc_range{i,1}(2,s) ||
range_upd{i,1}(1,2)<=Mc_range{i,1}(1,1))

                    if (range_upd{i,1}(2,1)<Mc_range{i,1}(1,1) &&
range_upd{i,1}(1,2)>Mc_range{i,1}(2,s))
                        Mc_range{i,1}(2,1)=0;
                    else
                        if (range_upd{i,1}(2,1)<Mc_range{i,1}(1,1) &&
range_upd{i,1}(1,2)<=Mc_range{i,1}(2,s))

Mc_range=s_function1(i,Mc_range,range_upd{i,1}(1,2));
                        else
                            if (range_upd{i,1}(2,1)>=Mc_range{i,1}(1,1) &&
range_upd{i,1}(1,2)>Mc_range{i,1}(2,s))

Mc_range=s_function2(i,Mc_range,range_upd{i,1}(2,1));
                            else

Mc_range1=s_function2(i,Mc_range,range_upd{i,1}(2,1));
```

```
Mc_range2=s_function1(i,Mc_range,range_upd{i,1}(1,2));

Mc_range{i,1}=cat(2,Mc_range1{i,1},Mc_range2{i,1});
                            end
                        end
                    end
                end
            end

            % eliminate the outside part/keep the inside part of a double XO
-- update a start and an end point of Mc_range
             if (XO_num(i,1)==2 && range_upd{i,1}(1,1)~=1)
                 %fprintf('keep inside %d\n',i);
                 %range_upd{i,1}
                if ~(range_upd{i,1}(1,1)<=Mc_range{i,1}(1,1) &&
range_upd{i,1}(2,1)>=Mc_range{i,1}(2,s))
                    if (range_upd{i,1}(2,1)<Mc_range{i,1}(1,1) ||
range_upd{i,1}(1,1)>Mc_range{i,1}(2,s))
                        Mc_range{i,1}(2,1)=0;
                    else
                        if (range_upd{i,1}(1,1)>Mc_range{i,1}(1,1) &&
range_upd{i,1}(2,1)>=Mc_range{i,1}(2,s))

Mc_range=s_function1(i,Mc_range,range_upd{i,1}(1,1));
                        else
                            if (range_upd{i,1}(1,1)<=Mc_range{i,1}(1,1) &&
range_upd{i,1}(2,1)<Mc_range{i,1}(2,s))

Mc_range=s_function2(i,Mc_range,range_upd{i,1}(2,1));
                            else

Mc_range=s_function1(i,Mc_range,range_upd{i,1}(1,1));

                            %update range_chr
                            if ~(Mc_range{i,1}(2,1)==0)

Mc_range=s_function2(i,Mc_range,range_upd{i,1}(2,1));
                            end
                            end
                        end

                    end
                end
             end

        end
    end


%update range_chr after updating Mc_range's one chromosome

for m=1:s_chr(2)
    i=range_chr_dup(1,m);
 if Mc_range{i,1}(2,1)==0
```

```
            range_chr=range_chr(range_chr~=i);
            %fprintf('\neliminate %d\n',i);
 end
end

end



%result reporting
%count #chr in Mc_range
rch_si=size(range_chr);
rch_num=rch_si(2);

%count #bp in Mc_range
rMc_bp=0;
for n=1:rch_num

    i=range_chr(1,n);
    si=size(Mc_range{i,1});
    s=si(2);

    for j=1:s
        rMc_bp=rMc_bp+(Mc_range{i,1}(2,j)-Mc_range{i,1}(1,j)+1);
    end
end
end
```

- Function 3:

```
%% eliminate the former part/keep the latter part of a single XO -- update a
start point of Mc_range
function Mc_range=s_function1(i,Mc_range,range_upd_value)

j=1;
y=0;
si=size(Mc_range{i,1});
s=si(2);

%if i==2
                    %fprintf('s1-update beginning\t');
                    %fprintf('j=%d\t',j);
                    %fprintf('s=%d\t',s);
                    %fprintf('i=%d\n',i);
                    %Mc_range{i,1}
                    %fprintf('\n');
                     %end

                while (j<s && y==0)

                      if (Mc_range{i,1}(1,j)<range_upd_value &&
Mc_range{i,1}(2,j)>=range_upd_value)
                          Mc_range{i,1}(1,j)=range_upd_value;
                          Mc_range{i,1}=Mc_range{i,1}(1:2,j:end);
```

```
                        y=1;
                    else
                        if (Mc_range{i,1}(2,j)<range_upd_value &&
range_upd_value<=Mc_range{i,1}(1,j+1))
                            Mc_range{i,1}=Mc_range{i,1}(1:2,j+1:end);
                            y=1;
                        end
                    end
                    j=j+1;
                end

                while (j==s && y==0)

                    if (Mc_range{i,1}(1,j)<range_upd_value &&
Mc_range{i,1}(2,j)>=range_upd_value)
                        Mc_range{i,1}(1,j)=range_upd_value;
                        Mc_range{i,1}=Mc_range{i,1}(1:2,j);
                    else
                        Mc_range{i,1}(2,1)=0;
                    end
                    j=j+1;
                end
end
```

- Function 4:

```
% eliminate the latter part/keep the former part of a single XO -- update an
end point of Mc_range
 function Mc_range=s_function2(i,Mc_range,range_upd_value)

 si=size(Mc_range{i,1});
 s=si(2);
 j=s;
 y=0;
   %if i==2
                    %fprintf('s1-update ending\t');
                    %fprintf('j=%d\t',j);
                    %fprintf('s=%d\t',s);
                    %fprintf('i=%d\n',i);
                    %Mc_range{i,1}
                    %fprintf('\n');
                     %end

                    while (j>1 && y==0)

                        if (Mc_range{i,1}(2,j)>range_upd_value &&
Mc_range{i,1}(1,j)<=range_upd_value)
                            Mc_range{i,1}(2,j)=range_upd_value;
                            Mc_range{i,1}=Mc_range{i,1}(1:2,1:j);
                            y=1;
                        else
                            if (Mc_range{i,1}(1,j)>range_upd_value &&
Mc_range{i,1}(2,j-1)<=range_upd_value)
                                Mc_range{i,1}=Mc_range{i,1}(1:2,1:j-1);
                                y=1;
```

```
                        end
                    end
                    j=j-1;
                end

                while (j==1 && y==0)
                    if (Mc_range{i,1}(2,j)>range_upd_value &&
Mc_range{i,1}(1,j)<=range_upd_value)
                        Mc_range{i,1}(2,j)=range_upd_value;
                        Mc_range{i,1}=Mc_range{i,1}(1:2,1);
                    else
                        Mc_range{i,1}(2,1)=0;
                    end
                    j=j-1;
                end
 end
```