DEVELOPMENT OF MUTATION-BASED BREEDING TECHNOLOGY IN FOREST TREE SPECIES

by

Dua'a Riyal
Bachelors in Science (First Class Honors) in Cell and Molecular Biology, Simon Fraser University 2007

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the
Department of Biological Sciences

SIMON FRASER UNIVERSITY

Spring 2011

APPROVAL

Name: **Dua'a Riyal**

Degree: **Master of Science**

Title of Thesis: **Development of mutation-based breeding technology in forest tree species**

Examining Committee:

Chair: **Dr. Tony D. Williams, Professor**

_____

Dr. J. Mattsson, Associate Professor, Senior Supervisor
Department of Biological Sciences, S.F.U.

_____

Dr. S. Bisgrove, Assistant Professor
Department of Biological Sciences, S.F.U.

_____

Dr. P. Unrau, Associate Professor
Department of Molecular Biology and Biochemistry, S.F.U.

_____

Dr. Z. Punja, Professor
Department of Biological Sciences, S.F.U.

Date Defended/Approved: 15 March 2011

Declaration of
Partial Copyright Licence

ABSTRACT

Agricultural breeding based on phenotypic selection of desired variants in self-fertilized populations has been highly successful. This approach has not been applied successfully to forest trees due to the constraints from long generation times and difficulties with controlled large-scale self-fertilization. Novel technologies allow identification of mutants based on DNA sequence deviations in heterozygous mutants, potentially circumventing the need for identification based on phenotype. To test this approach, we generated >6000 poplar trees from mutagenized calli and screened for genetic variants in targeted genes using TILLING technology. While we found genetic variants, DNA sequencing suggested that they were due to natural rather than induced genetic variation. As an alternative approach, we tested mutagenesis of willow pollen and generated a mutant population. We thereafter used novel DNA sequencing technology to screen for genetic variants in targeted loci. This approach identified potential mutants that cannot be explained by background natural variation.

Keywords: Biofuels; ligno-cellulosic feedstock; TILLING; forest tree breeding; *Populus*; *Salix*; ultra-deep DNA sequencing.

DEDICATION

"And say: My Lord! Increase me in knowledge."
(The Holy Qur'an, 20:114)

I dedicate both my undergraduate and graduate achievements to my husband Hussam Kaddoura. Hussam was of great support both emotionally and financially.

In addition, I would like to dedicate my work to my family, especially my parents, who continuously encouraged me throughout my undergraduate and graduate studies.

ACKNOWLEDGEMENTS

I would like to thank Dr. Jim Mattsson for all of his support and guidance during my Master's studies. Jim was always there when I needed help and advice. I did learn a lot from him.

I would also like to thank Dr. Sherryl Bisgrove who gave me support and encouraged me throughout my degree.

I would like to acknowledge Ala'a Zaza for her work on the graphic design of some of the figures that were used.

I would like to thank everyone in Drs. Mattsson's and Bisgrove's labs for their support and friendship. I would like to thank Laura Glesson, Shannon Squires and Adam Foster for their support. I would also like to thank Afsaneh Haghighi-Kia for her help on tissue culture, Lotta Pellonpera for her help on DNA extraction and tissue culture and Muhammad Arshad for his help on bioinformatics.

I would like to acknowledge and thank Drs. Peter Unrau, Erin Gilchrist and George Haughn for their support and encouragement. I would like to acknowledge Dr. Arthur Stipanovic and Kun Cheng for their work on lignin analysis.

I would like to thank the friendly staff at SFU, especially SFU Biological Sciences, for their support and for making SFU a great and diverse learning environment that is open to the world.

Moreover, this work would not have been possible without the financial support from the Natural and Sciences and Engineering Research Council of Canada (NSERC), Simon Fraser University, Mansonville Plastics (B.C.) Ltd. and First Choice Manufacturing Ltd.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Plant Breeding

Plant breeding has been used by farmers for thousands of years to produce new varieties of plants harbouring improved traits compared to parents (Lamkey and Lee, 2006). Plant breeding is a very long process that requires many cycles of screening, selection, crossing and propagation aimed at finding and fixing certain traits in any given population. In most plant breeding programs, a process known as recurrent selection is usually applied during the breeding program (Ceccarelli et al., 2009; Xu, 2010). There are several well-developed breeding methods that have been used for breeding a wide range of plants. Different breeding programs share common steps that involve cycles of selecting parents with desired traits, controlled mating, testing and selecting progenies and seed production from the generated superior individuals (Dickman, 2001; Xu, 2010). Factors like the availability of genetic variation, the biological characteristics of the bred species, desired traits, the surrounding environment, the mode of reproduction (self-pollination versus cross-pollination) and the ability to reproduce sexually and asexually all determine the ideal breeding method to be used (Lamkey et al., 2006; Xu, 2010).

1.1.1 Traditional Plant Breeding

The earliest form of plant breeding was to carefully observe the phenotype and performance of plants under various conditions and then select plants with favourable traits as the source for seeds for the next generation (Ceccarelli et al., 2009; Xu, 2010). Individuals that showed superior characteristics were selected and crossed to produce inbred lines carrying the same favoured traits (Xu, 2010). Such continuous selection under different biotic and abiotic stresses eventually led to the domestication of many of our current agricultural crops, like rice, wheat, maize and potato, that have become major sources for food around the world (Ceccarelli et al., 2009; Xu, 2010). The first evidence was the domestication of cereals in the Fertile Crescent region of the Near East from their wild grass parents that began around 11,000 years ago (Lamkey et al., 2006).

The number of crosses and the size of population in any given breeding program depends on several factors such as the number of traits to be selected for, the availability of genetic variations, the genomic characteristics of genes regulating the trait (such as quantitative versus qualitative) and the breeding method used (Ceccarelli et al., 2009). Plant breeding started with farmers selecting for visible phenotypes, which are generally simple qualitative traits that are controlled by one or few genes that are inherited according to the Mendelian rules of inheritance and that have strong effect on the plant's appearance or performance under various environmental stress conditions (Lamkey et al., 2006). Any changes in the genes could be detected after few generations as

plants are visually screened and then subjected to further controlled crossing and propagation for several generations to fix and capture such traits (Ceccarelli et al., 2009; Xu, 2010). Given the strong effect of qualitative genes on the plant's phenotype, any mutations or changes in these genes led in some cases to the generation of new lines of plants with dramatically different shapes and nutritional values compared to their parents. A good example is the generation of six different vegetables (six different cultivar groups) from a single species of *Brassica oleracea* that belong to the mustard family (Raven et al., 2003). The selection for altered leaf shape led to the generation of Kale, selection for altered stem and flower development led to broccoli, selection for stems let to kohlrabi, selection for terminal bud enlargement led to cabbage, selection for lateral bud led to Brussels sprouts and selection for flower cluster led to cauliflower (Raven et al., 2003). Selection for recessive qualitative genes as compared to dominant genes in diploid plants requires additional genetic crosses and requires working with larger number of individuals in order to find individuals harbouring both recessive alleles. In tetraploid and hexaploid plants, more sets of crosses, several generations and larger populations are required to capture the recessive traits in one plant. In addition, cross-pollinating plants requires additional generations in order to fix certain alleles (Ceccarelli et al., 2009). For example, it took 20-30 harvest seasons to fix the recessive alleles for non-shattering genes in wheat and barley (Ceccarelli et al., 2009).

1.1.2 Genetic Variation: natural versus induced

A key requirement in plant breeding is the availability of genetic variants in order to generate individuals harbouring new and better traits. Any changes to the genetic makeup of plants were based on natural variation and incidents such as natural out crossing, genetic drifts and random spontaneous mutations. It took a very long time, hundreds of generations and large populations to find such variants (Ceccarelli et al., 2009; Xu, 2010). In most cases, it took many years to successfully find and generate lines of plants with improved traits. Once a trait of interest is detected, plant breeders perform several genetic crosses of the individuals harbouring the desired trait, which is known as inbreeding, to fix specific traits. Inbreeding also reduces the available genetic variation in other loci. Thus, an alternative approach was needed to generate the desired genetic variation while maintaining the natural genetic variant in the background loci.

1.1.2.1 Mutational Breeding

In order to speed up the generation of various genetic variants, plant breeders used chemicals and radiation to induce mutations (Ceccarelli et al., 2009; Xu, 2010). Ultraviolet, gamma and X-rays are the most common radiation methods that are used for inducing mutations (Ceccarelli et al., 2009). Seeds are usually subjected to radiation for a short period and then used directly for the generation of the mutagenized population. The resulting plants are then screened to identify mutants (Ceccarelli et al., 2009). Gamma rays are the most commonly used radiation method. It was used for the generation of many

valuable cultivars of several plant species including rice, wheat and barley (Ceccarelli et al., 2009). Examples of gamma rays generated cultivars are the RD6, RD15 rice and Calrose 76 cultivars that were generated using gamma rays in the late 1970s (Ceccarelli et al., 2009). These rice cultivars led to the dramatic increase in rice production in several countries (Ceccarelli et al., 2009). These rice varieties were genetically crossed to several genetic backgrounds to generate cultivars that can survive and grow in different geographical areas. For example, Calrose 76 mutant rice was crossed to more than 20 different genetic backgrounds which resulted in varieties that were grown in China, Thailand and Australia (Ceccarelli et al., 2009). Beside radiation, plant breeders used chemicals to induce random mutations such as Ethyleneimine (EI), Dimethyl sulphate (DMS), sodium azide (NaN3) and Ethyl methane sulphonate (EMS) (Ceccarelli et al., 2009). Unlike radiation treatment, which causes chromosomal breaks or large scale genomic rearrangement, chemical mutagens cause less damage to the genomic DNA, such as causing point mutations and small deletions (Ceccarelli et al., 2009; Xu, 2010).

1.1.2.2 Transgenic Technology

The use of radiation and chemical mutagenesis can result in a collection of random mutations that could be silent, beneficial or detrimental. Sexual reproduction is essential for the transmission of the generated mutations into the desired individuals, which need to be sexually compatible (Lemaux, 2008). In many cases, the generated mutants have to be backcrossed several times with individuals harbouring desired genetic backgrounds to get rid of

undesirable background mutations that might have resulted from the mutagenesis. In addition, selection for desired traits also leads to indirect selection for other traits, which could be either beneficial or detrimental. For example, selection for attached glumes in cereals, also resulted in increased dormancy (Ceccarelli et al., 2009).Thus, methods that are more efficient were needed to allow researchers to specifically introduce or silence desired traits.

With advancements in molecular and genetic techniques and better knowledge of the genes that control traits of interest, plant breeders can now introduce or silence gene(s) of interest in the desired individuals or germplasm (Brown, 2001; Gilchrist and Haughn, 2010). Moreover, genes from different plant species, bacteria, fungi, viruses and mammals, can be introduced to generate novel varieties of plants that express new traits (Lemaux, 2008; Nicholl, 2002). Plants that are manipulated by the introduction of foreign DNA or RNA are known as transgenic plants or genetically modified organisms (GMOs) (Lemaux, 2008). Several techniques can be used to introduce or silence specific genes, such as gene transfer, virus-induced gene silencing and RNA mediated interference (RNAi) (Nicholl, 2002). Genes or small pieces of DNA or RNA could be delivered physically through microinjection, electroporation and particle bombardment (Somers et al., 2009). They could also be delivered biologically through a vector, such as viruses and bacteria (Gilchrist and Haughn, 2010; Nicholl, 2002). The most common approach used to introduce genes is by infecting the desired plant with *Agrobacterium tumefaciens* bacterium harbouring the desired gene(s) (Nicholl, 2002). The gene of interest

is cloned into a bacterial tumour inducing (Ti) plasmid and the bacterium is then used to infect the desired plant transmitting the T-DNA into the plant's genome (Brown, 2001). The infected cells are then used to generate a completely new plant harbouring the inserted genes (Brown, 2001). With the development of transgenic techniques, plant breeders were able to introduce genes for desirable traits into plants with good genetic backgrounds, thereby eliminating the need for sexual reproduction and the need for multiple genetic backcrosses (Xu, 2010). In addition, the introduced genes have a dominant effect and can be transmitted into the next generation according to the Mendelian rules of inheritance (Raven et al., 2003). Transgenic techniques are especially promising for the selection of quantitative traits where the introduction of genes of interest will decrease the number of crosses and the sizes of populations required to increase the frequency or fix quantitative trait loci (QTL) (Ceccarelli et al., 2009; Xu, 2010).

The generation of transgenic plants proved to be very effective in combating a wide range of biotic and abiotic stresses (Collinge et al., 2010; Lemaux, 2008). Several transgenic plants have been generated that carry herbicide genes, either bacterial or extra copy of the endogenous gene, generating herbicide resistant crops. This allowed farmers to treat weeds with herbicides without affecting their crops (Nicholl, 2002). Glyphosate is one of the most common herbicides used to control weeds (Nicholl, 2002). Glyphosate inhibits the action of 5-enolpyruvylshikimate-3-phosphate synthase enzyme, which catalyzes the synthesis of aromatic amino acids (Nicholl, 2002).

Glyphosate herbicide is commercially available as roundup and tumbleweed herbicides (Nicholl, 2002). A Roundup-ready soybean generated by Monsanto Company is one example of roundup herbicide resistant crop that is widely used in the US and Canada (Monsanto Canada Inc, 2006). Roundup-ready transgenic soybean accounts for about 60% of soybean planted in Eastern Canada (Monsanto Canada Inc, 2006). Other Roundup-ready crops include corn, cotton and canola (Schahczenski, 2006). Crops that make their own pesticide against viruses, bacteria, insects and fungi are another example of the successful application of transgenic technology in agricultural plants (Brown, 2001; Farre et al., 2010). Examples of pesticide producing plants include maize, potato and cotton where the *Bacillus thuringiensis* (BT) gene encodes toxins that specifically kill *Lepidopteran* larvae, moths, beetles and mosquitoes (Shimada et al., 2008). BT proteins are activated in the insect's gut where it results in holes in the digestive tract leading to its death (Lemaux, 2008). Transgenic technology was also used to generate plants that are resistant to abiotic stresses such as cold and drought. For example, the antifreeze gene from Arctic flounder was transformed into tobacco and tomato that made them more resistant to cold weather (Farre et al., 2010; Lemaux, 2008). Another example of transgenic plants where the natural biological process was manipulated to increases its performance is the FlavrSavr tomato. FlavrSavr transgenic tomato was the first genetically modified crop that was approved for commercial planting and human consumption in the US in 1994 (Kramer and Redenbaug, 1994; Shimada et al., 2008). The FlavrSavr transgenic tomato was

generated by the use of RNAi technology, where antisense mRNA of the polygalacturonase (PG) enzyme was transformed via *Agrobacterium* (Kramer and Redenbaug, 1994). The antisense mRNA binds to its complementary mRNA and inhibits the translation of the PG enzyme (Nicholl, 2002). These transgenic tomatoes had delayed ripening and thus prevented rotting before being delivered to consumers (Kramer and Redenbaug, 1994; Raven et al., 2003). Several attempts also focused on increasing the nutritional value of some agricultural crops through the use of transgenic technologies. For example, the transgenic golden rice was generated by the introduction of a construct containing cDNA of enzymes that participate in the biosynthesis of provitamin A (β-carotene), including two phytoene synthase gene from daffodil plant and a bacterial phytoene desaturase (Beyer et al., 2002). Endosperm specific promoter was present in the construct to specifically activate β-carotene biosynthesis in the endosperm, the edible portion of the rice plant (Beyer et al., 2002). The generated transgenic rice was able to produce provitamine A in the endosperm (Beyer et al., 2002). Based on a study done by Tang et al. (2009), golden rice was found to contain around 35 μg of β-carotene per gram of rice (Tang et al., 2009). They concluded that 100 grams of uncooked rice fulfilled 55-75% of the recommended daily allowance (RDA) set by the US National Academy of Science for both men and women (Tang et al., 2009). Another field where transgenic techniques are believed to be promising is to genetically alter the lignin content and composition in lignocellulosic biomass to increase its efficiency for pulp, paper and biofuel processing (Abramson et al., 2010).

Several studies showed that transgenically generated lignocellulosic biomass with reduced or altered lignin content were more commercially favourable for pulp, paper and biofuel processing (Simmons et al., 2010).

Due to the successful use of transgenic techniques, more than 25 countries commercially grow genetically modified crops (Collinge et al., 2010). US, Brazil, Argentina, India and Canada are the top five countries that commercially use transgenic crops, where 87% of the world's planted acreage of transgenic crops, are in the US (Schahczenski, 2006). The major transgenic crops grown are herbicide and pesticide resistant crops including soybean, maize, cotton and canola (International Service for the Acquisition of Agri-biotech Applications). Transgenic crops have increased the yield of many agricultural crops and resulted in less damage due to biotic and abiotic stresses, which in turn decreased the amount of fertilizers and pesticides used (Lemaux, 2008; Raven et al., 2003). The use of transgenic crops improved the farmers' profits and return on their investments (Lemaux, 2008). Nonetheless, there are several limitations to the widespread use of transgenic techniques that include technical, ecological and ethical challenges (Lemaux, 2008). Some of the technical difficulties include limited ability of the biological vectors to transform a wide range of plants, limited knowledge of genes or systems that will be manipulated and instability of the genetic constructs (Raven et al., 2003). For example, the limited ability of *Agrobacterium* to transform monocots, which include rice, wheat and barley, limited the generation of transgenic cereal crops (Raven et al., 2003). Ethical and ecological concerns regarding the use of

transgenically modified organisms are the most limiting factors for its commercialized use. The first successful transgenic plant was generated in 1983, but, it was not until 1994 that a transgenic plant, the FlavrSavr tomato, was commercially grown in the US, mainly due to strict regulation on the commercial use and consumption of transgenic crops (Lemaux, 2008). Some believe that GMOs might pose danger to humans and to the environment in the long run (Lemaux, 2008; Nicholl, 2002). For example, there are fears that the increase in the number of pest resistant crops might drive the evolution of pests to become more resistant to insecticides (Nicholl, 2002; Raven et al., 2003). Also, the bacterial kanamycin resistant gene that is usually used in transgenic constructs as a marker might be transmitted to the bacteria that live in human gut making them resistant to kanamycin based antibiotics and also resistant to closely related antibiotics (Brown, 2001). Another concern regarding transgenic crops is that the introduced pieces of DNA, or its translated protein, might act as food allergens (Lemaux, 2008). For example, a study done in 1991 showed that two out of 123 studied farmers had allergic reactions to inhaled BT pesticide (Bernstein et al., 1999). Nonetheless, BT products have been safely used for more than 40 years mainly due to the BT protein being rapidly inactivated by the human's digestive fluids (Lemaux, 2008). Several transgenic crops have been recalled after its commercial distribution and consumption by the public, such as FlavrSavr tomatoes and Starlink corn (Lemaux, 2008). The main reason for the Starlink corn recalls was possible food allergic reactions that may have resulted after consumption mainly due the fact that the BT Cry9C protein structure in the

Starlink corn is similar to known food allergens (Lemaux, 2008). However, to date, there is no clear scientific evidence to back up these fears (Lemaux, 2008). Several countries have passed strict regulations relating to the generation and use of GMO crops, that differ from one country to the other, adding to the complexity and limiting the use of GMO crops worldwide.

1.1.3 Marker Assisted Selection

Early selection processes relied merely on carefully observing and analyzing individuals and then selecting based on phenotype and performance under various environmental conditions (Kole et al., 2008; Xu, 2010). This type of selection required very long time, was labour intensive and required large populations and several generations (Xu, 2010). Advances in using molecular markers, sequencing and other molecular techniques significantly enhanced selection and reduced, or in some cases even eliminated, the need for extensive bioassays and the need for visible phenotypes to generate desired genetic combinations (Kole et al., 2008; Xu, 2010). This in turn allowed for selection during early stages of the plant's life thereby significantly speeding up the selection process and plant breeding programs (Kole et al., 2008; Somers et al., 2009; Xu, 2010). Selection based on molecular markers, known as marker-assisted selection (MAS), refers to the use of markers, mainly genetic markers, to detect the presence or absence of a genetic variant (Kole et al., 2008; Xu, 2010). To be used in molecular breeding, genetic molecular markers should have the following characteristics: neutral effect on the plant's phenotype, evenly distributed throughout the plant's organs and tissues, have high level of

polymorphism, can distinguish heterozygotes from homozygotes linked to the gene or QTL of interest and could be easily detected without considerable expense (Somers et al., 2009; Xu , 2010). Genetic markers could be one base pair long, such as single nucleotide polymorphisms (SNPs), or a series of bases, such as tandem repeats (Xu, 2010). Several molecular techniques are being used to detect and trace molecular markers such as restriction fragment length polymorphisms (RFLP), which was the first molecular technique used for detecting molecular markers in plant breeding programs (Kole et al., 2008), random amplified polymorphic DNA (RAPD), microsatellites and DNA amplification fingerprinting (DAF) (Kole et al., 2008; Xu, 2010). In the case were SNPs are used as molecular markers, Targeting Induced Local Lesions IN Genomes (TILLING) can be used as screening technique. The development of such molecular markers requires knowledge about the genetic makeup, physical map, and genes surrounding the markers (Kole et al., 2008; Xu, 2010). Thus, the development of molecular markers is an expensive, labour intensive and time consuming process, but once it becomes available, its application in further breeding programs on the long term becomes cost efficient and significantly enhances the efficiency of plant breeding programs (Kole et al., 2008; Xu, 2010).

MAS can be applied at various stages of plant breeding programs including selection of parents for crosses, backcrosses for the recovery of recurrent parents with genes of interest, and recurrent selection programs (Lamkey and Lee, 2006; Somers et al., 2009). MAS enabled the detection of

genetic variants at early stages of the plant's life cycle and in heterozygotes rather than the need to obtain homozygotes for visible effect (Lamkey and Lee, 2006; Xu, 2010). MAS is especially useful for the selection of quantitative traits where it is very difficult to screen for visible phenotypes (Xu, 2010). MAS has been used in breeding for quantitative traits in several crops including wheat, rice and maize (Kole et al., 2008), mainly for traits that control the quality of crops such as grain protein content, high grain production, milling yields, seed composition and other quality traits (Kole et al., 2008). The high grain protein content gene (*Gpc-B1*) from *Triticum turgidum* wheat is one of the many traits where molecular markers are well established and have been used in breeding programs where RFLP, SSR and sequence specific PCR are used to identify and trace the *Gpc*-B1 gene (Kole et al., 2008). MAS has been gaining more popularity in recent decades. Emerging knowledge about the sequences of many plants along with the emergence of physical genomic maps and the development of molecular techniques are collectively aiding in the use of MAS and molecular aided breeding to further improve plant breeding. It is expected that the use of MAS in plant breeding will continue to grow in the future (Ceccarelli et al., 2009; Kole et al., 2008; Xu, 2010).

1.1.4 The Green Revolution

The continuous growth in the world's population is increasing the demand for food, land and water. This requires improvements in the efficiency of food production. The continuous selection process by plant breeders and farmers for plants with superior phenotypes and crossing different species

together drove the evolution of many plant species. This eventually led to the domestication of many of the current agricultural crops such as wheat, cereals, potato, rice and maize from their wild weedy ancestors (Ceccarelli et al., 2009; Xu, 2010). There is an increasing need to improve the yield and efficiency of food production worldwide. Continuous improvements in the field of agriculture to help meet growing needs were possible through the collaborative efforts of farmers and scientists as well as advances in technology (Raven et al., 2003). Advances in irrigation techniques, fertilizers, pesticides, the use of efficient and specialized machinery, the use of molecular tools, transportation systems and storage all together helped in improving the agricultural field and eventually resulted in what is known as “The Green Revolution” (Evenson and Gollin, 2003; Raven et al., 2003). Plant breeding played a major role in the success of the “Green Revolution” in the 20th century, especially in agricultural plants such as wheat, maize, and rice, that have helped in saving millions of lives around the world (Muangprom et al., 2005; Evenson and Gollin, 2003). The main benefit of the Green Revolution was increasing yield per unit of land area (Xu, 2010). For example, cereal production in developing countries increased by as much as 92% in 1990 as compared to 1961 (Xu, 2010).

It is believed that the generation of high yield crops obtained through various plant breeding techniques is responsible for one half of the increase in productivity of major crops between the 1960s and 1980s (Xu, 2010). For example, the introduction of semi dwarf genes into wheat and rice played a key role in the dramatic increase in their production (Evenson and Gollin, 2003;

Peter, 2002). The main objective was to increase yield, but, the problem that emerged was that increased yield meant heavier grains with tall thin stems, which forced the plant to bend due to its heavier weight, a process known as lodging, and eventually rot (Peter, 2002). The introduction of the reduced height gene (*Rht*) in wheat played a major role in the green revolution of wheat where high yields with limited loss have been achieved through decreases in the plant's height enabling it to withstand heavier grains. The semi dwarf trait originated from Japan where a semi-dwarf wheat cultivar known as Daruma was crossed with an American high yield cultivar (Peter, 2002). The genetic cross resulted in the generation of wheat cultivars that are semi dwarf with high yield (Peter, 2002). One of the generated progenies, known as Norin-10-Brevor 14, was then sent to Norman Borlaug, a famous plant breeder who won the Nobel prize in 1970 for his work in plant breeding and contribution to the Green Revolution, at the Centro International de Mejoramiento de Maize y Trigo (CIMMYT) in Mexico where he crossed it with a local wheat cultivar that is well adapted to growing in tropical and sub-tropical areas (Peter, 2002). Semi dwarf plants are achieved when heights are reduced by less than 50% in order to obtain shorter plants (Peter, 2002). The introduction of these dwarfing genes in addition to the use of fertilizers and other advanced agricultural techniques and machinery converted Mexico from an importing country in 1944 to not only being self sufficient but also exporting country by the year 1964 (Raven et al., 2003). These dwarfing genes are now found in 70% of the wheat cultivars worldwide (Peter, 2002). The introduction of semi dwarf traits into other crops,

like rice and maize, also lead to the increase of their production and collectively led to the increase in food production (Evenson and Gollin, 2003).

1.1.5 Forest Tree Breeding

The growth of human population on earth is continuously consuming space from our rainforests and exhausting its resources and wild life. Natural resources found in our forests, like wood products from trees, are being consumed at a faster rate than they are being replaced (Huntley et al., 2003). These natural resources play a very important role in many local economies around the world. For many years, research related to forest products has concentrated largely on forest management and sawmill, pulp, wood and paper technologies. Given the economic importance of forest trees as well as their environmental importance, there is now an increasing understanding that future efforts need to focus on tree breeding and the use of biotechnological advances to generate varieties that have traits that make them better suited for their intended use (Bhalerao et al., 2003). Those trees may need to be grown in tree farms to avoid affecting the natural population in our forests.

Several constraints like long generation time, long flowering time that might take up to 40 years in some tree species, the large size of populations that consume large spaces (Dickmann et al., 2001; Flachowsky et al., 2009), the difficulty in pollinating and the long time it takes to perform backcrosses to homogenize the desired traits limited the application of agricultural plant breeding techniques in forest trees (Flachowsky et al., 2009). In fact, backcrossing is not feasible in the majority of forest tree species mainly due to

the long generation cycle of forest trees (Flachowsky et al., 2009). In addition, inbreeding depression, that some tree species experience, and inter-species incompatibility have hampered breeding in forest trees (Dickmann et al., 2001). All of these difficulties hindered the application of conventional agricultural plant breeding techniques in breeding forest trees and made it very expensive and time consuming (Ellis et al., 2010; Flachowsky et al., 2009). In addition, unlike in agricultural and herbaceous plant breeding programs, more complex traits are usually of interest to breeders of forest trees such as wood quality, adaptability to various environmental stresses and long life (Ellis et al., 2010). Forest tree breeding programs usually involve “selection breeding” where a small number of elite trees are selected for controlled crossing and further propagation (Flachowsky et al., 2009). The elite trees are chosen mainly based on morphology and performance under a given environment (Flachowsky et al., 2009). The availability of genetic information, such as genome sequences, ESTs, and physical maps, have helped to apply existing molecular breeding techniques to forest trees (Flachowsky et al., 2009; Yang et al., 2009). For example, the use of MAS will help to detect and trace genetic variant at early growth stages of forest trees allowing for the selection of desired individuals without the need for additional crosses to observe the phenotype (Kole et al., 2008). In addition, there has been promising improvements in techniques used to generate transgenic forest trees where gene(s) of interest can be transformed into desired germplasm without the need to homogenize the gene(s) of interest through sexual reproduction, especially where genetic

outcross tend to dilute the desired trait (Ellis et al., 2010). However, to date, the generation of transgenic forest trees is only carried out in laboratories and greenhouses as no commercial application is yet allowed (Flachowsky et al., 2009; Lemaux, 2008). Despite the difficulties associated with forest tree breeding, plant breeding has been applied to some tree species, including pine, *Eucalypts* and *Populus* trees (Flachowsky et al., 2009) and there is no inherent reason to believe that trees cannot be shaped by breeding into various desired varieties considerably different from their wild relatives.

1.2 The Cell Wall

The plant cell wall is a rigid outer structure that encloses the plant's cells (Taiz and Zeiger, 2006). The cell wall is one of the major structures that differentiate plant cells from animal cells (Raven et al., 2003). The cell wall provides plant cells with support, rigidity and protection from the exterior environment, as well as connects plant cells together and gives plant cells their unique morphology (Keegstra, 2010; Raven et al., 2003). In addition, the cell wall plays an important role in protecting plant cells from bursting as a result of water uptake under osmotic pressure (Jarvis, 2009). The plant cell wall is made up of cellulose, polysaccharides, proteins and lignin that are linked together by both covalent and non-covalent bonds forming an organized and complex matrix (Taiz and Zeiger, 2006). Even though cell walls of different plant cells vary greatly in their composition, all cell walls share the main function, which is determining the cell's unique morphology and size and providing plant cells with strength and rigidity (Raven et al., 2003). In addition to its importance to plant

cells, cell walls are also a major raw material that is used for paper, textile, fibers, lumber and biofuel production (Wall et al., 2008).

Cellulose is the most abundant molecule in the plant's cell wall. Cellulose ($(C_6H_{10}O_5)_n$) is a homopolymer made up from glucose monomers that are bound together with 1,4 β-D-glycosidic bonds resulting in fibres of repeated glucose monomers (Raven et al., 2003; Taiz and Zeiger, 2006). Cellulose chains are bounded together by hydrogen bonds that form between the hydroxyl groups (-OH) that project from the glucose molecules forming a highly organized and tight structure known as cellulose microfibrils, as shown in figure 1-1 (Raven et al., 2003). The length and width of cellulose microfibers vary widely within different parts of the plant and among different plants. It ranges from as little as 10 nanometres in diameter to 40 nanometers (Raven et al., 2003; Taiz and Zeiger, 2006). Cellulose microfibers are cross-linked to other polysaccharides and proteins found in the cell wall, mainly hemicellulose and pectin, through hydrogen bonds resulting in a highly organized and stiff crystalline structure (David et al., 2010). Hemicellulose is made up of a mixture of six carbon (hexose) and five carbon (pentose) sugars (Cherubini and Stromman, 2010). Xylan is the major pentose sugar found in hemicellulose (Cherubini and Stromman, 2010). Other monosaccharides present in the hemicellulose include glucose, mannose, galactose, arabinose and uronic acids (David et al., 2010). Unlike cellulose, chains of hemicellulose do not tend to aggregate mainly due to their heterogeneous nature, where it is more difficult to aggregate different monosaccharides that have different chemical and physical

characteristics and different bonding types and positions (Burton et al., 2010). For example, the presence of 1,3-β-linkage in hemicellulose generates molecular kinks that greatly prevent hemicellulose chains from aggregating (Burton et al., 2010). Nonetheless, the heterogeneous nature of hemicellulose serves as glue that helps to cross link cellulose microfibrils together and binds to lignin to form a very stiff matrix (Umezawa, 2010).

Lignin is a complex highly branched phenolic polymer that is the second most abundant molecule in the plant's cell wall (Raven et al., 2003; Christiernin, 2006). Lignin is made up of three different phenylpropanoid alcohols known as monolignols: guaiacyl ($C_{10}H_{12}O_3$; G), syringyl ($C_{11}H_{14}O_4$; S) and p-hydroxyphenyl ($C_9H_{10}O_2$; H) (Hu et al., 1999; Vanholme et al., 2008). The major difference between these monolignols is the presence of methoxy (OCH_3) side chains that project from the aromatic rings. H monolignol has no methoxy group, while G monolignol has one methoxy group at the aromatic C3 position and S monolignol has two methoxy groups at the aromatic C3 and C5 positions (Umezawa, 2010). The proportion of these subunits varies greatly among different species and even within individuals from the same species and in some cases within the different layers of a single cell wall (Christiernin, 2006; Somerville et al., 2010). Since lignin is made up of three different monolignol subunits, lignin molecules greatly vary in composition, molecular weight and number of intermolecular bonds depending on the ratio of the three different monolignols presents (Hu, 2002). Monolignols form a very complex three dimensional macromolecule that is covalently bonded with mainly C-C and C-O-

C (β-O-4 linkage) bonds that can be organized in various ways resulting in a very complex structure that can posses different physical and chemical characteristics (Sazanov and Gribanov, 2009; Taiz and Zeiger, 2006). Like hemicellulose, lignin is an amorphous macromolecule, that is lignin does not tend to aggregate into a crystallized structure (David et al., 2010). Lignin covalently binds to different polysaccharides present in the cell wall, including the hemicellulose-cellulose complex, thus generating a very stiff and complex lignocellulosic matrix (David et al., 2010; Somerville et al., 2010). Lignin provides the plant with stiffness, compressive strength, defence, decay resistance, helps define the cell's final shape, and aids in the transportation of water (Somerville et al., 2010; Umezawa, 2010). For example, lignin present in the xylem facilitates the upward transport of water by protecting it from the negative pressure and that way water is transported without collapsing the plant's tissue (Raven et al., 2003).

1.2.1 Differentiation of Primary and Secondary Cell Walls

The plant cell wall is made up of two parts: primary cell wall and secondary cell wall (Raven et al., 2003). As the plant cell grows, the primary cell wall is deposited into the extracellular matrix between the plasma membrane and the middle lamella that connects two adjacent plant cells (Taiz and Zeiger, 2006). The primary cell wall is made up of cellulose microfibrils that are embedded in a matrix of polysaccharides consisting of hemicelluloses and pectin (Raven et al., 2003), as shown in figure 1-1. The primary cell wall is a very dynamic structure where new materials are deposited as the plant cell

grows and differentiates (Keegstra, 2010). Newly formed cellulose and polysaccharides are deposited near the plasma membrane pushing the existing cell wall outwards (Keegstra, 2010).

The secondary cell wall forms after the cells stop growing (Raven et al., 2003; Taiz and Zeiger, 2006). The secondary cell wall is deposited primarily between the plasma membrane and the primary cell wall and is organized in multiple layers (Taiz and Zeiger, 2006), as shown in figure 1-2. The primary cell wall is usually present as a thin layer with a diameter of less than one μm (Taiz and Zeiger, 2006). On the other hand, the secondary cell wall is organized into three distinct layers known as S_1, S_2 and S_3, where S_3 is the most inner and recent layer (Raven et al., 2003), as shown in the figure 1-3. Cellulose and hemicellulose are found in both the primary and secondary cell walls but they are more abundant and highly organized in the secondary cell wall (Raven et al., 2003). For example, the secondary cell wall has 40-80% more cellulose than the primary cell wall, where more than 80% of the cellulose microfibrils are located in the S_2 layer (Hing-arrasco and Johnsen, 2010). In addition, the different orientation of the cellulose microfibrils in the three layers greatly increases the rigidity of the secondary cell wall (Raven et al., 2003). Once cells reach their final size and shape, the secondary cell walls are deposited (Keegstra, 2010). This is followed by lignification of the secondary cell walls, where monolignols bind together forming a complex three dimensional lignin molecule which binds with the cellulose, hemicelluloses and other proteins found in the cell wall (Christiernin, 2006; Demura and Fukuda, 2007), as shown

in figure 1-2. The width and composition of the secondary cell wall varies greatly depending on the cell type and function (Burton et al., 2010). Plant cells that are involved in mechanical support and transportation have a thicker and more complex secondary cell walls, such as tracheary elements and fibres (Taiz and Zeiger, 2006). These types of cells usually undergo programmed cell death after the secondary cell walls are deposited and have completed lignification (Christiernin, 2006; Demura and Fukuda, 2007). Most of these types of cells are found in the stem where their main function is to transport solutes and water between different parts of the plant and provide posture and support for the plant (Raven et al., 2003). For example, cells in the xylem tissue, such as tracheids and vessel elements, undergo extensive thickening and lignification of their secondary cell walls and deposit several layers of secondary tissue (Taiz and Zeiger, 2006). Higher content of cellulose microfibrils and lignin in the secondary cell walls of these types of cells provides stiffness and rigidity to the cells which is essential for them to perform their function (Raven et al., 2003).

1.3 Biofuels

Global warming is one of the major challenges that humans are facing in recent decades. The huge amount of carbon dioxide emissions from burning of fossil fuels is believed to be the major contributor to global warming (Cherubini and Stromman, 2010; Ragauskas et al., 2006; Soetaert and Vandamme, 2009). It is estimated that fossil fuels currently accounts for 82% of the world's energy sources, mainly in the form of petroleum, natural gas and coal (Brown, 2003; Sivakumar et al., 2010). It is forecasted that fuel energy consumption will

increase by at least 50% by year 2025 (Ragauskas et al., 2006; World Energy Council, 2010). Only recently have we started to realize the tremendous danger that we are facing due to the continuous and surprisingly fast developing effects of global warming. The release of sulphur dioxide, nitrogen oxides and other pollutants from the combustion of fossil fuels resulted in their accumulation in the atmosphere and believed to have been contributing to ozone depletion and the generation of acid rain (Brown, 2003; www.doegenomestolife.org/biofuels/). Thus, there is a need to find alternative sources of energy that have less negative effects on the environment.

Beside the effect on the environment, there are concerns regarding long term availability and future prices of fossil fuels (Soetaert and Vandamme, 2009). The generation of fossil fuels takes millions of years and its rapid consumption will eventually deplete all available fossil fuels (Soetaert and Vandamme, 2009). Thus, there is a need to find alternative sustainable and renewable sources of energy (Soetaert and Vandamme, 2009). It is estimated that the current fossil fuel supply will vanish in 50 to 200 years if the pace of consumption remains the same and if no new reserves are found (Soetaert and Vandamme, 2009). In turn, prices of fossil fuels are expected to continuously rise as world reserves continue to decrease (Sivakumar et al., 2010; Soetaert and Vandamme, 2009). In addition, securing energy sources and supplies is threatening the security of many nations around the world (Soetaert and Vandamme, 2009; www.doegenomestolife.org/biofuels/). Despite the technical advances and lower cost of fossil fuel extraction and processing, attention and

efforts are being directed toward the use of alternative energy sources that are readily available, renewable, sustainable and more environmentally friendly such as solar, wind, water and renewable biomass (Soetaert and Vandamme, 2009; Vertès et al., 2010; World Energy Council, 2010).

Biomass, which is any organic material obtained from a biological origin, can be used for the production of energy, liquid fuels, chemicals and fibers (Brown, 2003; Wall et al., 2008). It is estimated that energy production from biomass can supply more than half of the world's energy needs by the mid of the 21st century (Soetaert and Vandamme, 2009; World Energy Council, 2010). Agricultural crops, trees, forest and agriculture crop residues, municipal waste and animal remains are all example of biomass that can be used for the generation of various forms of bioenergy (Brown, 2003; Soetaert and Vandamme, 2009). Solar energy stored in the form of chemical bonds in organic molecules, such as carbohydrates, lipids and proteins, can be utilized for the generation of various energy forms including heat, electricity and transportation fuels (Brown, 2003; Centre for Energy, 2010; Ragauskas et al., 2006).

The use of biomass to generate energy, mainly burning wood for generating heat and for cooking, was one of the first energy sources used and it is still being used to date (Centre for Energy, 2010). The focus is now shifting to the use of various biomass resources for the efficient generation of energy in its various forms in large scale to help meet the growing energy needs while reducing the negative impact on the environment (World Energy Council, 2010;

Centre for Energy, 2010). The goal is to expand the use of biofuels for the generation of liquid transportation fuels mainly in the form of ethanol, biodiesel, and biogas (Brown, 2003; Centre for Energy, 2010; Vertès et al., 2010). The net carbon balance, which refers to the amount of carbon dioxide released from fuel production and consumption minus the amount of carbon dioxide fixed by the biomass, is analysed to compare the effects of various types of fuels on the environment (Soetaert and Vandamme, 2009; Yuan et al., 2008). Burning fossil fuels releases carbon dioxide stored over millions of years into the environment (Brown, 2003). On the other hand, the combustion of biofuel releases the carbon dioxide that was recently consumed in the formation of organic molecules (Sivakumar et al., 2010). Depending on the type of biomass and the process used for bioenergy generation, the net amount of carbon dioxide released might be equal or even less than the amount of carbon dioxide fixed by the biomass, thus achieving a negative carbon balance (Yuan et al., 2008). It is believed that the use of biofuel will result in less greenhouse gas emissions into the environment (Soetaert and Vandamme, 2009; Vertès et al., 2010).

There are also many constraints associated with the use of biomass for energy production, especially for transportation fuels. One major constrain is the energy requirements for the various stages involved in the production of bioenergy, such as the energy required for biomass planting, harvest, fertilization, transport and processing (Vertès et al., 2010). In many cases, the overall process of bioenergy production might consume more energy than the energy actually produced, thus having a negative net energy balance

(Sivakumar et al., 2010). Continuous improvements and technological advances in various processes involved in bioenergy production is significantly decreasing energy input requirements. For example, it was found that ethanol production from corn biomass in the US during the 1970s to 1980s required energy inputs, mainly from fossil fuels, equal to the bioenergy finally produced (Soetaert and Vandamme, 2009). Later developments in corn planting, increased yield, improved cropping and processing, and the use of modified and improved enzymes for fermentation have all played a major rule in decreasing the total fossil energy required for ethanol production from corn in the US by 40% and increased bio-ethanol production by more than 20% since the 1980s (Soetaert and Vandamme, 2009). Another constraint to bioenergy production is that biomass are mainly solid which requires energy for it to be harvested, stored and transported and then ground or liquefied before it can be used for energy production (Brown, 2003). In addition, freshly cultivated biomass have high moisture content that adds to its weight which significantly increases the amount of energy required for handling and transport and requires larger spaces to store (Brown, 2003).

1.3.1 Current use of Biomass for Biofuel Production

The transportation sector is one of the major sectors that consume energy as more than 32% of the world's energy, mainly in the form of liquid fuels, is consumed by the transportation sector (Soetaert and Vandamme, 2009). Over the years, there has been considerable and continuous advancements in using plant biomass for biofuel production. Sugars, starches

and oils obtained from plants are among the first plant biomass to be successfully used for biofuel production, especially as transportation fuel; and hence are known as first-generation biofuels (Cherubini and Stromman, 2010; Yuan et al., 2008). Fermentation of agricultural crops is one of the first and now well-developed methods used for generating ethanol (C_2H_5OH), which is mainly used as transportation fuel (World Energy Council, 2010). Ethanol can be produced from sugars and starches, mainly from glucose sugars ($C_6H_{12}O_6$), found in many agricultural crops such as sugar cane, sugar beet, corn, wheat and potato (Soetaert and Vandamme, 2009; Vertès et al., 2010). Bio-ethanol production from plants requires physical and thermal treatment to separate the sugar molecules, in the form of glucose, from other cell components (Cherubini and Stromman, 2010). This is followed by an enzymatic reaction, which is aided by yeasts or bacteria, to finally ferment glucose releasing ethanol (Vertès et al., 2010). Glucose is fermented into ethanol releasing carbon dioxide and heat (Brown, 2003). Upon combustion, ethanol reacts with oxygen releasing carbon dioxide and heat which is used to power engines (Brown, 2003; Soetaert and Vandamme, 2009).

Brazil followed by the United States are among the first countries to produce and actually use ethanol from plant biomass in large scale (World Energy Council, 2010; Vertès et al., 2010). The ethanol produced is mixed with other fossil fuels, mainly gasoline, at different ratios and used as transportation fuel, mainly for automobiles (Soetaert and Vandamme, 2009). Brazil was the first country to use sugar cane for bioenergy production (Vertès et al., 2010;

World Energy Council, 2010). Ethanol produced from the fermentation of sugars in sugar cane has been successfully used as biofuel in Brazil for more than 30 years (Soetaert and Vandamme, 2009). Large-scale ethanol production from sugar cane started in 1975, and ever since, ethanol production increased by more than 28 times (Soetaert and Vandamme, 2009). The first step involved in ethanol production from sugar cane and other sugar feedstocks is grinding the stalk and stem via mechanical rollers (Vertès et al., 2010). The generated juice is heated and calcium hydroxide is added to the juice to precipitate the fibers and sludge and then the juice is filtered (Vertès et al., 2010). The crystallized sugar present in the filtrate is then centrifuged (Vertès et al., 2010). The extracted simple sugar is then fermented via the use of yeast mainly *Saccharomyces cerevisiae* yeast finally releasing ethanol (Soetaert and Vandamme, 2009). Ethanol is then distilled and purified (Soetaert and Vandamme, 2009). The generated ethanol is mixed with gasoline, mainly to serve as an oxygen and octane number enhancer, in different ratios and is used as automobile fuel without the need for engine adjustments (Vertès et al., 2010).

Brazil was the largest ethanol producer for 30 years until 2006 when the US became the largest producer of ethanol mainly from corn (Soetaert and Vandamme, 2009). Corn was chosen by the US for ethanol production due to several reasons, including suitable US climate for corn production, its abundance, low cost, richness in sugars in the form of starch and relatively long shelf life (Vertès et al., 2010). In addition, with starch comprising 70% of the

corn's dry weight, corn became the major agricultural crop used for ethanol production in the United States and the world's second largest ethanol feedstock (Soetaert and Vandamme, 2009; Vertès et al., 2010). Starch refineries for ethanol production from corn in the US are divided into two main groups, dry mills and wet mills, depending on the technique used for grinding and starch isolation (Soetaert and Vandamme, 2009). Dry mills are the major ethanol bio refineries in the US (Soetaert and Vandamme, 2009). In dry milling refineries, corn kernels are first ground and the generated powder is then sieved and cooked in hot water, while in wet milling refineries, corn kernels are first treated with sulphurous acid and then ground (Soetaert and Vandamme, 2009). The isolated starch molecules are then liquefied and hydrolyzed using alpha amylase enzyme, which breaks starch into smaller glucose polymers known as dextrins in a process known as starch saccharification (Vertès et al., 2010). Finally, glucose polymers are then fermented mainly via the treatment with yeasts, mainly *Saccharomyces cerevisiae* (Soetaert and Vandamme, 2009). Currently in the United States, ethanol is being mixed with gasoline in a ratio of around 10% and used as liquid biofuel in about 18% of vehicles (Centre for Energy, 2010; Vertès et al., 2010). In Canada, around 6% of energy needs are generated from biomass, mainly generating heat, steam and electricity from burning of forest bi-products, wood chips, sawdust and bark (Centre for Energy, 2010). Wheat is considered to be the main crop that can be used for ethanol production in Canada (Centre for Energy, 2010). Around 175 million litres of ethanol is annually produced in Canada from wheat and corn and that number

is expected to rise rapidly (Centre for Energy, 2010). Currently, ethanol is being used as a fuel additive in a ratio of 10%, ethanol to gasoline, (Centre for Energy, 2010).

1.3.2 Lignocellulosic Biomass

Lignocellulosic biomass refers to the plant's organs that are rich in cellulose, hemicelluloses and lignin polymers (Brown, 2003). Lignocellulose is mainly found in the plant's cell wall (Umezawa, 2010). Lignocellulosic biomass is known as second-generation biomass (Umezawa, 2010). Lignocellulose can be found in the cell wall of herbaceous and woody biomasses (Brown, 2003). Perennial grasses, such as switchgrass, and agricultural residues, are examples of herbaceous lignocellulosic feedstock (David et al., 2010; Yuan et al., 2008). Stems, stalks, corncobs, barns, sugarcane bagasses, and straws are all examples of agricultural residues that are rich in lignocellulose (Brown, 2003). Forest trees and woody plants, such as *Populus* and *Salix*, which are also known as short rotational woody crops, are also rich in lignocellulose (Umezawa, 2010). Forestry residues such as woodchips, tree bark, shaving, dead trees, as well as by-product of pulp and paper processing, logging and fire leftovers are all lignocellulosic biomass that can be used for energy production (Sivakumar et al., 2010). Based on a study done by the US Department of Energy (DOE) and the United States Department of Agriculture (USDA) in 2005, more than 1.3 billion dry tons of lignocellulosic feedstock could be available by mid 21st century for bioenergy production, where 368 million tons are expected

to come from forest biomass and 998 million tons from agricultural lignocellulose (www.doegenomestolife.org/biofuels/; Vertès et al., 2010).

Given the success achieved so far in bioenergy production from first-generation biomass, researchers are now examining the possibility of using lignocellulosic biomass for biofuel production in the form of ethanol, biogas and biodiesel (www.doegenomestolife.org/biofuels/). There is need to expand the use of biomass for bioenergy production to help meet the world's growing energy needs and in order to increase its market share as an alternative energy source to fossil fuels (Umezawa, 2010). Since lignocellulosic feedstock covers wider variety of plant biomass, including both herbaceous and woody plants and residues, lignocellulosic biomass are more readily available in a wider variety and quality as compared to first-generation biomass (Wall et al., 2008). It is expected that lignocellulosic feedstock will be able to replace up to 90% of traditional fossil fuels, unlike agricultural crops which so far only replace around 3% of traditional fossil fuels (Abramson et al., 2010). First generation biomass that are currently used for bioenergy production could then be used instead as sources for food and animal feed as there is limited supply of land to grow them (Abramson et al., 2010; Umezawa, 2010). If the use of agricultural crops for bioenergy production increases, prices of agricultural crops will increase and there might not be enough food supply to satisfy the rapidly growing human population (Cherubini and Stromman, 2010; Ragauskas et al., 2006). Attention is recently shifting from the use of agricultural crops to the use of inedible plant biomass, such as forest trees, for bioenergy production.

Cellulose is the most carbon sequestering molecules on earth (Taiz et al., 2006). Great amount of solar energy is stored in the bonds between the carbon molecules in cellulose and hemicelluloses, which makes lignocellulosic biomass an ideal biofuel (Cherubini and Stromman, 2010; Sivakumar et al., 2010). Carbon is also present in the building blocks of lignin (Hu, 2002). In fact, the energy stored in lignin polymer is higher than the energy stored in cellulose and hemicellulose (Hu, 2002; Novaes et al., 2010). On average, the energy stored in lignin molecules is 30% more than that stored in cellulose (Novaes et al., 2010). The higher energy content of lignin is mainly due to the higher carbon concentration (Novaes et al., 2010). The use of lignocellulosic feedstock as biofuels is believed to have a more favourable net energy balance compared to the use of sugar, starch and oils obtained from agricultural crops (Abramson et al., 2010). For example, the net energy balance from ethanol production from maize and sugarcane ranges between 10 to 80 gigajoules per hectare per year (Gj/ha/yr) while ethanol production from switchgrass and *Populus* ranges between 150 to 550 Gj/ha/yr (Yuan et al., 2008). A study conducted in Belgium showed that the net energy generated per hectare of poplar and willow, mainly as firewood, is higher than the energy generated from most biofuels obtained from first generation biomass (Soetaert and Vandamme, 2009). Biofuel yield from wheat, sugar beets and rapeseeds ranged between 1.28 to 4.84 trillion per hectare per year (t/ha/yr) while 10.8 t/ha/yr was generated from poplar and willow (Soetaert and Vandamme, 2009).

Researchers also believe that the net carbon dioxide balance of bioenergy production from lignocellulosic feedstock will most likely be negative, which means that the carbon dioxide released into the environment will be less than carbon dioxide fixed in the plant (Abramson et al., 2010). This will hopefully help to reduce or at least slowdown global warming. It is also believed that greenhouse emissions from the combustion of lignocellulosic feedstock are much lower than that from the combustion of first-generation biomass (Cherubini and Stromman, 2010; World Energy Council, 2010). It is estimated that the use of second-generation biomass for biofuel production for the transportation sector will help to lower greenhouse gas emissions by up to 85% as compared to the use of fossil based fuels (Abramson et al., 2010; U.S Department of Energy, 2010). Whereas, the reduction in greenhouse gas emissions from the use of biofuels generated from first generation biomass is estimated to only range between 30 to 60% (Vertès et al., 2010). The lower net energy balance and higher carbon balance obtained from first-generation biomass is due mainly to the high energy inputs required for the cultivation, propagation, maintenance and processing of first-generation biomass (Cherubini and Stromman, 2010; Keegstra, 2010). Most of this needed energy is fulfilled using fossil fuels, which in turn increases greenhouse emissions and results in a neutral or even positive carbon balance (Yuan et al., 2008). In addition, woody and some perennial grass feedstock usually have extensive root systems that help to sequester atmospheric carbon and store it deep into the soil (Carroll and Somerville, 2009). The extensive root system also helps in

fixing nitrogen in the soil (Carroll and Somerville, 2009). Since perennial grasses and woody biomass are not harvested every year, the carbon sequestered into the soil remains in the soil and this also helps to limit soil erosion and losses of minerals (Carroll and Somerville, 2009). This also decreases the amount of energy inputs and labour required for harvest (Soetaert and Vandamme, 2009; Carroll and Somerville, 2009). Most second-generation biomass does not require extensive care to cultivate, maintain, and require less fertilizers than first-generation feedstock (Carroll and Somerville, 2009).

The highly complex and inter-crossed cell wall in plants makes working with plants, especially lignocellulose rich plants, very difficult as it increases the number of steps and treatments required to process lignocellulosic feedstock (Sazanov and Gribanov, 2009). In addition, the presence of lignin makes the plant cell wall highly resistant to microbial and chemical attacks (Ragauskas et al., 2006). The separation and depolarization of lignin is the most expensive step in lignocellulosic feedstock processing where it accounts for 80% of the time and energy used for lignocellulose processing during pulp, paper and biofuel production (Cherubini and Stromman, 2010). The great amount of energy required for separating cell wall components and for cellulose and lignin hydrolysis and fermentation made the use of lignocellulose as biofuel very difficult and challenging, and to date hindered its large-scale use as biofuel for the transportation sector (Yuan et al., 2008).

In recent years, scientists have been working hard at efficiently producing biofuels, fibers, pulp and chemicals from lignocellulosic biomass (McCann and Rose, 2010). More than 44% of the lignocellulosic biomass is made up of cellulose, 30% is made up of hemicellulose and 20% is made up of lignin (www.doegenomestolife.org/biofuels/). The major obstacle faced during lignocellulosic biomass processing, for bioenergy production and pulping, is the tightly bounded and cross-linked nature of the lignocellulose matrix (Chen and Qiu, 2010). Cellulose, unlike starch, is highly organised into microfibrils and cross linked with hemicellulose and lignin resulting in a highly crystallized structure that resists hydrolysis into free glucose molecules (Umezawa, 2010). The resistance of cellulose to hydrolysis is known as cellulose recalcitrance (Brown, 2003; Umezawa, 2010). Lignin, present in the cell wall, masks and protects cellulose molecules from hydrolytic enzymes, thus preventing cellulose hydrolysis (Abramson et al., 2010). For example, 80% of the pores found in the lignocellulosic matrix are smaller than the sizes of the hydrolytic enzymes, cellulases, which are large proteins with molecular weight ranging between 30,000 and 60,000 (Brown, 2003). In addition, lignin can bind to cellulases, enzymes that hydrolysis cellulose, which decreases the number of free cellulases available for cellulose hydrolyses thereby decreasing hydrolysis efficiency and increasing the amount of cellulases required (Brown, 2003). Lignin is also resistant to most chemical and biological treatments which add to the complexity and difficulty of lignocellulose biomass processing (Brown, 2003). In addition, the presence of acids in the hemicellulose and lignin, such as

acidic acid and furifuric acid that are released during various pre-treatments, greatly hinders microbial activity during fermentation (Abramson et al., 2010).

Differences in the chemical, physical and thermal properties of cellulose, hemicellulose and lignin makes it very difficult to find a single method for lignocellulosic biomass processing (Mosier et al., 2005; Vertès et al., 2010). Moreover, lignin and hemicellulose are made up from a collection of different subunits that differ significantly between different biomass and even within the cells of one biomass (Abramson et al., 2010). Each different combination of the subunits in both hemicellulose and lignin affects their physical and chemical properties (Abramson et al., 2010). In order to ferment hemicellulose, a mixture of enzymes are required to hydrolyse the different monosaccharides present in the hemicellulose macromolecule (Mosier et al., 2005). Thus, there is no one strain of microorganism that can hydrolyse all different components, which in turn increases the number of treatments needed in lignocellulosic processing (Abramson et al., 2010).

Gasification, pyrolysis and enzymatic hydrolysis and fermentation of saccharides are the major methods used for generating energy from lignocellulosic biomass (Mosier et al., 2005; Soetaert and Vandamme, 2009). In gasification, lignocellulosic biomass is subjected to high temperatures with the presence of oxygen generating a mixture of gases, known as syngas, and heat (Mosier et al., 2005). The generated heat can be used to generate power (Regalbuto, 2009). In addition, the generated syngas that consist mainly of carbon monoxide and hydrogen can be used to generate hydrocarbon liquid

biofuels in the presence of metal catalysts through a reaction known as the Fischer-Tropsh reaction (Regalbuto, 2009; Sivakumar et al., 2010). During pyrolysis, the lignocellulosic biomass is subjected to high temperatures in the absence of oxygen generating very acidic and unstable liquid or solid intermediates that can be then converted into biofuels in the presence of catalysts (Regalbuto, 2009; Vanholme et al., 2010). Both gasification and pyrolysis require intense thermal and chemical treatments (Regalbuto, 2009; Umezawa, 2010). On the other hand, biological generation of energy requires much milder conditions and releases less pollutants (Soetaert and Vandamme, 2009)

Bio-ethanol production through biological conversion from lignocellulosic biomass is usually achieved through four steps: pre-treatment of the lignocellulosic biomass, saccharification of cellulose and hemicellulose into simple sugars, fermentation and finally ethanol purification and distillation (Abramson et al., 2010; Mosier et al., 2005). Lignocellulosic biomass needs to be pre-treated before cellulose and hemicellulose can be hydrolysed into monosaccharides (Abramson et al., 2010). The main purpose of pre-treatment is to break or loosen the bonds that link cellulose, hemicellulose and lignin together to allow the hydrolytic enzymes access to the cellulose and hemicellulose, finally releasing monosaccharides (Mosier et al., 2005; Umezawa, 2010). In addition, some pre-treatments remove the hemicellulose and lignin generating only cellulose that can be then easily hydrolysed into glucose molecules (Umezawa, 2010). Pre-treatment is the most costly step in

the generation of biofuels from lignocellulosic biomass, which accounts for more than 33% of the total cost of biofuel production (Brown, 2003). Nonetheless, without pre-treatment only less than 20% of the cellulose gets hydrolysed and fermented into biofuels (Mosier et al., 2005). In many cases, several pre-treatment procedures were successful in increasing sugar yield to 90% (Brown, 2003). Pre-treatment involves physical, chemical or biological treatments (Mosier et al., 2005). A combination of different treatments is usually applied depending on the final product needed (Mosier et al., 2005). The most common pre-treatment procedures used are thermal treatment, acidic treatment, steam explosion, ammonia explosion, lime treatment and biological treatment (Mosier et al., 2005; Umezawa, 2010). In heat treatment and steam explosion, exposure to very high temperatures (500-1000˚C) for few minutes reduces the size of the biomass, increases the pore's volume and releases acetyl groups from hemicellulose and lignin (Chen and Qiu, 2010; Mosier et al., 2005). The released acetyl group, mainly acetic acids, hydrolyzes the hemicellulose (Mosier et al., 2005). This is known as hemicellulose autohydrolysis, since the acetic acids that hydrolyzes hemicellulose are released from the lignocellulosic biomass (Mosier et al., 2005). Several acids can be used for pre-treatments, such as sulphuric acid, nitric acid, hydrochloric acid and phosphoric acid (Mosier et al., 2005). Acids degrade the hemicellulose into its monosaccharides, break the bonds that link lignin with hemicellulose-cellulose complex and break some of the inter lignin bonds thus loosening the tight matrix and allowing hydrolysis enzymes to reach the cellulose and hemicellulose (Abramson et al.,

2010; Mosier et al., 2005). Ammonia explosion and lime treatment help to depolymerise the lignocellulose matrix by cleaving the lignin-carbohydrate bonds that separate cellulose from lignin and hemicellulose (Abramson et al., 2010; Mosier et al., 2005). Acid, ammonia and lime treatments require much lower temperatures, ranging from 100-150˚C, as compared to thermal treatments (Chen and Qiu, 2010). In addition, these treatments help to depolymerise lignin, hydrolyse hemicellulose into monosaccharides and reduce cellulose crystallinity, which greatly aids in the next step which involves cellulose saccharification (Abramson et al., 2010; Mosier et al., 2005).

The next step following pre-treatment is to hydrolyse the polysaccharides into their building blocks in preparation for fermentation (Martinez et al., 2009). The two main methods that are used for cell wall polysaccharides hydrolysis are acid hydrolysis and enzymatic hydrolysis (Brown, 2003). In acid hydrolysis a combination of acid treatment, mainly H_2SO_4, acetic acids or furfural acids, and high temperatures, 160-220˚C, are used to decompose polysaccharides into their building blocks (Brown, 2003). Acid treatments require large quantities of acids, almost equal to the amount of sugars produced, have reduced sugar yields (55 to 60% of theoretical yield) and generate pollutants, such as $CaSO_4$ (Brown, 2003). Acid treatments hydrolyze both hemicellulose and cellulose, which results in a mixture of different monosaccharides (Brown, 2003). On the other hand, enzymatic hydrolysis hydrolyzes cellulose into glucose and is usually applied after pre-treatments (Brown, 2003). Enzymatic hydrolysis does not hydrolyse pentose sugars that result from hydrolysis of hemicellulose that

took place during pre-treatment (Brown, 2003). Pentose sugars can then be fermented separately from cellulose and this way hemicellulose and cellulose are more efficiently hydrolyzed (Brown, 2003). In addition, enzymatic hydrolysis requires milder temperatures and is more environmentally friendly with much less pollutants released. The enzymatic hydrolysis of cellulose and hemicellulose into its constituent monosaccharides is known as saccharification (Abramson et al., 2010). The three main enzymes that hydrolyze cellulose are 1,4-β-D-cellobiohydrolase, 1,4-β-D-endoglucanase and β-glycosidase, which are collectively known as cellulases (Abramson et al., 2010; Umezawa, 2010). 1,4-β-D-cellobiohydrolase enzyme breaks down the long glucose chains of cellulose into smaller oligosaccharides known as cellodextrins (Brown, 2003). Further digestion of the smaller oligosaccharides is done through the action of the 1,4-β-D-endoglucanase enzyme which hydrolyzes cellodextrins into di-glucose molecules known as cellobiose (Brown, 2003). Finally, cellobiose is digested releasing free glucose monomers by the action of the β-glycosidase enzyme (Brown, 2003). Cellulases can be produced from a variety of fungi and bacteria, mainly from T*richoderma* fungi and *Penicillium* (Brown, 2003).

The released glucose molecules can then be fermented into ethanol by microorganisms, such as *S. cerevisiae*, *S. pombe* and *Kluyveromyces lactis* yeasts, using the same fermentation procedure as first-generation biomass (Mosier et al., 2005; Umezawa, 2010). On the other hand, hemicellulose is hydrolyzed via hemicellulases into xylan and other pentose sugars (Umezawa, 2010). Most pre-treatments generate toxic by-products, such as free acetic acid,

that decrease fermentation efficiency or in some cases inhibit fermentation (Mosier et al., 2005). These toxic by-products may need to be removed before fermentation which adds further steps and costs to lignocellulosic biomass processing (Umezawa, 2010). On average, acetic acid fermentation above 0.5 g/l can inhibit fermentation (Brown, 2003). For example, pre-treatment of aspen results in 60-10 g/l of acetic acids that requires its removal before proceeding to fermentation (Brown, 2003). After fermentation, the generated ethanol needs to be distilled and purified from lignin, hemicellulose and cellulose that were not removed during the pre-treatment and hydrolysis steps (Mosier et al., 2005). The purified ethanol can then be used in a similar manner as ethanol produced from first-generation biomass, that is, it can be either used in combination with fossil fuels or used in its pure form (Sivakumar et al., 2010).

Given the progress made so far in using plant biomass for bioenergy production and pulp production, further improvements are needed in order to meet the world's growing energy needs. Further improvements in agricultural practices are essential to decrease the net energy required for biofuel generation (Soetaert and Vandamme, 2009). Disease and pest resistance, reduced or controlled dormancy, high rates of photosynthesis, and drought resistance are all traits that researchers are working on developing to improve plants' suitability as bioenergy sources (Ragauskas et al., 2006). A fast growing plant with relatively short rotational plantation cycle and with more biomass, known as energy crops, is ideal for use as biofuel (Ragauskas et al., 2006; Soetaert and Vandamme, 2009). Recent advances in biotechnology and

genetic engineering helped to produce varieties of lignocellulosic feedstock with less or altered lignocellulose content, which in turn decreased the amount of energy or chemicals required for lignocellulose processing thus improving its use as source of pulp (Huntley et al., 2003; Suckling et al., 2010). Most of the generated varieties are transgenic plants and are limited to laboratory scale. More research and biotechnical advances are needed to make lignocellulosic feedstock processing more efficient for large-scale application. The combined use of first and second-generation biomass would help to significantly increase the net energy value and decreases the net carbon balance. For example, if ethanol is produced from the starch present in corn as well as from corn residues, including cobs, straws and stover, this would at least double the amount of ethanol production from corn (Keegstra, 2010).

1.4 Pulp and Paper Manufacturing

Pulp and paper industries are considered to be one of the major industries in many countries. According to the PricewaterhouseCoopers global forest paper and packaging industry survey of the top 100 companies, PwC Top 100, total sales for year 2009 was 318 billion US dollars with 16 billion US dollars in operating profit (Suckling et al., 2010). Canada was the fourth country in pulp and paper manufacturing, where nine of the top 100 pulp and paper companies were Canadian with total sales of $21,631 million (Suckling et al., 2010).

Fibers are long cells with thickened secondary cell walls that are made up mainly from cellulose microfibrils and lignin (Wall et al., 2008). Plant fibers

are used for paper, tissue, and cardboard production (Hing-arrasco and Johnsen, 2010). Fibers are extracted from the plant's biomass, as separated dry fibers known as pulp, which are then used for paper production (Hu, 2002). Fibers could be obtained from herbaceous, mainly agricultural wastes, and from woody plants (Hu, 2002). About 90% of the raw materials used in pulp and paper industries come from woody fibers (Naqvi et al., 2010). Woody fibers are more favourable for pulp and paper industries due to their abundance, low-cost and higher density per biomass (Hu, 2002).

Pulp can be produced either mechanically or chemically (Suckling et al., 2010). In mechanical pulping, wood chips are subjected to mechanical forces, such as stone grinders or rotating metal discs, to physically separate fibers from each other (Brown, 2003; Suckling et al., 2010). In chemical pulping, inorganic chemicals, such as sodium sulphide (Na_2S) and sodium hydroxide (NaOH), are used to dissolve lignin by breaking the inter and intra-molecular bonds releasing cellulose fibers, a process known as Kraft pulping (Naqvi et al., 2010). The solubilization and subsequent removal of lignin from the plant's biomass is known as the Kraft delignification process (Suckling et al., 2010).

The main difference between mechanical and chemical pulping is in the removal of lignin from the fibers, which affects the quality of the generated fibers (Brown, 2003). The presence of lignin in the fibers generated from mechanical pulping decreases the bleaching quality of these fibers (Chang et al., 2010). A fiber rich in lignin tends to darken or turns yellow when exposed to light (Naqvi et al., 2010). The α-hydroxyl (CH-OH) and the α-carbonyl (C=O) groups present

in the lignin molecules absorb more than 80% of the ultraviolet (UV) light and reacts with oxygen through a series of radical reactions known as photo-oxidation (Chang et al., 2010; Hu, 2002). Photo-oxidation converts lignin phenols into yellow O-quinones which are responsible for the yellow coloring of fibers (Hu, 2002). In addition, photo-oxidation reduces the mechanical and chemical properties of the fibers as it disintegrates lignin and breaks the fibrous tissue (Chang et al., 2010). Therefore, fibers generated from mechanical or thermo-chemical pulping are used for low quality paper like newspapers, packaging and catalogues (Suckling et al., 2010). On the other hand, fibers generated through chemical pulping consists mainly of cellulose fibers and thus do not change its color (Brown, 2003). Cellulose fibers are used for producing high quality papers such as printing papers (Suckling et al., 2010). In addition, chemical pulping yields range from 40 to 69% of the initial biomass as compared to mechanical pulping which can yield more than 90% of initial biomass, mainly due to the fact that lignin is not removed (Naqvi et al., 2010). Nonetheless, 60% of the world's pulp production comes from chemical Kraft pulping due to the higher quality of the generated fiber (Naqvi et al., 2010). Recent advances in chemical pulping that include enhancing biomass digestion, the recycling of chemicals used, generating energy from leftovers and by-products and decreasing the release of environmental pollutants are all improving chemical pulping (Naqvi et al., 2010). For example, solid leftovers, such as the black liquor, undigested biomass and bark, which are rich in lignin, can be used for energy production which could reach up to 500 mega watt

(MW) per mill (Naqvi et al., 2010). Another approach for increasing the efficiency of chemical pulping is to genetically modify plants to decrease their lignin content or alter its composition. Several studies have shown that plants with decreased or altered lignin composition increased pulping efficiency and decreased the amount of pollutants generated (Suckling et al., 2010; Vanholme et al., 2008).

1.5 Alteration of Lignin Content and Composition

Lignin present in woody biomass makes wood processing for pulp and paper and biofuel production very challenging and costly (Simmons et al., 2010). Better knowledge about the enzymes involved in lignin biosynthesis will enable researchers to genetically engineer plants with desirable lignin content and quality. This will improve the efficiency of lignocellulosic feedstock for the wood, pulp and paper industries as well as make it more efficient as biofuels.

Lignin biosynthesis is a complex process which involves the function of many genes and takes place in different parts of the cell and involves a series of hydroxylation, methylation and reduction (Shi et al., 2010; Vanholme et al., 2008). Genes involved in lignin biosynthesis have been studied in a wide range of plants including *Arabidopsis thaliana*, *Zea mays* (maize), *Nicotiana tabacum* (tobacco) and *Populus* species (Vanholme et al., 2010). Most of these genes belong to small gene families (Hamberger et al., 2007; Shi et al., 2010), so it is possible that they have overlapping or partly redundant functions. There is evidence, however, that family members have specific expression patterns (Hamberger et al., 2007), explaining, at least in part, the need for gene families.

For example, Shi et al. (2010) were able to identify the presence of 95 complete gene models in *Populus trichocarpa* involved in lignin biosynthesis that belong to 10 enzyme families. They found that 23 genes out of the 95 genes identified were differentially expressed in xylem tissue during wood formation (Shi et al., 2010). It is believed that genes involved in lignin biosynthesis have evolved as plants were adapting to terrestrial environment (Hamberger et al., 2007). Lignin is believed to have helped terrestrial plants grow upright and overcome other challenges associated with living in terrestrial environment such as wind, competition for light and water, water transport from the soil to the upper parts of the plant and herbivores attacks (Hamberger et al., 2007; Vanholme et al., 2010).

Lignin is a polymer made up of three different phenolic monomers: guaiacyl (G), syringyl (S) and p-hydroxyphenyl (H) (Boudet, 2000). All three monolignols are synthesized from phenylalanine which is found in the cytoplasm (Vanholme et al., 2008). Figure 1-4 shows the steps involved in lignin biosynthesis and enzymes known to be involved in lignin monolignol biosynthesis. Phenylalanine is synthesized in the plastid through the shikimate biosynthesis pathway and then transported to the cytoplasm where it is used for monolignol biosynthesis, known as cinnamate monolignol pathway (Chen and Qiu, 2010; Umezawa, 2010). The first step in monolignin biosynthesis involves the removal of amine group (NH2) from phenylalanine and from tyrosine (Suckling et al., 2010). The deamination of phenylalanine is catalyzed by the phenylalanine ammonia-lyase enzyme (PAL) which converts phenylalanine into

cinnamic acid (Shi et al., 2010; Vanholme et al., 2008). Hydroxyl group (OH) is then added to the aromatic carbon at position 4 (C_4) by cinnamate-4-hydroxylase (C4H) forming 4-hydroxycinnamic acid (Whetten et al., 1998). At this point, the biosynthesis pathway is divided into two pathways where either 4-coumarate 3-hydroxylase (C3H) adds another hydroxyl group at the aromatic C_3 position generating caffeic acid, or 4-coumarate:CoA ligase (4CL) adds a Coenzyme A (CoA) group generating p-coumaroyl CoA (Umezawa, 2010). The generated p-coumaroyl CoA is used as a substrate for a wide range of secondary metabolite production including, H monolignin, flavonoid and soluble phenolics biosynthesis (Hamberger et al., 2007). For the generation of the G and S monolignols, a set of methoxylation is carried out by the caffeic acid 5-hydroxyconiferaldehyde O-methyltransferase (COMT), caffeoyl-CoA O-methyltransferase (CCoAOMT) and ferulate/coniferaldehyde 5-hydroxylase (F5H) enzymes (Umezawa, 2010). The methoxylation of caffeic acid at the aromatic C3 position by the COMT enzyme generates ferulic acid (Whetten et al., 1998). Ferulic acid can be used for G and S monolignol biosynthesis (Whetten et al., 1998). Further methoxylation of ferulic acid by the F5H and CCoAOMT enzymes at the aromatic C_5 position generates sinapic acid, which is the precursor for S monolignol biosynthesis (Umezawa, 2010). Next, the 4CL enzyme adds CoA to the caffeic acid, ferulic acid and sinapic acid (Umezawa, 2010). Cinnamoyl-CoA reductase (CCR) then removes the CoA group generating the corresponding aldehydes (Umezawa, 2010). The last step in monolignol biosynthesis is catalyzed by cinnamyl alcohol dehydrogenase

(CAD), which reduces the aldehyde precursors into the three monolignol alcohols: coniferyl (CA), sinapyl (SA) and p-coumaryl (Shi et al., 2010; Vanholme et al., 2008). In addition to the CAD enzyme, sinapyl alcohol dehydrogenase (SAD) enzyme specifically reduces sinapaldehyde into sinapyl alcohol (Umezawa, 2010). After monolignol alcohols are synthesized in the cytoplasm, they are transported in to the cell wall where they are oxidized via peroxidases or laccases enzymes and then coupled together forming lignin, a process known as lignification (Kärkönen and Koutaniemi, 2010). The region where lignification takes place in the cell wall is known as the lignifying zone (Shi et al., 2010). After oxidation and incorporated into the lignin molecule, the CA, SA and p-coumaroyl alcohols are then known as guaiacyl (G), Syringyl (S) and p-hydroxyphenyl (H) monolignol respectively (Kishimoto et al., 2010). Usually, lignification takes place after the cellulose microfibrils are deposited into the cell wall (Van Parijs et al., 2010).

Several attempts have been carried out by many researchers to decrease or alter the lignin content and composition. Several reviews summarise the findings of many researches in this field, such as reviews by Suckling et al. (2010), Vanholme et al. (2008) and Shi et al. (2010). For each of the genes involved in lignin biosynthesis a decrease in their expression resulted in lowering the lignin content to some extent, except for the F5H gene were alteration of its expression resulted in changing the S/G ratio (Vanholme et al., 2010; Vanholme et al., 2008; Suckling et al., 2010). Alterations in the expression and function of genes involved in the initial steps in lignin

biosynthesis had a greater effect on the total lignin content than lignin composition. For example, 85% reduction in PAL expression resulted in 52% reduction in total lignin content (Suckling et al., 2010) and the mutated plants had a range of phenotypic abnormalities including curled leaves, thinner cell walls and the plants were overall stunted (Vanholme et al., 2010). Also, 80% reduction in C4H expression, using RNAi silencing, in hybrid aspen (*Populus tremula x Populus tremuloides*) resulted in 30% reduction in lignin content with no change in the S/G ratio, 10% decrease in wood density and an increase in cellulose/hemicellulose content (Bjurhager et al., 2010). The down regulation of the C4H gene in *Arabidopsis*, tobacco and alfalfa plants also resulted in the decrease in total lignin content (Bjurhager et al., 2010; Suckling et al., 2010). Several studies have investigated the effect of the down regulation of the 4CL gene on lignin content due to its importance in the early biosynthesis of monolignols and secondary metabolites (Hu et al., 1999; Shi et al., 2010; Voelker et al., 2010). Voelker et al. (2010) and Hu et al. (1999) found that the down regulation of the 4CL gene in transgenic hybrid *Populus* had decreased lignin content of up to 50% compared to wild type and had no or very little effect on the H/G/S ratio (Hu et al., 1999; Voelker et al., 2010).

The down regulation of methylating proteins, such as CCoAOMT, COMT and F5H had little or no effect on lignin content, but, it had a great effect on lignin composition and in many cases resulted in the incorporation of unusual units (Vanholme et al., 2008; Meyermans et al., 2000). Huntley et al. (2003) found that the over-expression of the F5H gene in transgenic *Populus* (*Populus*

tremula x Populus alba) had no effect on the total lignin content and had no effect on the cellulose and other polysaccharides content. However, it had a great effect on the S/G ratio, where the lignin was made up mainly of S monolignol where the S/G ratio was 14:2 (Huntley et al., 2003). In addition, the over-expression of the F5H gene had no negative effect on the plant's phenotype (Huntley et al., 2003). Kraft pulping of the generated transgenic *Populus* required much less pulping time and chemicals for lignin removal as compared to wild type *Populus* (Huntley et al., 2003). In addition, the generated pulp had less residual lignin along with enhanced pulp bleaching (Huntley et al., 2003). Similar results were also obtained from a group of researchers (Marita et al., 1999) in transgenic *Arabidopsis* plants. Another study by Meyermans et al. (2000) found that 90% reduction in CCoAOMT expression resulted in 12% reduction in lignin content, 11% increase in S/G ratio and the incorporation of p-hydroxybenzoate unit (Meyermans et al., 2000).

The down regulation of the CCR gene, which is involved in the second to last steps of lignin biosynthesis, had great effect on the lignin content and resulted in the incorporation of monolignol precursors (Shi et al., 2010; Turner and Somerville, 1997). It was found that the down regulation of the CCR gene expression in transgenic *Populus* resulted in a total reduction of lignin content of up to 50% as compared to wild type (Leplé et al., 2007). Similar results were obtained from *Arabidopsis* where mutant plants had 50% reduction in total lignin content (Jones et al., 2001). In both studies, the decrease in lignin content was associated with an increase in cellulose content, which improved pulp

processing (Jones et al., 2001; Leplé et al., 2007). Mutants in monolignol biosynthesis, such as mutants in the 4CL and CCR genes, were characterised by their orange-brown discoloration of the xylem tissue due to the incorporation of monolignol precursors, such as ferulic acid, during lignification (Leplé et al., 2007; Voelker et al., 2010). The down regulation of the CAD gene, which catalyses the final step in monolignol biosynthesis, had little or no effect on total lignin content but greatly affected the structure of the lignin molecule (Lapierre et al., 2004; Shi et al., 2010). Lapierre et al. (2003) found that in *Populus* with significant CAD deficiency, with only 10% CAD activity, had moderate decreases in total lignin content, where the total lignin content decreased from 20% in wild type to only 17-18% of the total extract-free dry wood content (Lapierre et al., 2004). However, the lignin in the generated CAD deficient *Populus* had incorporated sinapaldehyde units, which is the direct precursor of S monolignol, through an 8-O-4 linkage (Lapierre et al., 2004). This in turn resulted in the presence of lignin as small patches with free phenolic groups, which can be easily solubilized at room temperature in an alkali environment (Lapierre et al., 2004).

Results from several studies sometimes contradicted each other mainly due to variation in gene expression, method of analysis, which gene in the family was studied or what plant was used for the study. For example, Jouanin et al. (2000) found that the down regulation of the COMT gene expression in transgenic *Populus* (*Populus tremula x Populus alba*) resulted in a 17% decrease in total lignin content (Jouanin et al., 2000). While in a previous study

conducted by the same group, they found that the down regulation of COMT activity of up to 90% in transgenic *Populus* did not have an effect on lignin content, but only altered lignin composition and structure (Jouanin et al., 2000; Lapierre et al., 1999). Nonetheless, both studies showed that the lignin structure in the generated transgenic *Populus* was affected (Jouanin et al., 2000). For example, in their Jouanin et al. (2000) study, they were able to obtain a *Populus* line with almost no COMT activity and they analyzed the lignin content in the generated 6 month old transgenic *Populus* (Jouanin et al., 2000). The lignin in the transgenic *Populus* lacked S monolignols and 5-hydroxyguaiacyl units were incorporated into the lignin (Jouanin et al., 2000). This generated *Populus* construct yielded more pulp, around 10% more, during Kraft pulping, but, the generated pulp had some lignin residues which decreased the pulp's bleaching quality (Jouanin et al., 2000).

1.6 Populus

Populus are angiosperm forest trees that are from the Salicaceae family (Dickmann et al., 2001). There are more than 30 species of *Populus* found in the Northern Hemisphere, 12 of the 30 species are native to North America (Dickmann et al., 2001). They are dioecious trees where the flowers of one tree could be either male or female (Yang et al., 2009). *Populus* trees are all deciduous as they shed their leaves seasonally. *Populus* trees can reproduce both sexually and asexually (Jansson and Douglas, 2007). Pollen from male trees is carried by wind until it reaches ovules of female trees where it fertilizes the mature eggs, which is commonly known as wind pollination (Jansson and

Douglas, 2007). The fertilized eggs then develop into fruits and eventually mature and then split open releasing cottony seeds which are carried by wind (Dickmann et al., 2001). The use of wind for pollination and seed dispersal helped to spread *Populus* over vast area which in turn helped to produce hybrids that carry unique allelic combinations that resulted in the generation of a wide range of *Populus* with various genetic varieties and thus wide range of characteristics (Dickmann et al., 2001). In addition to sexual reproduction, *Populus* have a great ability to regenerate asexually from cut stem, bud, burned tree or from roots that had their aboveground section removed by either being cut or burned (Yadav et al., 2009). Beside their environmental and ecological roles, *Populus* are important sources for the pulp, paper, building boards, oriented strand board and plywood industries in North America (Huntley et al., 2003). Six of the native North American *Populus* species are considered to be of economical importance (Dickmann et al., 2001). Pulp and Fibers from *Populus* can be generated through mechanical, semi-mechanical and chemical pulping (Dickmann, 2001).

Unlike other forest trees, *Populus* have a relatively short life span (Dickmann et al., 2001) compared to other forest trees, such as oak and Douglas fir trees. Their rapid growth allows *Populus* to reach heights of up to 10 m tall and width of up to 7 cm within the first 5 years (Dickmann et al., 2001). Thus, *Populus* are recognized by their tall, straight and slender stems (Dickmann, 2001). In some *Populus* species, wood traits can be assessed in 1 to 3 years-old trees (Brunner et al., 2004), compared to 10 to 20 years-old trees

in other species. Because of their rapid growth, wood from *Populus* can be harvested within couple of years from their plantation (Hinchee et al., 2009) with minimal requirements for nutrients and fertilizers (Flachowsky et al., 2009). *Populus* species are believed to be a viable renewable source for lignocellulosic biomass that can be used for pulp, paper, timber, and bioenergy production in temperate climates (Hinchee et al., 2009; Jansson and Douglas, 2007). In addition, several *Populus* species were subjected to extensive breeding programs (Flachowsky et al., 2009). The fast growing *Populus* enable plant breeders to observe phenotypes in a reasonable amount of time as compared to other forest trees. On average, *Populus* trees reach sexual maturity in 7 to 15 years from time of germination as compared to 40 years in other forest trees (Taylor, 2002). Nonetheless, since trees take longer time to flower, compared to annual and biannual plants, that gives scientists more time to develop transgenic trees with desired traits and consume them before they flower thus limiting its contamination of the environment. As a result, researches can overcome one of the major limitations of using transgenic plants while still benefiting from its efficiency. In addition, *Populus* ability to reproduce vegetatively enabled breeders to capture desired gene combinations and to clonally reproduce them without shuffling genes during sexual reproduction, which might change the desired genotype (Dickmann et al., 2001). Several breeding programs involved the generation of hybrids through interspecies genetic crosses among individuals from distinct geographical areas (Flachowsky et al., 2009). For example, the genetic crosses between the

Japanese *P. deltoids* and the European *P. nigra (Populus x canadensis)* generated fast growing hybrids (Flachowsky et al., 2009). There are also prospects of inducing early flowering in *Populus* species through mechanical, chemical, hormonal and genetic approaches (Flachowsky et al., 2009). Several studies showed that the over expression of the genes involved in inducing flowering and sexual maturity, including *FLOWERING LOCUS T (FT)*, LEAFY (LFY) and APETALA I (API), accelerated the flowering time of *Populus* from 10 to 15 years to only a few weeks (Brunner and Nilsson, 2004; Nilsson et al., 2006; Zhang et al., 2010).

Several characteristics of *Populus* species make them also suitable as models for genetic studies. Rapid growth, small genome (around 550 mega base pairs), ease to clonally propagate and ability to generate transgenic *Populus* via *Agrobacterium* (Bhalerao et al., 2003; Dickmann et al., 2001), are among the characteristics that make *Populus* suitable for genetic studies. In fact, *Populus* were the first forest trees to be successfully transformed by *Agrobacterium* (Yang et al., 2009) in 1984 and most of the generated transgenic trees were stable (Ellis et al., 2010). Several transgenic approaches have been successfully applied to *Populus* trees including, *Agrobacterium* mediated transformation, RNAi interference and viral induced gene silencing (Jansson and Douglas, 2007). For example, several transgenic *Populus* lines with altered lignin content and composition have been successfully generated (Vanholme et al., 2010). Several *Populus* species can be propagated via tissue culture from various tissue sources and from various ages (Ellis et al., 2010). Moreover, its

fast growth rate helps to minimize changes in its characteristics, like wood characteristics, during tissue culturing (Ellis et al., 2010).

To further facilitate research on *Populus*, several genomic tools and resources have been developed. There are more than 445784 expressed sequence tags (EST) that have been generated under various environmental conditions and from different *Populus* species that are publically available in the National Centre for Biotechnology Information (NCBI) database and in the ForestTree database (Jansson and Douglas, 2007; Pavy et al., 2007; http://Foresttree.org/ftdb). The majority of the generated EST were from wood forming tissue (Jansson and Douglas, 2007). The genome of a female *Populus trichocarpa* tree named Nisqually-1, which has been used in several *Populus* breeding programs, has been sequenced to approximately 8.5 times depth using shotgun sequencing of small insert libraries (Tuskan et al., 2003; http://genome.jgi-psf.org/Poptr1/Poptr1.home.html). *P. trichocarpa* is the first and only fully sequenced tree to date. *P. trichocarpa* genome is made up of 403 megabases arranged in 19 chromosomes, where around 40668 loci contain protein coding transcripts (http://www.phytozome.net/poplar#D) and possessing more than 45 thousand proposed gene models (Tuskan et al., 2003), which is the largest set of genes among the sequenced plants to date (Jansson and Douglas, 2007). The genome size of a *P. trichocarpa* tree is four times bigger than that of *Arabidopsis thaliana* (http://www.phytozome.net/poplar#D). *Populus* and *Arabidopsis* are more phylogenically related, as compared to other forest trees such as conifers (Jansson and Douglas, 2007). Being evolutionally closely

related to *Arabidopsis*, where both *Populus* and *Arabidopsis* are in the same Eurosid I clade of Eudicotyledonous plants, makes *Populus* ideal for comparative functional, genomic and evolutional studies (Jansson and Douglas, 2007). In fact, several gene functions were found to be conserved between *Arabidopsis* and *Populus* (Jansson and Douglas, 2007). In addition to the availability of a fully sequenced genome, many cDNA libraries, chromosomal maps, DNA microarrays and QTL markers are also publically available in several databases (Yang et al., 2009). As a result, and due to the many advantages of *Populus*, many plant biologists have chosen *Populus* as a model for genetic studies.

1.7 Arabidopsis thaliana

Arabidopsis Thaliana is the most well-studied flowering plant (Raven et al., 2003). It is a small weed that belongs to the *Brassicaceae* family, which includes some cultivated crops such as canola, cabbage and radish (Raven et al., 2003). It has gained popularity because of its short generation time and high fecundancy. *Arabidopsis* can be both self pollinated or cross pollinated, where homozygosity can be generated in mutant loci by allowing the plant to self pollinate and novel and controlled combination of alleles can be obtained through crossing two different plants (Raven et al., 2003; Taylor, 2002). Its small size, that requires minimal space, and ease to propagate and maintain under various conditions facilitated the generation of a wide range of mutants and enabled extensive and large-scale studies to be undertaken (Raven et al., 2003). *Arabidopsis* has a relatively small genome of around 125 megabases

arranged over five chromosomes and with little repetitive DNA compared to other plants (Raven et al., 2003). In addition, *Arabidopsis* can be easily transformed by *Agrobacterium tumefaciens*. All of these advantages made it ideal for molecular studies (Raven et al., 2003). To further facilitate more genetic and molecular studies, its genome was completely sequenced by the year 2000, and thus it was the first plant to be fully sequenced (Raven et al., 2003). Both molecular and biological information are publically available in a number of databases, such as The *Arabidopsis* Information Resource (TAIR), The Institute for Genomic Research (TIGR) and others. Even though *Arabidopsis* is a small annual weed, it develops woody stem, enabling the study of the genetic basis of wood differentiation in this plant. It was found that the same genes that participate in lignin biosynthesis and wood formation in *Arabidopsis* have homologies in forest trees such as *Populus* and pine trees (Hamberger et al., 2007; Taylor, 2002).

1.8 TILLING Technology

Targeting Induced Local Lesions IN Genomes (TILLING), is a high-throughput reverse genetic approach that is used for the rapid identification and large scale screening for the presence of single nucleotide polymorphisms (SNP) and point mutations in natural and mutagenized populations (Gilchrist and Haughn 2005). TILLING provides an efficient method to identify the presence of mutations in selected genes based on molecular lesions independent of phenotypic screening (Stemple, 2004). TILLING has been applied to different organisms such as *Arabidopsis thaliana* (MacCallum et al.,

2000), maize (Till et al., 2004), wheat (Slade et al., 2005), barley (Cladwell et al., 2004), *Brassica oleracea* (Till et al., 2004; CAN-TILL, http://www.botany.ubc.ca/can-till/), soybean (Cooper et al., 2008), rice (Till et al., 2007), zebrafish (Stemple, 2004), *Caenorhabditis elegans* (Gilchrist et al., 2006) and as a possible disease diagnostic tool in humans (Till et al., 2006).

TILLING was first developed to rapidly detect mutations in *Arabidopsis thaliana* (MacCallum et al., 2000). The *Arabidopsis* TILLING Project (ATP) was established in 2001 by Comai and colleagues in Seattle, WA (http://tilling.fhcrc.org/). Since then, TILLING has gained more popularity and has been applied to a wide variety of organisms including a wide range of agriculturally and commercially important plants. One of the recognized TILLING initiatives in North America is the CAN-TILL initiative at the University of British Columbia in Vancouver, Canada (http://www.botany.ubc.ca/can-till/). CAN-TILL is a facility where researchers are working to identify mutations in genes from a wide variety of different organisms. So far, the TILLING projects at CAN-TILL included screens of mutagenized populations of *Arabidopsis thaliana*, maize, *Caenorhabditis elegans, Brassica oleracea* and natural populations of black cottonwood (*Populus*) *(*http://www.botany.ubc.ca/can-till/).

In TILLING, a randomly mutagenized population is generated through chemically mutagenizing seeds or pollen (Gilchrist and Haughn 2005). Ethyl methane sulfonate (EMS) is the most commonly used mutagen used for the generation of mutagenized populations to be screened through TILLING (Till et al., 2004). Figure 1-5 summarizes steps commonly used in TILLING. The

genomic DNA is extracted from individuals from either F1, F2 or F3 population and are pooled several folds and arrayed in 96-well microtiter plates (Colbert et al., 2001). Genes of interest are then amplified using infrared-dye labelled PCR primers (Gilchrist and Haughn 2005; Colbert et al., 2001). The amplified products are denaturated by heating and then gradually cooled to permit renaturation. The gradual renaturation allows the binding of wild type alleles with EMS-induced mutant alleles resulting in the formation of heteroduplexes at the position of EMS-induced point mutations. The products are then digested using an endonuclease enzyme, such as CEL I isolated from celery, that specifically cleaves the heteroduplexes at single base pair mismatches that have resulted from the reannealing of a wild type with an EMS-induced mutated DNA strand. The resulting DNA fragments are then denatured and separated on polyacrylamide denaturing gels. The presence of a mutation within the amplified region is indicated by novel shorter bands on the gel that add up to the total PCR product. The DNA from the individuals in the identified pool is then screened using the same procedure and the amplicons from the identified individuals are sequenced to identify the molecular mutation that has occurred.

1.8.1 Advantages of TILLING

TILLING has many advantages over forward and other reverse genetic techniques used for the screening and identification of genetic mutations in large populations. Unlike forward genetic approaches that depend on the identification of mutations based on a detectable phenotype, TILLING can detect the presence of mutations at early stages of the organism's development

independent of the phenotype, thereby reducing the time required to identify mutations from years to only few weeks of growth (Stemple, 2004). This is potentially a great benefit, especially in studying forest trees that usually require many years of growth to show a phenotype. Once the mutation of interest is detected based on genotype in heterozygous F1 generation during early stages of growth, the mutant individuals can then be set aside for further analysis and for generation of F2 population while the rest of the population can be discarded. This eliminates the need to obtain homozygous F2 generations from self-fertilization of F1 progenies. This is an important issue when dealing with plants that suffer from inbreeding depression, especially in forest trees. To work around this, genetic cross between heterozygous plants, for different mutant alleles of the target gene, can generate homozygous individuals for mutations in a desired gene while avoiding inbreeding depression by maintaining all other mutant loci heterozygous. This greatly reduces the effort and cost that is needed for nursing large populations to reach F2 or F3 generations. Moreover, in forward genetic approaches homozygous individuals are often required to express recessive mutations. In such cases, mutations in essential genes might be lethal to the organism making many mutations undetectable in typical screens for phenotypes (Colbert et al., 2001).

The growth in the number of sequenced genomes of a wide variety of organisms has facilitated the emergence of reverse genetic techniques in which mutations can be targeted to a known gene. There are many reverse genetic techniques that are well developed and have been successfully used in the

identification of gene functions in different organisms. Examples of reverse genetic techniques that are commonly used include homologous recombination, RNA interference silencing (RNAi) and insertional mutagenesis (T-DNA and transposon) (Colbert et al., 2001; Stemple, 2004). All of these techniques require the generation of transgenic organisms. Unfortunately, many of these techniques are organism specific, lengthy, labour intensive and require the development of transformation technology. In addition, the use of the generated transgenic organisms in field trials is highly regulated which in turn restricts the use of transgenic organisms (International Service for the Acquisition of Agri-biotech Applications; Lemaux, 2008). Moreover, risks of vandalism further discourage the use of transgenic organisms (Brunner et al., 2004). TILLING, on the other hand, does not require the formation of transgenic organisms and can be applied to any organism regardless of the genome size and complexity. The EMS mutagen used in TILLING induces a large number of point mutations in the genome, mainly causing a transition of G/C to A/T (Tuskan et al., 2006). The high frequency of mutations relative to other mutagen, such as x-ray and insertional mutagen, requires a considerably smaller population (Stemple, 2004). Unlike insertional and radiation mutagenesis that tend to result in a complete gene knockout (loss of function) (Martín et al., 2009), TILLING can be used to generate a series of allelic variations from complete to different degrees of partial loss of function alleles thereby allowing mutations in essential genes to survive (Colbert et al., 2001). In cases where a complete loss of function results in lethality, partial loss of function may be more informative (Stemple, 2004) and

highly desired in breeding efforts. In addition, EMS mutagenized individuals might carry multiple point mutations, including missense and nonsense, in a number of genes, unlike other types of mutagens that result in a small number of mutations (Martín et al., 2009). This in turn minimizes the number of mutagenized individuals needed to find the desired mutation.

1.8.2 Limitations of TILLING

Beside its recent success and the many advantages that the TILLING technique offers, some difficulties and limitations could be encountered when using the TILLING technique. Like any other reverse genetic technique that depends on some molecular knowledge, TILLING requires previous knowledge of the DNA sequence of the selected genes or regions of interest. This limits large-scale application of TILLING to only organisms with sequenced genomes or extensive cDNA sequence information. Nonetheless, the continuous sequencing projects of more organisms are providing more sequence information which is facilitating the use of many reverse genetic techniques including TILLING. In addition, the presence of model organisms from different species such as bacteria, D*rosophila*, C*.elegans*, *Arabidopsis*, *Populus* trees and many more, is facilitating the process of finding homologues or targeting genomic regions that are conserved among different organisms. This in turn makes it possible to screen in organisms where limited sequence information is available.

Another limitation in applying TILLING is the availability and generation of mutant populations. Since TILLING is used to detect point mutations, it is

critical to generate a saturated mutant population which will depend on the species, mutagen and dosage used (Martín et al., 2009). EMS mutation frequency and tolerance varies greatly from one species to another (Martín et al., 2009). For instance, EMS mutation rate in *Arabidopsis* (Columbia background) is 1 in every 300 kb, in *Brassica oleracea* 1 in every 447 kb and in *Drosophila* 1 in every 91 kb (Martín et al., 2009). In addition, the differences in the genotypic makeup and sets of chromosomes present affect the frequency of mutations. For example, mutation frequency in diploid organisms is less than that in tetraploid and hexaploid organisms. For instance, the EMS induced mutation rate in *Arabidopsis* is around 1 in every 300 kb while in *Triticum durum* (a tetraploid wheat) is 1 in every 25 kb (Martín et al., 2009). Point mutations are the most difficult types of mutations to detect because most of the point mutations generated do not result in severe or noticeable phenotype. In addition, most of the generated point mutations are silent, where the base changed does not affect the amino acid or where the amino acid is changed to a very closely related amino acid, which eventually does not affect the protein encoded and thus no detectable or visible phenotypes are obtained. Therefore, a very specific and delicate technique is required to find point mutations based on molecular lesions and not phenotype.

In addition, the presence of high degree of allelic polymorphism in some organisms results in the formation of heteroduplexes due to natural SNPs, which can be recognized and cleaved by the CEL I endonuclease enzyme used in TILLING, thereby making it difficult to distinguish point mutations from allelic

polymorphisms. Nonetheless, this problem can be overcome by using inbred lines as background to facilitate the detection of allelic polymorphisms that are caused by EMS-induced mutations.

1.9 Haploinsufficiency

In diploid organisms, there are two alleles of each gene due to the presence of two sets of chromosomes. Generally, the presence of two alleles provides an additional spare copy that acts as a supplement in case one copy is mutated (Veitia, 2005). Therefore, the presence of one functional wild type allele is usually sufficient to result in a normal phenotype (Deutschbauer et al., 2005; Veitia, 2005). In general, the presence of one allele (dominant) masks the effect of the other allele (recessive) and determines the phenotype. Two copies of recessive alleles are needed to imprint the phenotype (Veitia, 2005). In some cases, two functional wild type copies are needed to obtain a wild type phenotype. When one copy is mutated the resulting phenotype will be an intermediate phenotype (between wild type and mutant phenotype) (Deutschbauer et al., 2005; Veitia, 2005) an effect known as haploinsufficiency.

Haploinsufficiency is believed to play a role in many human genetic diseases where it has been observed that individuals harboring just one mutated copy of the gene showed an intermediate phenotype, where the severity of the disease was less than that with individuals with two mutated alleles (Alkuraya et al., 2006; Dang et al., 2008; Veitia, 2005). Most human genetic autosomal disorders were found to be haploinsufficient related (Dang et al., 2008). Smith–Magenis syndrome, Ehler-Danlos syndrome, Retinitis

Pigmentosa, Sotos syndrome and Williams syndrome are among many human genetic disorders that do show haploinsufficiency in which heterozygote individuals show either mutant or intermediate phenotype (Dang et al., 2008; Veitia, 2005). In addition, it was also found that some haploinsufficiency in certain genes might increase the susceptibility of heterozygote individuals to disease and developing medication side effects with higher rates than individuals with both normal copies of the genes (Dang et al., 2008). New studies also showed that some haploinsufficiency in certain genes, such as the ATM and BLM genes, increased the chance of developing cancer (Deutschbauer et al., 2005). To date, there are 299 haploinsufficient genes identified in the human genome, but, the number of haploinsufficient genes is believed to be higher (Dang et al., 2008).

Haploinsufficiency has also been detected in other organisms including yeasts, plants, zebrafish and other diploid organisms (Dang et al., 2008; Deutschbauer et al., 2005; Krieger et al., 2010). Examples of haploinsufficiency in plants include the ERECTA-LIKE 2 (ERL2) gene in *Arabidopsis,* where plants with one copy had less severe defect in the plant's floral and ovule development (Pillitteri et al., 2007), and the SINGLE FLOWER TRUSS gene in tomato, where heterozygotes had 60% increase in yield (Krieger et al., 2010). Recent extensive study showed that there are more than 180 haploinsufficient genes in the yeast which accounts for 3% of the yeast's genome (Deutschbauer et al., 2005). Around 58% of those haploinsufficient genes in yeast have homologues in the human genome, especially the genes that encode protein subunits

(Deutschbauer et al., 2005). New results and findings are rapidly emerging with haploinsufficiency identified in new genes, which in turn is expanding our knowledge about haploinsufficiency and helping to better understand the mechanisms involved in haploinsufficiency.

Most of the haploinsufficiency identified so far was in genes that encode proteins that are involved in signaling, subunits of a multiunit protein complex and transcription factors (Dang et al., 2008; Deutschbauer et al., 2005). Several mechanisms and models have been proposed to explain haploinsufficiency. The mechanism by which a gene results in haploinsufficiency varies depending on the gene's function. For instance, if the gene encodes a subunit of a multisubunit protein, then having a lower number of the right subunit might affect the total number of functional and correctly assembled three dimensional protein complexes, which will eventually affect its availability and hence its function (Deutschbauer et al., 2005; Veitia, 2010). Also, genes that encode transcription factors are most likely to be haploinsufficient genes because usually transcription factors regulate a number of targets and their concentration highly affects the number of targets being regulated and whether enough of its target is being regulated (Deutschbauer et al., 2005; Veitia, 2010). Gene products that are part of signaling pathways or part of metabolic pathways are likely to be haploinsufficient genes as their concentration will affect the pathways, especially when the concentration of a certain component in the pathway affects the downstream or feedback loops in the pathway (Deutschbauer et al., 2005). In general, all the proposed models focus on

protein levels where in heterozygotes the level of protein synthesized is half the normal level which in turn has implications on the phenotype.

1.10 Research Objectives

The main goal of our research is to induce random EMS mutations in a population of hybrid *Populus* and then screen for single nucleotide mutations that result from the EMS treatment in genes that are involved in lignin biosynthesis, followed by analysis of the effect on lignin content and composition. The first step is to generate an EMS-mutagenized poplar population. Since no established protocols are available for mutagenesis of calli, we had to develop a method to generate this EMS-mutagenized poplar population. The next step is to identify a suitable DNA extraction protocol based on the following criteria: yield, purity and PCR productivity. We then have to extract DNA from the generated EMS-mutagenized poplar plants and PCR amplify genes of interest. Also, to screen for EMS-induced mutations, we have to establish a TILLING protocol that is simple and affordable in our laboratory. The final step is to find a suitable high-throughput lignin analysis protocol that is accurate, fast, and affordable.

Several research studies done on genes involved in lignin biosynthesis showed that knocking out the function of some of the genes resulted in reducing lignin content and in some cases changed monolignols ratio. To our knowledge, all of these studies involved the use of transgenic plants where the genes' expression have been either completely disrupted or highly reduced. Here, we studied the effect of EMS treatment, which is a non-transgenic approach, in

order to achieve a reduction in lignin content or change in lignin composition. The question that we are trying to answer through our project is will knocking out one of the two functional copies of genes known to be involved in lignin biosynthesis be sufficient to reduce lignin content or change its composition to a limited extent in *Arabidopsis* plants. We are trying to investigate whether haploinsufficiency that is observed in some human genetic diseases is also present in plants. Our goal is to generate lines of plants with desired phenotype that will balance between the benefits from certain mutations and at the same time limit the side effects of such mutations. For example, transgenic *Populus* with defects in CCR gene were found to have reduced growth rate and had orange-brown xylem (Leplé et al., 2007). Even thought the main goals of reducing lignin content and energy required for wood processing were achieved, it was also accompanied by negative effects on the plant. So if we are able to obtain an intermediate phenotype where lignin content is lower while plant is still healthy, then we would have achieved the goal of decreasing energy required for wood processing while maintaining a healthy plant. To test this hypothesis we obtained several mutant lines of *Arabidopsis thaliana* and generated a heterozygous population and then analyzed their lignin content.

1.11 Figures

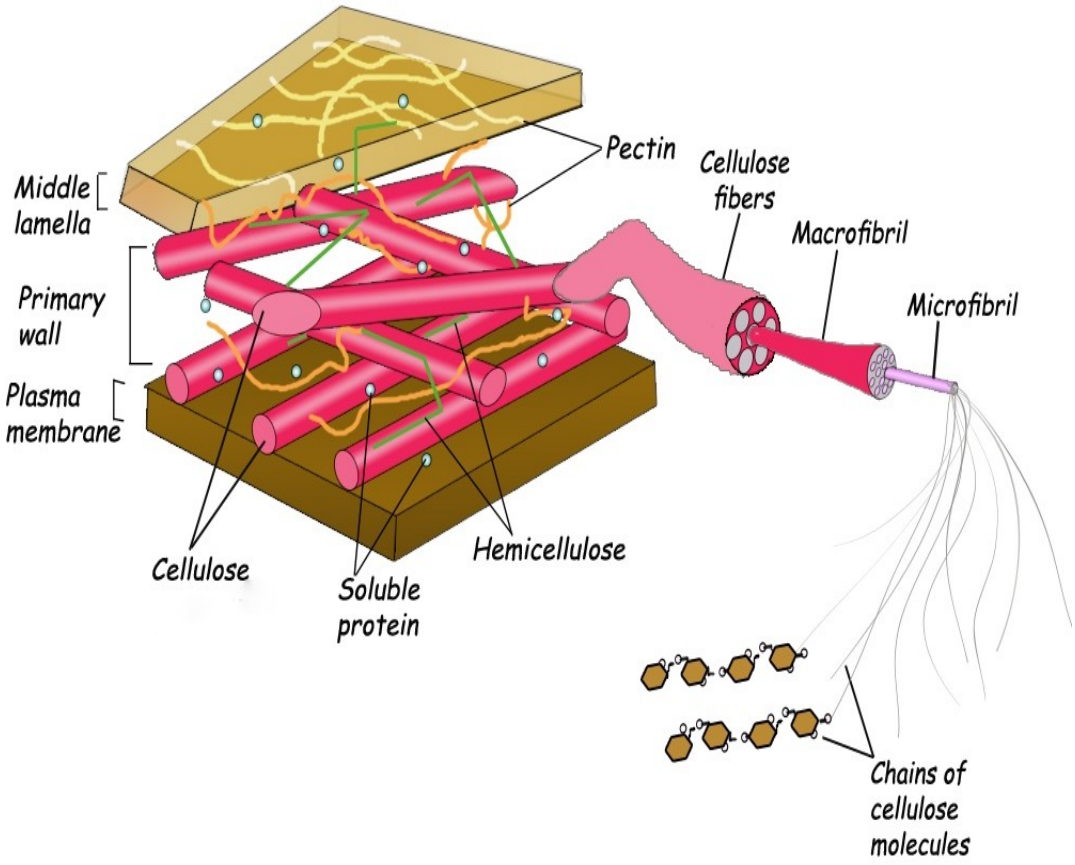


Figure 1-1 Primary plant cell wall.

Primary plant cell wall consists mainly of cellulose, hemicellulose and pectin. Cellulose is made up of chains of glucose polymer organized into microfibrils and the fibers are organized into a more organized structure known as macrofibril. The macrofibrils in turn are organized into fibers known as cellulose fibers or simply known as cellulose. *Note.* Adapted from "Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol," by M.B. Sticklen, 2008, *Nature reviews. Genetics*, *9*(6), p.434. Copyright 2008 by Nature Publishing Group. And Adapted from "Carbohydrates," by http://nutrition.jbpub.com/resources/chemistryreview9.cfm, Copyright 2006 by Jones and Bartlett Publishers.

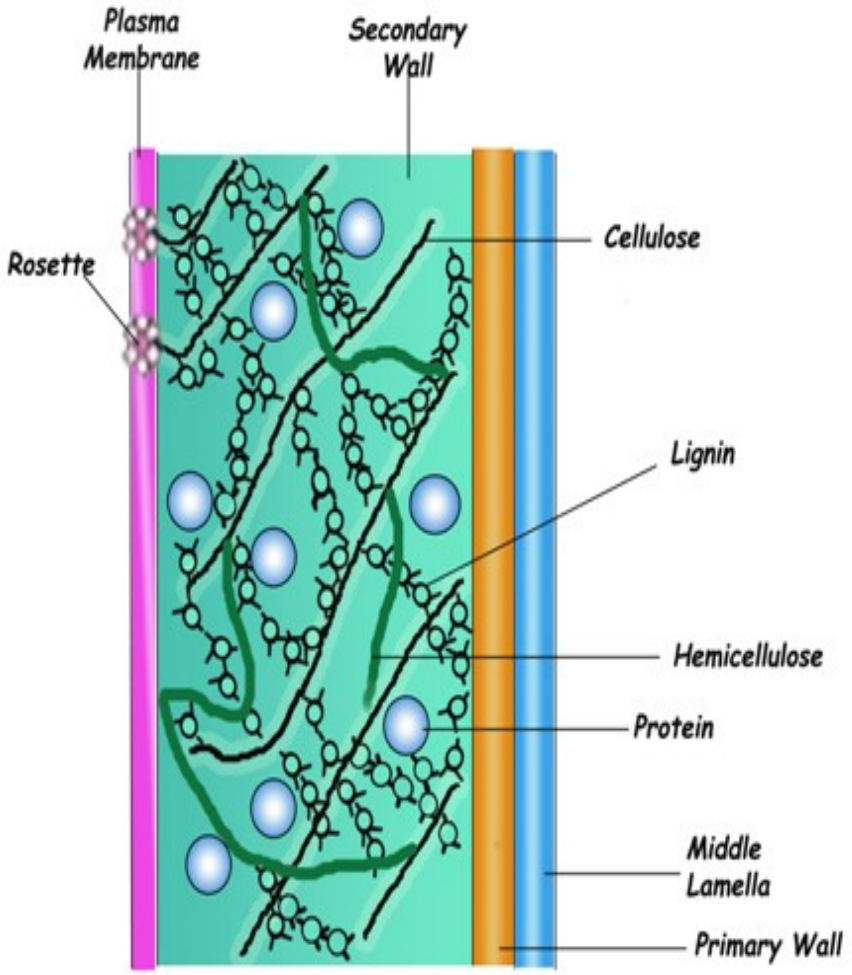


Figure 1-2 Secondary plant cell wall.

Secondary cell wall consists mainly of cellulose, hemicellulose and lignin molecules. Lignin links cellulose together and fills the spaces in the secondary cell wall adding to its strength. *Note.* Adapted from "Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol," by M.B. Sticklen, 2008, *Nature reviews. Genetics*, *9*(6), p.434. Copyright 2008 by Nature Publishing Group.

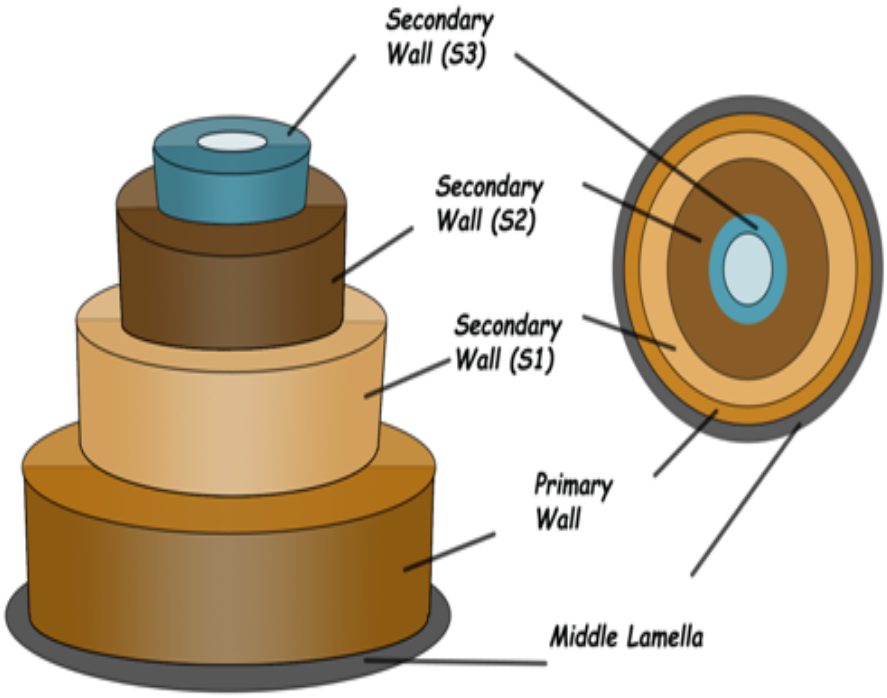


Figure 1-3 Secondary cell wall layers.

Secondary cell wall is made up of three different layers: S1, S2 and S3. *Note.* Adapted from "Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol," by M.B. Sticklen, 2008, *Nature reviews. Genetics*, *9*(6), p.434. Copyright 2008 by Nature Publishing Group.

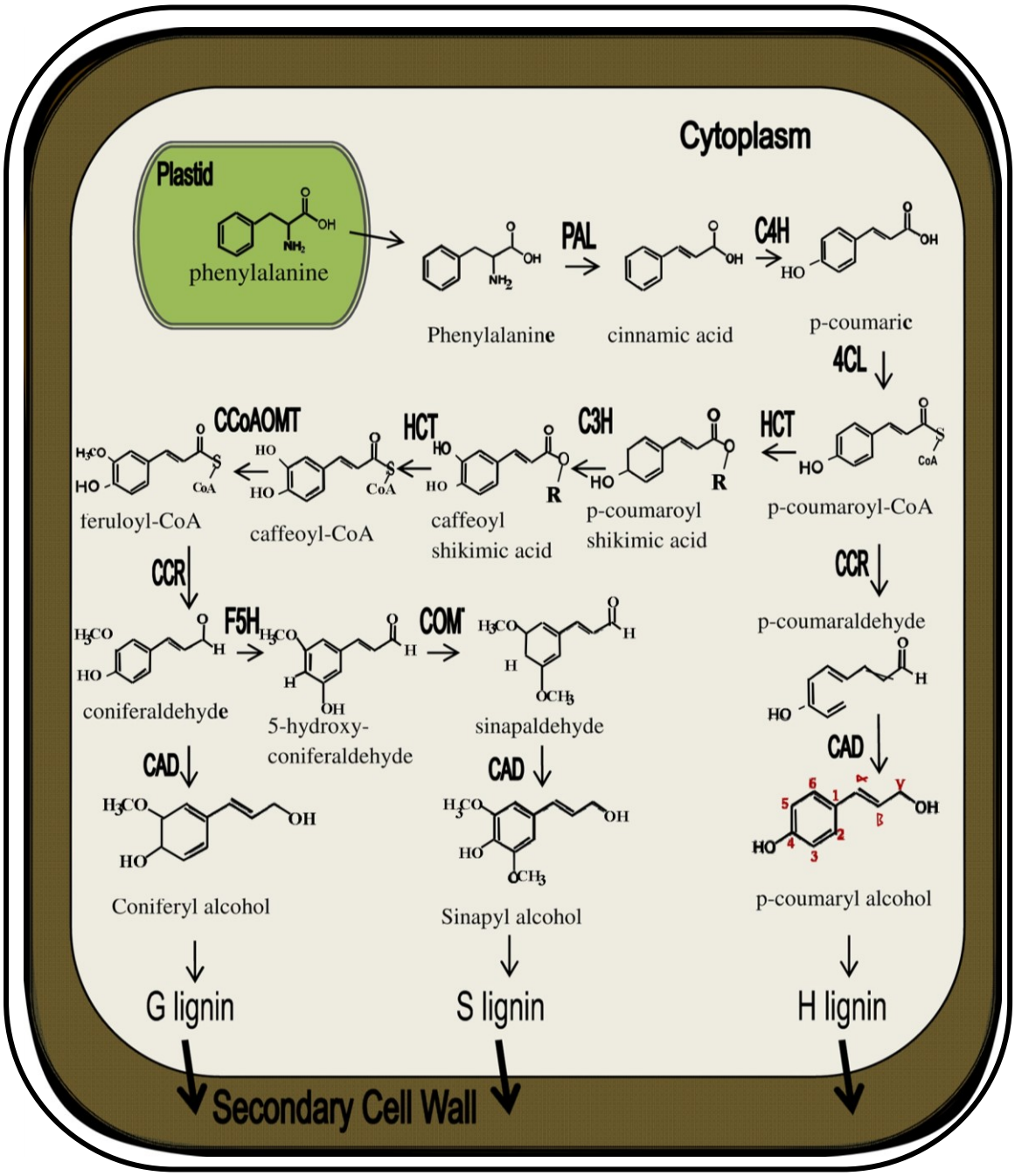


Figure 1-4 Lignin biosynthesis and enzymes known to be involved in lignin biosynthesis.

The figure shows the common pathway involved in lignin biosynthesis which takes place mostly in the cytoplasm. The monolignols are then transported to the cell wall where they bind together forming a complex lignin molecule. *Note*. Adapted from "Lignin and biomass: a negative correlation for wood formation and lignin content in trees," by E. Novaes, M. Kirst, V. Chiang, H. Winter-Sederoff, and R. Sederoff, 2010. *Plant physiology, 154*(2), pp. 555-561. Copyright 2010 by Plant physiology.

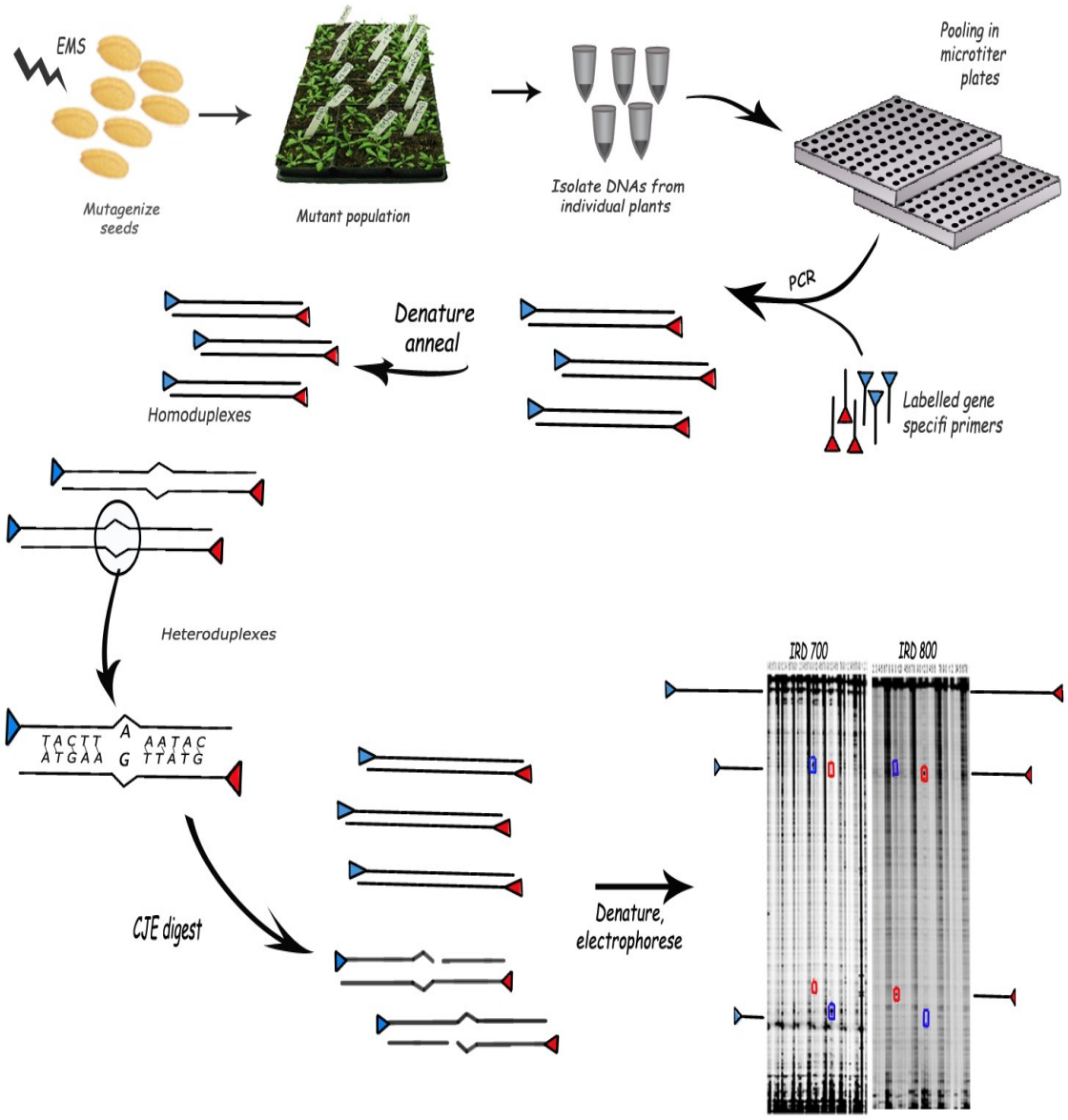


Figure 1-5 Steps commonly involved in TILLING.

First step is to generate a mutagenized population followed by generating F1 and then F2 generation. Genomic DNA is then extracted from F2 plants and pooled in pools of 4 to 8. The region to be screened is amplified using labelled gene specific primers. The PCR products are then heated and allowed to re-anneal forming heteroduplexes. The presence of a mutation is indicated by presence of novel shorter bands on the gel. The DNA of the individuals in the identified pool is then screened again and the identified individuals are sequenced to identify the molecular mutation that has occurred. *Note*. Adapted from "TILLING to detect induced mutations in soybean," by J.L. Cooper, B.J. Till, R.G. Laport, et al., 2008, *BMC Plant Biology, 8*, p.3. Copyright © 2008 by Cooper et al; licensee BioMed Central Ltd.

1.12 Tables

Table 1-1 Comparison of net energy and CO2 balances from different energy crops.

Energy Crop	Source of Energy	Net Energy Balance (GJ/ha/yr)	Net CO2 Balance
Maize	Starch	10-80	positive
Sugarcane	Sucrose	55-80	positive
Soybean	Biodiesel	-20-10	positive
Switchgrass (Alamo)	Lignocellulose	150-500	possible negative
Populus	Lignocellulose	150-250	possible negative

Note. Adapted from “Plants to power: bioenergy to fuel the future ,” by J.S. Yuan, K.H. Tiller, H. Al-Ahmad, et al., 2008, *Trends in plant science*, 13(8), p.421-429.

Table 1-2 Comparison of cell wall composition between several energy crops.

Energy Crop	Cellulose (%)	Hemicellulose (%)	Lignin (%)	Reference
Maize	18	40	1	www.doegenomestolife.org/biofuels/
Switchgrass (Alamo)	30.97	24.4	17.56	Carroll & Somerville, 2009
Corn stover	37.69	22.6	18.59	Carroll & Somerville, 2009; www.doegenomestolife.org/biofuels/
Wheat straw	38.2	24.7	23.4	www.doegenomestolife.org/biofuels/
Surgarcane bagasse	39.01	23	23.09	Carroll & Somerville, 2009; Vertès et al., 2010
Populus	39.23	20.4	25.18	Carroll & Somerville, 2009; www.doegenomestolife.org/biofuels/

1.13 References

Abramson, M., Shoseyov, O., and Shani, Z. (2010). Plant cell wall reconstruction toward improved lignocellulosic production and processability. *Plant Science, 178*: 61-72.

Brown, T.A. (2001). *Gene Cloning and DNA Analysis: An Introduction*. London: Blackwell Science Ltd.

Bernstein, I.L., Bernstein, J.A., Miller, M., Tierzieva, S., Bernstein, D.I., Lummus, Z., Selgrade, M.K., Doerfler, D.L., and Seligy, V.L. (1999). Immune responses in farm workers after exposure to Bacillus thuringiensis pesticides. *Environmental health perspectives, 107*(7), 575-82.

Beyer, P., Al-babili, S., Ye, X., Lucca, P., Schaub, P., Welsch, R., and Potrykus, I. (2002). Golden Rice : Introducing the β-Carotene Biosynthesis Pathway into Rice Endosperm by Genetic Engineering to Defeat Vitamin A Deficiency. *Symposium : Plant Breeding : A New Tool for Fighting Micronutrient Malnutrition*, 506-510.

Bjurhager, I., Olsson, A.M., Zhang, B., Gerber, L., Kumar, M., Berglund, L.A., Burgert, I., Sundberg, B., and Salmén, L. (2010). Ultrastructure and mechanical properties of populus wood with reduced lignin content caused by transgenic down-regulation of cinnamate 4-hydroxylase. *Biomacromolecules*, *11*(9), 2359-2365.

Boudet AM. (2000). Lignins and lignification: Selected issues. *plant physiology and biochemistry*, *38*(1-2), 81-96.

Brown, R.C. (2003*). Biorenewable resources: engineering new products from agriculture* (1st ed.). Iowa: A blackwell Publishing Company.

Brunner, A.M., and Nilsson, O. (2004). Revisiting tree maturation and floral initiation in the poplar functional genomics era. *New Phytologist*, *164*(1), 43-51.

Burton, R.A., Gidley, M.J. and Fincher, G.B. (2010). Heterogeneity in the chemistry, structure and function of plant cell walls. *Nature chemical biology, 6*(10), 724-732.

Carroll, A., and Somerville, C. (2009). Cellulosic biofuels. *Annual review of plant biology, 60*, 165-182.

Ceccarelli S., Guimaraes E.P., and Weltzien E. (Eds.) (2009). Plant Breeding and farmer participation. Rome: Food and Agriculture Organization of the United Nations.

Center for Energy. (2010). *Biomass:Overview*. Retrieved from http://www.centreforenergy.com/AboutEnergy/Biomass/Overview.asp?page1-11

Chen, H., and Qiu, W. (2010). Key technologies for bioethanol production from lignocellulose. *Biotechnology advances, 28*(5), 556-562.

Collinge, D.B., Jørgensen, H.J.L., Lund, O.S., and Lyngkjær, M.F. (2010). Engineering Pathogen Resistance in Crop Plants: Current Trends and Future Prospects. Annual review of phytopathology, 48, 269-291.

Chang, T., Chang, H., Wu, C., Lin, H., and Chang, S. (2010). Stabilizing effect of extractives on the photo-oxidation of Acacia confusa wood. *Polymer Degradation and Stability*, *95*(9), 1518-1522.

Cherubini, F., and Stromman, A. (2010). Production of Biofuels and Biochemicals from Lignocellulosic Biomass: Estimation of Maximum Theoretical Yields and Efficiencies Using Matrix Algebra. *Energy and Fuels, 24*(4), 2657-2666.

Chinga-Carrasco, G., Johnsen, P.O., and Øyaas, K. (2010). Structural quantification of wood fibre surfaces--morphological effects of pulping and enzymatic treatment. *Micron*, *41*(6), 648-659.

Christiernin, M. (2006). Lignin composition in cambial tissues of poplar. *Plant physiology and biochemistry, 44*(11) 700-706.

Dang, V.,Kassahn K., Marcos, A. and Ragan, M. (2008). Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *European journal of human genetics, 16*, 1350-1357.

David, K. and Ragauskas, A.J. (2010). Switchgrass as an energy crop for biofuel production: A review of its ligno-cellulosic chemical properties. *Energy and Environmental Science, 3*(9), 1182-1190.

Demura, T., and Fukuda, H. (2007). Transcriptional regulation in wood formation. *Trends in plant science, 12*(2), 64-70.

Deutschbauer, A., Jaramillo, D., Proctor, M., Kumm, J., Hillenmeyer, M., Davis, R., Nislow, C., and Giaever, G. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics, 169*,1915-1925.

Dickmann D., Isebrands J., Eckenwalder J. and Richardson J. (Eds.). (2001). *Poplar Culture in North America.* Ottawa, Canada: National Research Council (Canada) Research Press.

Ellis, B., Jansson, S., Strauss, S. H., and Tuskan, G. A. (2010). Genetics and Genomics of Populus. In S. Jansson, R. Bhalerao, and A. Groover, (Eds.), *Genetics and Genome of Populus* (pp. 3-14). New York, NY: Springer New York. Retrieved from http://www.springerlink.com.proxy.lib.sfu.ca/content/n42160t625t40208/fulltext.pdf.

Evenson, R.E., and Gollin, D. (2003). Assessing the impact of the Green Revolution, 1960 to 2000. *Science*, *300*(5620), 758-762.

Farre G., Ramessar K., Twyman R.M., Capell T., and Christou P. (2010). The humanitarian impact of plant biotechnology: recent breakthroughs vs bottlenecks for adoption. Current opinion in plant biology, *13*(2), 219-225.

Flachowsky, H., Hanke, M.V., Peil, A., Strauss, SH., and Fladung, M. (2009). A review on transgenic approaches to accelerate breeding of woody plants. *Plant Breeding*, *128*(3), 217–226.

Gilchrist, E., and Haughn, G. (2010). Reverse genetics techniques: engineering loss and gain of gene function in plants. Briefings in functional genomics, 9 (2), 103-110.

Hamberger, B., Ellis, M., Friedmann, M., Azevedo Souza, C. de, Barbazuk, B., and Douglas, C. J. (2007). Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. *Botany*, *85*(12), 1182-1201.

Hinchee, M., Rottmann, W., Mullinax, L., Zhang, C., Chang, S., Cunningham, M., Pearson, L., and Nehra, N. (2009). Short-rotation woody crops for bioenergy and biofuels applications. *In vitro cellular and developmental biology-Plant*, *45*(6), 619-629.

Hu, TQ. (Ed.) (2002). *Chemical Modification, Properties, and Usage of Lignin*. New York, USA: Kluwer Academic/Plenum Publishers.

Hu, W.J., Harding, S.A., Lung, J., Popko, J.L., Ralph, J., Stokke, D.D., Tsai, C.J., and Chiang, V.L. (1999). Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nature biotechnology*, *17*(8), 808-812.

Huntley, S.K., Ellis, D., Gilbert, M., Chapple, C., and Mansfield, S. (2003). Significant increases in pulping efficiency in C4H-F5H-transformed poplars: improved chemical savings and reduced environmental toxins. *Journal of agricultural and food chemistry*, *51*(21), 6178-6183.

International Service for the Acquisition of Agri-biotech Applications (ISAAA). (2011). ISAAA Brief 41-2009: Executive Summary, *Global Status of Commercialized Biotech/GM Crops: 2009, The first fourteen years, 1996 to 2009*. Retrieved from http://www.isaaa.org/resources/publications/briefs/41/executivesummary/default.asp.

Jarvis, M.C. (2009). Plant cell walls: Supramolecular assemblies. *Food Hydrocolloids*, *25*(2), 257-262.

Jansson, S., and Douglas, C. J. (2007). Populus: a model system for plant biology. *Annual review of plant biology*, *58*, 435-58.

Jones, L., Ennos, a R., and Turner, S. R. (2001). Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of *Arabidopsis*. *The Plant journal: for cell and molecular biology*, *26*(2), 205-216. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11389761.

Jouanin, L., Goujon, T., de Nadaï, V., Martin, M.T., Mila, I., Vallet, C., Pollet, B., Yoshinaga, A., Chabbert, B., Petit-Conil, M., and Lapierre, C. (2000). Lignification in transgenic poplars with extremely reduced caffeic acid O-methyltransferase activity. *Plant physiology*, *123*(4), 1363-74.

Kärkönen, A., and Koutaniemi, S. (2010). Lignin biosynthesis studies in plant tissue cultures. *Journal of integrative plant biology*, *52*(2), 176-185.

Keegstra, K. (2010). Plant cell walls. *Plant physiology 154*(2), 483-486.

Kishimoto, T., Chiba, W., Saito, K., Fukushima, K., Uraki, Y., and Ubukata, M. (2010). Influence of syringyl to guaiacyl ratio on the structure of natural and synthetic lignins. *Journal of agricultural and food chemistry*, *58*(2), 895-901.

Kole C., and Abbott A. (Eds.) (2008). Principles and Practices of Plant Genomics, Volume 2: Molecular Breeding. South Carolina, USA: Science Publishers.

Kramer, M.G., and Redenbaug, M.G. (1994). Commercialization of a tomato with an antisense polygalacturonase gene: The FLAVR SAVR^{TM} tomato story. *Euphytica*, *79*(3), 293-297.

Krieger, U., Lippman, Z. B., and Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature genetics*, *42*(5), 459–463.

Lamkey, K., and Lee, M. (Eds.). (2006). *Plant Breeding: The Arnel R. Hallauer International Symposium*. Victoria, Australia: Blackwell Publishing Ltd.

Lapierre, C., Pilate, G., Pollet, B., Mila, I., Leplé, J., Jouanin, l., Kim, h., and Ralph, J. (2004). Signatures of cinnamyl alcohol dehydrogenase deficiency in poplar lignins. *Phytochemistry*, *65*(3), 313-321.

Lapierre, C., Pollet, B., Petit-Conil, M., Toval, G., Romero, J., Pilate, G., Leple, J., Boerjan, W., Ferret, V., De Nadai V., and Jouanin, L. (1999). Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant physiology*, *119*(1), 153-164.

Lemaux, P.G. (2008). Genetically Engineered Plants and Foods: A Scientist's Analysis of the Issues (Part I). Annual review of plant biology, 59, 771-812.

Leplé, J., Dauwe, R., Morreel, K., et al. (2007). Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *The Plant cell*, *19*(11), 3669-3691.

Li, X. and Chapple, C. (2010). Understanding lignification: challenges beyond monolignol biosynthesis. *Plant physiology* 154(2), 449-452.

Marita, J.M.,Ralph, J., Hatfield, R.D., and Chapple, C. (1999). NMR characterization of lignins in *Arabidopsis* altered in the activity of ferulate 5-hydroxylase. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(22), 12328-12332.

Martín, B., Ramiro, M., Martínez-Zapater, J. M., and Alonso-Blanco, C. (2009). A high-density collection of EMS-induced mutations for TILLING in *Landsberg erecta* genetic background of *Arabidopsis*. *BMC plant biology*, *9*, 147.

Martínez, A.T., Ruiz-Dueñas, F.J., Martínez, M.J., Del Río, J.C. and Gutiérrez, A. (2009). Enzymatic delignification of plant cell wall: from nature to mill. *Current opinion in biotechnology, 20*(3), 348-357.

McCann, M., and Rose, J. (2010). Blueprints for building plant cell walls. *Plant physiology*, 153(2), 365.

Meyermans, H., Morreel, K., Lapierre, C., Pollet, B., De Bruyn, A., and Busson, R. (2000). Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis. *Journal of biological chemistry, 275*(47), 36899-36909.

Monsanto Technology LLC. (2006). *Roundup Ready Soybeans: Your Greatest Advantage*. Monsanto Canada Inc., licensee. Retrieved from http://www.monsanto.ca/_pdfs/seeds_traits/rrsoybeanbro.pdf

Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., and Ladisch, M. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource technology, 96*(6), 673-86.

Naqvi, M.,Yan, J., and Dahlquist, E. (2010). Black liquor gasification integrated in pulp and paper mills: A critical review. *Bioresource technology*, *101*(21), 8001-8015.

Nicholl, D. S. (2002). *An introduction to Genetic Engineering.* cambridge: Cambridge university press.

Novaes, E., Kirst, M., Chiang, V., Winter-Sederoff, H. and Sederoff, R. (2010). Lignin and biomass: a negative correlation for wood formation and lignin content in trees. *Plant physiology, 154*(2), 555-561.

Pavy, N., Johnson, J.J., Crow, J.A., Paule, C., Kunau, T., MacKay, J., and Retzel, E.F. (2007). ForestTreeDB: a database dedicated to the mining of tree transcriptomes. *Nucleic acids research*, *35*, D888-894.

Peter H. (2002). The genes of the green revolution. *Trends in Genetics*, *19*(1), 5-9.

Pillitteri, L., Bemis, S., Shpak, E. and Torii, K. (2007). Haploinsufficiency after successive loss of signalling reveals a role for ERECTA-family genes in *Arabidopsis* ovule development. *Development, 134*(17), 3099-3109.

Ragauskas, A., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R., and Tschaplinski, T. (2006). The path forward for biofuels and biomaterials. *Science* 311: 484-489.

Raven P., Evert R., and Eichhorn S. (2003). Biology of Plants, 6th edition. New York, USA: W.H. Freeman and Company.

Regalbuto, J.R. (2009). Cellulosic Biofuels--Got Gasoline? *Science, 325*(5942), 822 – 824.

Sazanov, Y., and Gribanov, A. V. (2009). Thermochemistry of lignin. *Russian Journal of Applied Chemistry, 83*(2), 175-194.

Schahczenski J. (2006). Transgenic Crops. *The National Sustainable Agriculture Information Service*. Retreived from http://attra.ncat.org/attra-pub/PDF/geneticeng.pdf

Shi, R., Sun, Y., Li, Q., Heber, S., Sederoff, R., and Chiang, V.L. (2010). Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant and cell physiology*, *51*(1), 144-163.

Shimada N., Murata H., and Miyazaki S. (2008). Safety Evaluation of Bt Plants. *Jarq-japan agricultural research quarterly*, *42*(4), 251-259.

Simmons B.A., Logue D., and Ralph J. (2010). Advances in modifying lignin for enhanced biofuel production. *Current opinion in plant biology*, *13*(3), 313-320.

Sivakumar, G., Vail, D., Xu, J., Burner, D., Lay, J., Ge, X. and Weathers, P. (2010). Bioethanol and biodiesel: Alternative liquid fuels for future generations. *Engineering in Life Sciences, 10*(1), 8-18.

Soetaert, W., and Vandamme, E.J. (2009). *Biofuels* (Eds.). Great Britain: John Wiley and Sons. Ltd.

Somers, D. J., Langridge, P., and Gustafson, J.P. (Eds.). (2009). *Plant Genomics: Methods and Protocols.* New York: Springer-Verlag.

Somerville, C., Youngs, H., Taylor, C., David, S. C., and Long, S. P. (2010). Feedstocks for Lignocellulosic Biofuels. *Science, 329*(5993), 790-792.

Stemple, D.L. 2004. TILLING—a high-throughput harvest for functional genomics. *Nature Reviews Genetics, 5,* 145–150.

Suckling, C., Campbell, C., Mcintyre, B., et al. (2010 Edition). *Global Forest , Paper and Packaging Industry Survey*. In PricewaterhouseCoopers Global Forest LLP. Retrieved from http://www.pwc.com/gx/en/forest-paper-packaging/publications/2010-fpp-survey.jhtml

Taiz, L., and Zeiger, E. (2006). *Plant Physiology* (4th edition). Sunderland, Massachusetts: Sinauer Associates, Inc.

Tang, G., Qin, J., Dolnikowski, G.G., Russell, R.M., and Grusak, M.A. (2009). Golden Rice is an effective source of vitamin A. *American Journal of Clinical Nutrition*, *89*(6), 1776-1783.

Taylor, G. (2002). Populus: *Arabidopsis* for Forestry. Do We Need a Model Tree? *Annals of Botany*, *90*(6), 681-689.

Till, B. J., Reynolds, S. H., Weil, C., Springer, N., Burtner, C., Young, K., Bowers, E., Codomo, C. A., Enns, L. C., Odden, A. R., Greene, E. A., Comai, L., and Henikoff, S. (2004). Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biology, 4*, 12.

Till, B. J., Zerr, T., Bowers, E., Greene, E. a, Comai, L., and Henikoff, S. (2006). *Nucleic acids research*, *34*(13), e99.

Trigiano, R., and Gray, D. (2005). *Plant Development and Biotechnology* (Eds.). Florida: CRC press LLC.

Turner, S. R., and Somerville, C. R. (1997). Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary cell wall. *The Plant cell*, *9*(5), 689-701.

Tuskan, G.A., Difazio, S., Jansson, S., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr and Gray). *Science*, *313*(5793), 1596-1604.

Umezawa, T. (2010). The cinnamate/monolignol pathway. *Phytochemistry Reviews*, *9*(1), 1-17.

U.S. DOE.2006. *Breaking the Biology Barriers to Cellulosic Ethanol: A Joint Research Agenda*, DOE/SC-0095, U.S Department of Energy Office of Science and office of Energy Efficiency and Renewable Energy.

Van Parijs, F.R.D., Morreel, K., Ralph, J., Boerjan, W., and Merks, R.M. H. (2010). Modeling lignin polymerization. I. simulation model of dehydrogenation polymers. *Plant physiology*, *153*(3), 1332-1344.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant physiology, 153*(3), 895-905.

Vanholme, R., Morreel, K., Ralph, J., and Boerjan, W. (2008). Lignin engineering. *Current opinion in plant biology*, *11*(3), 278-285.

Vertès, A., Qureshi, N., Blaschek, H., and Yukawa, H. (2010). *Biomass to biofuels; strategies for global industries* (Eds.). Great Britain: John Wiley and Sons. Ltd.

Veitia, R.A (2006). The biology of genetic dominance. Georgetown, Tex. : Landes Bioscience: Eurekah.com. Retrieved from http://www.netlibrary.com.proxy.lib.sfu.ca/Reader/

Veitia, R. a. (2010). A generalized model of gene dosage and dominant negative effects in macromolecular complexes. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, *24*(4), 994-1002.

Voelker, S. L., Lachenbruch, B., Meinzer, F. C., Jourdes, M., Ki, C., Patten, A. M., et al. (2010). Antisense Down-Regulation of 4CL Expression Alters Lignification, Tree Growth, and Saccharification Potential of Field-Grown Poplar. *Plant physiology*, *154*(2), 874-86.

Wall, JD., Harwood, C.S., and Demain, A. (Eds.). (2008). *Bioenergy*. Washington: ASM press. Retrieved from http://site.ebrary.com/sfu/Doc?id=10346892&ppg=4

Whetten, R.W., MacKay, J.J., and Sederoff, R.R. (1998). Recent Advances in Understanding Lignin Biosynthesis. Annual Review of Plant Physiology and Plant Molecular Biology, *49*, 585-609.

World Energy Council. (September 2010). *Biofuels: Policies, Standards and Technologies*. Retrieved from http://www.worldenergy.org/publications/2837.asp

Xu Y. (2010). Molecular Plant Breeding. CAB international.

Yadav, R., Arora, P., Kumar, D., Katyal, D., Dilbaghi, N., and Chaudhury, A. (2009). High frequency direct plant regeneration from leaf, internode, and root segments of Eastern Cottonwood (*Populus deltoides*). *Plant Biotechnology Reports*, *3*(3), 175-182.

Yang, X., Kalluri, U., DiFazio, S., Wullschleger, S., Tschaplinski, T., Cheng, M.Z., and Tuskan, G. (2009). Poplar Genomics: State of the Science. *Critical Reviews in Plant Sciences*, *28*(5), 285-308.

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

Zhang, H., Harry, D.E., Ma, C., Yuceer, C., Hsu, C., Vikram, V., Shevchenko, O., Etherington, E., and Strauss, S.H. (2010). Precocious flowering in trees: the FLOWERING LOCUS T gene as a research and breeding tool in *Populus*. *Journal of experimental botany*, *61*(10), 2549-2560.

CHAPTER 2: DEVELOPING AN EMS-MUTAGENIZED POPULATION OF HYBRID *POPULUS* BY CLONAL MICROPROPAGATION

2.1 Abstract

The purpose of this study was to generate a population of poplar plants that had been subject to chemical mutagenesis and that therefore might contain individuals with mutations in genes of interest to us. To avoid problems associated with traditional seed mutagenesis, we choose to mutagenize tissue from culture-derived calli rather than seeds. For this purpose, we tested a number of hybrid poplar varieties and identified four that performed well in tissue culture. We thereafter mutagenized calli with a series of concentrations of ethyl methane sulfonate (EMS) ranging from standard 25 mM up to unusually high concentrations of 200 mM EMS. Higher concentrations, 100 and 200 mM, resulted in extensive cell death indicative of the strong effects of the mutagen. We regenerated >6,000 plants from mutagenized calli and grew them in a greenhouse. Plants with abnormal phenotypes were observed in the generated individuals including leaf variegation (albino), abnormal leaf shapes and dwarf plants. Based on this evidence, we concluded that this population may contain mutations in various genes.

2.2 Introduction

While transgenic technology is currently the method used for altering the activity of specific genes in poplar, there are several limitations with this approach. First, genetic transformation through *Agrobacterium tumefaciens* infection is primarily carried out on a *Populus alba* variety, but has proven technically difficult on North-American species and hybrids. Secondly, introduced constructs are frequently silenced over time. More important from an applied research perspective, the forest industry does not want to associate themselves with transgenic trees due to fear of a public backlash. An alternative method that may address the above limitations is the use of induced mutations to generate required genetic variation in a relatively small population compared to natural populations, followed by identification of desired mutants in the following generation(s). While this method has been highly successful in agri- and horti-cultural species, to date it has not been applied to forest tree species. There are several reasons for this. The long generation time in forest tree species prevent a timely progression to F2 and F3 generations, which is required to reveal deviating phenotypes resulting from homozygosity of recessive alleles in a mutant locus. For example, it takes 10 to 30 years for a cottonwood poplar tree to reach sexual maturity (Dickman, 2001; Flachowsky et al., 2009) and in aspen it takes 8 to 20 years to become sexually mature (Giri et al., 2004). This means we need a minimum of 20 years to generate an F2 poplar population. In addition, many tree species, including those in the *Populus* genus, are dioecious, i.e. have separate female and male individuals,

preventing a simple self-fertilization of F1 individuals to generate F2 plants that are homozygous in mutant loci. As a consequence, homozygous mutants can only be generated by parent-offspring mating or mating of siblings, by controlled crosses of F2 plants with parental F1 plants, or crosses of F2 sibling plants, to generate an F3 population containing homozygous mutants. The undertaking of such projects is beyond the realm of forest tree breeding programs around the world. It is possible, however, that if mutations in desired genes could be detected in heterozygous mutant plants, those plants can be subjected to conditions that shorten the time to flowering, and male and female individuals carrying mutant alleles in the same locus can be crossed, generating a stable homozygous F2 mutant. TILLING technology (see chapter 1 and 3) enables screening of mutant populations based on molecular lesions in heterozygous individuals. To test this mutagenesis, we pursued the generation of a pilot mutant population in hybrid poplar.

While mutagenesis is traditionally performed on seeds, it comes with a serious drawback in that generated plants are chimeric, harbouring sectors of cells with different patterns of induced mutations. Usually this is rectified by generating selfed F2 plants, that are no longer chimeric and that may be homozygous for induced mutations. This approach is not practical in poplar plants due to long generation time (> 12 years for females, about 5 years for males). In addition, open-air pollinated seeds from poplar trees harbour a potentially large number of allelic variants in each locus, making it difficult to distinguish induced single nucleotide polymorphisms (SNPs) from natural SNPs

by TILLING. A potential alternative to seeds as a source for mutagenesis are axenic cell cultures. There is evidence that shoot organogenesis from cell cultures initiates from single cells (Broertjes and Keen, 1980), which, if true, would result in the regeneration of non-chimeric plants from mutagenized and chimeric cell cultures. In addition, clonal propagation through cell culture generates plants that are genetically identical, thereby reducing the available genetic variation to that present in the two alleles of the diploid starting material. This advantage is potentially very important in the identification of mutants by TILLING (see chapter 3).

Plant tissue culture is a technique used to artificially grow, maintain and propagate plant cell, tissue, or organ, which dates back to the late 19th century to the work done by Gottlieb Haberlandt (Evans et al., 2003; Raven et al., 2003). Advances in plant physiology, especially the discovery and better understanding of plant hormones, significantly improved tissue culture techniques and allowed researchers to successfully generate and maintain various types of plants in tissue culture (George et al., 2008). Auxin and cytokinin are the most important hormones used in many tissue culture media (Evans et al., 2003). Media used in tissue culture should contain four main elements: water, nutrients both organic (sucrose, vitamins and cofactors) and inorganic (minerals), plant hormones and a supporting media such as agar (Evans et al., 2003). It relies on the ratio of auxin and cytokinin present in the media to direct organogenesis. In general, equimolar amounts of auxin and cytokinin initiates callus induction, higher cytokinin concentration induces shoot

formation and higher auxin concentration induces root formation (Evans et al., 2003). Plant tissue culture has been used for many different reasons including the generation of genetically modified plants (transgenic plants), crop improvement, propagation of disease free plants and rapid large-scale propagation of genetic stocks especially when sexual reproduction is not feasible or when the plant's life cycle is very long (Evans et al., 2003).

Clonal micropropagation is one of the techniques used in plant tissue culture where hundreds of plants are generated by vegetative reproduction from one parent under controlled aseptic environment (Evans et al., 2003; George et al., 2008). The part of the mother plant that is used in plant tissue culture is known as the explant (Trigiano et al., 2005). The explant can be obtained from various parts of the mother plant such as meristems, buds, leaves or developing embryos (George et al., 2008; Raven et al., 2003). The main steps involved in plant micropropagation include preparation of explants, which involves excision and sterilization, followed by inducing organ formation through culturing in different media in sterile environment and finally the generation of mature plants (George et al., 2008). Clonal micropropagation can be divided into three broad classes: meristem cultures, organogenesis and embryogenesis (Evans et al., 2003; George et al., 2008). In meristem cultures, the shoot tips containing the shoot apical meristem or shoots containing auxiliary buds are placed directly into various growth media for induction and propagation of lateral shoots (George et al., 2008). In propagation by organogenesis, the explant is first induced to produce growth of undifferentiated cells, either as a suspension

culture or as calli growing on solid medium. The generated calli are then typically placed on shoot-inducing media, followed by transfer of shoots to root-inducing media to finally generate new plants via organogenesis (George et al., 2008). In some species, cell suspension or calli cultures can be induced to undergo somatic embryogenesis, often simply by washing off exogenously applied hormones and plating cells at low density (Evans et al., 2003). The generation of new plants through somatic embryogenesis occurs through stages that are similar to zygotic embryogenesis, including globular and heart stages (Evans et al., 2003). The origin of the explants, plant species, the reason for tissue culturing and the availability of resources are the main factors that determine which type of micropropagation technique to use (Trigiano et al., 2005).

Plant micropropagation has been successfully used for the clonal propagation of fruit trees, such as apple trees, as well as forest trees such as spruce, poplar and some species of pine (Dickman, 2001; Ernst, 1993; Evans et al., 2003; Yadav et al., 2009). *CellFor Inc*., located on Vancouver Island, British Columbia, Canada, is one of the companies that use micropropagation, mainly embryogenesis, to generate conifer seedlings with improved traits such as wood quality, disease resistance, and higher yield (http://www.cellfor.com/home.html).

Besides all the promising advantages of clonal micropropagation, there are also some disadvantages of this technique. Plant tissue culture, especially micropropagation, is labour intensive and expensive (Evans et al., 2003). In

addition, the ability to culture plants artificially depends on the species, age, and what part of the plant used as explant (Trigiano et al., 2005). For example, even though the majority of members of *Populus* species can be propagated using plant tissue culture, *Populus trichocarpa*, clone Nisqually-1 (whose genome has been sequenced) is one of *Populus* species that is very difficult to propagate (Kang et al., 2009; Rutledge and Douglas, 1988). In addition, contamination with bacteria or fungi easily result in the loss of axenic material (Evans et al., 2003). Proper hygiene practices, which includes periodic sterilization of tools, media, and working bench, significantly prevents contamination and increases the chance of generating healthy plants (George et al., 2008). In addition, from a final deployment perspective, plants generated using micropropagation are genetically identical, which makes the whole population sensitive to the same pathogens (Trigiano et al., 2005).

Here we tested the possibility of generating a poplar mutant population by ethyl methane sulfonate (EMS)-induced mutagenesis of cell cultures followed by organogenesis. We generated ~ 6,700 poplar plants through this procedure.

2.3 Materials and Methods

2.3.1 Plant Material and Lines

Twenty lines of hybrid poplar were obtained from Dr. Michael Carlson at the BC Ministry of Forests. An EMS-mutagenized *Populus* population was generated from V5, K5, K8, K9, and K10 hybrid *Populus* lines, which are

different clones from a genetic cross between female *Populus trichocarpa* and a male *Populus deltoides.*

2.3.2 The Generation of EMS-mutagenized Hybrid *Populus* Population

All of the steps in micropropagation were carried under axenic conditions (George et al., 2008). Petioles of young leaves and immature buds of lines V5, K5, K6, K8, K9 and K10 hybrid *Populus* were used for the generation of an EMS-mutagenized population. Murashige and Skoog (MS) media (Phyto Technology Laboratories) was used as the basic medium for all of the tissue culture steps. For calli induction, 1 cm long petiole segments were sterilized by soaking them in a 70% ethanol for 2 minutes and then rinsed four times using sterile water. The subsequent tissue culture and organogenesis was carried out as described in Noël et al. (2002). The sterilized petioles were then placed on a callus initiation medium (CIM) containing 0.5 µM Benzylaminopurine (BAP), 0.5 µM Zeatin, 5 µM NAA and 5 µM 2,4-Dichlorophenoxyacetic acid (2,4-D), for two weeks under 16 hours day light and 8 hours dark. The generated calli were then propagated to increase the amount of generated calli. The resulting calli were then chopped into small pieces and soaked in different EMS concentrations (0, 25, 50, 100 and 200 mM) for 1.5 hours at 30°C on a rotary shaker. The EMS-treated calli were then allowed to proliferate on CIM for approximately two weeks under 16 hours day light and 8 hours dark. Four to five distant regions of each callus were cut off and transferred to shoot initiation media (SIM) containing 0.25 µM Thidiazuron (TDZ) and 0.01% pluronic F68. Upon the

production of elongated shoots, the shoots were cut off and plated on Root Induction Media (RIM) containing 0.5 µM of 3-Indolebutyric acid (IBA). The rooted plantlets were transferred into small pots containing western PRO-MIX HP with Mycorise Pro soil mix (Evergro Canada Inc.) covered with plastic cover, and left for 2-3 days to acclimate in a growth chamber under 16-hour daylight, 24˚C and 70% humidity conditions. When they reached a size of ~ 5 cm, the plantlets were transferred to larger pots in the greenhouse at SFU for further growth.

2.4 Results

We tested a range of EMS concentrations from 25 mM to a maximum of 200 mM. Effects on viability were scored after two weeks on CIM. Pictures of representative petri-dishes are shown in figure 2-1. As can be seen, many of the small calli had grown into large and rounded calli after exposure to 0 and 25 mM EMS. At 50 mM, the frequency of large, rounded calli had dropped considerably. After treatment with 100 or 200 mM EMS, the great majority of calli remained small and showed obvious cut surfaces, indicative of no growth. In addition, these small and cut calli did not initiate growth at a later point and remained gray to brown in comparison to the bright yellow colour of growing calli, suggesting that all cells were dead. Figure 2-2 shows higher magnification pictures of calli exposed to mock treatment (A) or 200 mM EMS. The arrow in B points at the only calli that shows obvious growth after treatment with 200 mM EMS. We estimate that ~ 20/40 of the calli were growing in A, whereas ~1/40 was growing in B, translating roughly into a 95% drop in viability. From this

strong effect of EMS on viability, we concluded that we had identified a suitable interval of EMS concentrations. We thereafter focused on regenerating plants from all the surviving calli from the 100 and 200 mM EMS treatments, but also many of the calli from the 50 mM treatment and a limited number from the 25 mM treatment. Figure 2-3 illustrates the main steps that were used for the micropropagation of the EMS-mutagenized *Populus* population.

The first population of hybrid poplar that survived all the steps during micropropagation and that were successfully established in the greenhouse consisted of ~1,700 rooted trees generated from the hybrid poplar line V5. Since morphometric measurements at the BC Ministry of Forests showed that the hybrid line V5 showed poor performance relative to newer hybrids, we generated a second population from better performing hybrids, abbreviated as K5, K8, K9, and K10 hybrid consisting of ~ 5,000 trees (Figure 2-4).

We screened the generated populations for morphological defects. Abnormal pigmentation (albino leaf colour), abnormal leaf shapes and dwarfism were the major types of morphological abnormalities that were detected in the generated populations (figure 2-5). In 622 plants that were generated from 100 mM EMS treated calli, 12 were found to have an abnormal phenotype, that is around 1.9% of the generated 100 mM population. In 226 plants generated from 200 mM EMS treated calli, 9 were found to have an abnormal phenotype, that is around 3.4% of the plants generated from the 200 mM population. Pigmentation mutant was the major abnormal phenotype observed which accounted for more than 50% of abnormal phenotypes observed.

2.5 Discussion

Here we have generated a large population of putative poplar mutants through micropropagation of mutagenized cells. To the best of our knowledge, no such population has been generated before in any forest tree species. Typically, mutant populations are generated by mutagenizing seeds followed by growth and self-fertilization of F1 plants. The F2 plants come from a single egg cell and thus are not chimeric with respect to induced mutations (Neuffer, 1994). In addition, F2 plants derived from self-fertilization will be homozygous for mutations, and may therefore reveal phenotypes deviating from the wild type norm (Weil and Monde, 2009). The frequency of mutants in an F2 population is the ultimate measure of the efficiency of the initial mutagenesis. What is considered an acceptable frequency varies considerably and depends on how rapidly a population can be screened for desired phenotypes. For example, hundreds of thousands of F2 wheat plants growing in a field can be visually screened for dwarfed plants in a short time, omitting a need for high-frequency mutagenesis. On the other hand, screens for mutants defective in internal structures or metabolite profiles are much more arduous and require populations with high frequency of mutations.

Numerous experiments on seed mutagenesis have shown that the effective dose of EMS varies between species, probably due to obvious differences such as seed coat thickness and accessibility to the stem cells in the shoot apical meristem that generate the subsequent shoot. For the purpose of TILLING, it has been found that a dose of EMS that results in approximately

50% loss in viability also results in a frequency of mutations suitable for TILLING (Haughn and Gilchrist, 2006; Weil and Monde, 2009). The 50% value includes seed lethality, early growth lethality, and inability to set seeds. Usually a suitable dose of EMS is identified empirically by testing the effect of a range of concentrations on lethality. While mutagenesis of cell cultures may not be directly comparable to seed mutagenesis, we also aimed at reducing the "viability" with at least 50%.

To address the problem of genetic chimeras, we mutagenized a large number of tissue culture-derived calli followed by regeneration of shoots and roots (organogenesis). Published data suggests that shoots generated by shoot induction in tissue culture are derived from single cells or from a small number of adjacent cells (sectors), thereby potentially avoiding the creation of chimeras (Broertjes and Keen, 1980). As explained by Broertjes and Keen (1980) a single cell in a callus will generate sectors of identical cells as a result of cell division, thus callus is made up of sectors where each sector is made up of identical cells. When mutagenized, a number of cells will become mutated and upon further cell division, will generate a sector of mutagenized cells that are genetically identical. After mutagenesis, the shoot arising from a callus could arise from cells that are either genetically identical (either mutated or un-mutated sectors) or from genetically different cells, where one of these sectors could be mutated and thus results in chimeric plant with respect to mutations (Broertjes and Keen, 1980). There is insufficient evidence that shoots arise from one single cell or from few cells in callus, but the percentage of chimeras from

callus cultures is always low, less than 5% (Broertjes and Keen, 1980), and the majority of the generated plants are genetically identical (Broertjes and Keen, 1980). TILLING could be used to detect the presence of chimeras in the generated population. Once a mutant plant is identified, DNA obtained from different parts of the mutant plant, for example from leaves, stem, buds and roots, can be re-screened using TILLING. If screening showed the same mutation in all tested samples from the same plant, then that is strong evidence that the generated mutant plant is non-chimeric. The combined use of tissue culture and EMS mutagenesis could therefore enable us to generate a non-chimeric mutant population in a fraction of the time needed if seeds were used as source material. To assess the almost insurmountable problem of identifying poplar mutants in desired genes based on phenotypic deviations, we sought to identify mutants based on genotype deviation in the generated plants rather than phenotype deviation in F3 plants (see chapter 3).

Several approaches have been taken for mutagenesis in relation to subsequent detection based on genotype deviations. For example, hundreds of thousands of *Agrobacterium tumefaciens* T-DNA insertion lines have been generated in *Arabidopsis thaliana* and the disrupted genes have been identified based on PCR and sequencing of flanking sequences (available at www.arabidopsis.org). This is possible in *A. thaliana* because *Agrobacterium* transformation by floral dipping in this species is much easier than transformation of in vitro cultured material that is still the norm in other species. Similar populations have been generated in *A. thaliana* and a few other species

after transposon insertional mutagenesis. In addition to the technical difficulties in the introduction of T-DNAs or the initiation of transposon jumping, these approaches are also hampered by the low frequency of resulting mutations. For example, a population of around 500,000 F1 plants is required to obtain a saturated mutagenized *Arabidopsis* population using a transposon as a mutagen (Maple and Moller, 2006). For that reason, such projects have been undertaken as well-funded large-scale projects to serve a larger research community, such as the *Arabidopsis* research community. Projects in individual laboratories generally rely on chemical or radiation mutagenesis, because these mutagens can result in much higher mutation frequency. An extreme example is an EMS-mutagenized *A. thaliana* population in which 10 or more mutant alleles are typically identified for each target gene in a mutant population with no more than 3,000 individuals (Till et al, 2003). The downside of these mutagens is that the lesions they induce can be difficult to detect. For example, EMS generally cause G to T single nucleotide transitions that are hard to detect because they are easily confused with similar natural variation and also requires highly specialized detection methods (see chapter 3). The relatively low number of required mutant plants after efficient EMS mutagenesis is due to the large number of mutations per mutated cellular genome (Stemple, 2004). In addition, the degree of polyploidy and heterozygosity are other factors that affect the degree of mutagenesis that results from EMS treatment (Table 2-1). For example, EMS treatment can result in one mutation in every 25 kb in *Triticum durum* (tetraploid wheat), while in *Arabidopsis* thaliana (Columbia) it results in

one mutation in every 300 Kb (diploid) (Martín et al., 2009). In polyploidy species, the higher number of mutations generated by EMS treatment is due to the masking of deleterious effects from mutagenesis due to the presence of multiple copies, thus plants with higher number of mutations can survive the mutagenesis. Unlike insertional and radiation mutagenesis that tend to result in a complete gene knockout (Martín et al., 2009), EMS can be used to generate a series of allelic variations from complete to different degrees of partial loss of function alleles thereby allowing mutations in essential genes to survive (Colbert et al., 2001). In cases where a complete loss of function results in lethality, partial loss of function may be more informative (Stemple, 2004). In addition, EMS mutagenesis results in fewer incidents of chromosomal rearrangement including deletion and inversions, thus it is less deleterious than radiation and insertional mutagenesis (Kim et al., 2006).

As mentioned above, it is not possible to assess accurately the efficiency of mutagenesis in populations of heterozygous mutants such as our poplar population, since recessive mutations require homozygosity to reveal deviating phenotypes. For this reason, it is not possible for us to assess the frequency of induced mutations based on phenotypes. We did, however, observe other indications of mutagenesis in our mutant population. First, we observed a dramatic reduction of growth from calli that had been exposed to higher concentrations of EMS, with up to 95% reduction in viability after exposure to 200 mM EMS. Although possibly not directly comparable, this frequency is much higher than the lethal dose of 50% often used as a measure for efficient

mutagenesis in seeds (Haughn and Gilchrist, 2006; Weil and Monde, 2009). In addition, we also observed a low frequency of deviant phenotypes in the mutant population, in particular white and green leaf variegation. Such variegation has been used before as a morphological marker for EMS mutagenesis (Sakamoto, 2003; Koh and Davis, 2001; Martín et al., 2009). Albino leaf sectors is mainly due to the presence of mutant (white) and wild type (green) chloroplast in different cells that make up the plant (Miura et al., 2007). The cytoplasmic inheritance of chloroplasts, where chloroplasts are distributed unequally to daughter cells after mitosis, results in the colour variegation morphology (Taiz and Zeiger, 2006). It is possible that observed phenotypes in the mutant population are due to a phenomenon known as somaclonal variation (Evans et al., 2003). Somaclonal variation could have an effect on the genetic, physical and biochemical level of the generated plants (Trigiano et al., 2005). Some of the abnormal phenotypes that are associated with somaclonal variation include leaf variegation, abnormal leaf shape, plant height and number of flowers (Jain, 2001). Somaclonal variation has been attributed primarily to tissue culture-induced chromosomal duplications, rearrangements, deletions and epigenetic modifications such as DNA methylation (Jain, 2001; Trigiano et al., 2005). While the ultimate causes of somaclonal variation are unknown, the high concentration of auxin hormones in tissue culture is one of the main suspects, as well as the potential production of free radicals in tissue culture (Trigiano et al., 2005). We have reason to believe, however, that somaclonal variation is an unlikely explanation for the phenotype deviations that we observed. Several

hundred plants were regenerated from calli that had undergone the same tissue culture procedures, except for the exposure to EMS, and none of them showed the leaf variegation phenotypes observed in EMS-exposed plants. Many of them were also grown in soil for an extended period of time, without signs of variegation or dwarfism.

In summary, we were successful in generating >6,000 poplar trees from several hybrid lines using clonal micropropagation. In addition, we have reason to believe that the obtained population harbours mutants. The big question is, at what frequency. To address this issue we turned to molecular detection of lesions in target genes using TILLING (see chapter 3).

2.6 Figures

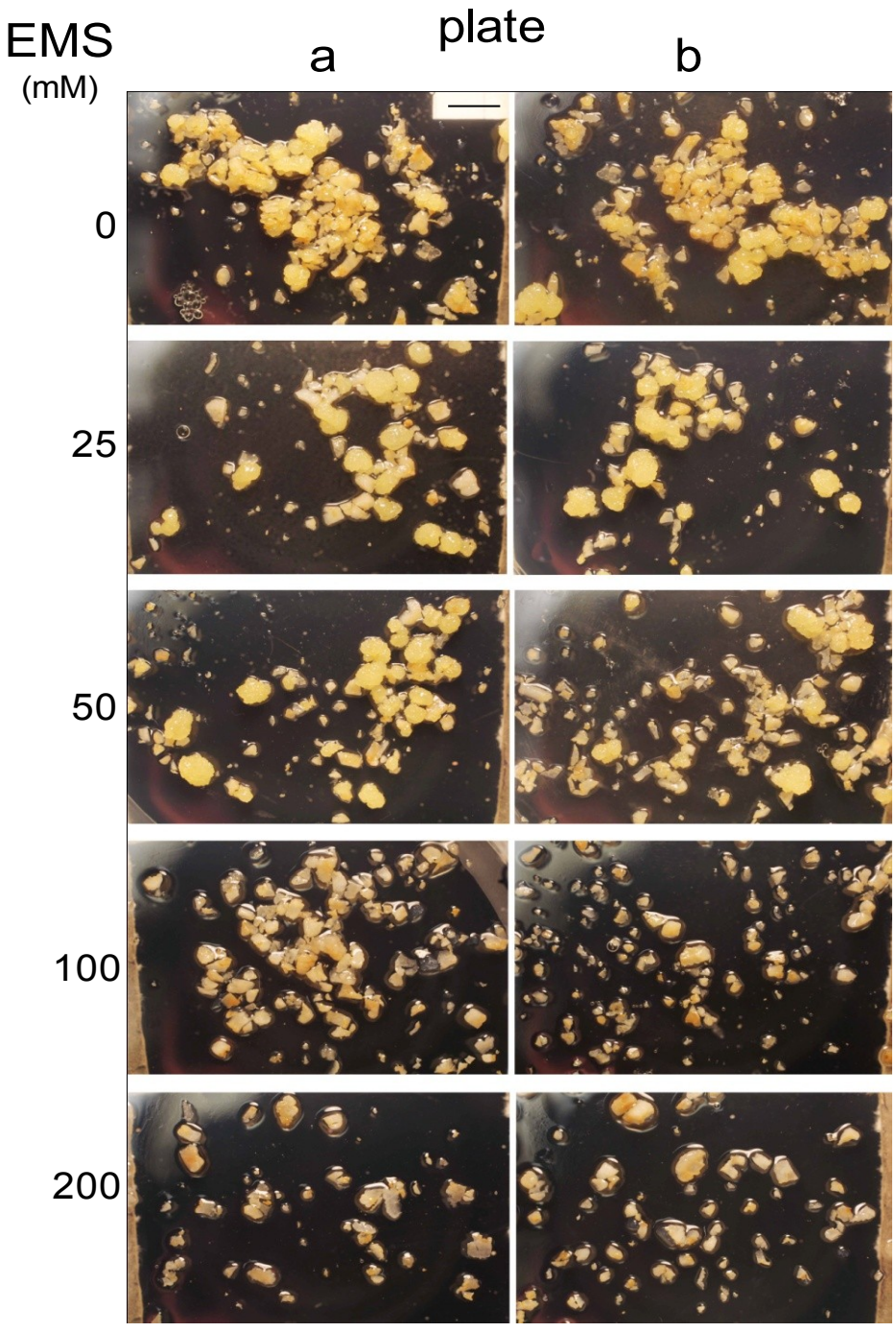


Figure 2-1 The effect of EMS treatment on calli growth.

After treatment with the indicated concentrations of EMS, the calli were grown for two weeks on CIM. Panel a and b are independent replicas. Size bar = 10 mm.

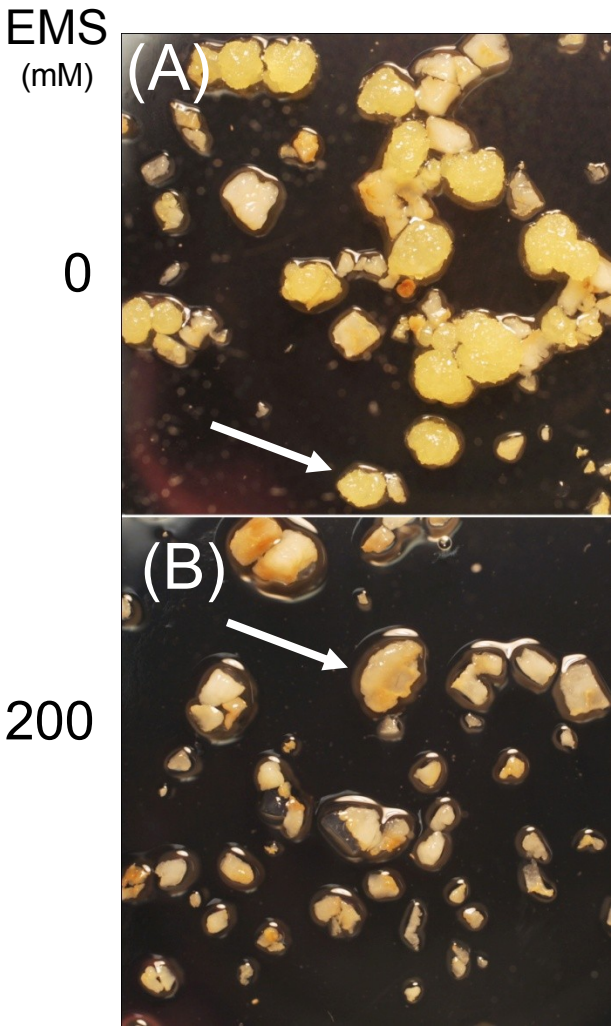


Figure 2-2 Treatment with 200 mM EMS result in an almost complete inhibition of calli growth.

Calli grown on CIM for two weeks after (A) mock treatment without EMS, (B) after exposure to 200 mM EMS. Calli were chopped with a scalpel before exposure to mock or EMS treatment, resulting in flat sides. The increase in size as well as the rounded surfaces in A indicates growth after mock treatment. In (B), the majority of calli remain small and have flat surfaces, indicating a strong effect of the 200 mM EMS treatment. Arrows point at some of the calli that show obvious signs of growth. Identical magnification in both pictures.

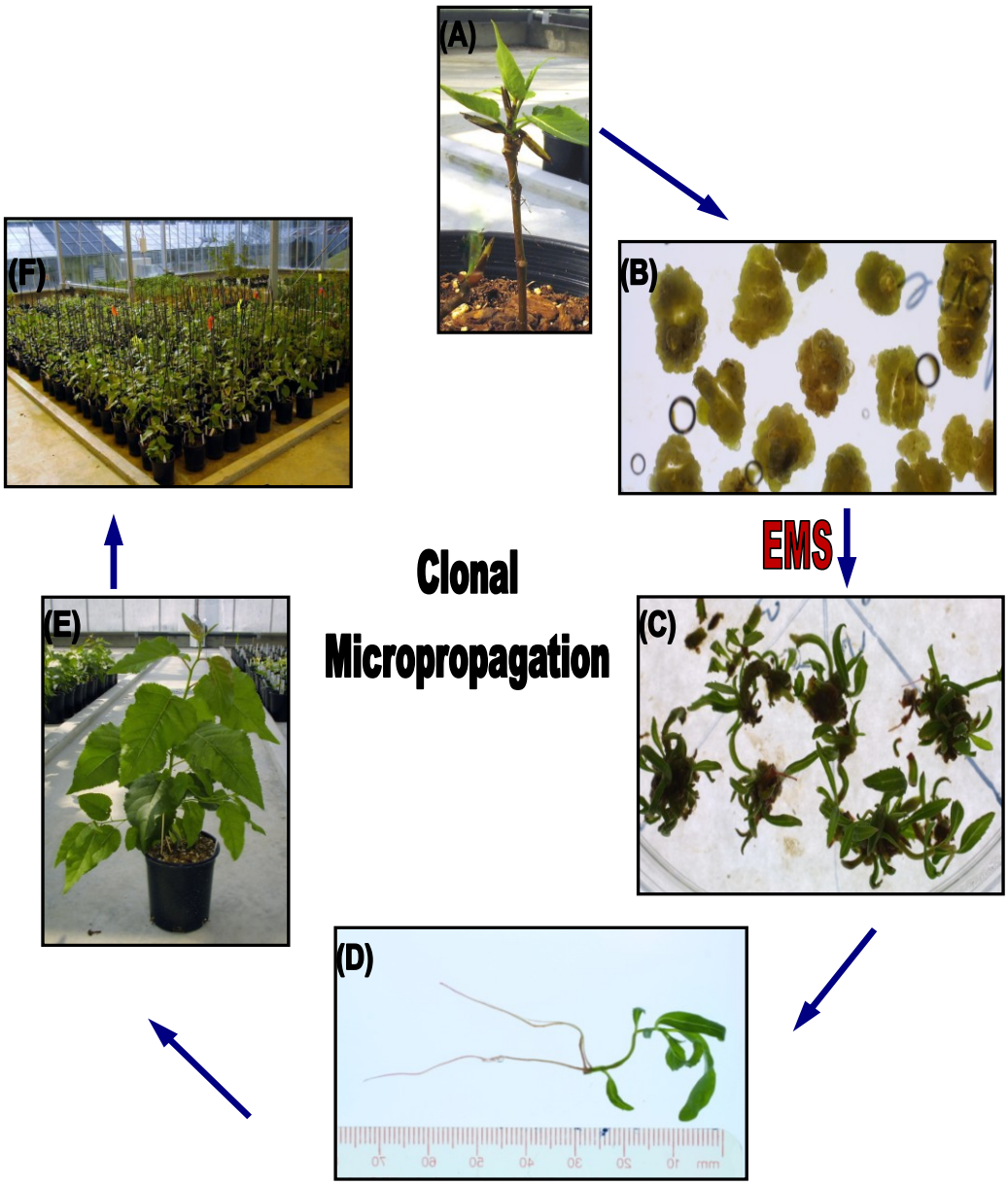


Figure 2-3 Generation of an EMS-mutagenized hybrid *Populus* population using clonal micropropagation.

(A) One of the original hybrid *Populus* (line V5) that was used to generate the mutagenized population. The petiole was used as the explant. (B) Calli on a CIM. The generated calli were mutagenized using EMS. (C) Shoots emerging from calli when placed on SIM. (D) A sample of rooted plantlet obtained from RIM. (E) Resulting poplar plant after approximately 2 months of growth. (F) Plants transferred to pots at SFU's greenhouse.

(A)

(B)

Figure 2-4 The generated EMS-mutagenized hybrid *Populus* population grown at SFU's greenhouse.

(A) EMS mutagenized *Populus* after two month of growth in the greenhouse. (B) After seven months of growth in the greenhouse.

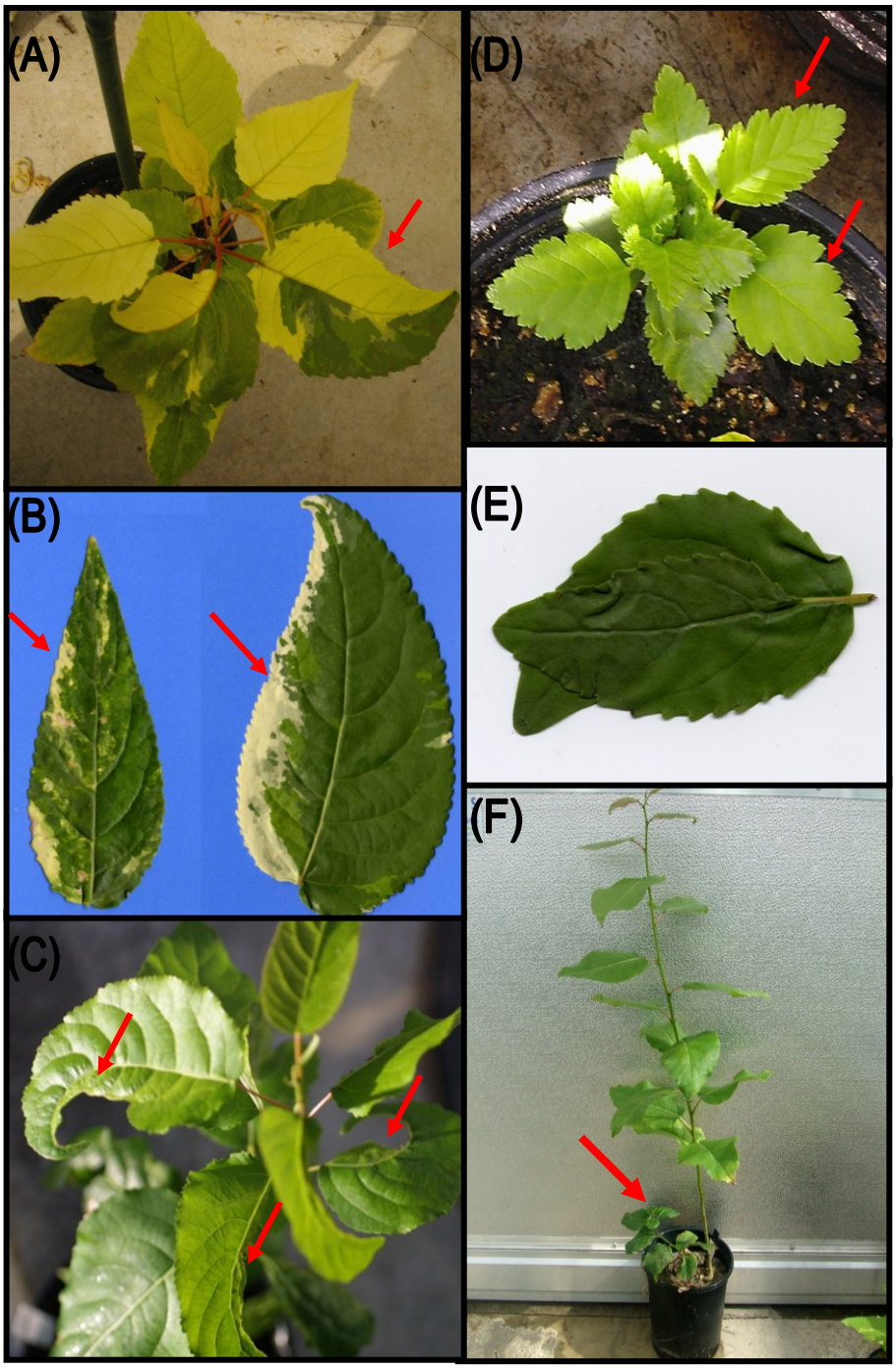


Figure 2-5 Morphological defects present within the generated EMS-mutagenized *Populus* population.
Several abnormal phenotypes were detected in the generated EMS-mutagenized *Populus* population. Abnormal phenotypes included leaf pigmentation (A) and (B), (C) asymmetric leaves, (D) abnormal leaf edges, (E) bifurcated leaf and (F) dwarf. The arrows point to the morphological defects.

2.7 Tables

Table 2-1 Frequency of EMS induced point mutations as detected by TILLING in several species.

Species	Ploidy level	Mutation density per line
Arabidopsis Thalaiana	**Diploid**	**1/300 kb**
Oryza sativa	**Diploid**	**1/294 kb**
Zea mays	**Diploid**	**1/485 kb**
Caenorhabditis elegans	**Diploid**	**1/293 kb**
Triticum durum	**Tetraploid**	**1/25 kb**
Brassica napus	**Tetraploid**	**1/41.5 kb**
Triticum aestivum	**Hexaploid**	**1/40 kb**

Note. Adapted from "A high-density collection of EMS-induced mutations for TILLING in *Landsberg erecta* genetic background of *Arabidopsis*," by Martín, B., Ramiro, M., Martínez-Zapater, J. M., and Alonso-Blanco, C., 2009, *BMC plant biology*, *9*, 147.

2.8 References

Ahloowalia, B.S., 1975. Regeneration of ryegrass plants in tissue culture. Crop Science, *15*, 449–452.

Ahloowalia, B.S., 1976. Chromosomal changes in parasexually produced ryegrass. In K. Jones and P Brandham (Eds.), *Current Chromosome Research* (pp. 115–122). North Holland, Amsterdam: Elsevier.

Broertjes, C. and Keen, A. (1980). Adventitious Shoots: Do They Develop from One Cell? *Euphytica*, *29*, 73-87.

Chun, Y.W. (1993). Clonal propagation in non-aspen poplar hybrids. In M.R. Ahuja (Ed.), *Micropropagation of Woody Plants* (pp. 209-222). Netherlands: Kluwer Academic Publishers.

Dickmann D., Isebrands J., Eckenwalder J. and Richardson J. (Eds.). (2001). *Poplar Culture in North America.* Ottawa, Canada: National Research Council (Canada) Research Press.

Ernst, S.G. (1993). In vitro culture of pure species non-aspen poplars. In Ahuja, M.R. (Ed.), *Micropropagation of Woody Plants* (pp. 195-205).Netherlands: Academic publishers.

Evans, D.E., Coleman, J.O.D., and Kearns, A. (2003). *Plant Cell Culture*. London: BIOS Scientific Publishers.

Flachowsky, H., Hanke, M.V., Peil, A., Strauss, SH., and Fladung, M. (2009). A review on transgenic approaches to accelerate breeding of woody plants. *Plant Breeding*, *128*(3), 217–226.

George, E.F., Hall, M.A., and De klerk, G-J. (Eds.). (2008). *Plant Propagation by Tissue Culture: Volume 1. The Background* (3rd edition). Dordrecht, The Netherlands: Springer. Retrieved from http://www.springerlink.com.proxy.lib.sfu.ca/content/n5tm30/#section=164347andpage=2&locus=43

Giri, C.C., Shyamkumar, B., and Anjaneyulu, C. (2004). Progress in tissue culture, genetic transformation and applications of biotechnology to trees: an overview. *Trees Structure and Function, 18*(2), 115-135.

Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., et al. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics*, *164*(2), 731-40.

Haughn, G.W., and Gilchrist E.J. (2006). TILLING in the Botanical Garden: a reverse genetic technique feasible for all plant species. In: Floriculture, Ornamental and Plant Biotechnology: Advances and Topical Issues (1st Edition), JA Teixeira da Silva (Ed), Global Science Books, Ltd., London, UK.

Jain, S. M. (2001). Tissue culture-derived variation in crop improvement. *Atomic Energy*, *118*(2), 153-166.

Kang, B., Osburn, L., Kopsell, D., Tuskan, G.A., and Cheng, Z. (2009). Micropropagation of Populus trichocarpa 'Nisqually-1': the genotype deriving the Populus reference genome. *Plant Cell, Tissue and Organ Culture*, 99(3), 251-257.

Koh, Y. and Davis Jr., F.T. (2001). Mutagenesis and in vitro culture of *Tillandsia fasciculata Swartz* var. fasciculata (Bromeliaceae). *Scientia Horticulturae*, *87*(3), 225-240.

Kole, C., and Abbott, A. (Eds.) (2008). *Principles and Practices of Plant Genomics, Volume 2: Molecular Breeding*. South Carolina, USA: Science Publishers.

Maple, J., and Moller, S.G. (2006). In E. Rosato (Ed.), *Methods in Molecular Biology, Vol. 362: Circadian Rhythms: Methods and Protocols* (pp.197-205). Totowa, NJ, USA: Humana Press.

Martín, B., Ramiro, M., Martínez-Zapater, J. M., and Alonso-Blanco, C. (2009). A high-density collection of EMS-induced mutations for TILLING in *Landsberg erecta* genetic background of *Arabidopsis*. *BMC plant biology*, *9*, 147.

Miura, E., Kato, Y., and Sakamoto, W. (2010). Comparative transcriptome analysis of green/white variegated sectors in *Arabidopsis* yellow variegated2: responses to oxidative and other stresses in white sectors. *Journal of Experimental Botany, 61*(9), 2433-2445.

Miura, E., Kato, Y., Matsushima, R., Albrecht, V., Laalami, S., and Sakamoto, W. (2007). The balance between protein synthesis and degradation in chloroplasts determines leaf variegation in *Arabidopsis* yellow variegated mutants. *The Plant cell*, *19*(4), 1313-1328.

Neuffer, M.G. (1994). Mutagenesis. In M. Freeling, and V. Walbot (Eds.), The Maize Handbook (pp. 212-219). New York: Springer-Verlag.

Noël, N., Leplé, J.-C., and Pilate, G. (2002). Optimization of in vitro micropropagation and regeneration for *Populus* × *interamericana* and *Populus* × *euramericana* hybrids (*P. deltoides , P . trichocarpa , and P . nigra*). *Plant Cell Reports*, *20*(12), 1150-1155.

Qibin, Y., Mäntylä, N., and Salonen, M. (2001). Rooting of Hybrid Clones of Populus tremula L. × P. tremuloides Michx. by Stem Cuttings Derived from Micropropagated Plants. *Scandinavian Journal of Forest Research*, 16(3), 238-245.

Raven, P., Evert, R., and Eichhorn, S. (2003). *Biology of Plants* (6th edition). New York: W.H. Freeman and Company.

Rutledge, C.B., and Douglas, G.C. (1988). Culture of meristem tips and micropropagation of 12 commercial clones of poplar in vitro. *Physiologia Plantarum*, *72*(2), 367-373.

Sakamoto, W. (2003). Leaf-variegated mutations and their responsible genes in *Arabidopsis thaliana*. *Genes and genetic systems*, *78*(1), 1-9.

Stemple, D.L. 2004. TILLING—a high-throughput harvest for functional genomics. *Nature Reviews Genetics, 5,* 145–150.

Taiz, L., and Zeiger, E. (2006). *Plant Physiology* (4th edition). Sunderland, Massachusetts: Sinauer Associates, Inc.

Till, B. J., Reynolds, S. H., Weil, C., Springer, N., Burtner, C., Young, K., Bowers, E., Codomo, C. A., Enns, L. C., Odden, A. R., Greene, E. A., Comai, L., and Henikoff, S. (2004). Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biology, 4*, 12.

Trigiano, R., and Gray, D. (2005). *Plant Development and Biotechnology* (Eds.). Florida: CRC press LLC.

Weil, C. F., and Monde, R.-A. (2007). Getting the Point-Mutations in Maize. *Crop science*, *47*(S1). Crop Science Society of America. Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=20162538.

Weil, C.F., and Monde, R-A. (2009). EMS mutagenesis and Point Mutation Discovery. In A.L. Kriz and B.A. Larkins (Eds.), *Molecular Genetics Approaches to Maize Improvement: Biotechnology in Agriculture and Forestry, Vol. 63* (pp. 161-170). Berlin, Heidelberg: Springer-Velag.

Yadav, R., Arora, P., Kumar, D., Katyal, D., Dilbaghi, N., and Chaudhury, A. (2009). High frequency direct plant regeneration from leaf, internode, and root segments of Eastern Cottonwood (*Populus deltoides*). *Plant Biotechnology Reports*, *3*(3), 175-182.

CHAPTER 3: SCREENING FOR MUTANTS USING TILLING

3.1 Abstract

The purpose of this study was to screen our poplar population of putative mutants for lesions in genes involved in lignin biosynthesis to identify mutants of interest to pulp and paper as well as biofuels industries. First, we tested several DNA extraction protocols for yield, purity and suitability of obtained DNA for PCR. ChargeSwitch® gDNA Plant Kit provided the best results and was used to extract DNA from the generated poplar population. Secondly, we tested several protocol variants of TILLING (Targeting Induced Local Lesions IN Genomes) for ease of use and efficiency in the detection of point mutations and natural single nucleotide polymorphisms (SNPs). We found that a modified TILLING protocol, using an enzyme that generates double stranded breaks in heteroduplex mismatches, best suited our technical resources at SFU. Finally, we applied these methods to screen our population for lesions in a limited number of genes in the lignin biosynthesis pathway. We identified the presence of 25 unique SNPs in individuals of this population. Upon DNA sequencing, we observed however, that the detected SNPs were of natural rather than induced origin. In any case, we conclude that the mutant population appears to harbour a low frequency of mutants, rendering it impractical for screens by resource intensive TILLING technology.

3.2 Introduction

TILLING, which stands for Targeting Induced Local Lesions IN Genomes, has proven an effective reverse genetic approach to screen for point mutations and single nucleotide polymorphisms (SNPs) (reviewed in Comai and Henikoff, 2006). TILLING enables the detection of mutations based on molecular lesions rather than phenotypic effects, which has many potential advantages. Since its first introduction in the year 2001, the list of species subjected to TILLING is continuously expanding, and include well known species such as *Arabidopsis thaliana*, wheat (Dong et al., 2009), zebrafish, *Caenorhabditis elegans* (Gilchrist et al., 2006), Drosophila as well as humans (Till et al., 2006). Despite the many examples of successful applications of the TILLING technique, it is still a difficult technique requiring considerable resources and expertise. There have been several attempts to modify the current TILLING procedure to simplify its use and to apply it to different organisms. The basic steps of TILLING include the selection of genetic stock with limited heterozygosity, generation of a heavily mutagenized population, amplification of genomic regions of interest, heteroduplex formation, enzymatic cut at heteroduplex points and finally a method for detecting the novel digested products indicative of an induced single nucleotide polymorphism (Dong et al., 2009, Gilchrist et al., 2006; Till et al., 2004).

In this chapter, the focus will be on the detailed steps that were used for screening the generated EMS mutant poplar population for point mutations in genes of interest. We are screening for point mutations in genes that are known

to be involved in the biosynthesis of lignin monomers. To our knowledge, we are the first group to attempt to generate an EMS mutated tree population and to screen them for mutations using TILLING. As a result, we went through many trials and modifications to adapt TILLING and to successfully and efficiently use it in screening of the generated EMS mutagenized poplar population.

3.3 Materials and Methods

3.3.1 DNA extraction and pooling

Leaves were frozen with liquid nitrogen and then ground using mortar and pestle. An amount of 0.1 g of the ground material was used for total genomic DNA extractions using different DNA extraction protocols. The following five extraction protocols were then tested:

Protocol 1: GenElute™ Plant Genomic DNA Miniprep Kit (SIGMA)

The steps for extracting DNA were performed according to the guide provided with the GenElute™ Plant Genomic DNA Miniprep Kit (SIGMA).

Protocol 2: DNAzol Genomic DNA Isolation Buffer (Molecular Research Center, Inc.)

One ml of DNAzol buffer (lysis buffer) was added to the 0.1 g of the ground tissue and left at room temperature for 10 minutes. The mixture was centrifuged for 10 minutes at 10,000 x g to sediment the plant tissue. An amount of 0.5 ml of 100% cold ethanol was added to the supernatant to precipitate the DNA, and then mixed several times by inverting the tube and centrifuging for 5 minutes at 10,000 x g. The resulting pellet was washed twice with 0.75 ml 75% cold ethanol and dissolved in 100 µl of 8 mM NaOH. After the

DNA had dissolved, 10 µl of 0.1 M of 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) was added to adjust the pH to a final pH of 8.

Protocol 3: Kirby Mix Buffer (Covey, 1981)

One milliliter of Kirby mix (1% Sodium dodecyl sulphate (SDS), 6% P-aminosalicylates, 5% phenol and 50 mM of Tris-HCL pH 8.5) was added to the 0.1 g of the ground tissue. The mixture was incubated in a water bath for 30 minutes at 65ºC. The DNA was then extracted using 0.5 ml of chloroform and precipitated with 1/10 volume Sodium acetate (NaOAc) and 1 volume of isopropanol. The resulting pellet was re-suspended with 0.5 ml of TE buffer (10 mM Tris pH 8 and 1 mM EDTA). The resulting DNA mixture was treated with RNase A and re-extracted using the same procedure. The resulting pellet was washed with 70% cold ethanol and then dissolved in 100 µl TE buffer.

Protocol 4: Cetyltrimethylammonium Bromide (CTAB) Isolation Buffer (Stewart, 2004)

Initial trials were performed as described in the protocol (Stewart, 2004). Due to the very low yield and impure DNA obtained, some modification to this protocol were performed to both increase yield and purity and to apply it to a large number of extractions with minimal materials required. The modified CTAB protocol was performed as follows: an amount of 0.1 g of the ground tissue was thoroughly mixed with 500 µl of preheated CTAB buffer (100 mM Tris pH 8, 20 mM EDTA, 2% w/v CTAB, 1.42 M NaCl, 2% w/v polyvinylpyrrolidone (PVP), 5 mM ascorbic acid and 0.2% v/v 2-Mercaptoethanol and 10 µg/ml RNase A). The mixture was incubated at 65ºC

for 60 minutes with occasional shaking by gently inverting the tubes twice every 15 minutes. The DNA was then extracted using 700 µl of chloroform. The DNA was then precipitated using cold 100% ethanol. The resulting pellet was washed twice with 70% ethanol. The pellet was then resuspended in 60 µl of Tris-EDTA buffer (TE).

Protocol 5: ChargeSwitch® gDNA Plant Kit (Invitrogen Life Science Technologies, CS18000)

Young leaves from *Populus*, 5 cm blade length, were collected, ground and 100 mg was used for extracting DNA. The following steps were followed according to the manual provided by the kit. Briefly, 551 µl reagent A ($CaCl_2$, Polyvinylpyrrolidone, ChargeSwitch® Lysis Buffer (L18) and RNase A was added into the frozen ground tissue. The frozen tissue was then thawed in water bath at 65ºC for 5 minutes. The steel balls were then removed from the tube and the mix was centrifuged at 13000 x g for 5 minutes. The supernatant was then transferred into a new centrifuge tube and 50 µl of detergent D1 and 20 µl magnetic beads were added and mixed thoroughly by continuous pipetting up and down. The tube was then placed in the magnetic rack (MagnaRack™). The magnetic beads that the DNA is binding to are pulled to the magnet and the mixture is aspired and discarded. The DNA that bonded to the beads was then washed twice with 500 µl of ChargeSwitch® Wash Buffer (W12). Finally, the DNA was eluted using 75 µl of ChargeSwitch® Elution Buffer.

Determining DNA Concentration and Purity

The concentration and the A260/A280 ratios of the extracted DNA were determined by spectrophotometry. One microgram of the extracted DNA

obtained from each of the protocols were loaded into a 0.8% tris-borate EDTA (TBE) agarose gel, along with 1 µg of *Hind* III digested λ genomic DNA as a reference (ladder), to assess the integrity of the extracted DNA. The gel was run for two hours at 50 volts. The extracted DNAs were then normalized, pooled in pools of two and four and arrayed in a 96-well microtiter plate. The final concentrations of the pooled DNA were 10 ng/µl.

Protocol Used for Extracting DNA from the EMS-mutagenized Poplar Population

Leaves, 30-50 mm long, were used for the extraction. The leaves were washed with water and were placed into scintillation tubes and then were freeze-dried for 48 hours. Six steel balls with different sizes (1 mm, 2 mm and 5 mm) were added to each tube containing the dried leaves. The tubes were placed horizontally into an empty paint container. The tubes were secured in place with pieces of rubber and foam. The assembled paint container was placed in the paint shaker and was turned on for 5 minutes. DNA was then extracted from 0.1 g of the ground leaves as previously described. To further simplify the grinding process we minimized the grinding scale to 15 ml falcon tubes. Each leaf was placed in a 15 ml falcon tube and six steel balls from different sizes (2 of 2 mm, 2 of 1 mm and 2 of 5 mm sized balls) were added to the tube. Liquid nitrogen was then added and the tubes were vortexed at maximum speed for 30 seconds. In most cases, this was sufficient to grind the leaves into fine powder. The frozen ground tissue was then used for the extraction. We determined that this was the most efficient method to use for

grinding, therefore, we used it for extracting DNA from around 5,000 *Populus* plants.

3.3.2 Primers design and Polymerase Chain Reaction (PCR)

Gene model for each of the selected genes was obtained (intron/exon position) by aligning cDNA sequences from the TIGR *Populus* Gene Index (PpIGI) website with the genomic DNA sequence from the Joint Genome Institute (JGI) *Populus trichocarpa* v1.0 website. BioEdit software was used for alignment. Primers were designed to amplify 1 to 2.5 kilo base pair (Kbp) regions using the Primer3 software according to the following parameter: primer length 23-35 bases, melting temperature of 65-70ºC and GC content of ~50%. Unlabeled primers were mainly purchased from Sigma-Aldrich Co. and some from Operon MWG biotech, whereas infrared dye-labeled primers were purchased from Operon MWG Biotech, Inc. Table 3-1 lists the names and sequences of the generated primers and the amplicon sizes.

In order to find the optimal conditions for the Polymerase Chain Reaction (PCR), several sets of PCR were performed at different conditions. Variable annealing temperatures (65ºC to 70ºC) and different amounts of template DNA (500 ng, 50 ng, 10 ng and 5 ng) were tested. All of the PCRs were performed in 15 or 20 µl reactions using different *Taq* polymerase (*Thermus aquaticus*) which included: *TaKaRa Ex Taq*™, Fermentas Taq DNA Polymerase, SharkaTAQ™ and Phusion™ high-fidelity DNA polymerase. Aliquots of 10 µl from each of the PCR products, along with a 5 µl of 0.1 µg/µl 100 bp plus DNA ladder, were

loaded onto a 1.5% agarose gel stained with ethidium bromide and then visualized using ultraviolet (UV) gel documentation.

3.3.3 Optimising TILLING

Several TILLING protocols and kits were tested on the generated EMS mutagenized *Populus*. Table 3-2 summarizes the different protocols that were tested along with a brief comparison.

TILLING using LI-COR machine (Biosciences, NE, USA)

The TILLING steps using LI-COR machine were performed as described by Gilchrist et al., 2006. Infrared dye-labelled primers as well as unlabeled primers were purchased from MWG Biotech, Inc,. The infrared dye-labelled primers had the 5' end of the forward primers labelled with infrared dye700 (IRD700) and the 5' end of the reverse primers labelled with infrared dye800 (IRD800). The primers were dissolved at a concentration of 100 µM in TE buffer (pH 8.0). A mixture of 0.2 µM of 3:2 (labelled:unlabelled) for the forward primers and a mixture of 0.2 µM of 4:1 (labelled:unlabelled) for the reverse primers were used for PCR amplification for TILLING. PCR was performed in a 20 µl reaction with 10 ng genomic DNA in 96-well PCR plates using 0.02 units of *TaKaRa Ex Taq*™ polymerase, 0.2 mM dNTPs and 1.5 mM MgCl2. The PCR program was designed as follows using a BioRad PCR machine: 95°C for 2 minutes, eight cycles of 94°C for 20 seconds, 73°C for 30 seconds (decrementing 1° C per cycle) and 72°C for 1 minute, 45 cycles of 94°C for 20 seconds, 65°C for 30 seconds and 72°C for 1 minute, 72°C for 5 minutes, 99°C for 10 minutes (to denature DNA heteroduplexes and to inactivate the Taq polymerase enzyme),

70 cycles of 20 seconds starting at 70°C and 0.3°C decrease per cycle, for the random re-annealing of the heteroduplexes, followed by a hold at 4°C.

The resulting heteroduplexes were digested with 20 µl of crude celery juice extract (CJE) and a buffer mix (100 mM $MgSO_4$, 100 mM HEPES, 300 mM KCL, 0.02% Triton X-100, 0.002 mg/ml BSA and 0.2% to 0.3% crude celery extract) and incubated at 45ºC for 15 minutes. Homemade crude celery juice extract (CJE) was extracted as described in (Gilchrist et al., 2006) from celery. An aliquot of 2.5 µl of 0.5 M EDTA was added to stop the reaction. The DNA samples were then purified by gel filtration through G50 Sephadex matrix in 96-well Millipore Multiscreen filtration plates and concentrated for 30 minutes at 90ºC followed by separation on a LI-COR sequencing gel with a 0.4 mm thick 96-well sharkstooth comb. Finally, the gel images were analyzed using the GelBuddy software (Henikoff, 2005).

TILLING using Agarose gel Kits (SURVEYOR and Sniper)

PCR reactions were performed in 20 µl reactions with 10 ng extracted DNA, 1.5mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM unlabeled primers and 0.5 Units of Fermentas *Taq* DNA Polymerase or SharkaTAQ™ polymerase (Frontier Genomics). The PCR program was designed as follows using BioRad PCR machine: 95°C for 3 minutes, 35 cycles of 95°C for 30 seconds, 65-68°C for 40 seconds and 72°C for 2 minute, 72˚C for 10 minutes and finally hold at 4°C. The double stranded PCR products were then separated by heating to 99 ºC for 5 minutes and then were allowed to re-anneal slowly as the temperature was decreasing 0.6ºC per cycle from 70ºC to 31ºC. The resulting heteroduplexes

were then digested with SURVEYOR (Transgenomic, Inc.) or Sniper (Frontier Genomics Inc.) endonuclease that were provided by the kits. The steps were performed as specified by each Kit. The digested samples were run on special agarose gel mixes, Stellar™ Gel Mix (Agarose A and Agarose B), that were provided by the Stellar kit from Frontier Genomics Inc. Ethidium bromide (EtBr) dye was added to the agarose mix to stain the DNA bands. UV documentation was used to visualize the gel.

Modified TILLING protocol

In the modified TILLING protocol, PCR reactions were performed in 15 µl reactions with 10 ng extracted DNA, 0.2 µM unlabelled primers, 0.2 mM dNTPs, 1.5 mM $MgCl_2$ and 0.075 units Phusion® High-Fidelity DNA Polymerase (Finnzymes). The PCR reactions were performed in 96-well PCR plates. The PCR program was designed as follows: 98°C for 30 seconds, 35 cycles of 98°C for 20 seconds, 70°C for 30 seconds and 72°C for 40 seconds, 72°C for 5 minutes, and finally hold at 4°C. The resulting double stranded PCR products were then denatured by heating to 99ºC for 5 minutes and then were allowed to re-anneal slowly as the temperature was decreasing 0.6ºC per cycle from 70˚C to 35˚C. The formed heteroduplexes were then digested by adding 10 µl of Sniper DecoMix™ (0.1 µl of Deco mix containing Sniper enzyme and 9.8 µl of 2x reaction buffer) at 45˚C for 1 hour. The reaction was stopped by adding 2.5 µM EDTA. Two microliters of loading dye (40% sucrose and 0.25% bromophenol blue) was added to the digestion mix and 20 µl of the digestion mixture was loaded directly into 1.6% TBE high resolution grade (BioShop

Canada Inc.) agarose gel stained with SYBR® Safe DNA gel stain. The identified possible mutant pools were then independently tested several times for the presence of shorter bands in the pool as well as the individuals that made up the pool. Three positive controls with known point mutations were tested using the modified TILLING protocol before it was used for large scale screening of the EMS-mutagenized *Populus* (Table 3-3).

3.3.4 Sequencing and bioinformatics analysis

The PCR products were purified (QIAGEN PCR purification kit) and sent to MWG Operon for sequencing. To further improve detection of SNPs, we cloned some of the mutants and then sent ten positive colonies for sequencing. TOPO TA Cloning® kit from Invitrogen was used for cloning. The region to be cloned was first amplified using PCR in a 15 µl reaction with 10 ng extracted DNA, 0.2 µM unlabelled primers, 0.2 mM dNTPs, 1.5 mM $MgCl_2$ and 0.075 units Phusion® High-Fidelity DNA Polymerase (Finnzymes). The PCR program was designed as follows: 98°C for 30 seconds, 35 cycles of 98°C for 20 seconds, 70°C for 30 seconds and 72°C for 40 seconds, 72°C for 5 minutes, and finally hold at 4°C. The resulting amplified products were then purified using Qiagen PCR purification kit. An A' overhang was added to the purified PCR reaction by adding 1 unit of Fermentas *Taq* DNA Polymerase and 0.5 mM dATP for 10 minutes at 72˚C. An amount of 2 ul of the A' overhang PCR product was then used for cloning, which was performed following the procedures provided by the TOPO TA Cloning® kit. Plasmids from ten confirmed positive colonies were

then purified using plasmid Qiagen kit and sent to MWG Operon for sequencing.

The sequence data obtained from MWG Operon was then used for further analysis. The first step was to look for double peaks in the sequence's chromatogram, which indicates the presence of SNPs or heterozygous point mutations. FinchTV software version 1.4 (Geospiza, Inc.) was used to view the chromatogram. Sequences from multiple individuals, both mutant and wild type, were aligned using CLC Sequence Viewer 6.3 developed by CLC bio A/S.

3.4 Results

Several DNA extraction protocols were tested in order to find the most suitable protocol for the large number of DNA extractions required from the generated mutant population. Table 3-4 compares the yield and purity of the extracted DNA using the different tested protocols. The DNA extracted using ChargeSwitch® plant kit and the modified CTAB protocol had the highest yields and purity. Upon further analysis, we found that the yield and purity of the DNA extracted using the ChargeSwitch® plant kit were more consistent (Figure 3-1) and the protocol was easier and faster than the other protocols. The extracted DNA was also tested for PCR efficiency and it was found that the DNA extracted using ChargeSwitch® plant kit had the highest PCR quality in terms of yield, reproducible and high purity (Figure 3-1, D). Thus, the ChargeSwitch® plant kit was used for extracting DNA. On average, the DNA was extracted from 24 to 48 individual plants daily.

Based on initial screening, which included 1,700 plants in pools of four, and after experimenting with screening positive controls with known SNPs in pools of various ratios, we concluded that pools of two is most appropriate for our screening and thus most of the screening was then done in pools of two.

The first set of the generated EMS-mutagenized hybrid poplar, which was made up of 1,700 plants grouped in pools of four individuals, were screened for point mutations by TILLING using a LI-COR machine. A total of 40 TILLING plates, each with 96 wells, were prepared and ran using LI-COR machines and were screened for point mutations using the following primers: COMT-1, COMT-2, F5H-1, F5H-2, GAOX-1, GAOX-2, 4CL3-1 and 4CL3-2 (see table 3-1). Seven putative mutant pools were detected during the initial screening: three putative mutant pools in COMT-1, two in COMT-2 and two in 4CL3-1. Figure 3-2 shows a sample of the polyacrylamide TILLING gels (COMT-1 plate1). Two pools in the gel showed two shorter bands that added up close to the full length PCR product indicating that the pools may contain mutant individuals. All seven putative mutant pools (including ones in figure 3-2) were re-screened more than 5 times for confirmation and only one pool from COMT-1 showed the shorter bands during the re-screening process. Thus, we concluded that the other six putative pools were false positives. The DNA was re-extracted from the individuals that made up the re-screened COMT-1 putative pool (except for one individual that was found dead) and the individuals were then screened using TILLING. No shorter bands were detected. We also sequenced each of the individuals, and the sequencing results showed no point

mutations. We have screened more than 3 x 10^7 bases (40 TILLING plates x 96well/plate x 8 chromosomes (pool of 4) x 1000 base per PCR) for point mutations in our mutant population in the V5 hybrid background and no mutations were found. Therefore, we decide to stop further screening of this population.

A larger population of EMS-mutagenized *Populus* consisting of 5,000 individual plants was generated. Before we started screening the generated mutagenized population, different TILLING protocols were tested using positive controls (G/C, GA3 and a positive control from stellar kit). Table 3-1 lists the three TILLING protocols that were tested and the results obtained from each. To increase efficiency and minimize cost, we have designed a modified TILLING protocol that combined steps from the different tested protocols. Figure 3-3 shows the results of tests on positive controls using three different CEL I endonucleases that we have carried out to decide on which one to use with our modified TILLING protocol. Sniper CEL I was chosen for being cheaper than SURVEYOR and more pure than homemade CJE.

Twenty five putative mutants were detected in the first half of screening (of 2,479 samples divided into eight plates and using 13 sets of primers) using the modified TILLING protocol. Figures 3-4, 3-5, 3-6 and 3-7 summarize the steps involved in detecting and analyzing 5 of the 25 putative mutant pools. Each putative mutant pool was re-screened, at least two times, using TILLING. Individuals that make up each of the putative pools were also screened using TILLING. The putative pools as well as the individuals that make up these pools

were then sequenced. In addition, the individuals that make up the putative pools were cloned and then sent for sequencing. In order to distinguish between SNPs caused by EMS and naturally occurring SNPs, individuals generated from calli that were treated with 0 mM EMS from each of the hybrid lines used (K5, K8, K9 and K10) were sequenced. Results of our analysis proved that the SNPs detected were natural pre-existing SNPs in the different hybrid *Populus* lines and did not result from the EMS treatment. The majority of the natural SNPs detected were silent. To eliminate the possibility of detecting natural SNPs present between the different *Populus* lines, we thereafter made sure that further pools were made up from individuals that belong to the same hybrid line. After doing so, four new putative mutants were detected, but, further screening and sequencing revealed that they were false positives. Figure 3-8 shows one of these false positives.

3.5 Discussion

Initial DNA quality is crucial for successful screening by TILLING. The quality and purity of the extracted DNA affects almost all steps involved in screening using TILLING. Therefore, we invested considerable time, money and effort to optimize DNA extraction. We tested five different DNA extraction protocols and compared the results based on yield, efficiency, time and quality of the extracted genomic DNA. The extraction protocols were tested on different samples, using different parts of the plants and from different developmental stages. Samples tested included leaves of fully mature *Populus trichocarpa* trees growing at SFU's Burnaby mountain campus, from mutagenized and un-

mutagenized calli of hybrid poplar lines, and from young leaves from the generated EMS mutagenized poplar trees after they were transferred to SFU's greenhouse. The yield, purity and PCR re-productivity of the extracted DNA obtained from each of the tested protocols were used to determine the protocol with optimal results to be used for the large scale DNA extraction of the generated *Populus* population.

EMS, $C_3H_8O_3S$, chosen as a mutagen in our study, is known to induce point mutations through alkylation, addition of ethyl group, mainly to Guanine (G) nucleotides (Kim et al., 2006). Alkylation changes the binding ability of G nucleotides so that they pair with Thymine (T) instead of Cytosine (C) nucleotides, eventually replacing G/C to A (Adenine)/T (Kim et al., 2006). More than 99% of EMS induced mutations in the majority of organisms that were studied are G/C → A/T transitions that are randomly distributed throughout the genome (Greene et al., 2003; Kim et al., 2006). For example, in *Arabidopsis*, 99% of the EMS-induced mutations were G/C → A/T where 53% were G→A and 47% were C→T on the coding strand (Greene et al., 2003). In some organisms, the percentage of G/C → A/T transitions is slightly lower. For example, in *Caenorhabditis elegans*, Gilchrist et al. (2006) found that 96% of the induced mutations were G/C → A/T and 3% were A/T → T/A and G/C → T/A. The induced base substitution could result in missense, complete knockout (nonsense) or silent mutations (Greene et al., 2003). On average, around 48% of EMS-induced mutations are missense, 3.4% to 5% are complete knockout and 45% are silent (Greene et al., 2003; Kim et al., 2006; Till et al., 2004).

The first population of the generated EMS-mutagenized poplar, made up of around 1,700 individuals, is considered to be relatively small for screening using TILLING and is possibly too small to identify desired mutants, especially in a limited number of genes. Therefore, the next step was to generate a larger mutagenized poplar population. Since we could not find published data regarding EMS mutation frequency in poplar or other forest tree species, we compared the mutation frequencies and corresponding sizes of EMS populations from different diploid organisms (Table 2-1). EMS treatments ranging from 30 mM to 100 mM have been used to induce mutations in a wide range of organisms to generate TILLING populations (Haughn and Gilchrist, 2006). Thus, in generating our mutagenized *Populus* population we have used 25, 50, 100, and 200 mM EMS concentrations. We focused regeneration efforts on calli that survived treatment with 100 and 200 mM as these concentrations lead to high rates of calli lethality, potentially indicating heavy mutagenesis. In plants regenerated from calli exposed to these high concentrations of EMS, we also observed an elevated frequency of morphological phenotypes that are known to be associated with EMS mutagenesis, including leaf variegation (albino) and abnormal leaf shapes (Figure 2-3). As suggested by Haughn and Gilchrist (2006), a typical TILLING population should consist of less than 10,000 plants in the F2 population. The size of the required population depends on the species, ploidy level and EMS dosage (Haughn and Gilchrist, 2006; Martín et al., 2009). For example, the *Arabidopsis* TILLING Project (ATP) screened an EMS-mutagenized population made up of 3,072 *Arabidopsis* plants from the F2

generation, grouped in pools of eight, for point mutations in 1 kb segments of genes of interest (Greene et al., 2003). Other examples of generated TILLING populations include *Caenorhabditis elegans* population made up of 1,500 individuals (Gilchrist et al., 2006), soybean population made up of 768 individual (Cooper et al., 2008) and maize population made up of 750 individuals (Till et al., 2004). Thus, in the second population of our mutagenized *Populus* we generated a relatively larger population consisting of around 5,000 plants.

In spite of all the precautions and signs we used during the generation of the EMS-mutagenized *Populus* population, which included the use of a range of EMS concentrations that reached up to 200 mM (higher than any published data) and screening a fairly large population (more than 5,000 plants), we were unsuccessful in identifying or confirming mutations in the selected genes. Since we have evidence, both in form of detection of positive controls, as well as detection of natural SNPs, we have reason to believe that we were unable to detect induced mutations not because of a pronounced problem in our screening technology, but rather because of a low frequency of mutations in the desired genes in the population. This is rather surprising, since the EMS had strong effect on cell viability, which usually correlates well with high frequency of mutations. Since little is known about the effects of mutagenesis in calli cultures, it is possible that calli somehow responds to EMS differently from seeds, and that a loss of viability is a poor indicator of mutagenesis in calli. There are a limited number of studies on applying EMS to calli cultures, including lavender (Falk et al., 2009) and sweet potato (Luan et al., 2007). They

detected, based on abnormal phenotypes, very low frequencies of mutants. In this context, however, phenotypic analysis is an inefficient tool as most mutant alleles will be recessive, and therefore not result in a detectable phenotype in heterozygous mutant plants regenerated from mutagenized calli. Mutagenizing relatively big clumps of calli might have led to heavily mutagenizing the cells in the surface hindering their growth and the generation of shoots. This in turn might have protected the inner cells from being heavily mutagenized and thus were more vigorous and generated shoots much faster and thus were more dominant in the generated population. One way to overcome this effect is to cut the calli into much smaller sizes or to use suspension tissue culture. Another reason that might have lead to diluting the effect of mutagenesis was that the generated shoots from calli may have arisen either from one single cell or from a group of neighbouring cells. In the case where the shoot arose from multiple cells, there is a likely chance that the generated plant might have consisted of both mutated and un-mutated cells, thus might have generated chimera. Screening and isolating point mutations in chimeric plants is very difficult. One way to overcome all of these difficulties is to use cell suspension cultures. In cell suspension culture, calli can be agitated to separate the clumps into single cells or a group of few cells in liquid media (Evans et al., 2003). Mutagenesis can then be performed in liquid media and the mutagenized cells can then be propagated into callus. Another possible mutagenesis strategy might be to mutagenize protoplasts, which are individual plant cells lacking cell wall. The cell wall of the plant's cells can be hydrolyzed mainly through the use of a mix of

enzymes, including cellulases, hemicellulases and pectinase, that digest the cell wall (Evans et al., 2003). The mutagenized protoplasts can then be stimulated to re-synthesize cell wall and undergo further cell division generating callus (Evans et al., 2003). To our knowledge, this strategy has not been previously applied in poplar trees.

Many researchers working with EMS treatments have had difficulties in generating TILLING populations, (Jim Mattsson, personal communication). Besides EMS as a mutagen, several groups of researchers used other alkylating agents for mutagenesis, including Sodium Azide (NaN_3) and Methyl Nitrosourea (MNU). For example, Suzuki et al. (2008) generated a TILLING population of *Japonica* rice that was mutagenized using MNU. Their screening showed that mutation frequency was higher using MNU, where MNU resulted in one mutation in every 135 kb while EMS resulted in one mutation in every 295 kb (Suzuki et al., 2008).

The modified TILLING protocol used for screening the EMS-treated *Populus* population proved to be effective in detecting SNPs. All of the detected SNPs turned out to be natural SNPs that are present between the different hybrid poplar lines used in our study. In addition, several SNPs detected were natural SNPs found in the same individual due to the fact that each of the lines used was hybrid. Members of the genus *Populus* are known to have high levels of heterozygosity. This is mainly due to the nature of fertilization, which is typically through wind pollination, and their dispersal over large geographic areas (Dickmann et al., 2001; Ingvarson, 2010). For example, *Populus*

tremuloides posses a high level of genetic variation compared to other plants (Ingvarson, 2010). In addition, interspecies crosses carried out by humans have generated a large number of hybrid clones that are being used for breeding, plantation and cultivation (Dickmann et al., 2001). In a study done by Gilchrist et al. (2006), where they screened a natural population of *Populus trichocarpa,* which consisted of trees collected from Northern British Columbia through southern Oregon, they identified on average of one naturally occurring SNP every 130 bp. Thus, it is not surprising for us to find high level of naturally occurring SNPs that are present in the different hybrid poplar lines used in our study. These natural SNPs might be used as markers for future applications such as for marker assisted breeding.

3.6 Figures

Figure 3-1 Comparison between five different DNA extraction protocols tested for DNA extraction from *Populus* based on yield, purity and PCR productivity, showing that the ChargeSwitch® kit provided the best results.

Five different DNA extraction protocols were tested on samples of *Populus* obtained from the EMS-mutagenized population generated by clonal micropropagation. One microgram (µg) extracted DNA from each extraction protocol was loaded into 0.8% TBE agarose gel stained with ethidium bromide (A and B). The gel in (A) was loaded as follows: lambda λ (digested with hind III) in lane 1 and 1 µg of the extracted DNA using CTAB in lanes 2 and 3, DNAzol in lanes 4 and 5, mini-scale CTAB in lanes 6 and 7, modified CTAB in lanes 8 and 9 and Kirby mix in lanes 10 and 11. The arrow in (A) refers to 23 kbp bands under all protocols except under DNAzol in lanes 4 and 5. The gel in (B) is a 0.8% TBE agarose gel loaded with 1 µg of extracted DNA form 16 random *Populus* individuals obtained using ChargeSwitch® kit along with 100 bp plus DNA ladder. The arrow in (B) corresponds to a 3 kbp band of 100 bp plus DNA ladder showing that all bands from the 16 *Populus* individuals are higher than 3 kbp (around 23 kbp). (C) Shows a sample of the NanoDrop reading of one of the samples extracted using ChargeSwitch® kit, which shows a peak at 260 nm wavelength with a A260/280 ratio of 1.89, reflecting the presence of pure DNA, and yield of 0.7 ng/µl, which is considered to be of good quantity. The gel in (D) shows aliquots of 5 µl of the PCR using 5 ng of the extracted DNA from 5 random *Populus* individuals using ChargeSwitch® kit as template loaded into a 1.5% TBE agarose gel stained with ethidium bromide along with 100 bp plus DNA ladder. The arrow in (D) refers to 1.2 kbp band and shows the presence of one thick distinct band in each lane with the right amplicon size, 1.2 kbp, which proves that the template used is of good quality and as a result is suitable for PCR amplification.

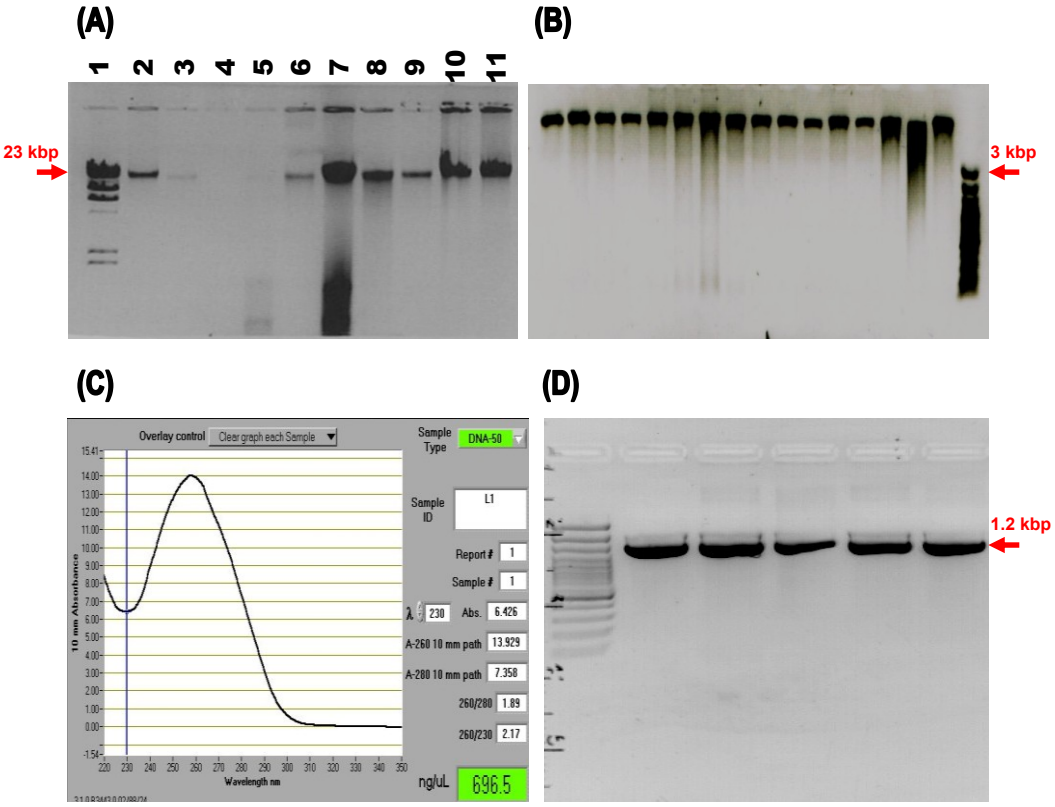


Figure 3-1 Comparison between five different DNA extraction protocols tested for DNA extraction from *Populus* based on yield, purity and PCR productivity, showing that the ChargeSwitch® kit provided the best results.

Figure 3-2 Images of a polyacrylamide TILLING gel, obtained using LI-COR machine, from the first population of the EMS-mutagenized *Populus* (COMT-1 plate1 containing 96 pools, with four individuals in a pool), showing two possible mutant pools in lanes 60 and 72.

The figure shows images of one polyacrylamide TILLING gel, from around 40 TILLING gels, obtained using LI-COR machine and analyzed using the GelBuddy software. Image (A) is obtained from the IRD700 channel and image (B) from the IRD800 channel provided by the LI-COR machine. The presence of a mutant pool is detected by the presence of two or more distinct shorter bands that appear in either of the two channels and where the sizes of the bands from both channels add up to the full length (uncut) PCR product. In lane 60 of the gel (A and B), image from the IRD700 channel (A) shows a distinct band of around 350 bp in size (red box) while image from the IRD800 channel (B) shows a distinct band of around 700 bp in size (blue box). When we add up the sizes of the distinct shorter bands in lane 60 from both channels we get 1050 bp (350 bp + 700 bp), which is fairly close to the full length (uncut) PCR product of around 1010 bp (for COMT-1), indicating that the pool (2A) in lane 60 may contain mutant individuals. The same results are also observed in lane 72 of the gel (A and B), where the sizes of the distinct shorter bands in lane 72 from both channels (A and B) fairly add up close to the full length (uncut) PCR product, indicating that the pool (12B) in lane 72 may contain mutant individuals.

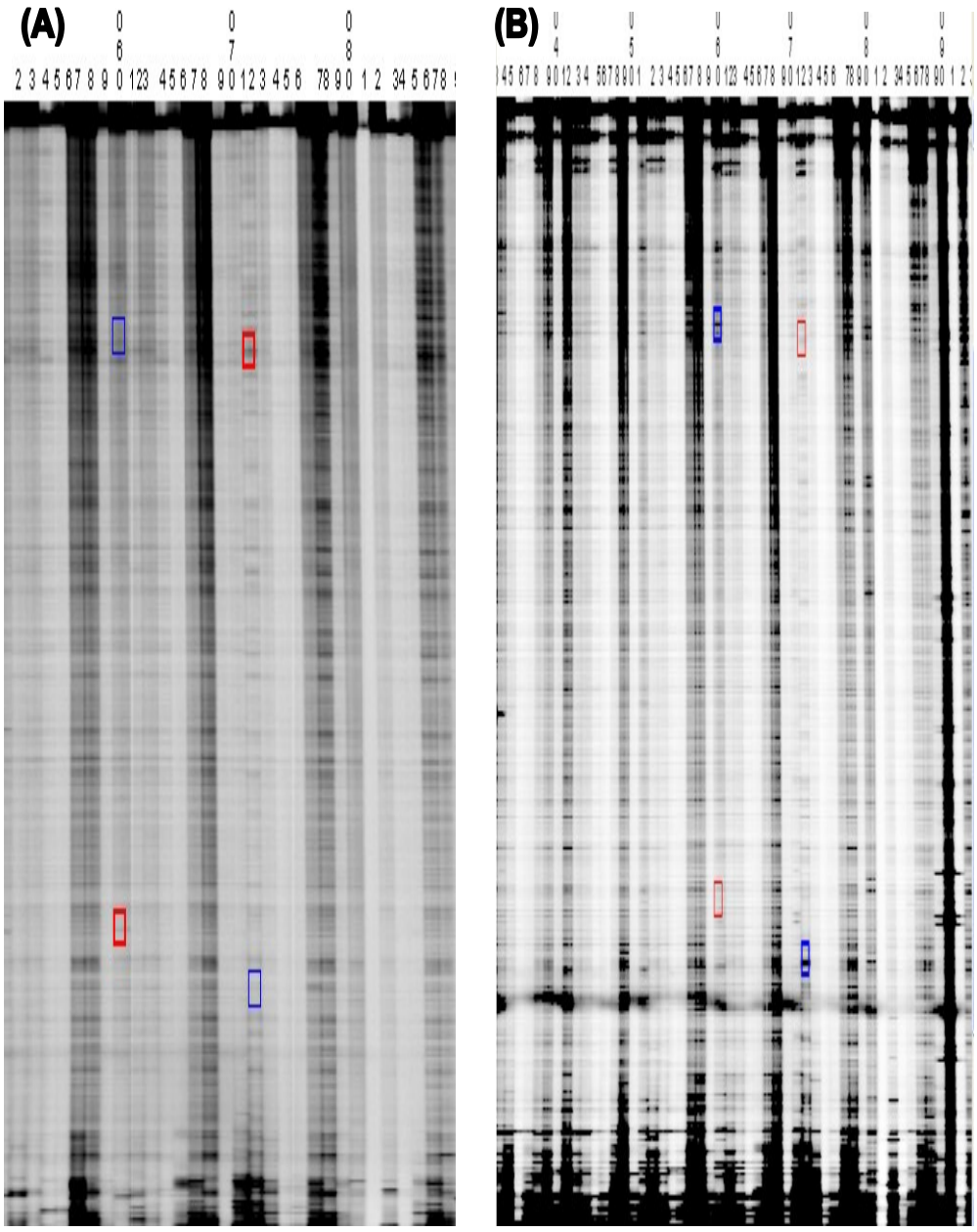


Figure 3-2 Images of a polyacrylamide TILLING gel, obtained using LI-COR machine, from the first population of the EMS-mutagenized *Populus* (COMT-1 plate1 containing 96 pools, with four individuals in a pool), showing two possible mutant pools in lanes 60 and 72.

Figure 3-3 Using positive controls with previously known point mutations to design a modified TILLING protocol, that is both more efficient and economical, to be used for the screening of the second population of the EMS-mutagenized *Populus,* by testing its ability to detect pre-existing point mutations in the positive controls

Celery endonuclease (CEL I) from three different sources, SURVEYOR (A), homemade (B and C) and Sniper (D), were tested for their ability to induce double strand cuts and the shorter bands were visualized using high resolution gels (A, B, C and D). Three positive controls were used: G/C, GA3 and positive control from a stellar kit (+ve). All four gels (A, B, C and D) are 1.6% high resolution grade (BioShop Canada Inc.) TBE agarose gels stained with SYBER safe and loaded with 100 bp plus ladders. The samples GA3 1:1 and GA3 1:3 were made by mixing PCR products from homozygous mutant with homozygous wild type in 1:1 and 1:3 ratios, while the samples Mix 1:1 and Mix 1:3 were made by mixing the template DNA, before PCR, from homozygous mutant and homozygous wild type in1:1 and 1:3 ratios. G was used as a negative control (A and D). The three celery endonucleases used were able to induce double strand cuts and the shorter bands (A, B, C and D) did add up to the full size of the undigested PCR product. Even though the results using SURVEYOR CEL I (A) were the clearest of the three, we decided to use the Sniper CEL I (D) after further optimization due to its low cost and it was used in further screening using TILLING (our modified TILLING protocol) .

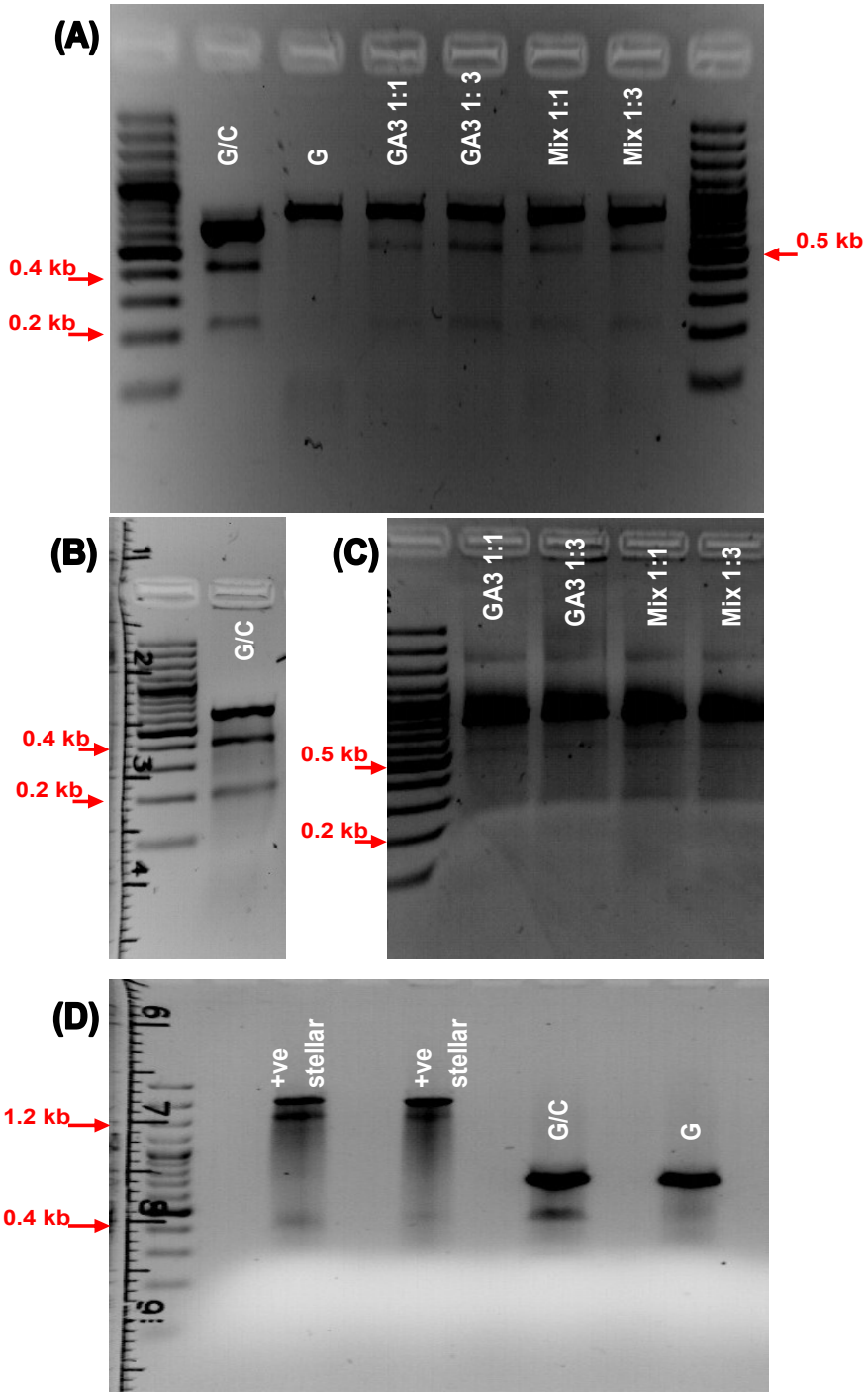


Figure 3-3 Using positive controls with previously known point mutations to design a modified TILLING protocol, that is both more efficient and economical, to be used for the screening of the second population of the EMS-mutagenized *Populus,* by testing its ability to detect pre-existing point mutations in positive controls.

Figure 3-4 Two putative mutant pools, 2C and 2H, of the EMS-mutagenized *Populus* from 4CL3-1 plate3 were initially detected using the modified TILLING protocol, but, further analysis proved that they were natural pre-existing SNPs in the different *Populus* lines and did not result from the EMS-treatment.

(A) Image of gel, 4CL3-1 plate3, obtained using the modified TILLING protocol of the EMS-mutagenized *Populus* with two pools showing shorter bands. (B) Close-up view of the gel image containing the two putative mutant pools, pools 2C and 2H. Pool 2C is made up of two *Populus* individuals, 793-K5 and 888-K9. Pool 2H is made up of two *Populus* individuals, 853-K8 and 948-K9. Pools 2H and 2C along with pool-1D, used as a negative control, were re-screened several times. (C) Samples of the images obtained from the re-screens, where the undigested PCR product with one distinct band is shown on the left gel and the digested PCR product on the right. Only pools 2C and 2H had shorter bands after digestion confirming the existence of SNPs. Pools 2C and 2H were sequenced and both showed double peaks in the expected positions, around 300 bp, on the chromatogram. (D) Chromatogram of pool 2C showing the double peaks, pointed by the red arrows, and the locations match the sizes of the shorter bands in the right gel in (C). The same chromatogram results were also obtained from pool 2H. The individuals from each pool were then cloned and 10 positive colonies from each cloned individual were sequenced. (E) Multiple sequence alignment analysis showing the detected SNPs and its effect on the amino acid encoded. The first SNP was AAG→AAA, both encode lysine, while the second SNP was TGC→TAC, which changed the amino acid from Cysteine (with thiol side chain) to Tyrosine (an aromatic amino acid). Both SNPs are located in the first exon. By comparing the multiple sequence alignments with the un-mutagenized original lines (K5, K6, K8, K9 and K10) we found that the first SNP is also common to lines K5, K6 and K8 while the second SNP is also present in line K8. Thus, the detected SNPs are actually natural pre-existing SNPs found in the *Populus* lines that were treated with EMS.

.

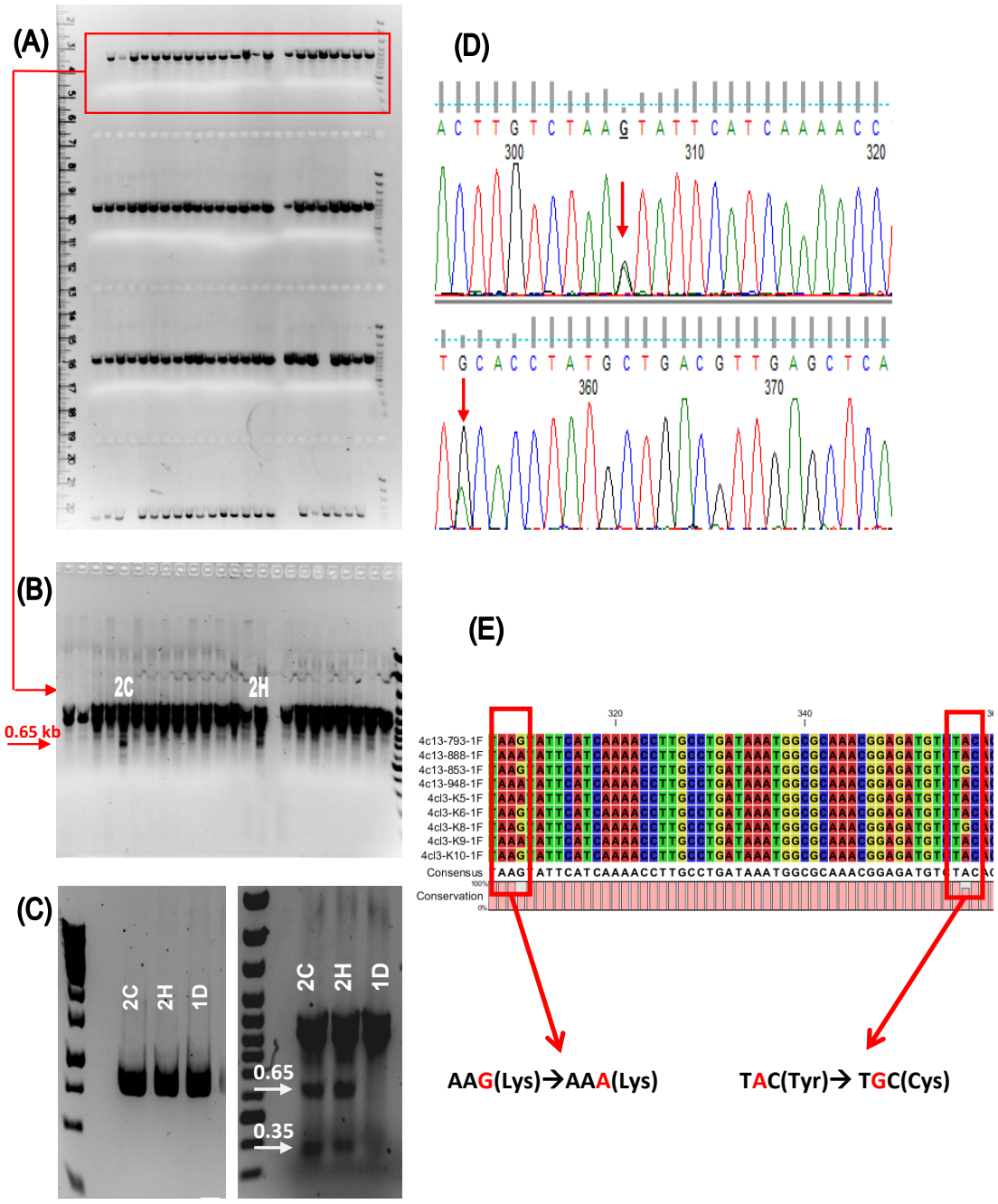


Figure 3-4 Two putative mutant pools, 2C and 2H, of the EMS-mutagenized *Populus* from 4CL3-1 plate3 were initially detected using the modified TILLING protocol, but, further analysis proved that they were natural pre-existing SNPs in the different *Populus* lines and did not result from the EMS-treatment.

Figure 3-5 One putative mutant, 6E, of the EMS-mutagenized *Populus* from CCR2-plate1 was initially detected using the modified TILLING protocol, but, further analysis proved that it was natural pre-existing SNP present in the different *Populus* lines and did not result from the EMS treatment.

(A) Image of gel, CCR2-plate1, obtained using the modified TILLING protocol of the EMS mutagenized *Populus* with one pool, 6E (red circle), showing shorter bands. (B) Close up view of the gel image showing the putative mutant pool, 6E (red arrow). Pool 6E is made up of two *Populus* individuals, 450-K8 and 438-K5. Pool 6E and the individuals, 450-K8 and 438-K5, were re-screened more than eight times. (C) Image of gel from the rescreening of pool 6E, the individual samples, G/C positive control and three negative control (5E, 6D and 6F). Only pool 6E shows the shorter bands after digestion. (D) Picture of plants 450-K8 and 438-K5, which showed that sample 438-K5 was dwarf. Both individuals were cloned and 10 positive colonies from each were sequenced. (E) Chromatogram of pool 6E showing the double peak, highlighted, and the location match the size of the shorter band in (B and C). The multiple sequence alignment of the cloned individuals showed that the same SNP is present but all 10 colonies from each individual were identical. (F) Multiple sequence alignment analysis of original lines showing that K8 has the same SNP detected (G→A). Thus, the detected SNP is actually a natural pre-existing SNP found in the *Populus* line K8. The detected SNP is silent as it encodes the same amino acid.

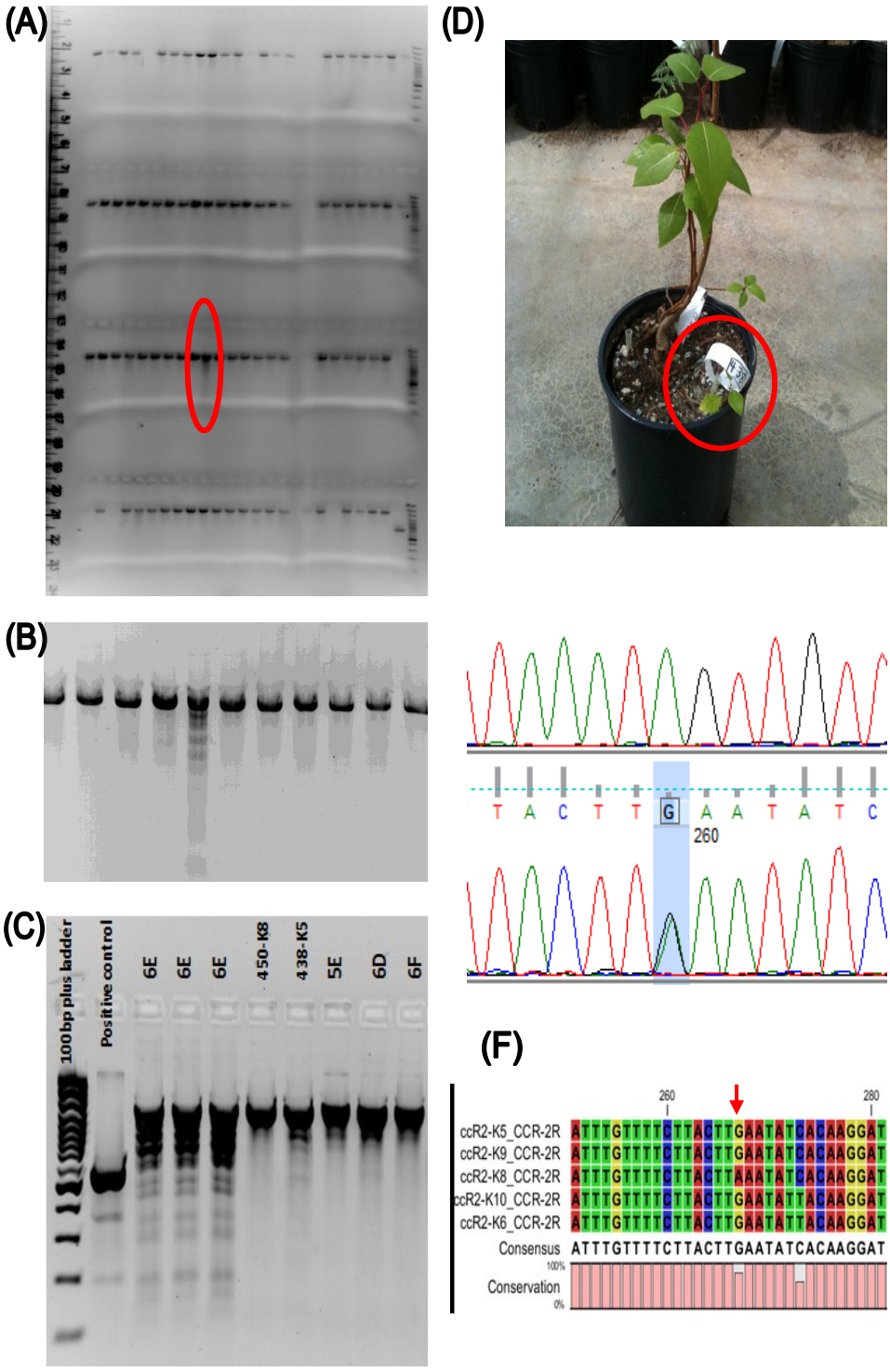


Figure 3-5 One putative mutant, 6E, of the EMS-mutagenized *Populus* from CCR2-plate1 was initially detected using the modified TILLING protocol, but, further analysis proved that it was natural pre-existing SNP present in the different *Populus* lines and did not result from the EMS treatment.

Figure 3-6 One putative mutant pool, 12G, of the EMS-mutagenized *Populus* from CAD1 plate1 was initially detected using the modified TILLING protocol, but, further analysis proved that the SNPs were natural pre-existing SNPs in the different *Populus* lines and in the hybrid *Populus* and did not result from the EMS-treatment.

(A) Image of gel, CAD1 plate1, obtained using the modified TILLING protocol of the EMS-mutagenized *Populus* with one pool (red arrow) showing shorter bands. (B) Close -up view of the gel image containing the putative mutant pool, 12G. Pool 12G is made up of two individuals, 468-K8 and 564-K5. (C) Image of gel from the rescreening of pool 12G, where both undigested and digested PCR products of the pool and the individual samples were loaded to ensure that the shorter bands are due to digestion. Both Mix1 and Mix2 (C) are made up of 1:1 ratio of the two individuals, where Mix1 is before PCR and Mix2 after PCR. (D) Image of gel from rescreening of sample 564-K5 showing the shorter bands after digestion. Pool 12G and 10 colonies from each sample were sequenced. (E) The multiple sequence alignment of the sequenced colonies along with original lines and pool 12G showing that the SNP is due to different lines. In addition, two colonies from 564-K5 had SNPs (A→T) and (T→C) that are similar to line K8 and K10. Thus both SNPs detected are natural pre-existing SNPs, the first is between two different lines, K8 and K5, while the other is most likely due to sample 564-K5 being hybrid. Upon further analysis of other putative mutants in CAD1, the same SNPs were found within samples from the hybrid line K5. Both SNPs are in the upstream region and did not result from the EMS-treatment.

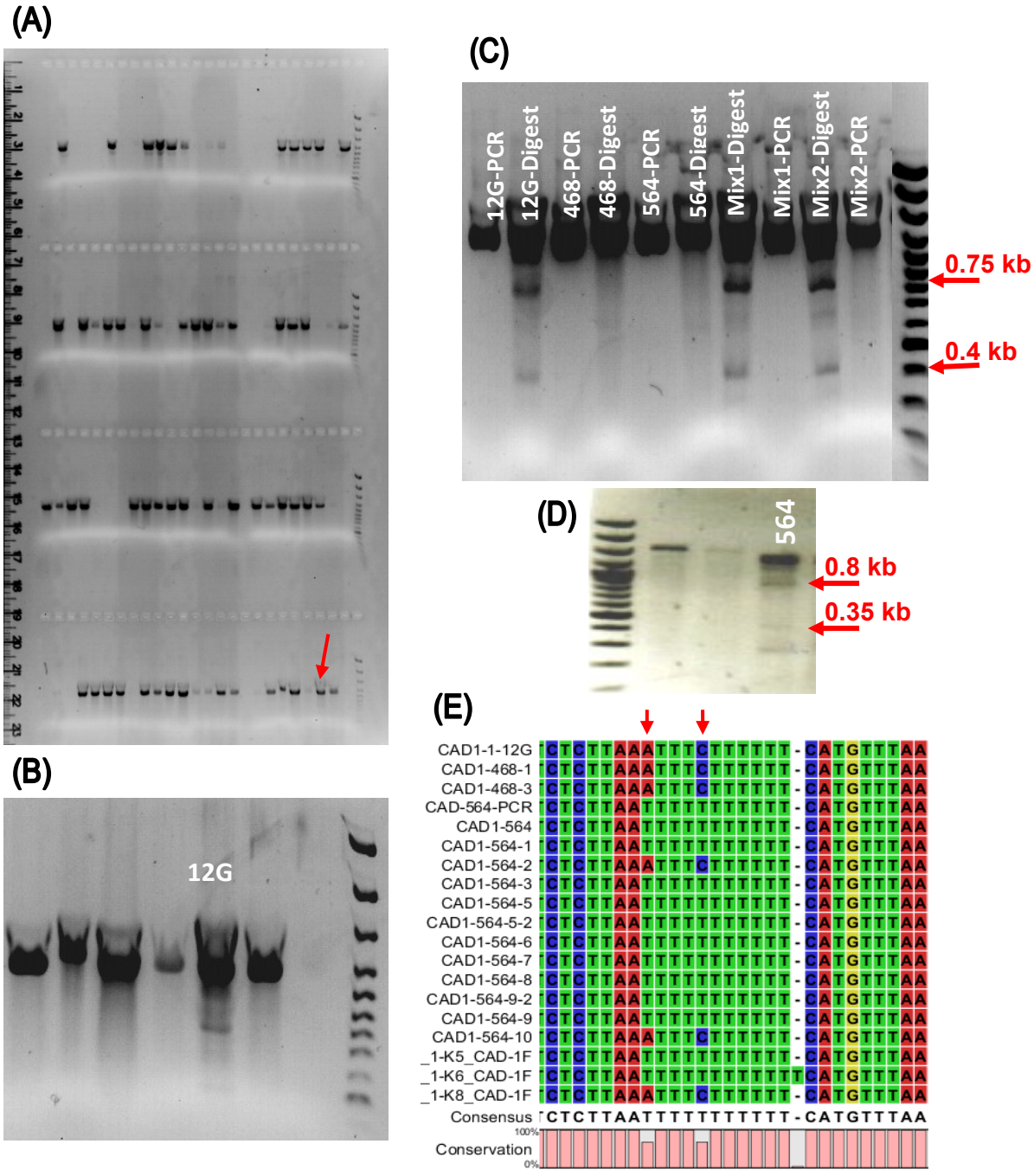


Figure 3-6 One putative mutant pool, 12G, of the EMS-mutagenized *Populus* from CAD1 plate1 was initially detected using the modified TILLING protocol, but, further analysis proved that the SNPs were natural pre-existing SNPs in the different *Populus* lines and in the hybrid *Populus* and did not result from the EMS-treatment.

Figure 3-7 One putative mutant pool, 2A, of the EMS-mutagenized *Populus* from CAD1 plate2 was initially detected using the modified TILLING protocol, but, further analysis proved that the SNPs were natural pre-existing SNPs in the different *Populus* lines and in the hybrid *Populus* and did not result from the EMS-treatment.

(A) Close-up view of the gel image, CAD1 plate2 obtained using the modified TILLING protocol of the EMS-mutagenized *Populus* with one pool (2A) showing shorter bands. (B) Images of gels from the rescreening of pool 2A and the individuals that make up the pool, 578-K10 and 674-K8. Only sample 578-K10 had shorter bands when the individuals were rescreened individually. (C) Picture of plants 578-K10 and 674-K8, which showed that sample 578-K10 was dwarf. Both individuals were cloned and 10 positive colonies from each were sequenced. (D) Section of the multiple sequence alignment of the sequenced colonies along with original lines and pool 2A showing the presence of SNP (T→C) at position 285, which matches the double peak showing in the chromatogram of pool 2A. (E) Another section of the multiple sequence alignment analysis showing that two colonies from 578-K10 had SNPs (A→T) and (T→C) that are similar to line K8 and 674-K8, which also matches the double peaks in the chromatogram of pool 2A (red arrows). Thus both SNPs detected are natural pre-existing SNPs, the first (D) between two different lines, K8 and K10, while the other (E) is most likely due to sample 578-K10 being hybrid. Both SNPs are in the upstream region and had no effect on the amino acids and did not result from the EMS-treatment.

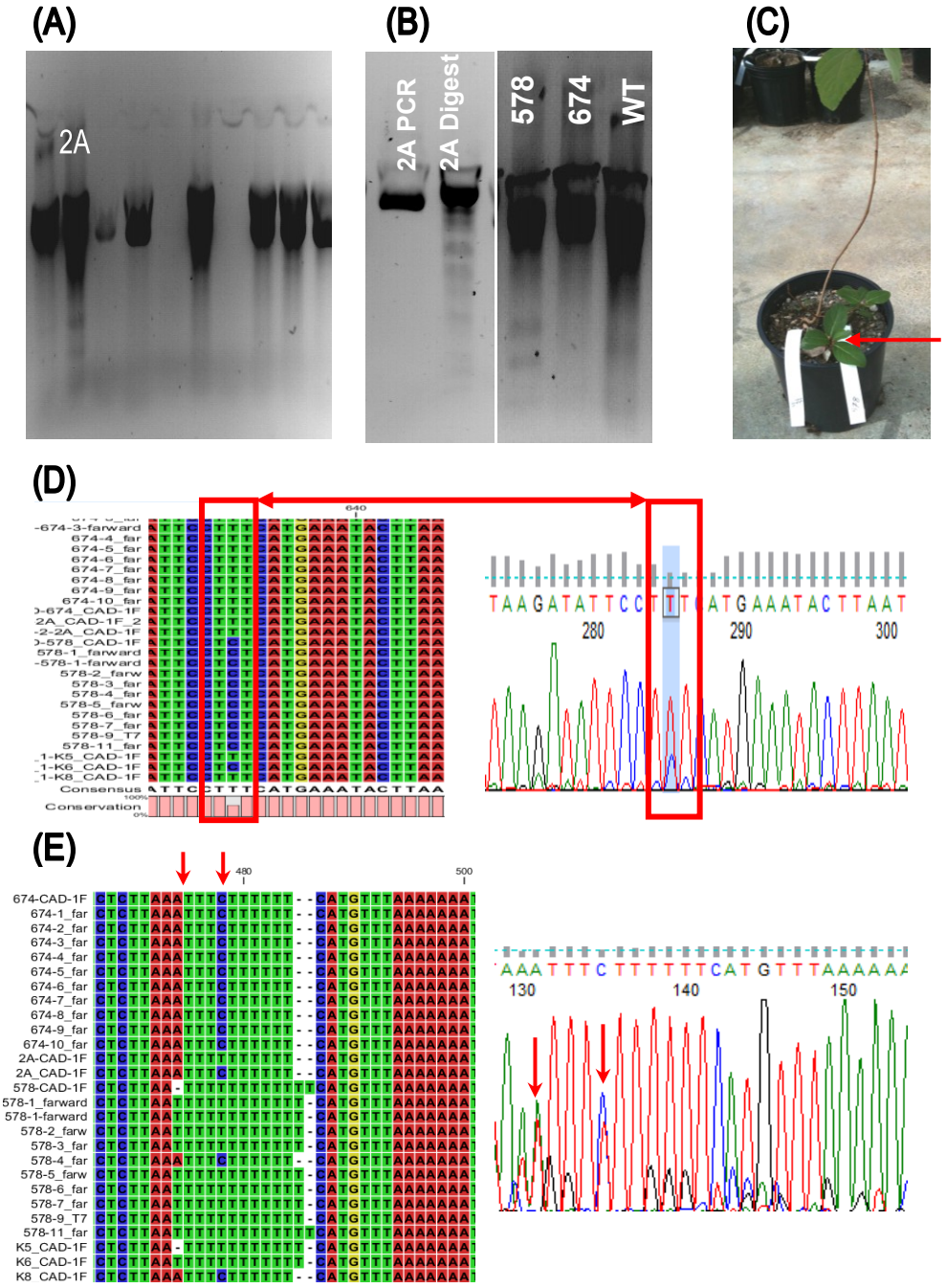


Figure 3-7 One putative mutant pool, 2A, of the EMS-mutagenized *Populus* from CAD1 plate2 was initially detected using the modified TILLING protocol, but, further analysis proved that the SNPs were natural pre-existing SNPs in the different *Populus* lines and in the hybrid *Populus* and did not result from the EMS-treatment.

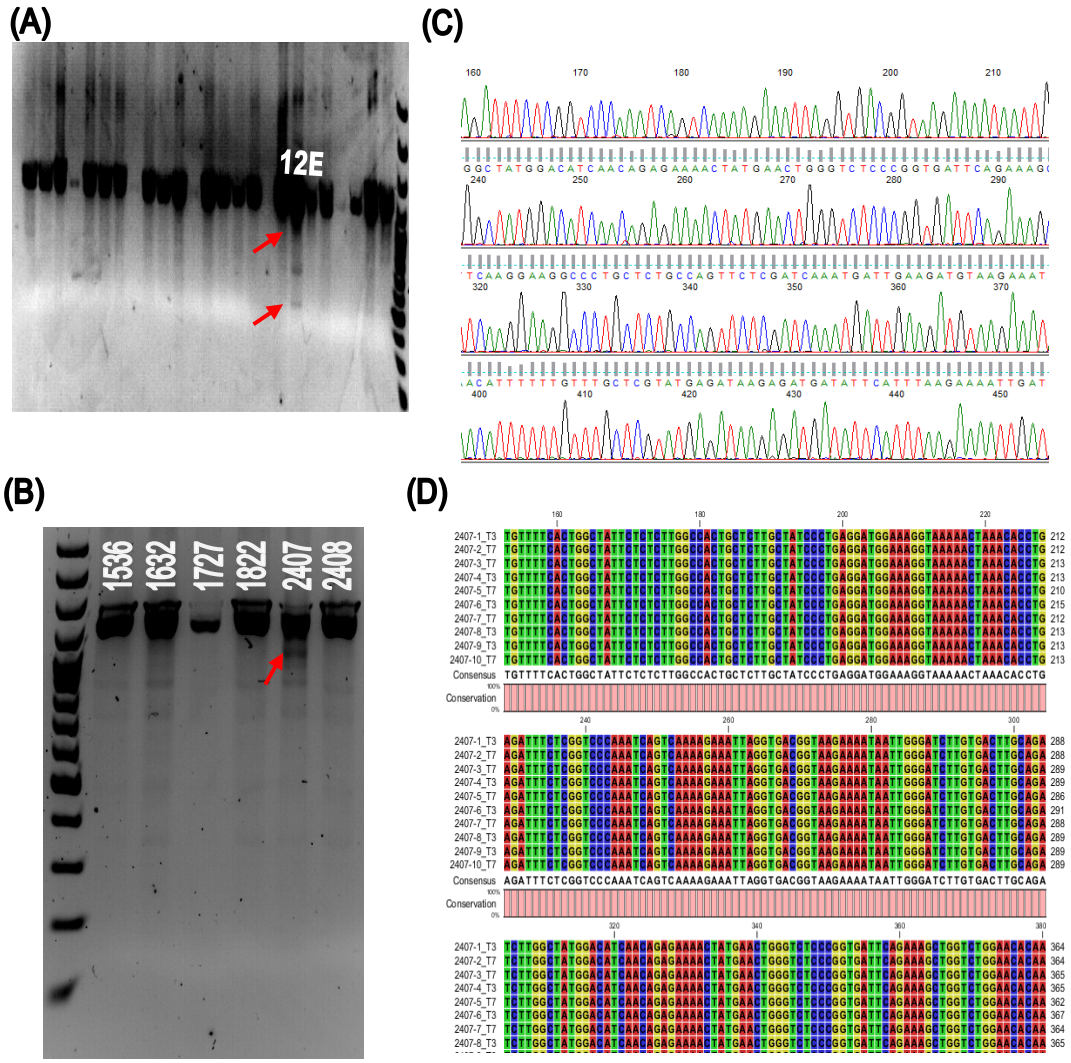


Figure 3-8 One putative mutant pool, 12E, of the EMS –mutagenized *Populus* from CCoAMT2 plate8 was initially detected using the modified TILLING protocol, but, further analysis proved that it was a false positive.

(A) Close-up view of the gel image, CCoAMT2 plate8, obtained using the modified TILLING protocol of the EMS-mutagenized *Populus* with one pool (12E) showing shorter bands (red arrows). (B) Image of a gel from the re-screening of the individuals that make up the pool, 2407-K9 and 2408-K9, along with random samples that belong to K9 (1536, 1632, 1727 and 1822) used as negative controls. Only Sample 2407 had a distinct shorter band (red arrow). (C) Chromatogram of Pool 12E, no double peaks were detected. (D) Multiple sequence alignment of 10 positive colonies from sample 2407, where no SNPs were detected. Thus, the shorter bands detected might be due to PCR artifact and thus pool 12E is a false positive.

3.7 Tables

Table 3-1 List and sequences of primers used for screening the generated EMS mutagenized *Populus* population using TILLING.

Name	Forward Sequence	Reverse Sequence	Amplicon Size (bp)
F5H-1	GGGGTGGGGGAGGTCA CAAAAA	TCGTAAAGCATCCGCACC ATTTAGTGT	1358
F5H-2	CAGCGCATGTGACACCA TCCAATAG	GCAAATGAGCAACAGCCA AATCAAGG	1003
COMT-1	CATCATGTTCCAATAGTC TCCCCTGAAACA	ATGCCCATTTGGACTGGT CGATAAA	1010
COMT-2	CCTTTTTGTCGTGAAACA CTAACCAATCAC	CCCTCAAATTCCTTTTCG GTCCTCTC	1025
SPINDLY	TGGGATCTGATTGAGGG TTGGATGA	CTCGCGCCAACATTGTGA GCATGTA	1354
4CL3-1	TTCACCAAACGCAACCC ATTTTTCA	TGGCATTATCAAAATCGA GGCACCA	1000
4CL3-2	TGCGAAAAGGGGTTGGA AACCAATA	TTTCATCTTCGGTGGCCT GAGACTT	1300
GA2ox-1	ACACCAAATCACACGGG AATATGGAA	TGAGGGTCAGAATGCTCT CCAAATCCT	1011
GA2ox-2	CTTTGACATGACCCTTTG CTCAGTCC	CTGATTGCTCATTGGAGC GCATGAC	1051
CCoAMT-1	CCATCCTGTCATGGAGA AAGATTCC	GGATGCTTGGCAGTCAAC TCTCTAA	1341
CCoAMT-2	CTGATGTGAATGATGGG AAACAGG	GCTGCATGTACAGAAAGT CAGAACG	1211
2CCoAMT-1	GCCAGAGGAGGTGATAT TTTGA	TAGAAATTGGGTGAGGGA TAGGTG	1209
2CCoAMT-2	CAGGTTAGGCATCACTA CCATCATC	GAGCTCCAAAACAAAGTC CCTGTAG	1242

Name	Forward Sequence	Reverse Sequence	Amplicon Size (bp)
CCR-1	TACATCCACTCCACCATG TCTTTCC	CCTTATTGGGATCCATGT ACACAGC	1016
CCR-2	ACCGAGTAACAGGTGAC ACAAAAGC	GGTCAACAAACCTCTGGT GAAACAC	1455
CAD-1	GTCATGGGTTTCAACGT GTTAGTCC	ACTTTGTCACATCTGATC CCACCTC	1275
CAD-2	CAAGGAGGCTTTGCTGA ATCCAT	TAAGCTTGCTACCAGCAA CATCGAC	1263
PAL4	TCACCAAGAATGGCTAC CAAAATGG	TGTAAGGACAGAAAAGCC GTCGAAA	3093
C3H	CTCGCAGAGACCAACTC CATTCCTA	CCCCAATTAACAAAGTGA AGGTTGGA	3038
C4H-1	TTGGGTTCTTTTGTTGCC ATTCTTG	TGCACGTCAACATCACTT GTTTCTT	3436
C4H-2	TTTTTATGTTGTTTTGATG GGTTGA	TTCACATGGTCTAGGGGC AAAGCTA	2208
MYB003-1	CATGATGATTACCTTGTG AGTGAAG	ATAGACAATTCCTTAAAG CCTCTGC	1500
MYB003-2	CACCCATAAATCATAATG TCTACCC	CTAGCTAATCTCCACTTT GCAATTC	1350
MYB192-1	GGTTACTGGAGTGGGAT AGTCTGCA	TTATGGATTTATTCCCCTC CAATCG	1400
MYB192-2	AGAGGACTAGAAGGGAG GGAAGATT	GCCAAAGCTACAGTTTTA AGCCAGT	1200
MYB021-1	CACCAATTATGTGGTCCA TTGAAC	AAAGAACTTGCCTTTGTT CTCCTAG	1500
MYB021-2	CAAAGGCAAGTTCTTTCA GAGAGTA	GATCAGCAGCTATGCTAT CTGTCTT	1300

Table 3-2 Different approaches used in TILLING for screening for point mutations and SNPs in an EMS-mutagenized *Populus* population.
We combined steps from different approaches to finally come up with a modified TILLING approach that gave us the best result for our screening.

TILLING Steps	LI-COR	Agarose gel	Stellar Kit
PCR reaction volume	10 µl	20-50µl	25 µl
Polymerase	Sharka/Phusion	Fermentas *Taq*	Sharka
Primers	IRD labelled	Unlabelled	Unlabelled
Amplicon	~1 kb	Up to 2.5 kb	~1.5 kb
Endonuclease	CJE	SURVEYOR® or Homemade CJE	Sniper
Detection	LI-COR machine	Agarose gel	Special gel mix supplied by the kit
Digest	Single stranded cut	Double stranded cut	Double stranded cut
PCR purification	Sephadex purification	No purification required	No purification required
Gel running time	5 hours/1kb	25 minutes/ 1kb	40 minutes/1kb
Cost	Expensive	Expensive	Relatively cheap

Table 3-3 Positive controls with known one base substitutions.

Control	Substitution	Position	Source	Reference
G/C	G→C	415	Plasmid	SURVEYOR® kit
GA3	G→A	200	DNA from *Arabidopsis*	EMS TILLING population from Seattle TILLING Project
Positive control	G→A	475	N/A	Stellar™ kit

Table 3-4 The yields and ratios of the extracted genomic DNA using different extraction protocols.

Results shown are average of three or more extractions.

Protocol	Yield (mg DNA/g tissue)	Ratio (A260/A280)
GenElute column	7.3×10^{-3}	2.4
DNAzol	0.30	2.0
Kirby mix	0.07	1.7
CTAB	0.57	2.2
CTAB modified miniprep scale (RNase A and 2x washing)	0.75	2.0
ChargeSwitch®	0.70	1.9

3.8 References

Comai, L., and Henikoff, S. (2006). TILLING: practical single-nucleotide mutation discovery. *The plant journal, 45*(4), 684-694.

Cooper, J. L., Till, B. J., Laport, R. G., Darlow, M. C., Kleffner, J. M., Jamai, A., El-mellouki, T., Liu, S., Ritchie, R., Nielsen, N., Bilyeu, K. D., Meksem, K., Comai, L., and Henikoff, S. (2008). TILLING to detect induced mutations in soybean. *BMC Plant Biology, 8*, 9.

Dickmann D., Isebrands J., Eckenwalder J. and Richardson J. (Eds.). (2001). *Poplar Culture in North America.* Ottawa, Canada: National Research Council (Canada) Research Press.

Dong, C., Dalton-Morgan, J., Vincent, K. And Sharp, P. (2009). A Modified TILLING Method for Wheat Breeding. *The Plant Genome, 2*(1), 39-47.

Gilchrist, E.J., O'Neil, N.J., Rose, A.M., Zetka, M.C., Haughn, G.W. (2006). TILLING is an effective reverse genetics technique for *Caenorhabditis elegans*. *BMC Genomics, 7*, 262.

Gilchrist, E. J., Haughn, G. W., Ying, C. C., Otto, S. P., Zhuang, J., Cheung, D., Hamberger, B., Aboutorabi, F., Kalynyak, T., Johnson, L., Bohlmann, J., Ellis, B. E., douglas, C. J. and Cronk, Q. C. B. (2006). Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa*. *Molecular Ecology*, 15(5), 1367–1378.

Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., et al. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics*, *164*(2), 731-40.

Evans, D.E., Coleman, J.O.D., and Kearns, A. (2003). *Plant Cell Culture*. London: BIOS Scientific Publishers.

Falk, L., Biswas, K., Boeckelmann, A., Lane, A., and Mahmoud, S. S. (2009). An Efficient Method for the Micropropagation of Lavenders: Regeneration of a Unique Mutant. *The Journal of essential oil research*, *21*(3), 225-228.

Haughn, G.W., and Gilchrist E.J. (2006). TILLING in the Botanical Garden: a reverse genetic technique feasible for all plant species. In: Floriculture, Ornamental and Plant Biotechnology: Advances and Topical Issues (1st Edition), JA Teixeira da Silva (Ed), Global Science Books, Ltd., London, UK.

Ingvarson, P.K. (2010). Nucleotide Polymorphism, Linkage Disequilibrium and Complex Trait Dissection in *Populus.* In Jansson, S., Bhalerao, R., and Groover, A. (Eds.), volume 8, Plant Genetics and Genomics: Crops and Models (pp. 91-111). London: Springer.

Kim, Y., Schumaker, K.S., and Zhu, J.K. (2006). EMS mutagenesis of *Arabidopsis*. In Salinas, J. and Sanchez-Serrano, J. (Eds.), *Methods in Molecular Biology, vol 323: Arabidopsis Protocol*, Second edition (pp. 101-103). Totowa, NJ: Humana Press Inc.

Luan, Y.-S., Zhang, J., Gao, X.-R., and An, L.-J. (2006). *Plant Cell, Tissue and Organ Culture*, *88*(1), 77-81.

Martín, B., Ramiro, M., Martínez-Zapater, J. M., and Alonso-Blanco, C. (2009). A high-density collection of EMS-induced mutations for TILLING in *Landsberg erecta* genetic background of *Arabidopsis*. *BMC plant biology*, *9*, 147.

Suzuki, T., Eiguchi, M., Kumamaru, T., Satoh, H., Matsusaka, H., Moriguchi, K., et al. (2008). MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. *Molecular genetics and genomics, 279*(3), 213-23.

Till, B. J., Cooper, J., Tai, T. H., Colowit, P., Greene, E. A., Henikoff, S., et al. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol.*, *11*(1), 7-19.

Till, B. J., Zerr, T., Bowers, E., Greene, E. a, Comai, L., and Henikoff, S. (2006). High-throughput discovery of rare human nucleotide polymorphisms by Ecotilling. *Nucleic acids research*, *34*(13), e99. doi: 10.1093/nar/gkl479.

Wilkins, O., Nahal, H., Foong, J., Provart, N. and Campbell, M. (2009). Expansion and Diversification of the Populus R2R3-MYB Family of Transcription Factors. *Plant physiology, 149*, 981-993.

CHAPTER 4: ASSESSING HAPLOINSUFFICIENCY IN GENES INVOLVED IN LIGNIN BIOSYNTHESIS IN *Arabidopsis thaliana*

4.1 Abstract

It is increasingly clear that many genetic diseases in humans are based on a genetic phenomenon known as haploinsufficiency. Haploinsufficiency occurs when one copy of a gene cannot fully make up for the loss of the second copy in a diploid genome. Haploinsufficiency could provide a novel avenue for breeding of certain plant species. For example, in poplar trees, haploinsufficiency in the heterozygous F1 plants may result in desirable phenotypes without waiting for the 10-20 years that it would take to generate homozygous mutants. At the same time, poplar trees are easily propagated vegetatively by stem cuttings, avoiding the need for homozygosity to maintain the desired trait. Here we have initiated a study to assess whether haploinsufficiency can be observed in genes that regulate lignin content, a trait of interest for the biofuels and pulp and paper industry. We were able to generate heterozygous populations of *Arabidopsis* plants from lines that are mutated in genes that are homologous to the genes selected in our *Populus* study. Based on lignin quantification, results obtained to date showed that mutation in the CAD-1 gene did not result in haploinsufficiency. Based on morphological defects, GA3, *IRX4*, *TT5-1, GA5-1* and *FAH1-5* genes did not

result in haploinsufficiency. Further lignin quantification of plant with mutations in these genes is still to be carried out.

4.2 Introduction

There are weekly reports associating heritable human diseases with genetic lesions in specific genes. Among them, an unexpected number of mutant alleles cause disease in heterozygous rather than homozygous form. This effect, where one gene copy does not fully make up for the loss of the second gene copy in a diploid individual is known as haploinsufficiency. Upon closer scrutiny, it is often revealed that homozygosity results in stronger defects than heterozygosity, including embryo lethality (Alkuraya et al., 2006; Dang et al., 2008; Veitia, 2005). In fact, it has been claimed that most human genetic autosomal disorders have a haploinsufficiency component (Dang et al., 2008). Smith–Magenis syndrome, Ehler-Danlos syndrome, Retinitis Pigmentosa, Sotos syndrome and Williams syndrome are among many human genetic disorders that show haploinsufficiency (Dang et al., 2008; Veitia, 2005). To date, there are 299 haploinsufficient gene loci identified in the human genome (Dang et al., 2008).

Haploinsufficiency has also been detected in plants (Deutschbauer et al., 2005; Krieger et al., 2010). Examples of haploinsufficiency in plants include the *ERECTA-LIKE 2 (ERL2)* gene in *Arabidopsis,* where plants with one copy had less severe defects in flower and ovule development (Pillitteri et al., 2007), and the *SINGLE FLOWER TRUSS* gene in tomato, where heterozygotes had 60% increase in yield (Krieger et al., 2010). In plant breeding, mutant alleles that

result in desirable changes are usually fixed in homozygous conformation in order to avoid segregation of the trait in subsequent generations. Similarly, studies assessing gene functions in plants tend to focus on the strong phenotypes of homozygous mutant individuals, usually ignoring a potential weaker phenotype in heterozygous individuals. As a consequence, the extent to which haploinsufficiency exists also in plants is unknown. There are situations, however, where haploinsufficiency could be useful for plant breeding. Fruit trees are tedious to breed as they have long generation times. Similarly, *Populus* and *Eucalyptus* species, used extensively across the world as fast-growing sources of wood fiber, also have long generation times that prevent the generation of homozygous mutants in a timely manner. In addition, *Populus* species are generally dioecious, which in practice puts the generation of homozygous individuals beyond the scope of most breeding programs. In these cases, haploinsufficiency in key loci may provide sufficient alteration of traits of interest. For example, *Populus* and *Eucalyptus* varieties with modest reductions in lignin content due to haploinsufficiency may be of considerable interest to paper as well as biofuels industries. Identified heterozygous individuals could essentially be deployed immediately. Since both *Populus* and *Eucalyptus* species can, and often are, propagated vegetatively, the heterozygous condition will be maintained.

Here we analyzed the effect of haploinsufficiency on several genes that are involved in lignin biosynthesis. Previous studies showed that complete loss of function in most of the studied genes involved in lignin content had negative

side effects on the plant (Leplé et al., 2007; Li et al., 2010; Turner and Somerville, 1997) that often out-weighed the benefits obtained from decreasing or altering lignin content. Lignin provides much of the strength required for upright growth of plants, as well as the strength required of tracheary elements to withstand the strong tension caused by the transpiration stream. The chemically resistant lignin is also important for decay resistance (Somerville et al., 2010; Umezawa, 2010). Therefore, completely disrupting lignin biosynthesis or altering its composition would have negative effects, beside the benefits that could be gained from decreasing energy and chemicals needed for wood processes or bioenergy production. Our hypothesis is that instead of completely knocking out both copies of genes of interest, we will knock out one of the two functional copies and analyze its effect on lignin biosynthesis, by analysing heterozygous plants. To assess this hypothesis, we have initiated a study to assess haploinsufficiency in *Arabidopsis thaliana*, as mutants defective in lignin biosynthesis genes are readily available in this species.

4.3 Materials and Methods

4.3.1 Generation of Mutant *Arabidopsis* Population

Seeds from several lines of mutant *Arabidopsis* plants that are known to either have T-DNA inserts or EMS mutations in genes of interest were obtained from The *Arabidopsis* Biological Resource Centre (ABRC). Table 4-1 provides a complete list of the obtained lines with the type and location of the mutation. Seeds were gas sterilized in a closed chamber where a mixture of 100 ml of

bleach and 3 ml hydrochloric acid (HCL) was used for gas sterilization for two hours. The sterilized seeds were then plated on an *Arabidopsis thaliana* salt (ATS) media and were placed at 4˚C for three days for cold treatment. ATS media was prepared as described by Lincolin et al. (1990). The plates were then transferred into a growth chamber (16 hours day light and 8 hours night, 70% humidity and temperature of 24˚C) for 5-7 days for seeds to germinate. The seedlings were then transferred into pots and were allowed to grow in the growth chamber for 4 month to reach maturity and finally dry out. After the plants completely dried out, the seeds were collected and then used to generate an F2 population. During early stages of seedling growth, at around the fourth week of growth, the first or second cauline leaves were used for DNA extraction to genotype the generated plants.

4.3.2 Genotyping and genetic crosses

Genomic DNA was extracted as described in Dellaporta et al. (1983) with minor modification. Briefly, the first or second cauline leaf was placed in a 1.5 ml centrifuge tube. An amount of 100 µl of extraction buffer (200 mM Tris-CL pH 8, 250 mM NaCl, 25 mM EDTA and 0.5% SDS) was added into each tube. Cauline leaf was ground using plastic pestle that was attached to an electric drill. After grinding, 400 µl of the extraction buffer was added (total volume 500 µl) and then centrifuged at 3,000 rpm for 10 minutes. The supernatant was then transferred into a new 1.5 ml centrifuge tube and 500 µl of cold 100% isopropanol was added to precipitate the DNA and then centrifuged for 15

minutes at maximum speed. The pellet was then washed with 500 µl of cold 70% ethanol and the DNA pellet was then suspended with 100 µl Tris buffer.

Genotyping T-DNA mutant plants was performed as described by Salk Institute Genomic Analysis Laboratory (SIGnAL) by using specifically designed primers that amplify the T-DNA insert and the gene of interest using PCR. In addition to PCR amplification, antibiotic and herbicide resistance of plants carrying T-DNA inserts was also used to genotype the T-DNA mutant plants. *Arabidopsis* plants were plated on ATS media supplemented with 50 µg of Kanamycin or 25 mg of BASTA depending on the resistant gene present in the T-DNA insert. Wild type homozygous plants that lack T-DNA inserts also lack the resistant gene and thus grow very weak, yellow coloured, due to the presence of antibiotic or herbicide in the media, as shown in figure 4-1. Heterozygous and homozygous mutant plants will grow normally. EMS mutated plants were genotyped using the modified TILLING protocol as described in chapter 3 of this thesis or genotyped by sequencing. Plants generated from ga5-1 and GA3 lines were genotyped using the modified TILLING protocol, as shown in figures 4-2 and 4-7(A).

Some of the ordered *Arabidopsis* lines were only homozygous mutants, while others were heterozygous. Heterozygous plants were directly used to generate the F2 population, where it should generate a population composed of 25% homozygous mutant, 25% homozygous wild type and 50% heterozygous plants. Lines that had only homozygous plants were crossed with homozygous wild type, either *Columbia-0* (CS60000) or *Landsberg erecta-20* (CS20)

depending on the lines background. Pollen from wild type plants was used to pollinate homozygous plants that had their stamens removed before fertilization. The resulting heterozygous seeds were then used to generate the F2 population. The generated F2 population was then used for molecular and phenotypic screening.

4.3.3 Molecular and phenotypic screening

The generated F2 populations were screened throughout their developmental stages for any morphological abnormalities. Morphological phenotypes that were screened for included plant's height, silique sizes, seed colour, posture and unusual flowers or leaves. The total heights of the plants were measured using ruler and the siliques' sizes were measured using calibre. The colour of the seeds was analyzed both visually and by taking pictures using a Canon digital camera set to magnify 3.5 times, with maximum light and speed of detection of ISO-100.

Molecular screening of the generated *Arabidopsis* plants involved measuring lignin content and composition. We tried several approaches in order to find the most effective and efficient measurement method. Few samples, consisting of homozygous wild type, homozygous mutant and heterozygous plants from the CCR lines, were first sent to the Canadian Light Source Inc. at the University of Saskatchewan to help in analyzing total lignin content. Dried *Arabidopsis* stems were ground into fine powder using liquid nitrogen and Ultra Turrax and then freeze-dried for four days. The dried powder was then sent to The Canadian Light Source Inc. for lignin analysis using infrared

spectromicroscopy and soft X-ray spectromicroscopy. Another set of samples were also sent to the Pulp and Paper Research Institute of Canada (PAPRICAN) for lignin analysis using acetyl bromide-based method. Micro-scale quantification lignin was performed on three homozygous wild type plants to test whether the technique used by PAPRICAN to analyze lignin in woody tissue can be also applied to *Arabidopsis*. Lignin analysis was performed as described in Chang, et al. (2008). Lignin was extracted from each sample and lignin content was measured in triplicates. In addition to using the traditional chemical approach in lignin analysis used by PAPRICAN, we also collaborated with a group of scientists from the Environmental Science and Forestry Department at The State University of New York (SUNY) College where they are using a novel approach for lignin and carbohydrate content analysis in woody plants, known as High-Resolution Thermogravimetric Analysis (HR-TGA). Stems from fully mature and dried *Arabidopsis* plants were ground using a Mini-Mill with 60 mesh sieve. The ground tissue from each plant was divided into two equal sections and they were analyzed independently. The ground tissue was then analyzed using Thermogravimetric Analyzer 2950 as described in Serapiglia et al. 2009.

4.4 Results

T-DNA mutagenized *Arabidopsis* lines, listed in table 4-1, were genotyped and the plant lines for the following genes were found to be heterozygous: *CAD-1, CAD-2, CAD-3, F5H-1, F5H-2, F5H-4, CCOAMT* and *GA1*, and thus were used for generating the F2 population. *F5H-3* and *GA2*

lines had no inserts as confirmed by PCR and by selection markers, antibiotic and BASTA resistance, and were not pursued further. The *ccr, ga3, fah1-7, irx4, ga5-1, tt5-1* and *comt* lines were homozygous mutant plants, and therefore were crossed with wild type *Columbia-0 (col-0)* or *Landsberg erecta-20 (ler-20),* depending on their genetic background, to obtain heterozygous seeds. The obtained seeds were used to create the F2 generation. The generated F2 populations were genotyped and were stored for further lignin analysis.

Six samples from the CCR line were sent to The Canadian Light Source Inc. for lignin analysis using infrared spectromicroscopy and soft X-ray spectromicroscopy. The samples consisted of two homozygous mutants, two homozygous wild types and two heterozygotes. The results obtained showed variation despite coming from the same genotype as explained in figure 4-3. We then tried the service of PAPRICAN for lignin analysis. Three wild type plants were used as a trial. Even though the results were accurate and consistent, the cost was going to be extremely high, especially as we needed to screen a large number of plants. As a result, we decided to look for other alternatives. We then were able to initiate a new collaboration with the Environmental Science and Forestry Department at SUNY and we sent five wild type plants to be analysed using HR-TGA to test the accuracy and reproducibility of results obtained using HR-TGA. Each tested sample was run twice and the results from both readings were almost identical, as shown in figure 4-4. Since the readings were reproducible, we decided to send a set of 30 samples from the CAD-1 line for analysis. The samples were genotyped using both PCR and the election

marker, as explained in figure 4-1. One homozygous mutant was lost during the grinding process. The results obtained showed no significant differences between the three different genotypes tested, as shown in figure 4-5. To further investigate whether differences between samples of the same genotype were significant, we superimposed the results obtained from homozygous wild type samples and it showed a variation of about 40%, as shown in figure 4-6. All of the samples tested were from fully mature plants. Further analyses were carried on plants harvested during early stages of development. Set of 100 individuals from the *fah1-7, ga5-1 and irx4* lines were harvested at middle stage of their development and only the first 10 cm of the main inflorescence will be analyzed, at a later date surpassing this thesis, using HR-TGA.

We also screened for the presence of visible morphological abnormalities in all lines obtained for this study. Results from lines *ga3, fah1-7, irx4, tt5-1* and *ga5-1* showed visible morphological defects associated with the mutations present in these lines. Homozygous mutant *ga3* plants were dwarf and had short and thick siliques. To test whether the mutation in this gene is haploinsufficient, we measured siliques length from homozygous and heterozygous plants and compared their heights. A significant difference was found between homozygous mutant and homozygous wild type plants, whereas, no significant difference was found between heterozygous and homozygous plants, as shown in figures 4-7(B) and 4-8. Thus, we concluded that heterozygosity for a mutant allele in the gibberellin 2-oxidase-6 gene (Ga2ox6), line *ga3*, does not result in haploinsufficiency. The same results were

obtained for *irx4* (figures 4-9 and 4-10) and in *ga5-1* (figures 4-11 and 4-12) where the homozygous mutants were dwarf and the heights of heterozygous plants were similar to homozygous wild types. Homozygous mutants in *tt5-1* had very pale seeds, as shown in figure 4-13(A). Heterozygous and wild type plants in *tt5-1* both had normal brown seed colours, and thus, we concluded that *tt5-1* did not result in haploinsufficiency, as shown in figure 4-12.

Since it is stated in the TAIR website that *col-0* is the genetic background for line *fah1-7*, our initial genetic cross and analysis was done with *col-0*. The screening results showed significant differences between the different genotypes with *col-0* where homozygous mutant was dwarf and heterozygous plants were taller than homozygous wild type *col-0*. Due to the similarity between the homozygous *fah1-7* mutant and wild type *ler-20,* we decided to perform another genetic cross with *ler-20* wild type. Results showed no significant morphological differences between the three different genotypes, as shown in figures 4-14 and 4-15. Thus, we concluded that heterozygous *fah1-7* mutants do not show haploinsufficiency.

4.5 Discussion

This study is still in its early stages, and we have as of yet not identified any heterozygous mutants that show obvious haploinsufficiency. While *Arabidopsis* is particularly suitable for genetic studies, it is poorly suited for studies of deficiencies in lignin content, especially when assessing a large number of individuals. The reason for this is the small amount of dry stem material, generally less than 300 mg that each plant generates. This amount is

too small for the assays typically used for lignin quantification. We have been engaged in identifying a collaborative partner that can handle many small samples. Here we will first explain our choice of method for lignin quantification, followed by a discussion that relates targeted loci with available literature.

Several methods, such as Klason, acid detergent, acetyl bromide and Nuclear Magnetic Resonance (NMR), have been developed to measure lignin content and composition in different woody and herbaceous plant species (Hatfield and Fukushima, 2005). Klason is the most commonly used method to measure total lignin content (Demura & Fukuda, 2007; Hatfield and Fukushima, 2005; Suckling et al., 2010). Fukushima and Hatefield (2004) compared different methods for measuring lignin content and found that none of the methods tested was superior to others. Several methods have been used with *Arabidopsis thaliana*. For example, Klason procedure has been used to measure total lignin content (Goujon et al., 2003) and NMR has been used to analyze lignin composition and structure (Marita et al., 1999). In our study, we tested several methods to measure total lignin content. We were searching for a sensitive technique that will allow us to identify minor differences in lignin content that might be present between heterozygous and homozygous plants (haploinsufficiency). This is especially crucial when analysing the effect of mutagenesis in heterozygous plants where the effect on total lignin content might be minimal. Since we are analysing 17 different populations (17 mutant lines, table 4-1) each consisting of around 100 individuals, we needed a high-throughput technique that can handle small quantities, is affordable, accurate

and fast. Serapiglia et al. (2008) and Serapiglia et al. (2009) showed that High-Resolution Thermogravimetric Analysis (HR-TGA) was suitable for high-throughput lignin analysis of willow (*Salix*). After testing several methods for lignin analysis, we decided to use HR-TGA to analyze our *Arabidopsis* plants. We found that HR-TGA was the most efficient, fast and reproducible method.

The last step in monolignol biosynthesis is catalyzed by cinnamyl alcohol dehydrogenase (CAD), which reduces the cinnamaldehydes precursors into the three monolignol alcohols: coniferyl (CA), sinapyl (SA) and p-coumaryl (Shi et al., 2010; Vanholme et al., 2008). In general, the down regulation of CAD genes in several plants results in a change in lignin composition (Baucher et al., 2003; Lapierre et al., 2004). On the other hand, the effect of CAD down regulation on total lignin content is not clear due to contradicting results obtained from different plant species (Baucher et al., 2003; Vanholme et al., 2008). For example, Lapierre et al. (2004) found that poplar lines with severely down-regulated CAD gene results in a moderate decrease in total lignin content and less S monolignols. They found that poplar lines with 10% or less active CAD resulted in a decrease in total lignin content ranging from 20%, in poplar lines with 40-100% active CAD, to 17-18% of the extract-free dry wood. Shi et al. (2010) reported that in several studies of *P. tremula x alba* with down regulated CAD there were no changes in total lignin content or S/G content. However, all studies showed that the lignin polymer in CAD deficient plants had unusual phenolics, such as cinnamaldehyde and dihydroconiferyl alcohols which are believed to be responsible for the generation of red to brown stem xylem

(Baucher et al., 2003). As indicated by the TAIR website, the T-DNA insertion in the CAD-1 line (At4g34230) is located in the first exon. There are nine CAD genes in *Arabidopsis* (Hamberger et al., 2007; Sibout et al., 2003). The CAD-1 gene that we are studying is *At*CAD-D, which belongs to a class of CAD genes that are highly similar in alfalfa, tobacco, poplar and aspen, and is highly expressed in lignified stem tissues (Sibout et al., 2003; Sibout et al., 2005). Atcad-D can reduce both sinapaldehyde and coniferaldehyde into S and G monolignols in lignified tissues (Sibout et al., 2003). Sibout et al. (2003) showed that null *Atcad*-D resulted in lower Klason lignin content (14.23% compared to 15.20% of dry weight in control) but significantly altered lignin composition where it was higher in G content along with the presence of sinapaldehyde (Sibout et al., 2003). In our study, there were no significant differences in lignin content between homozygous wild type, homozygous mutant and heterozygous *Arabidopsis* plants, figure 4-5. Thus, our results are consistent with other published studies on *Medicago sativa* (Baucheret al., 1999), tobacco (Halpin et al., 1994; Ralph et al., 1998) and poplar (Shi et al., 2010). In addition, as shown by Sibout et al. (2003), the decrease in total lignin content in *Arabidopsis* is limited. Further analysis of lignin structure and composition is required in order to examine the effect of haploinsufficiency in CAD-1 in *Arabidopsis*.

The analysis of five wild type samples, as shown in figure 4-6, suggested the presence of discrepancies between the tested wild types. This could be due to differences in the amounts of tissue analyzed. For our analyses, each plant was allowed to grow to maturity and then the main inflorescence was ground

and analyzed. During the grinding process thin stems, sometimes long thin stems, passed through the sieve without being ground. In addition, the samples tested were from mature dried stems of *Arabidopsis (Columbia)*. Based on a study conducted by Brown et al. (2005) the amount of cellulose in *Arabidopsis* varied during different plant growth stages, from 15% in early stages to up to 50% in later stages. The study also showed that immature material tended to show less variation than mature material. This is because in later stages the carbohydrate content is high, 50-60% of dry weight, which hinders lignin detection. This is evident in our study, where the cellulose and lignin contents in all samples that were analyzed using HR-TGA, including 5 wild type and 29 plants from CAD-1, were around 64% and 21% of the total weight, respectively. Therefore, for further studies we have collected immature stem samples. The samples are 10 cm long obtained, from the base of the main inflorescence, but, 2 cm above the root-stem transition level. Five populations were generated from lines *irx4, fah1-7 col, fah1-7 ler, tt5-1,* and *ga5-1.* The stems have been collected, plants have been genotyped, and are ready for lignin analysis. Further analysis of lignin content will be conducted and reported in future studies.

Several mutant lines had a visible phenotype that was associated with the presence of mutations. Lines *ga3, ga5-1, irx4* and *tt5-1* showed visible morphological defects associated with mutations present in these lines. The g*a3* and *ga5-1* lines contained EMS induced G→A substitution in genes involved in gibberellin biosynthesis. The *ga3* line comes from an EMS-mutagenized

Arabidopsis TILLING population (TAIR). The *ga5-1* line contains a G→A substitution in the second exon of a gene encoding gibberellin 20-oxidase (GA 20-oxidase1). The *ga3* line contains a G→A substitution in the first exon of the gibberellin 2-oxidase6 gene (GA 2-oxidase6). Gibberellin (GA) hormone is considered to be of great importance because it is known to regulate a number of essential processes involved in plant development that include: seed germination, stem elongation, biomass accumulation, lignin biosynthesis, flowering stimulation and wood development (Biemelt et al., 2004; Busov et al., 2003; Taiz and Zeiger, 2006). Several studies of enzymes that are known to be involved in GA biosynthesis and catabolism showed that increased GA biosynthesis resulted in an increase in plant growth, biomass and length of xylem fibers in various plant species (Beimelt et al., 2004; Busov et al., 2003). Plants are known to have several forms of GA, some of which are active while others are inactive (Yamaguchi, 2008). This mixture between active and inactive GA is controlled by enzymes that convert inactive GA, such as GA20 and GA8, into an active form of GA, such as GA1 and GA4, and vice versa (Taiz and Zeiger, 2006). GA 20-oxidase and GA 2-oxidase belong to the dioxygenases enzyme family, as they add a hydroxyl group (-OH) to the GA molecule (Taiz and Zeiger, 2006). Information obtained from various published studies, which includes studies conducted both on poplar trees (Busov et al., 2003) and other organisms such as *Arabidopsis thaliana* (Rieu et al., 2008; Schomburg et al., 2003), suggest that GA 2-oxidase and GA 20-oxidase affects the level of active GA present within the plant. GA 20-oxidase along with GA 3-

oxidase, oxidize the inactive GA9 and GA20 generating the biologically active G1 and G4 (Rieu et al., 2008). Thus, the down regulation of GA 20-oxidase will result in a decrease in bioactive GA resulting in dwarf phenotype (Rieu et al., 2008). Rieu et al. (2008) showed that GA 20-oxidase1 *Arabidopsis* mutants were semi-dwarf as compared to double GA 20-oxidase mutants. Our *ga5-1* homozygous mutants were dwarf, as shown in figure 4-11, which is consistent with published results. To further investigate the degree of difference between heterozygous and homozygous wild type, we generated a population of 52 samples from heterozygous plant and the generated samples were genotyped and the total height of each plant was measured (Figure 4-12). Our results showed that the presence of one functional copy of *ga5-1* is sufficient to restore the morphology (no haploinsufficiency) of the plant, as shown in figure 4-12.

GA 2-oxidase regulates the final conversion of active GA4 to the inactive form GA34 and the conversion of the active GA1 to the inactive form GA8 (Taiz and Zeiger, 2006; Yamaguchi 2008). Thus, the down regulation of GA 2-oxidase will increase the level of active GA (Yamaguchi 2008). Schomburg et al. (2003) showed that the over expressed GA2-oxidase7 and GA2-oxidase8 generated a dominant dwarf phenotype in the F1 generation in *Arabidopsis* and tobacco. Similar results were obtained by Huang et al. (2010) where transgenic rice plants with over expression of GA 2-oxidase6 were semi-dwarf. Busov et al., (2003) also showed that over expression of *Pta*GA 2-oxidase1 in *Populus tremula x Populus alba* resulted in dwarf trees. Wall et al., (2008) reported that there are no known phenotypes associated with mutated GA 2-oxidase

Arabidopsis. Rieu et al., (2008) studied the effect of loss of function in five of the GA 2-oxidase genes in *Arabidopsis*, including GA 2-oxidase6. They reported that a T-DNA mutation in each gene alone did not result in visible phenotypic changes while a generated line with all five genes mutated, quintuple mutant, generated taller plants. Plants with mutations in the five GA 2-oxidase genes were 8% taller than wild type (Rieu et al., 2008). They also reported that the siliques length was reduced as a consequence of reduction in seed number per silique in the quintuple mutants but not in GA 2-oxidase6 mutants. In peas, mutation in *Ps*GA 2-oxidase1 gene resulted in thin and taller plants (Martin et al., 1999). Unlike results from other studies, our homozygous mutant GA3 plants were dwarf and had short siliques. Several samples from our GA3 homozygous plants were sent for sequencing and the results confirmed that the mutations were present in the GA 2-oxidase6 gene. Our results contradict published results. One possible reason might be due to the type of induced mutation, EMS mutation versus T-DNA, were multiple mutations might be present and contributing to this phenotype.

The *IRX* gene in *Arabidopsis* encodes a cinnamoyl CoA reductase (CCR1) (Jones et al., 2001). Turner and Somerville (1997) reported that *Arabidopsis* plants with mutations in three loci in the *irx* (*irx1*, *irx2-1* and *irx3*) gene had a collapsed xylem due to decrease in lignin down to one eighth compared to wild type. *Irx* mutants showed little reduction in growth rate and were slightly shorter than wild type (Turner and Somerville, 1997). A study done by Jones et al. (2001) showed that homozygous *irx4* mutants had 50%

reduction in total lignin content with no effect on cellulose content. The presence of other CCR genes might have been responsible for 50% of lignin biosynthesis in homozygous *irx4* mutants (Jones et al., 2001). The *irx4* mutants also had a collapsed xylem, were unable to grow upright and were dwarf (Jones et al., 2001). The average height of homozygous *irx4* mutants was reported to be 10-12 cm (TAIR). The G→A substitution in *irx4* mutants was in the second intron, which affected the proper splicing of mRNA and resulted in the presence of stop codon, which in turn resulted in a non-functional protein (Jones et al., 2001). The phenotypes of our homozygous *irx4* mutants are consistent with results from Jones et al., (2001), but, our homozygous *irx4* mutants were shorter (Figure 4-9). Our results showed that the presence of one functional copy of *irx4* is sufficient to restore the morphology (no haploinsufficiency) of the plant, as shown in figures 4-9 and 4-10.

Yellow seeds and densely packed transparent siliques of our homozygous *tt5-1* mutants were consistent with *transparent testa* (*tt*) mutants from other studies (Buer et al., 2009; Shirley et al., 1995). The phenotype results from the absence of tannin pigments that give the testa (seed coat) its brown colour (Sagasser et al., 2002; Shirley et al., 1995). The yellow cotyledons beneath the testa are what give the seed its yellow colour (Shirley et al., 1995). Our results showed that *tt5-1* gene did not result haploinsufficiency (Figure 4-13).

The *fah1-7* line is defective in the *F5H* gene, which is involved in lignin biosynthesis. Huntley et al. (2003) found that the over-expression of the *F5H*

gene in transgenic *Populus* (*Populus tremula x Populus alba*) had no effect on total lignin content and had no effect on cellulose and the content of other polysaccharides. However, it did have a great effect on the S/G ratio; mutants showed a S/G ratio of 7/1 whereas wild type have a S/G ratio close to 1/1 (Huntley et al., 2003). In addition, the over-expression of the F5H gene had no negative effects on the plant's visible phenotype (Huntley et al., 2003). Similar results were obtained from *Arabidopsis* where Marita et al. (1999) found that lignin from *fah1-2* deficient mutants was almost lacking S monolignols. Our study suggests that there are no obvious repercussions, morphological or growth rate, from altering the S/G ratio in *fah1-7* mutant lines (Figures 4-14 and 4-15).

4.6 Figures

Figure4-1 Genotyping T-DNA mutagenized *Arabidopsis* (in CAD-1 gene) using PCR and antibiotic resistance.

T-DNA mutagenized *Arabidopsis* lines that were obtained from ARBC were genotyped by both PCR and antibiotic resistance. Sample of the genotyping of T-DNA mutagenized *Arabidopsis* in CAD-1 gene is shown in this figure. (A) Genotyping using PCR, where each sample was PCR amplified using two sets of primers that differ in the forward primers (one was gene specific and the other was T-DNA specific). Samples that showed one band (with T-DNA specific primer set) are homozygous mutant (Mut), samples that showed one band (with gene specific primer set) are homozygous wild type (WT) and samples that showed two bands (one with each primer set) are heterozygous (Het). (B) Genotyping using selection markers, in this case BASTA (herbicide) resistance. Homozygous wild type plants that were placed in agar plates containing BASTA either did not germinate or grew poorly with distinct yellow color (B, I). Homozygous mutant and heterozygous plants carrying T-DNA inserts grew normally in agar plates containing BASTA (B, II). (B, III) to insure that the parent used to generate the F2 population was heterozygous, agar plate containing BASTA was planted with seeds from a heterozygous parent (confirmed using PCR genotyping). The presence of seedlings that either grew normally, poorly or did not grow all in a single agar plate (B, III) further proved that the parent was heterozygous.

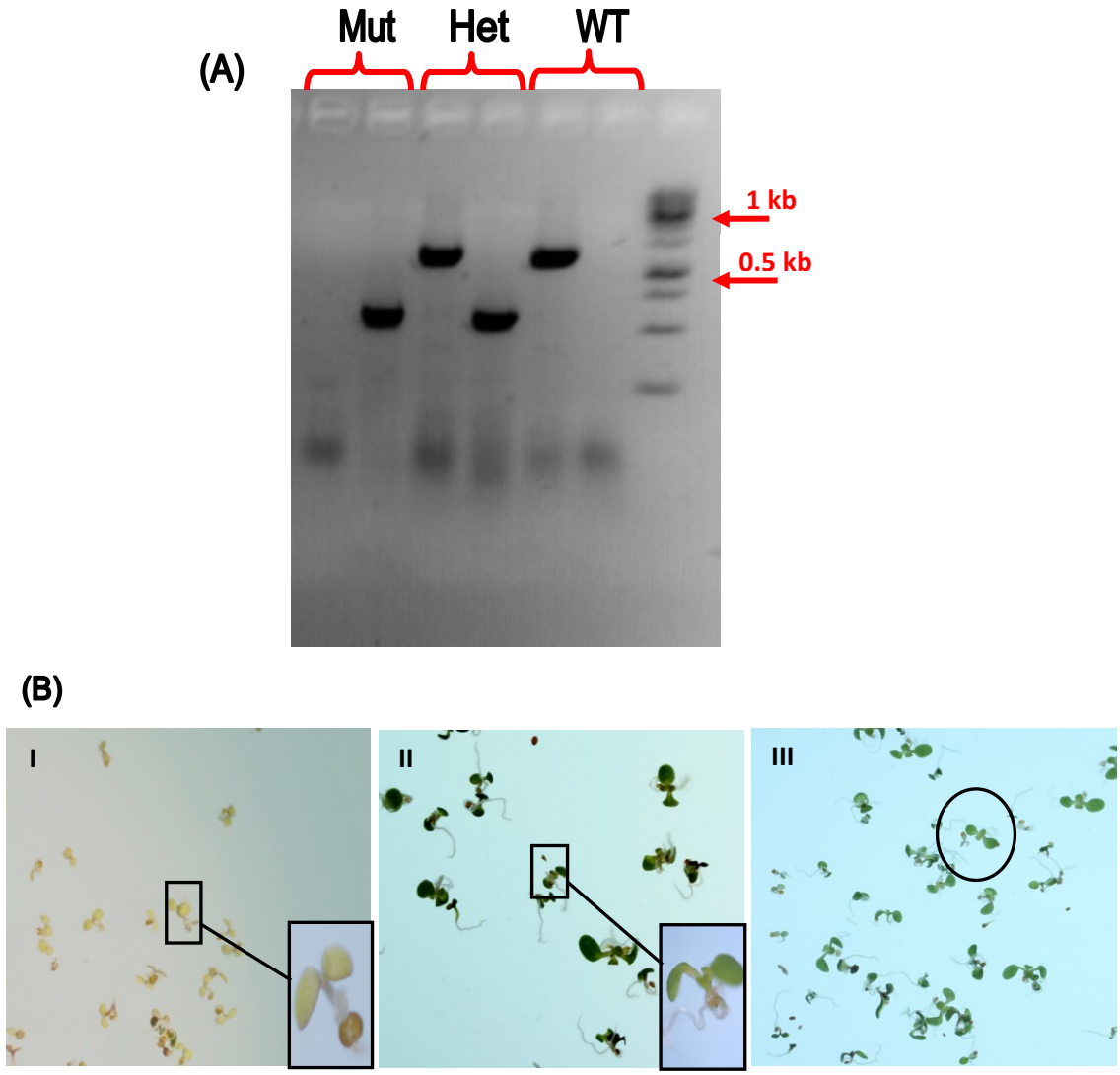


Figure 4-1 Genotyping T-DNA mutagenized *Arabidopsis* (in CAD-1 gene) using PCR and antibiotic resistance.

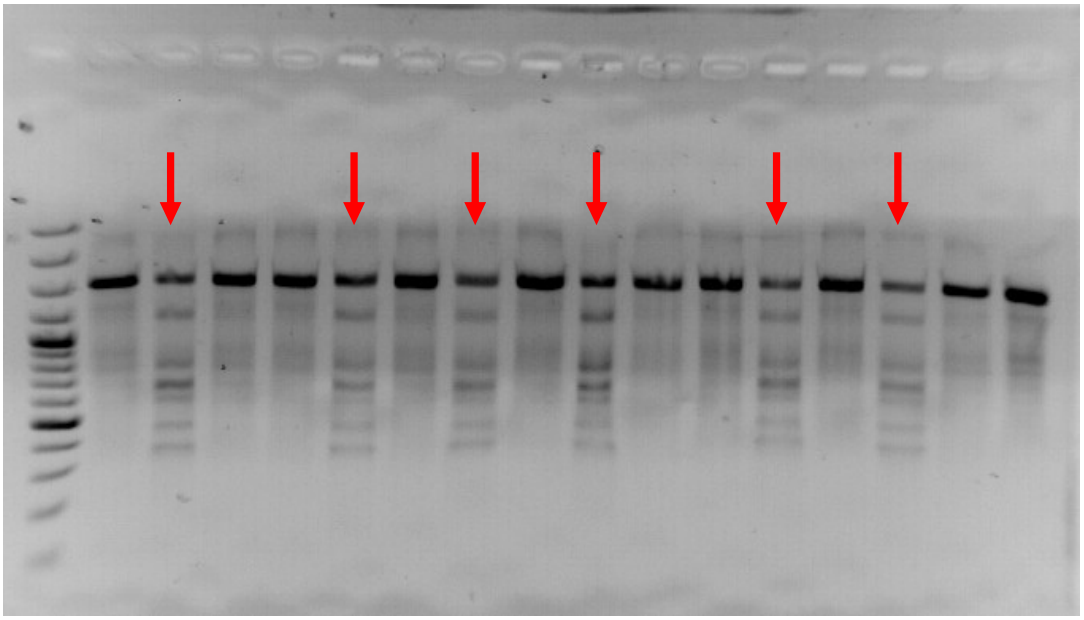


Figure 4-2 Genotyping ga5-1 using the modified TILLING protocol.

The generated population from heterozygous ga5-1 parent was genotyped using the modified TILLING protocol. The image shows a sample of the TILLING gels, where heterozygous plants resulted in shorter bands after digestion (red arrows).

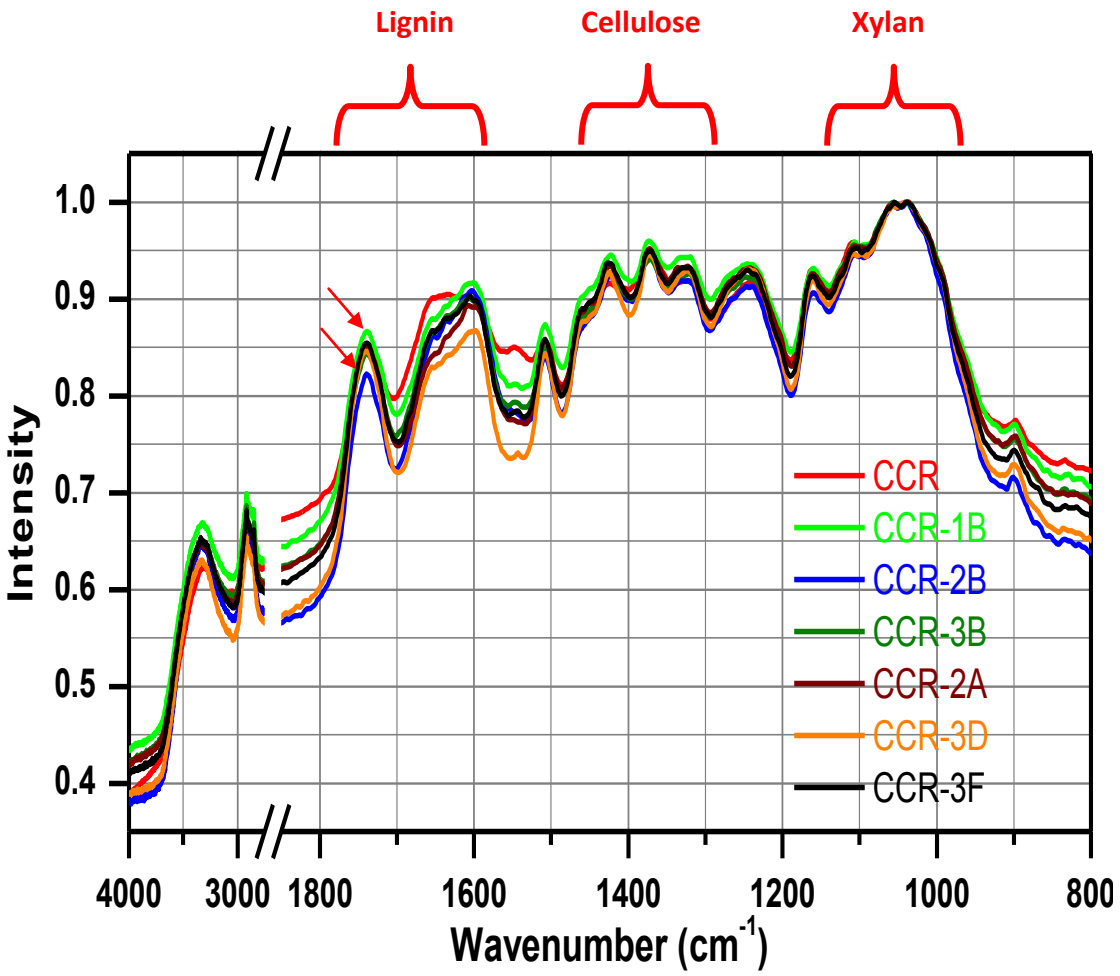


Figure 4-3 Cell wall analysis of CCR T-DNA mutagenized *Arabidopsis* using infrared spectromicroscopy and soft X-ray spectromicroscopy done by The Canadian Light Source Inc.

1B and 2B are homozygous mutant, 3F and 3D are heterozygous and 2A and 3B are homozygous wild type. Results do not provide realistic differences in the lignin content between the different samples. For example, CCR-1B and CCR-2B homozygous mutant samples (arrows) showed different lignin content despite the fact that they have the same genotype (one with highest lignin content and the other with the lowest among all the samples).

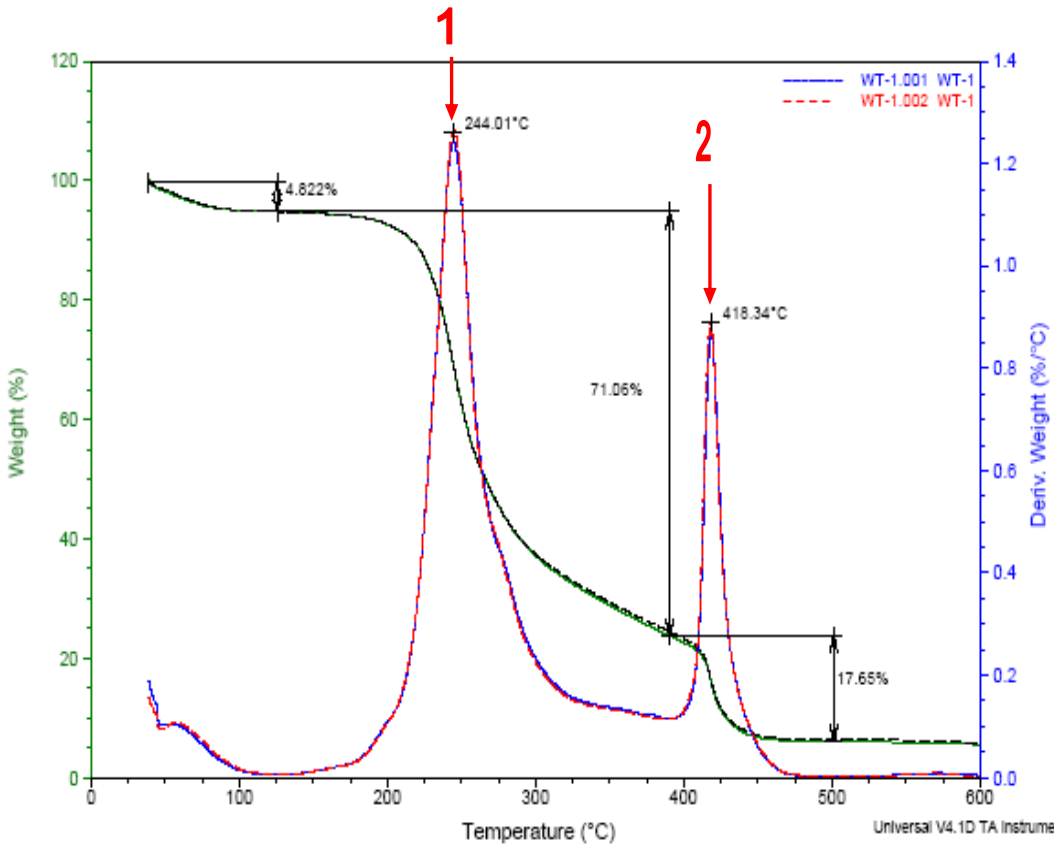


Figure 4-4 Analysis of total carbohydrates and lignin content in samples of *Arabidopsis* using High-Resolution Thermogravimetric Analysis (HR-TGA) showed that HR-TGA was accurate.

The total lignin content of the main inflorescence (stem) of *Arabidopsis* plants was analyzed using HR-TGA. Each sample was run twice and the average of the two readings was used for final analysis. HR-TGA graphs show two curves: weight loss vs. Temperature and first derivative of weight loss vs. Temperature. The first peak in the first derivative of weight loss vs. Temperature curve corresponds to carbohydrate decomposition (arrow 1) and the second peak corresponds to lignin and degraded carbohydrates decomposition (arrow 2). The graph shows the results of two readings from one of the *Arabidopsis* samples (WT-1). The two readings (WT-1 001 and WT-1 002) were superimposed on top of each other on the graph, which showed that they were fairly identical. The graph also showed that around 70% of the total stem weight is made up of cellulose/hemicellulose.

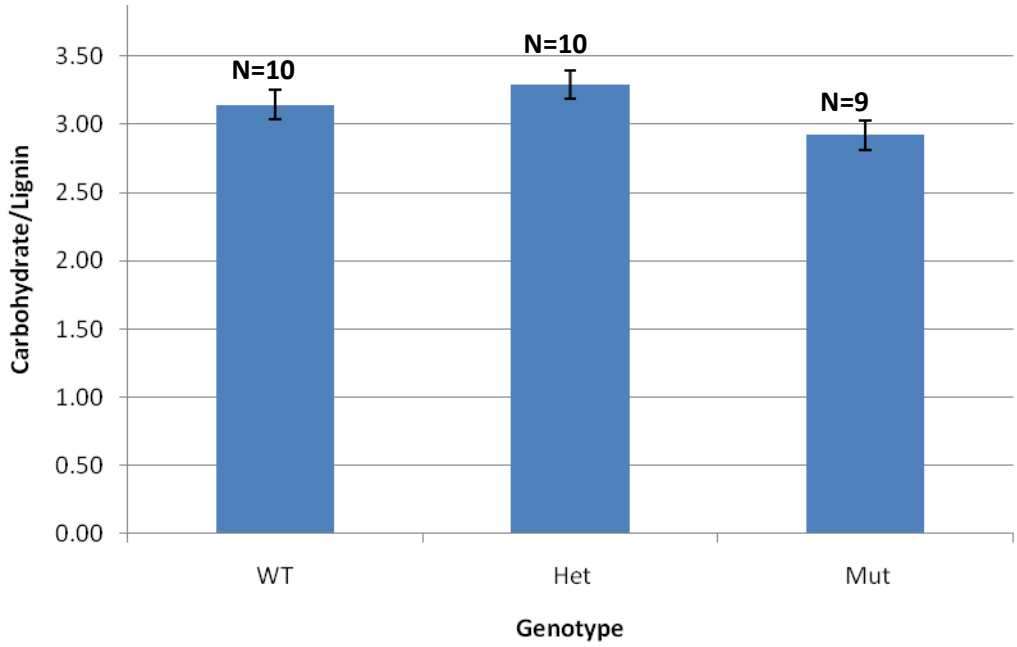


Figure4-5 Average Carbohydrate/Lignin ratios obtained using HR-TGA of homozygous wild type, heterozygous and homozygous mutant CAD-1 T-DNA *Arabidopsis* show no significant difference between the different genotypes.

Twenty nine samples, ten homozygous wild type (WT), ten heterozygous (Het) and nine homozygous mutants (Mut), of CAD-1 T-DNA *Arabidopsis* were analyzed using HR-TGA. Carbohydrate/Lignin ratios were calculated and the average from each of the three genotypes is shown on the graph (bars). Error bars shown on the graph represent the standard error. Student's t-test showed that there were no significant differences between the samples, with a probability of null hypothesis of <0.47 between WT and Het and of <0.23 between WT and Mut.

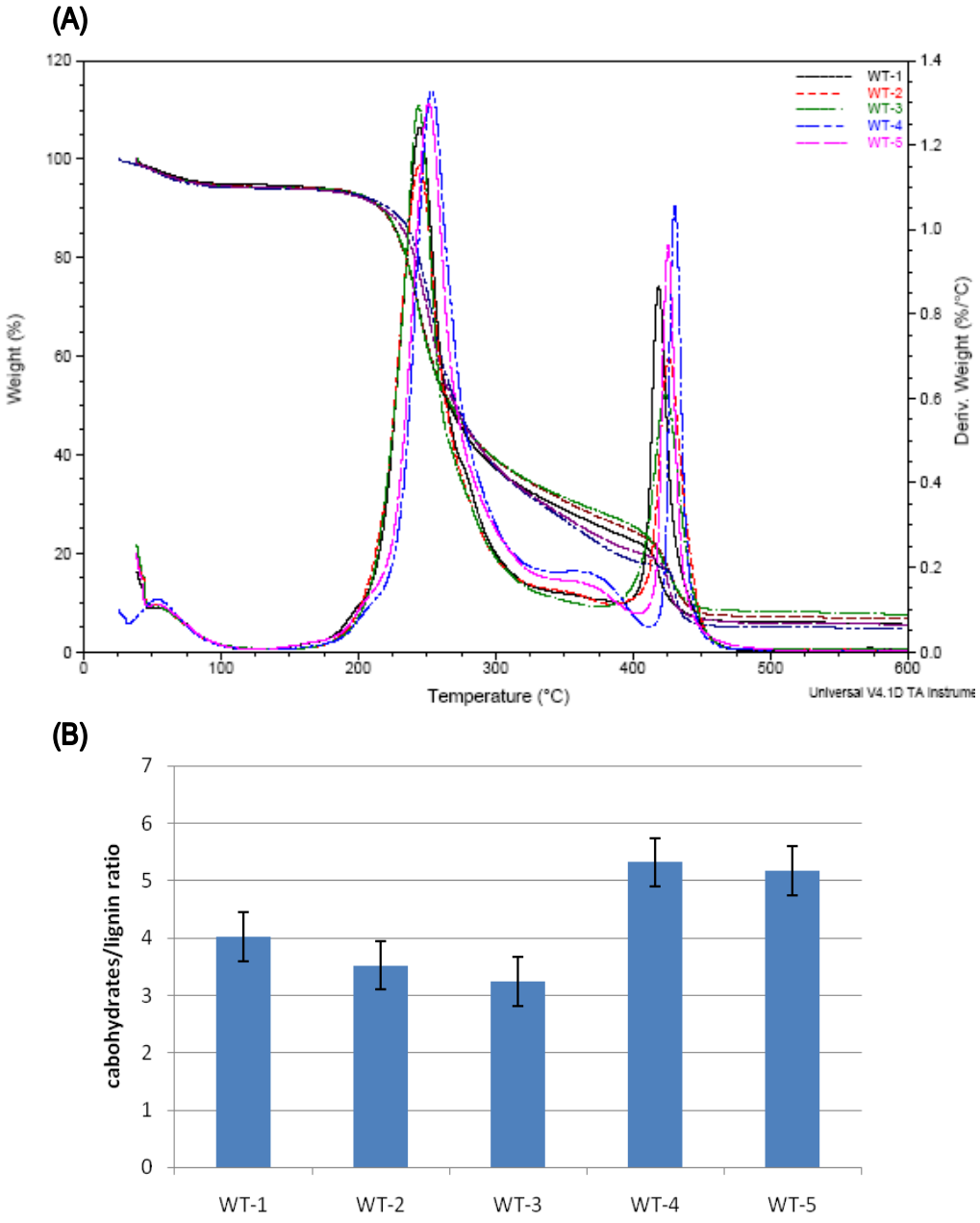


Figure 4-6 Analysis of total carbohydrate/lignin ratios in homozygous wild type samples of *Arabidopsis* using High-Resolution Thermogravimetric Analysis (HR-TGA) showing differences in results between the tested samples.

Five homozygous wild type *Arabidopsis* plants were analyzed using HR-TGA. Each wild type was run twice and the average of the two readings was used for final analysis. Graph in (A) shows the HR-TGA results of five readings from samples WT-1-5 superimposed on top of each other for comparison. (B) carbohydrates/lignin ratio for each wild type plant (average of two readings). Error bars in the graph represents the standard error

Figure 4-7 Visible phenotype comparison between homozygous and heterozygous GA3 *Arabidopsis* showing that GA3 gene does not show haploinsufficiency.

To test whether the presence of one wild type copy of the GA3 gene will rescue the mutant phenotype, we crossed a homozygous mutant GA3 plant to a wild type plant and the progenies were observed for any phenotypic difference. The modified TILLING protocol was used to genotype the GA3 *Arabidopsis* population generated from a heterozygous parent. (A) Sample of the TILLING gels used in genotyping GA3 *Arabidopsis* plants showing that homozygous plants (both homozygous wild type and homozygous mutant) did not result in shorter bands after digestion, while heterozygous plants did result in shorter bands that added up to the right size of 0.72 kb (arrows). (B) GA3 homozygous mutants are easily identified by their height (10 to 12cm tall) and short and thick siliques. The heterozygous plants look very similar to homozygous wild type plants, in both the plant's total height and siliques' size and thickness, and thus do not show haploinsufficiency in the GA3 gene.

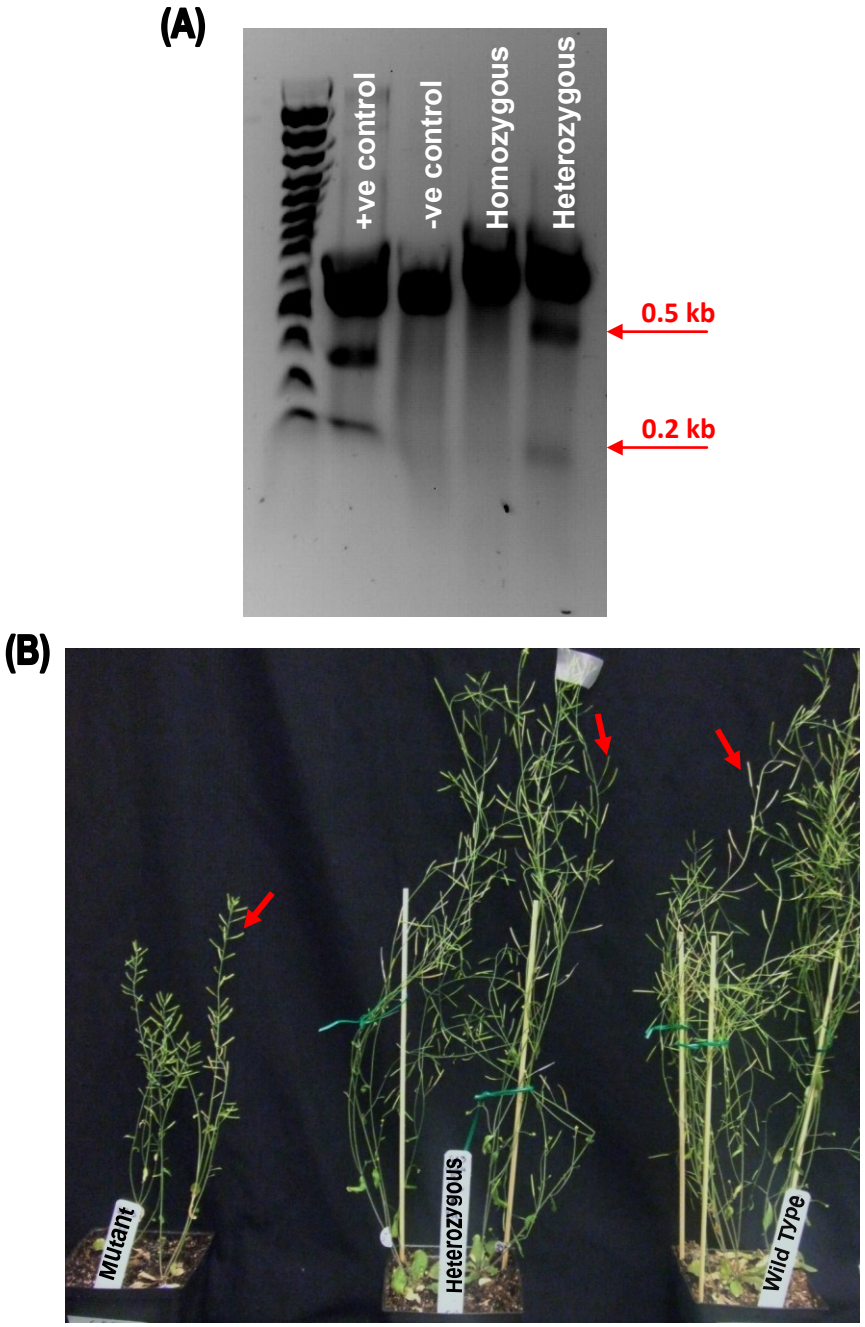


Figure 4-7 Visible phenotype comparison between homozygous and heterozygous GA3 *Arabidopsis* showing that GA3 gene does not show haploinsufficiency.

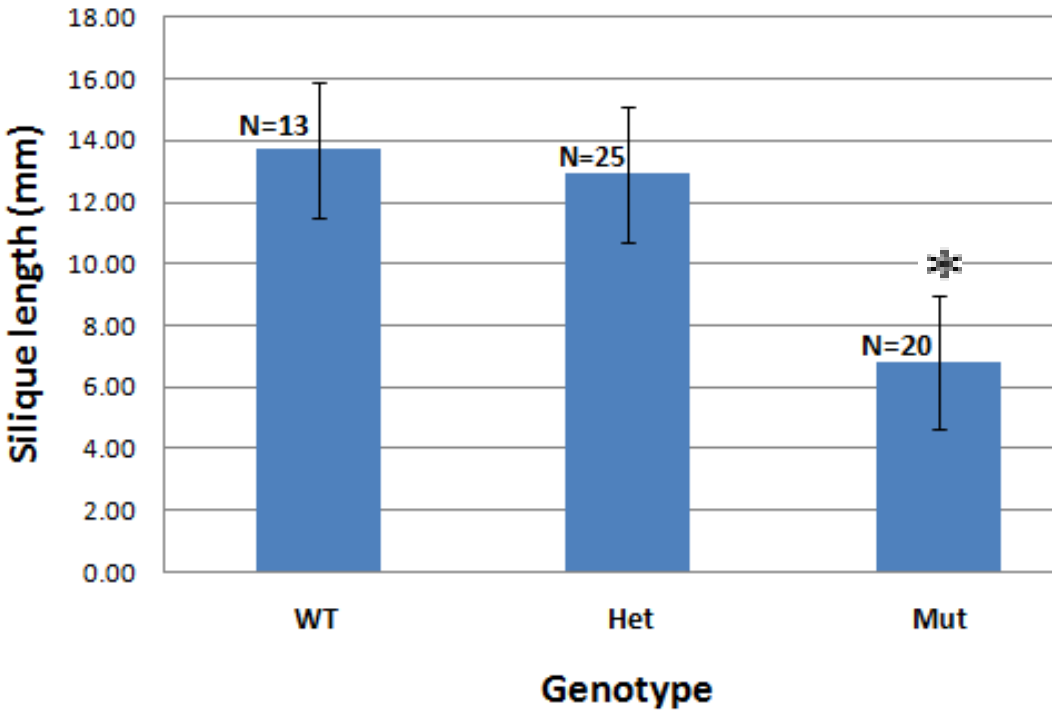
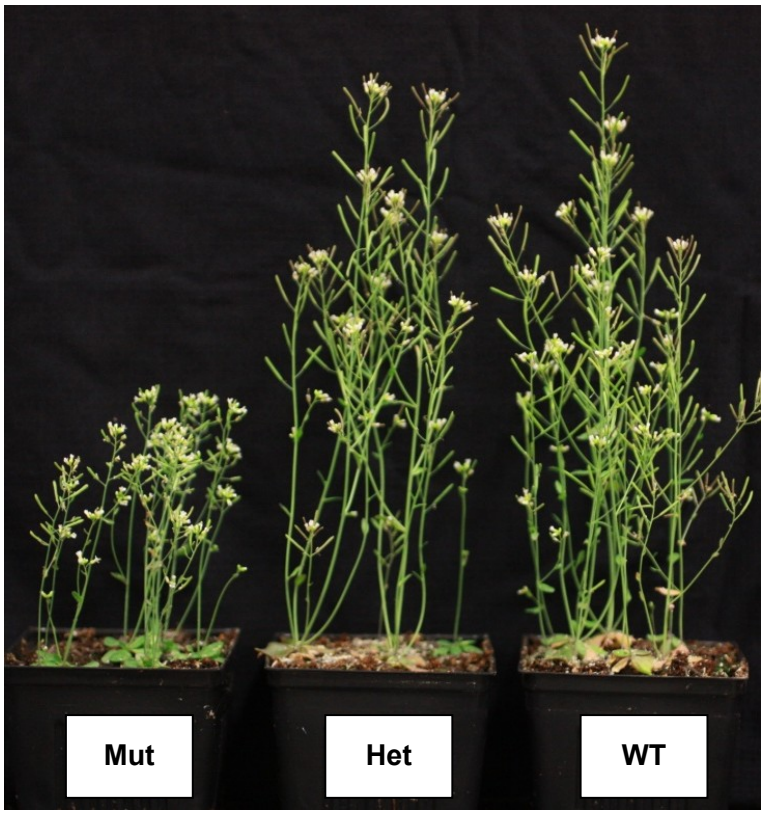


Figure 4-8 Average silique length of samples from F1 population generated from heterozygous GA3 *Arabidopsis* show no significant difference between heterozygous and homozygous wild type plants.

The lengths of the first five siliques were measured and the average was used for further analysis. We measured 13 homozygous wild type (WT), 26 heterozygous (Het) and 22 homozygous mutant (Mut) plants. A student's t-test was performed between the different genotypes and the results showed that there is a significant difference between homozygous wild type and homozygous mutant ,with a probability of null hypothesis of <0.0001 (Star on top of the bar). On the other hand, the student's t-test showed that there is no significant difference between heterozygous and homozygous wild type, with null hypothesis of <0.076. The error bars on the graph represent the standard error.

(A)



(B)

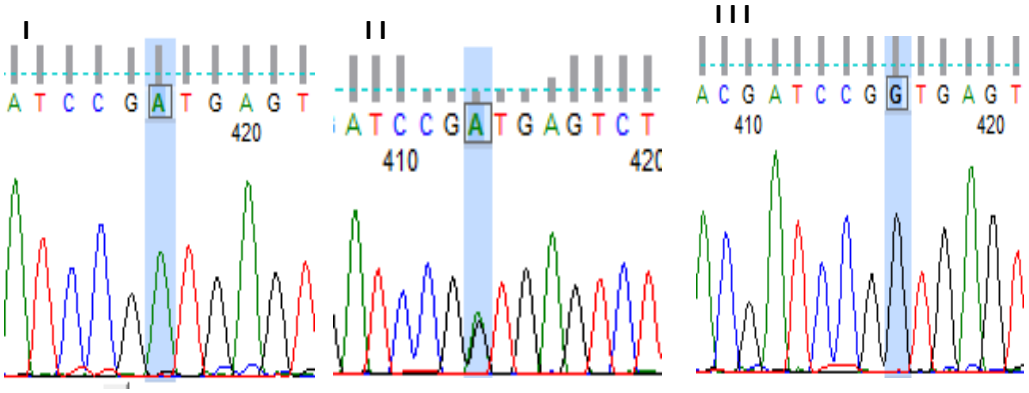


Figure 4-9 Visible phenotypic comparison between different *irx4* genotypes suggested that the *irx4* gene did not result in haploinsufficiency based on morphology.

A) Phenotypic comparison between homozygous mutants (Mut), heterozygous (Het) and homozygous wild type (WT) in the *irx4* gene. Homozygous mutants are dwarf and have reduced fertility compared to homozygous wild type. Visible phenotypic comparison suggested that heterozygous plants in *irx4* look similar to homozygous wild type. (B) *irx4* plants were genotyped by sequencing: (I) Homozygous mutant (A), (II) Heterozygous and (III) homozygous wild type (G).

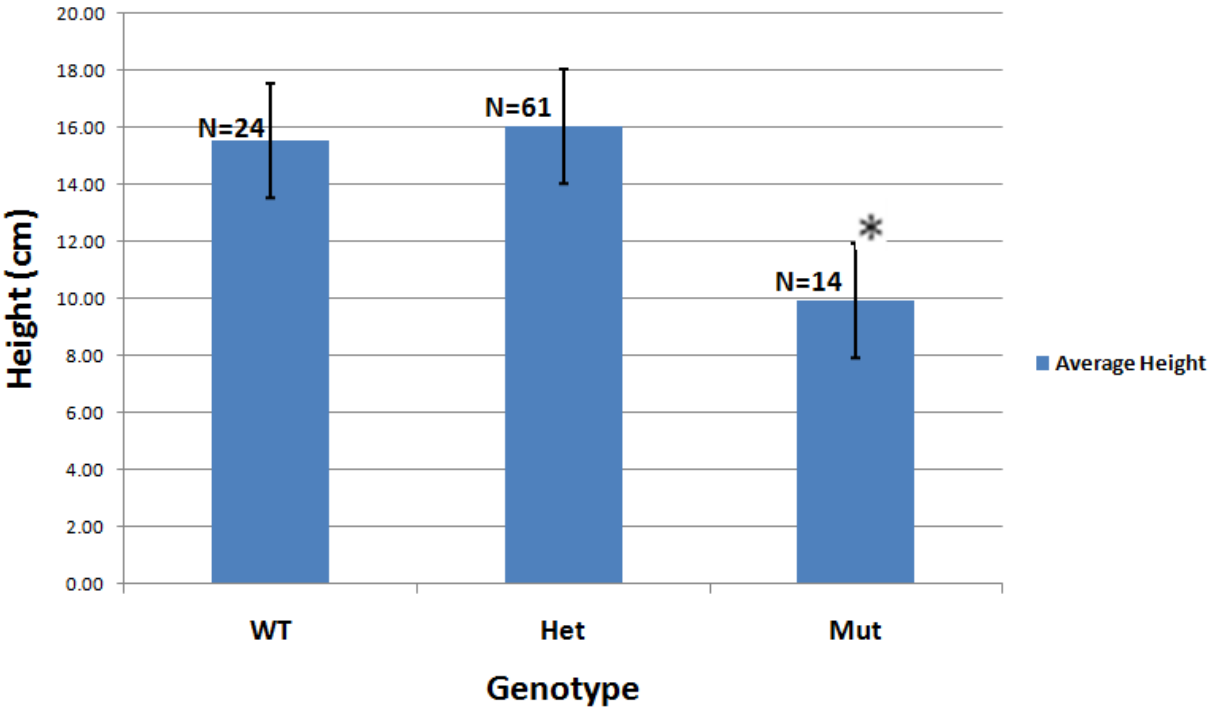
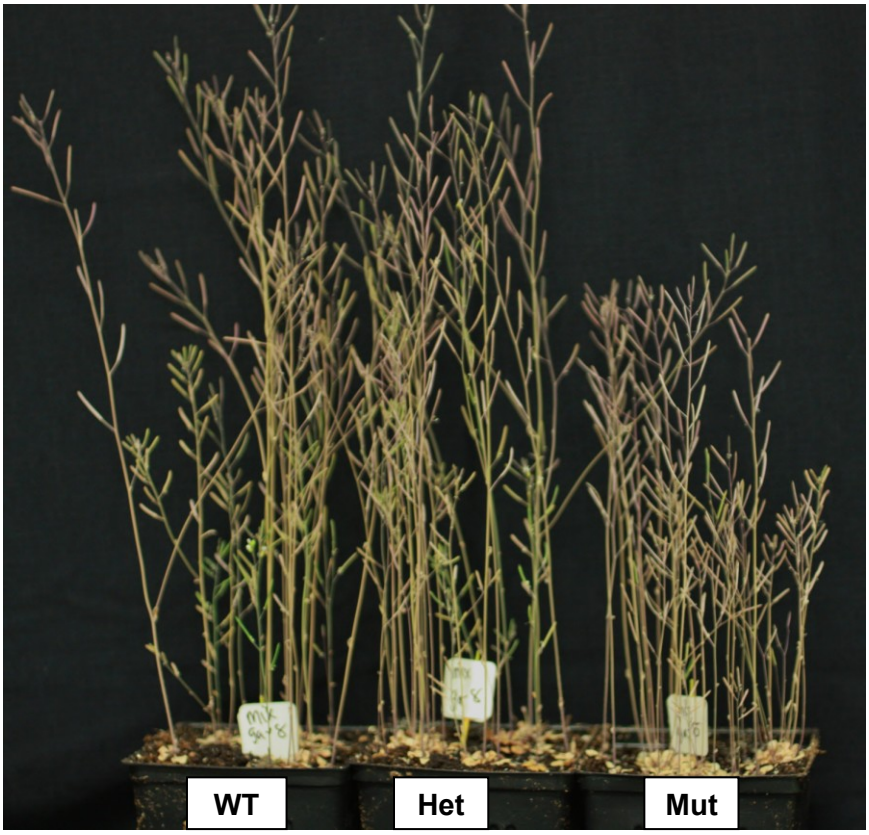


Figure 4-10 Average height from F1 population generated from heterozygous *irx*4 *Arabidopsis* show no significant difference between heterozygous and homozygous wild type plants.

A population made up from 24 homozygous wild type, 61 heterozygous and 14 homozygous *irx*4 mutants was generated. The total height was measured and the average (Bars on the graph) from each genotype was used for comparison. A student's t-test was performed between the different genotypes and the results showed that there is a significant difference between homozygous wild type and homozygous mutant and between heterozygous and homozygous mutant, with a probability of null hypothesis of <0.0001 (Star on top of the bar). On the other hand, the student's t-test showed that there is no significant difference between heterozygous and homozygous wild type, with null hypothesis of <0.37. The error bars on the graph represent the standard error.

(A)



(B)

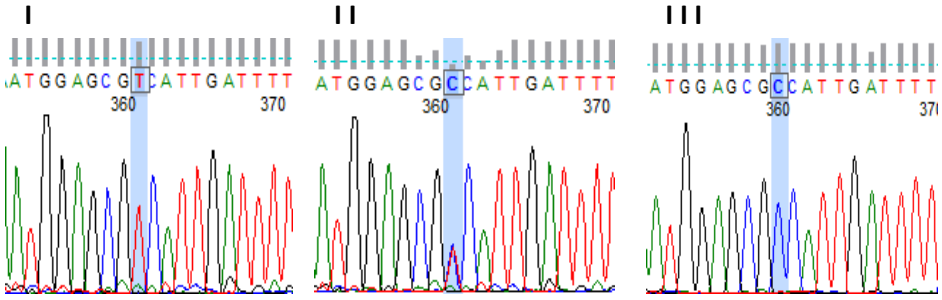


Figure 4-11 Visible phenotypic comparison between different ga5-1 genotypes suggested that the ga5-1 gene did not result in haploinsufficiency based on morphology.

(A) Phenotypic comparison between homozygous mutant (Mut), heterozygous (Het) and homozygous wild type (WT) in the ga5-1 gene. Homozygous mutants are dwarf compared to homozygous wild type. Visible phenotypic comparison suggested that heterozygous plants in ga5-1 did not look similar to (taller than) homozygous mutants. (B) ga5-1 plants were genotyped by sequencing: (I) Homozygous mutant (T), (II) Heterozygous and (III) homozygous wild type (C).

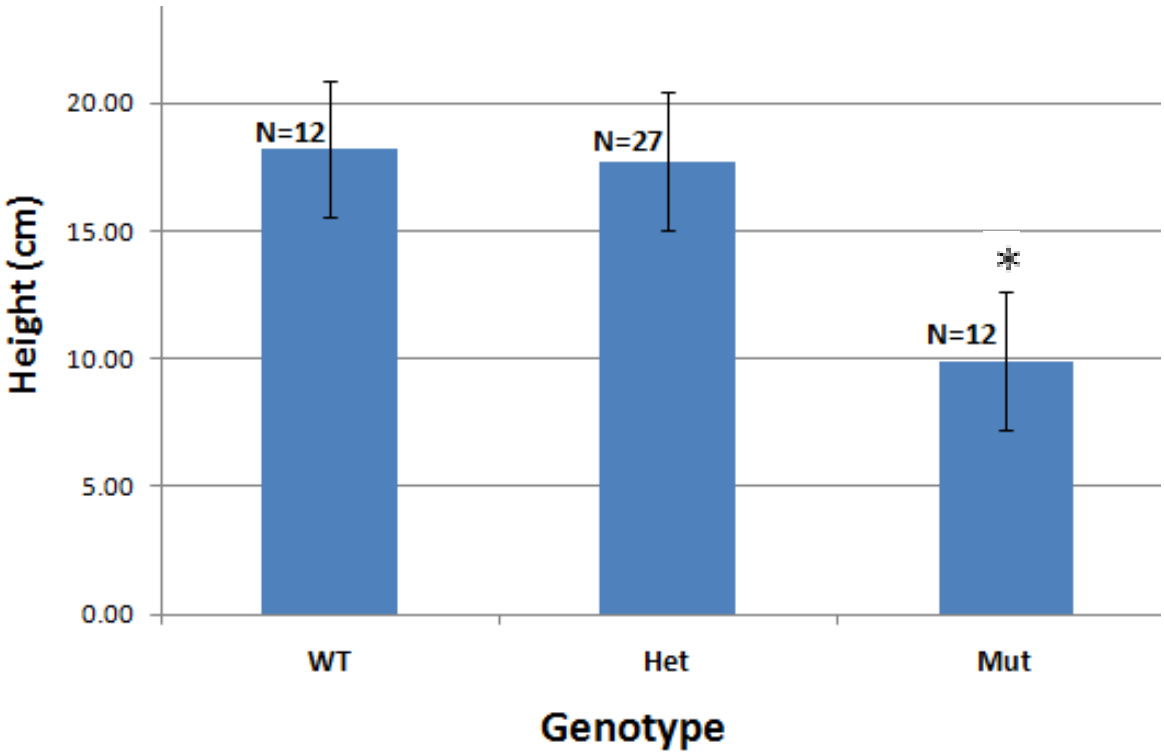


Figure 4-12 Average height from F1 population generated from heterozygous *ga5-1 Arabidopsis* show no significant difference between heterozygous and homozygous wild type plants.

A population made up from 12 homozygous wild type, 27 heterozygous and 12 homozygous mutants ga5-1 was generated. The total height was measured and the average (Bars on the graph) from each genotype was used for comparison. A student's t-test was performed between the different genotypes and the results showed that there is a significant difference between homozygous wild type and homozygous mutant and between heterozygous and homozygous mutant, with a probability of null hypothesis of <0.0001 (Star on top of the bar). On the other hand, the student's t-test showed that there is no significant difference between heterozygous and homozygous wild type, with null hypothesis of <0.57. The error bars on the graph represent the standard error.

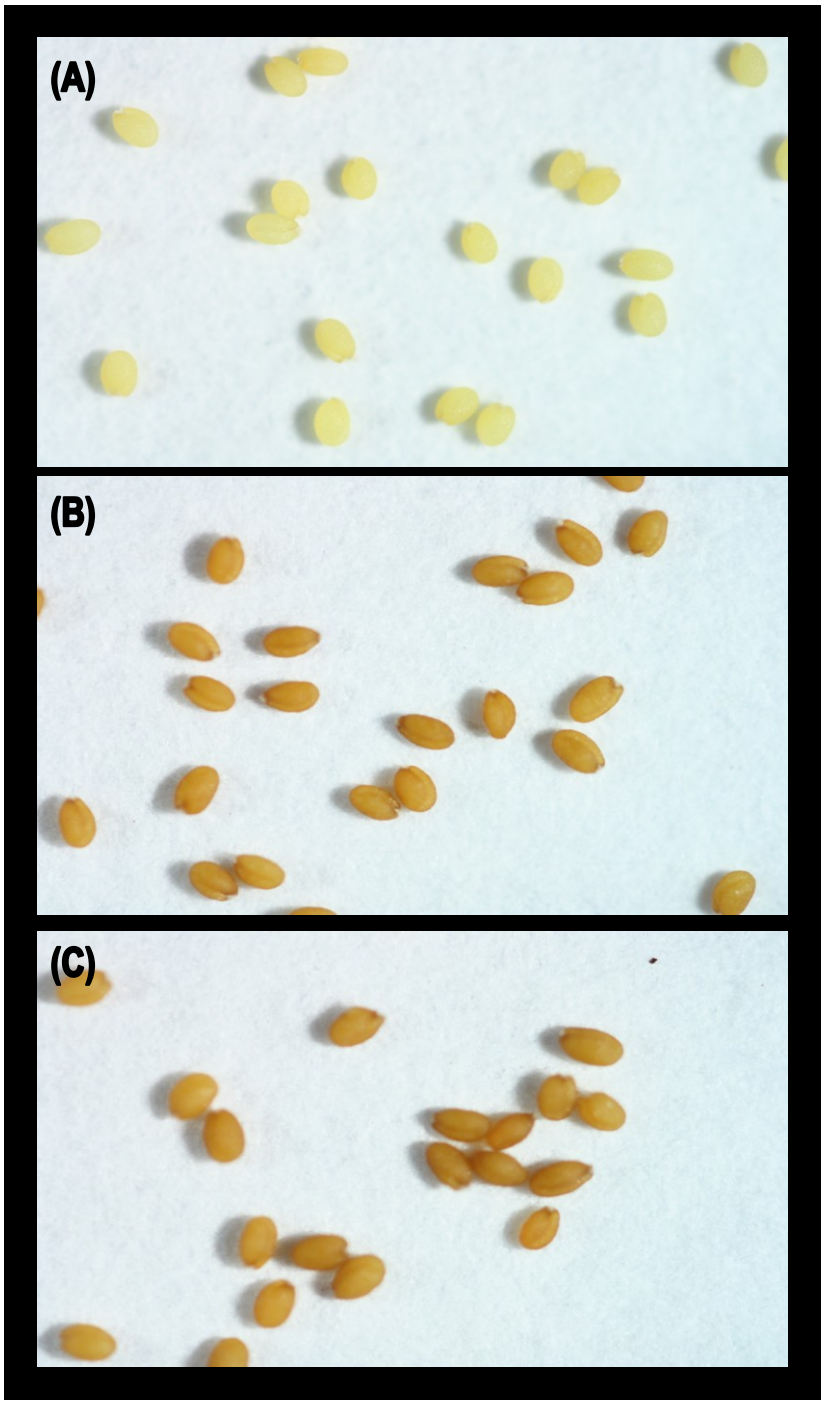
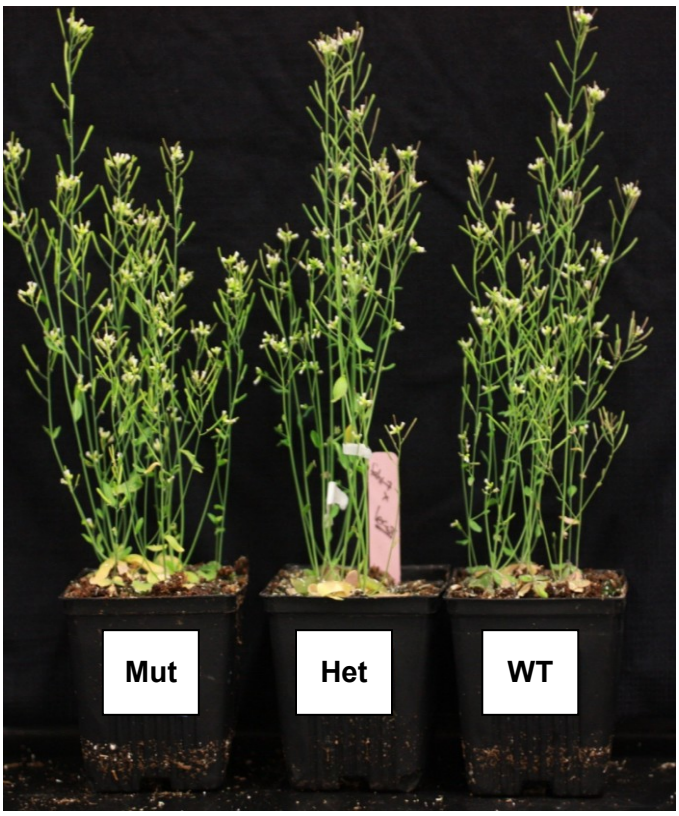


Figure 4-13 Seed color comparison between different tt5-1 genotypes suggested that the tt5-1 did not result in haploinsufficiency.

Seed color comparison between homozygous mutant (A), heterozygous (B) and homozygous wild type (C) in thett5-1 gene. Homozygous mutants have a pale seeds, due to the absence of anthocyanidins, brown pigmentation, in the seeds compared to homozygous wild type. Visible phenotypic comparison suggested that heterozygous plants in ga5-1look similar to homozygous wild type.

(A)



(B)

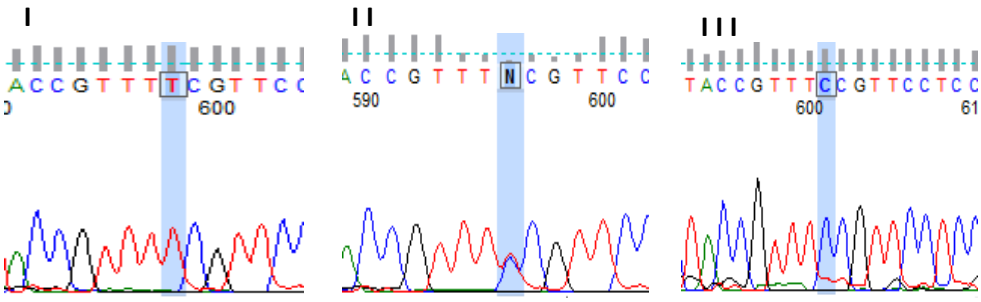


Figure 4-14 Visible phenotypic comparison between different *fah1-7* genotypes with Landsberg *erecta-20* background, showing that the mutation had no visible effect on the phenotype.

(A) Homozygous mutant (Mut), heterozygous (Het) and homozygous wild type (WT) showed no phenotypic differences. (B) *fah1-7* plants were genotyped by sequencing: (I) Homozygous mutant (T), (II) Heterozygous and (III) homozygous wild type *Landsberg erecta-20* (C).

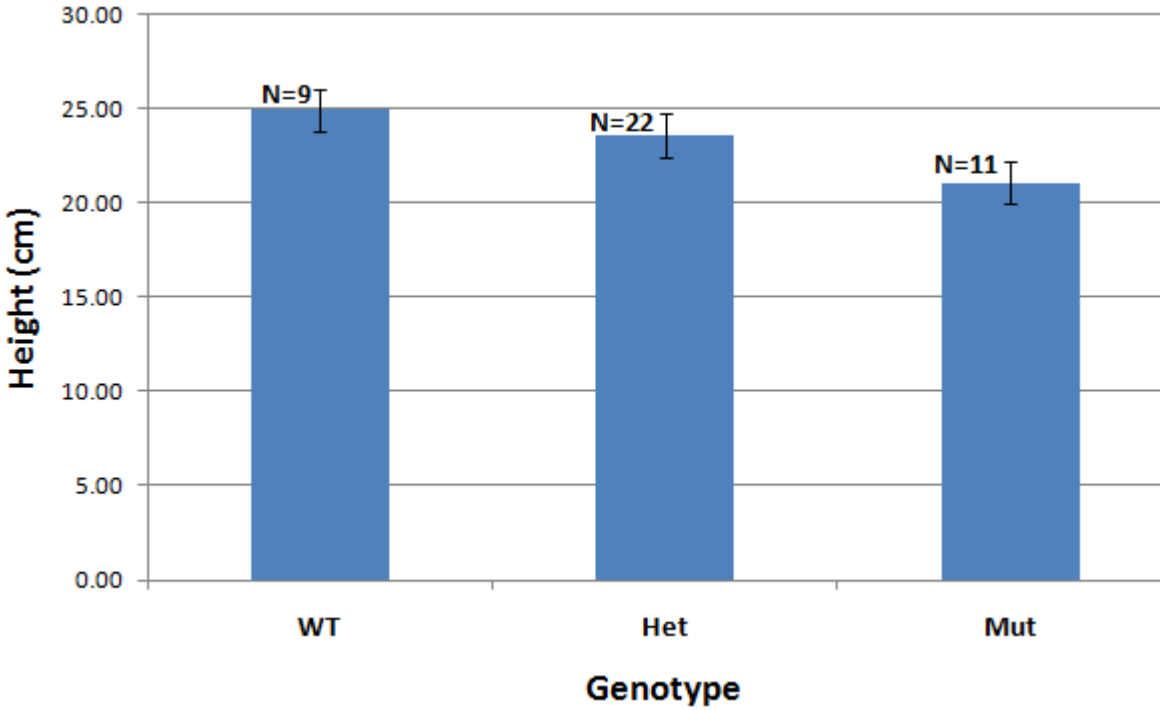


Figure 4-15 Average height from F1 population generated from heterozygous *fah1-7* genotype with *Columbia-0* background show no significant difference between heterozygous and homozygous plants.

A population made up from 9 homozygous wild type, 22 heterozygous and 11 homozygous mutants *fah1-7* was generated. The total height was measured and the average (Bars on the graph) from each genotyped was used for comparison. A student's t-test was performed between the different genotypes and the results showed there is no significant difference between heterozygous and homozygous, with null hypothesis of <0.51 between homozygous wild type and heterozygous and <0.26 between homozygous mutant and heterozygous. The error bars on the graph represent the standard error.

4.7 Tables

Table 4-1 Mutant *Arabidopsis* lines obtained from ABRC

Seed ID	Line	Gene Accession	Type of Mutation	Location of Mutation
COMT	SALK_070655	AT1G51990	T-DNA	Promoter
F5H-1	SALK_063792	AT4G36220.1	T-DNA	Exon
F5H-2	SALK_075014	AT3G11550	T-DNA	Promoter
F5H-3	SALK_090580	AT3G11540	T-DNA	Promoter
F5H-4	CS6172 (fah1-2)	AT4G36220	EMS	Exon→292bp from ORF
CAD-1	SAIL_776_B06	AT4G34230	T-DNA	Exon
CAD-2	SALK_094890	AT4G34240	T-DNA	Promoter
CAD-3	SALK_047151	AT4G34240	T-DNA	Promoter
CCR	SALK_042996	AT1G80820	T-DNA	Exon
CCOAMT	SALK_094890	AT4G34240	T-DNA	Promoter
GA1	SALK_020228C	AT1G78440	T-DNA	Exon
GA2	SAIL_803_C04	AT1G78440	T-DNA	Exon
GA3	atga2ox6_150H2	AT1G02400	EMS	G→A substitution in Exon
Irx4	CS19	AT1G15950	EMS	G→A substitution of first nucleotide in second intron
Fah1-7	CS8601	AT4G36220	EMS	G→A substitution of 23rd nucleotide in 3rd exon
Ga5-1	CS62	AT4G25420	EMS	G→A of 265th nucleotide in 2nd exon
tt5-1	CS86	AT3G55120	fast neutrons	Un known

4.8 References

Alkuraya, F. S., Saadi, I., Lund, J. J., Turbe-doan, A., Morton, C. C., and Maas, R. L. (2006). SUMO1 Haploinsufficiency Leads to Cleft Lip and Palate. *Science*, *313*(5794), 1751.

Barakat, A., Bagniewska-Zadworna, A., Choi, A., Plakkat, U., DiLoreto, D.S., Yellanki, P., and Carlson, J.E. (2009). The cinnamyl alcohol dehydrogenase gene family in Populus: phylogeny, organization, and expression. BMC plant biology, *9*, 26.

Baucher, M., Halpin, C., Petit-Conil, M., and Boerjan, W. (2003). Lignin: genetic engineering and impact on pulping. *Critical Reviews in Biochemistry and Molecular Biology*, *38*(4), 305–350. Informa UK Ltd UK. doi: 10.1080/10409230390242443.

Brown, D. M., Zeef, L. A. H., Ellis, J., Goodacre, R., and Turner, S. R. (2005). Identification of Novel Genes in *Arabidopsis* Involved in Secondary Cell Wall Formation Using Expression Profiling and Reverse Genetics. *Cellulose*, *17*(August), 2281-2295.

Buer, C. S., and Djordjevic, M. a. (2009). Architectural phenotypes in the transparent testa mutants of *Arabidopsis* thaliana. *Journal of experimental botany*, *60*(3), 751-763.

Busov, V. B., Meilan, R., Pearce, D. W., Ma, C., Rood, S. B., and Strauss, S. H. (2003). Activation Tagging of a Dominant Gibberellin Catabolism Gene (GA 2-oxidase) from Poplar That Regulates Tree Stature 1. *Society*, *132*(1), 1283-1291.

Chang, X.F., Chandra, R., Berleth, T., and Beatson, R.P. (2008). Rapid, microscale, acetyl bromide-based method for high-throughput determination of lignin content in *Arabidopsis thaliana*. *Journal of agricultural and food chemistry*, *56*(16), 6825-6834.

Dang, V.,Kassahn K., Marcos, A. and Ragan, M. (2008). Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *European journal of human genetics, 16*, 1350-1357.

Dellaporta, S. L., Wood, J., & Hicks, J. B. (1983). A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter*, *1*(4), 19-21.

Deutschbauer, A., Jaramillo, D., Proctor, M., Kumm, J., Hillenmeyer, M., Davis, R., Nislow, C., and Giaever, G. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics, 169*,1915-1925.

Fukushima, R. S., and Hatfield, R. D. (2004). Comparison of the acetyl bromide spectrophotometric method with other analytical lignin methods for determining lignin concentration in forage samples. *Journal of agricultural and food chemistry*, *52*(12), 3713-20.

Goujon, T., Sibout, R., Pollet, B., Maba, B., Nussaume, L., Bechtold, N., Lu, F., Ralph, J., Mila, I., Ere, B.Y., Lapierre C., and Jouanin, L. (2003). A new *Arabidopsis thaliana* mutant deficient in the expression of O-methyltransferase impacts lignins and sinapoyl esters. *Plant molecular*, 51, 973-989.

Hajduch, M., Tanaka, H., Morinaka, Y., Otake, Y., Nakamura, H., Kayano, T., et al. (2005). Protein analysis of dwarfed transgenic rice plants overexpressing GA2-oxidase gene. *Biologia Plantarum*, *49*(4), 621-624.

Hamberger, B., Ellis, M., Friedmann, M., Azevedo Souza, C. de, Barbazuk, B., and Douglas, C. J. (2007). Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. *Botany*, *85*(12), 1182-1201.

Harris, D., and DeBolt, S. (2010). Synthesis, regulation and utilization of lignocellulosic biomass. *Plant biotechnology journal, 8*(3), 244-262.

Hatfield, R. and Fukushima, S. (2005). Can Lignin Be Accurately Measured? *Crop Science*, *45*(3), 832-839.

Huntley, S.K., Ellis, D., Gilbert, M., Chapple, C., and Mansfield, S. (2003). Significant increases in pulping efficiency in C4H-F5H-transformed poplars: improved chemical savings and reduced environmental toxins. *Journal of agricultural and food chemistry*, *51*(21), 6178-6183.

Jones, L., Ennos, a R., and Turner, S. R. (2001). Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of *Arabidopsis*. *The Plant journal :for cell and molecular biology*, *26*(2), 205-216. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11389761.

Krieger, U., Lippman, Z. B., and Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature genetics*, *42*(5), 459–463.

Leplé, J., Dauwe, R., Morreel, K., et al. (2007). Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *The Plant cell*, *19*(11), 3669-3691.

Lapierre, C., Pilate, G., Pollet, B., Mila, I., Leplé, J., Jouanin, l., Kim, h., and Ralph, J. (2004). Signatures of cinnamyl alcohol dehydrogenase deficiency in poplar lignins. *Phytochemistry*, *65*(3), 313-321.

Li, X. and Chapple, C. (2010). Understanding lignification: challenges beyond monolignol biosynthesis. *Plant physiology, 154*(2), 449-452.

Lincoln C, Britton J.H., and Estelle, M. (1990). Growth and development of the axr1 mutants of *Arabidopsis*. *Plant cell* 2,1071-1080.

Marita, J.M.,Ralph, J., Hatfield, R.D., and Chapple, C. (1999). NMR characterization of lignins in *Arabidopsis* altered in the activity of ferulate 5-hydroxylase. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(22), 12328-12332.

Martin, D. N., Proebsting, W. M., and Hedden, P. (1999). The SLENDER gene of pea encodes a gibberellin 2-oxidase. *Plant physiology*, *121*(3), 775-781.

Pillitteri, L., Bemis, S., Shpak, E. and Torii, K. (2007). Haploinsufficiency after successive loss of signalling reveals a role for ERECTA-family genes in *Arabidopsis* ovule development. *Development, 134*(17), 3099-3109.

Rieu, I., Eriksson, S., Powers, S. J., Gong, F., Griffiths, J., Woolley, L., et al. (2008). Genetic analysis reveals that C19-GA 2-oxidation is a major gibberellin inactivation pathway in *Arabidopsis*. *The Plant cell*, *20*(9), 2420-2436.

Rieu, I., Ruiz-Rivero, O., Fernandez-Garcia, N., Griffiths, J., Powers, S. J., Gong, F., et al. (2008). The gibberellin biosynthetic genes AtGA20ox1 and AtGA20ox2 act, partially redundantly, to promote growth and development throughout the *Arabidopsis* life cycle. *The Plant journal : for cell and molecular biology*, *53*(3), 488-504.

Sagasser, M., Lu, G.-H., Hahlbrock, K., and Weisshaar, B. (2002). A. thaliana TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. *Genes and development*, *16*(1), 138-149.

Schomburg, F. M., Bizzell, C. M., Lee, D. J., Zeevaart, J. A. D., and Amasino, R. M. (2003). Overexpression of a Novel Class of Gibberellin 2-Oxidases Decreases Gibberellin Levels and Creates Dwarf Plants. *Society*, *15*(1), 151-163.

Serapiglia, M.J., Cameron, K.D., Stipanovic, A.J., and Smart, L.B. (2008). High-resolution thermogravimetric analysis for rapid characterization of biomass composition and selection of shrub willow varieties. *Applied biochemistry and biotechnology*, *145*(1-3), 3-11.

Serapiglia, M.J., Cameron, K.D., Stipanovic, A.J., and Smart, L.B. (2009). Analysis of biomass composition using high-resolution thermogravimetric analysis and percent bark content for the selection of shrub willow bioenergy crop varieties. *BioEnergy Research*, *2*(1), 1–9.

Shi, R., Sun, Y., Li, Q., Heber, S., Sederoff, R., and Chiang, V.L. (2010). Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant and cell physiology*, *51*(1), 144-163.

Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, Goodman HM. (1995). Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. Plant J., *8*(5):659-6571.

Somerville, C., Youngs, H., Taylor, C., David, S. C., and Long, S. P. (2010). Feedstocks for Lignocellulosic Biofuels. *Science, 329*(5993), 790-792.

Sibout, R., Eudes, A., Pollet, B., Goujon, T., Mila, I., Granier, F., et al. (2003). Expression Pattern of Two Paralogs Encoding Cinnamyl Alcohol Dehydrogenases in *Arabidopsis* . Isolation and Characterization of the Corresponding Mutants 1. *Society*, *132*(June), 848-860.

Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapierre, C., Jouanin, L., et al. (2005). CINNAMYL ALCOHOL DEHYDROGENASE-C and -D Are the Primary Genes Involved in Lignin Biosynthesis in the Floral Stem of *Arabidopsis*. *Society*, *17*(July), 2059-2076.

Taiz, L., and Zeiger, E. (2006). *Plant Physiology* (4th edition). Sunderland, Massachusetts: Sinauer Associates, Inc.

The *Arabidopsis* Information Resource (TAIR), www.arabidopsis.org, (2011).

Turner, S. R., and Somerville, C. R. (1997). Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary cell wall. *The Plant cell*, *9*(5), 689-701.

Umezawa, T. (2010). The cinnamate/monolignol pathway. *Phytochemistry Reviews*, *9*(1), 1-17.

Vanholme, R., Morreel, K., Ralph, J., and Boerjan, W. (2008). Lignin engineering. *Current opinion in plant biology*, *11*(3), 278-285.

Veitia, R.A (2006). The biology of genetic dominance. Georgetown, Tex. : Landes Bioscience: Eurekah.com. Retrieved from http://www.netlibrary.com.proxy.lib.sfu.ca/Reader/

Wall, JD., Harwood, C.S., and Demain, A. (Eds.). (2008). *Bioenergy*. Washington: ASM press. Retrieved from http://site.ebrary.com/sfu/Doc?id=10346892andppg=4.

Yamaguchi, S. (2008). Gibberellin Metabolism and its Regulation. *Annual Review of Plant Biology, 59*, 225-251.

CHAPTER 5: IDENTIFICATION OF INDUCED MUTATIONS IN WILLOW (*Salix viminalis*) THROUGH ULTRA-DEEP DNA SEQUENCING

5.1 Abstract

Willow has emerged as the energy crop most suitable for the climates of Northern Europe and North America. It is also subject to intense breeding efforts, primarily in Sweden, UK and USA. While the short generation time of willow (2-3 years) makes it suitable for breeding, the dioecious nature of willow complicates recurrent selection, and the absence of transgenic technology prevents introduction of dominant loss or gain of function gene constructs. Here we have set out to develop an alternative breeding technology for willow. We mutagenized willow pollen using ethyl methane sulfonate (EMS) to induce lesions, which to our knowledge has not been applied by others in forest tree species, and produced a small population of putative mutants. Thereafter, we used ultra-deep DNA sequencing technology to screen five target genes for novel single nucleotide polymorphisms (SNPs). To date, we have identified eleven SNPs that are not present in either of the parents. If these SNPs can be confirmed, this approach can be scaled up and used to identify mutant alleles in loci of interest in male and female willow plants, followed by genetic crosses to generate homozygous mutant plants that may display desired phenotypes.

5.2 Introduction

The realities of global warming and increased oil prices have triggered a global movement to reduce the reliance on fossil fuel and replace it with renewable fuels (Ragauskas et al., 2006). The predominant liquid biofuel, ethanol, is presently produced primarily from corn and sugarcane. However, neither species can replace significant portions of petroleum without impacting food supplies (Hill et al., 2006). One of the most important predicted long-term solutions is to use lignocellulose from wood as feedstock. However, there are several problems with the use of wood-derived lignocellulosic biomass for ethanol production, in particular, lignin, an abundant aromatic polymer that is intricately associated with the cellulosic polymer, which limits enzymatic hydrolyzability. From a Canadian perspective, willow is one of the most attractive potential sources for lignocellulosic biomass production (Ball et al., 2005). *Salix* (willow) and *Populus* (poplars) genus belong to the Salicaceae family and share common characteristics (Dickmann et al., 2001). Several characteristics present in willow makes them suitable for biofuel production, such as their rapid growth, even when grown in high densities, ability to propagate vegetatively, readily resprouts and ability to survive the harsh winters common to Canadian geography (Keoleian and Volk, 2005). Willow is considered to be one of the woody plants that are commonly known as short-rotation woody crops as it can accumulate biomass rapidly and can be ready to harvest within couple of years from plantation (Keoleian and Volk, 2005). For example, according to a Canadian study willow can yield up to 20 tonnes per

hectare of biomass annually (Natural Resources Canada, 2010). Several willow species, including *Salix viminalis*, have been subject to breeding programs in Sweden, USA and UK for the past 15 years (Keoleian and Volk, 2005), which are in their early stages of development. These breeding programs were able to generate willow varieties with significantly increased biomass (Keoleian and Volk, 2005). However, the breeding required to adapt willow for biofuel production is not without problems. While willow can flower within two years after planting, an unusually short time among forest tree species, it is also dioecious. Their dioecious nature prevents the use of agricultural type breeding based on natural or induced genetic variation followed by self fertilization and large-scale screens for homozygous mutants based on phenotypic alterations. Similarly, it is currently not technically possible to genetically transform willow (Berlin et al., 2010). A potential strategy to circumvent these limitations would be to identify desired genetic variants based on genotype of heterozygous males and females, followed by crosses to generate homozygous mutant varieties with potentially interesting phenotypes. A technology for rapid genotyping of specific loci in populations has been developed. This technology referred to as Targeting Induced Local Lesions IN Genomes (TILLING), has received considerable attention as a tool to identify desired genetic variants of targeted genes in mutagenized populations of both plants and animals, including agricultural crops (reviewed in Gilchrist and Haughn, 2005 and in chapter 1). Although we have used TILLING successfully to identify SNP variation, we found that this technique is labour intensive and takes

considerable time before the frequency of novel SNP variation in a population can be established (Chapter 3). A potential alternative is to use recently developed next-generation DNA sequencing technology as a tool to identify SNP variation in a mutant population that differs from that of the parental material. Next generation DNA sequencing technology is beginning to have a huge impact on a variety of molecular genetic research fields. Commercial DNA sequencing platforms, provided by Roche, Illumina and Applied Biosystems, are now used extensively; primarily for re-sequencing of known genomes and first-time sequencing of smaller genomes. The primary appeal of these methods lies with their extensive throughput. For example, one run on an Illumina machine can generate up to 40 gigabases of DNA sequence (www.illumina.com). On the downside, average read lengths of these technologies are shorter than that of traditional Sanger sequencing.

Novel applications of next generation sequencing technologies are also emerging at a rapid pace (reviewed in Mardis 2008; Morozova and Marra, 2008). Since the throughput is enormous and the cost per bp sequence is low, these technologies have the potential to be used for rapid and cost-effective screening of large populations for individuals that harbour genetic variations of interest. This application has yet to be exploited, primarily because only a few samples can be processed at once; the 454/Roche machine processes up to four samples at a time, while the Illumina machine processes eight samples (channels) simultaneously. Several studies have explored the incorporation of short “indexes” of specific nucleotide sequences in the primers used to generate

libraries for sequencing. These indexes allow the separation of DNA sequences from different samples after sequencing. In 2008, Illumina announced a method for multiplexed sequencing on their machine in which indexes are sequenced separately from the cloned fragment and therefore do not reduce the length of quality sequence recorded for each read. In addition, improved methods have been developed for evaluation of sequence quality as well as improved versions of software that take quality scores for deduced DNA sequences into account for reliable identification of Single Nucleotide Polymorphisms (SNPs). Thus, it appears that the time is ripe for using next generation sequencing technology as a high throughput method not only for recording consensus sequences, but also for recording rare SNP variation, including mutant identification in large populations.

Here we aim to explore the application of multiplexed Illumina sequencing in the identification of genetic variants in willow trees. The primary target are genes involved in lignin biosynthesis, since lignin is a polymer that is inhibitory to both biofuel and paper production. To this end, we created a pilot population of ~ 700 EMS-induced mutants from a fast-growing *Salix viminalis* variety and carried out a first screen by sequencing amplicons generated from this population. Here we will describe our first and still preliminary assessment of obtained results.

5.3 Materials and Methods

Basket willow (*Salix viminalis*) female clone 78195 and male clone 81084 from the Swedish Agricultural University in Uppsala, Sweden was used as

source material (described in Berlin-Kolm et al., 2009). Flower-bearing twigs were collected from the field in March of 2010 and brought into a greenhouse for flowering. Pollen was collected, dried at room temperature in the laboratory and stored until use in a refrigerator. Pollen germination frequency was tested by plating on solid pollen germination medium (PGM) followed by at least 2 hours incubation at room temperature before scoring for appearance of pollen tubes on an inverted microscope.

For mutagenesis, ethyl methane sulfonate (EMS; Sigma-Aldrich) was added to liquid PGM solution (containing 15% sucrose, 10 mM MES pH 6.5, 1.6 mM boric acid, and 2 mM calcium chloride, incubated for 2 hours on ice) at 0.3, 0.4 and 0.6 % concentrations (V/V) and thoroughly mixed. Roughly, 30 mg of pollen was mixed with 1 ml of EMS-containing PGM, and left on ice for 2 hours to allow pollen hydration while suppressing pollen tube growth. Pollen was allowed to sediment, and thereafter applied as a thick suspension with a micropipette directly onto female flowers with open stigma. Seeds were collected after approximately four weeks and directly sowed on wet filter paper in petri dishes for germination. Seedlings were transferred to standard potting soil and grown in a growth chamber for ~ 4 weeks before transfer to the greenhouse. Plants were re-potted in 1 litter pots, and placed in trays housing 15 plants.

One leaf of approximately the same size was collected from each of the 15 plants in a tray, pooled, and freeze-dried until DNA purification. A total of 52 pools were collected. DNA was extracted using the ChargeSwitch® gDNA Plant

Kit (Invitrogen), quantified, and assessed for integrity by gel electrophoresis. Gene fragments of interest were amplified by polymerase chain reaction using the Phusion DNA polymerase as per manufacturer's instructions (Finnzymes). A total of 48 DNA pools were used for amplification. Since the genome of any willow species has not been sequenced to date, we used PCR primers designed against genes of interest in the closely related poplar genome to amplify orthologues in the *S. viminalis* genome. Three different genes (2CCoAMT-2, 4CL3-1 and CAD-2) were amplified from each of the 48 pools. The three amplicons were pooled and subjected to gel electrophoresis to separate amplicons from contaminating primers and residual genomic DNA. Agarose with amplicons were cut out and DNA extracted using QIAquick® gel extraction kit (Qiagen). The obtained DNA samples were pipetted into 48 wells of a microtiter plate in portions of 100 ng. The procedure was repeated from the step of PCR amplification to generate an additional 48 pools of arrayed DNA from 5 genes (2CCoAMT-2, 4CL3-1, CAD-2, COMT-2 and Ga2-Ox), three of which were identical to the previously amplified genes. Approximately 30 ng of positive control samples, consisting of either a 15 or 30 fold dilution of mutant with wild type amplicon of the *Arabidopsis* GA3 gene and of the poplar CCR-2 gene, were added to several pools. The microtiter plate was taken to the BC Genome Science Centre in Vancouver for library constructions and DNA sequencing on the Illumina platform. The libraries were constructed by robotics using adapters that included a unique sequence index for each of the 92 libraries as well as for four control samples provided by BC-GSC. The 96

samples were pooled and loaded on one flowcell lane for sequencing. The sequencing procedure was repeated three times, with primers reading 75 bases of sequence by synthesis from each end of the insert, and a third primer to sequence the index tags. The obtained DNA sequences were converted to FASTQ format, and subjected to template-driven assembly against amplicon sequences previously obtained by Sanger sequencing using the NGen software (DNASTAR). Several PCR amplicons from each gene were cloned using TOPO TA Cloning® kit (Invitrogen) and sequenced (Sanger sequencing). The generated sequences from each gene were aligned and the consensus sequences were then used as reference templates for the alignment of the generated sequences from Illumina sequencing. Quality score criteria for selection of reads were set as stringent as possible while still allowing detection of GA3 positive control SNPs. The NGen SNP reporting module was used to generate a table of detected SNPs in the range between 0.5 and 7% of total recorded coverage of each DNA base.

5.4 Results

5.4.1 Generation of pilot mutant population

Traditionally mutagenesis is performed on seeds and there are procedures developed for this purpose for a number of species (Weil and Monde, 2009). For our purposes, seeds are an unsuitable source since the F1 plants will be chimeric and carry a large number of cell sectors with different patterns of mutations. We choose instead to mutagenize pollen, which after

fertilization will generate non-chimeric plants that are heterozygous for induced mutations. There are few reports available describing pollen mutagenesis. In maize, however, this approach is commonly used to generate mutant populations, using a procedure developed by M.G. Neufer (1994). In this procedure, EMS is mixed into paraffin oil to a final concentration of 0.02%, followed by addition of pollen and application to female flowers by the use of fine brushes. We assessed whether application of pollen in paraffin oil will work also in willow. Unfortunately, the application of pollen in mineral oil resulted not only in no successful fertilizations, but also killed the inflorescences (Figure 5-1, C). We also tested a water-based pollen germination medium (PGM) as carrier for pollen application. This approach was successful, generating a large number of fertilization events as judged by growth of fruits (Figure 5-1, B). We thereafter tested application of pollen that was suspended in PGM with EMS at concentrations of 0.3%, 0.4% and 0.6%. The 0.3% and 0.4% concentrations resulted in a much reduced seed set, to the extent that fruits failed to open, and about 50% of seeds failed to germinate. In control seeds, the corresponding rate is close to 100%. The 0.6% concentration resulted in complete failure to set seeds, despite application to more than 50 inflorescences. The surviving seedlings were transferred to the greenhouse, and arranged in flats with 15 plants.

5.4.2 Development of appropriate coverage for detection of rare SNPs

In discussion with specialists at the BC Genome Science Centre, we were advised to limit pool sizes so that the frequency of induced SNPs would be above 1% of total reads. In addition, he recommended designing the experiment such that each position was sequenced up to 1000 times, for accurate reporting of low-frequency SNPs. There is no precedence for our approach, so we followed his advice. Based on an informal estimate of the yield of sequence regularly obtained on the Illumina platform at BC-GSC, we settled for a pooling strategy involving construction and sequencing of 92 indexed libraries. In the first 46 samples (A1-46) 46 pools of DNA, each with DNA from 15 plants (30 chromosomes), were used for PCR amplification of 3 amplicons of ~ 1.2 kb. The 3 PCR products from each of the 46 pools were pooled and gel purified (see example in Figure 5-2) followed by DNA quantification. The same approach was used to generate a second set of 46 samples (B1-46) for construction of indexed libraries, this time with 5 amplicons, three of them identical to those used in the previous set. The two rounds were amplified separately to avoid incorporation of identical PCR-generated errors. In addition, the Phusion enzyme with proof reading activity was used to reduce PCR-induced errors. Rough calculations (by Jim Mattsson) suggested that combined sequencing of the resulting 92 samples may generate ~ 1000 times coverage of amplicon sequences in pools A1-46, and ~ 600 times coverage of amplicon sequences in pools B1-46, with a combined coverage of ~ 1600 times for the 3 amplicons common to all pools. Obtained coverage was on average above

calculated estimates, but also highly variable across the sequences. Two representative graphs of coverage are shown in Figure 5-3, with the top graph (A) from a pool of 3 amplicons, and bottom graph (B) from a pool of 5 amplicons. Both graphs show an overall similar profile, with extremely high coverage of both ends, followed by a drastic dip in coverage, and another peak of high coverage approximately 200 bases in from each end. In a longer stretch across the centre of the sequences, the coverage is relatively even, and above calculated levels, with > 2000 times average coverage in the pool of 3 amplicons, and ~ 700 times coverage in the pool of 5 amplicons. The top graph shows that two regions of approximately 30 bases have coverage below calculated values, and the bottom graph shows that these regions are slightly larger, probably as a consequence of the increased number of amplicons in this pool. These regions may be unsuitable for detection of induced SNPs. Overall, the obtained coverage is better than calculated values, leading us to proceed with the evaluation of positive controls included in some of the pools.

5.4.3 Detection of positive control SNPs

To assess whether pre-existing SNPs could be detected at expected frequencies, we introduced mixes of amplicons that differed in single base positions. These controls were generated by PCR amplification of homozygous wild type and homozygous mutant in the GA3 gene from *Arabidopsis* (GA3) and in the CCR-2 gene from poplar (pop). Sanger sequencing confirmed that these sequences differed only in a single base. Amplicons were pooled at a 1:30 dilution of mutant with wild type amplicon to mimic the presence of a

heterozygous mutant in a pool of 15 plants, and at a 1:15 dilution to mimic the presence of a homozygous mutant in a pool of 15 plants. Pools in both the A and B series were subsequently spiked with samples of these controls before library manufacturing. The predicted frequencies are 6.7% for a 1:15 dilution and 3.3% for a 1:30 dilution. Table 5-1 shows that the observed frequencies of these positive control SNPs were close to predicted frequencies for the GA3 amplicon. The CCR-2 control, on the other hand, reported 1.3 and 1.9% frequencies when we expected 3.3%. Since great care was taken in the pipetting of these samples, we took note that potentially small errors can reduce SNP frequencies considerably.

Moreover, since we knew that these SNPs were the only SNPs present in the GA3 and CCR-2 genes, we used the obtained assemblies of these contigs to define stringent criteria that filters out background SNPs. We found that when the SNP detection module of the DNASTAR Ngen software was set to Specific match size = 21, Specific match threshold = 95%, Quality score threshold = 20, the only SNPs that were detected above a 0.4% threshold were the expected SNPs. These criteria therefore provided suitable settings to screen for novel SNPs that may indicate induced mutations.

5.4.4 Detection of novel SNPs

The above criteria was used to screen pools for novel SNPs. Some SNPs were detected at frequencies well above expected ratios, and in many pools (not shown). These SNPs are likely to be natural SNPs present in screened genes in either of the two parents, and were therefore ignored. Table

5-2 shows a summary of the SNPs that were observed at a frequency close to expected 3.3%. With the exception of one, they are all detected at levels below expected frequencies, but also well above the cut off level of 0.4% used to select against false SNPs. So far, we have identified 11 SNPs that were present in both libraries. If these SNPs were all confirmed, we would obtain a frequency of ~ 1 SNP per 150 kb which is in agreement with the frequency commonly obtained in many TILLING populations (Table 2-1). We plan to assess whether these SNPs are present in any of the plants that were used as source material for the corresponding pools.

5.5 Discussion

While mutation-based breeding has been immensely successful in agriculture, it has not been applied to forest tree breeding, primarily due to the immense resources and time it would take to generate the F2 and F3 populations required to screen for desired mutants based on expressed phenotype/trait. Detection of desired mutants based on genotype in F1 populations may, for the first time, enable the use of mutation-based breeding in tree species and unleash the enormous phenotypic plasticity likely to exist in tree species. Here we also took the novel approach of mutagenizing willow pollen rather than seeds, or calli, to avoid the generation of chimeric plants. There is a strong interest in breeding that is targeted at biofuel feedstock species. Here we present evidence in support of a novel approach to the breeding of an important biofuel crop, willow. We also have data, albeit

preliminary, suggesting that we have identified novel SNPs that are not present in parental material and may therefore be due to induced mutation.

In theory, mutant identification by multiplexed high-throughput sequencing promises to be (1) rapid, since it allows simultaneous screening of a population for mutants in multiple target genes, and (2) eventually cost-efficient, primarily by reduced salary costs and immediate identification of non- and miss-sense mutations. Here we have taken a first exploratory step to assess how this translates into reality. While we have been waiting for sequence data to arrive for a considerable amount of time, the effort that went into preparation of library material as well as subsequent software analysis of SNPs was much less onerous than what we have experienced with the classical TILLING technology.

The experiment described here generated sequencing coverages that in general were higher than predicted. At the same time, most of the identified putative mutants occurred at a frequency that was below the expected values. These issues may be rectified in future screens by reducing the size of pools to increase the frequency of putative mutant SNPs, and at the same time screen a larger number of amplicons for mutations to reduce coverage. Another issue is the uneven coverage of sequence reads across amplicons, with double peaks of unnecessarily high coverage close to sequence ends, with a dip in coverage in between the two peaks with insufficient coverage. This uneven distribution is probably an artefact from the fragmentation step during the library manufacturing process. The ends of the amplicons will naturally be over represented in the final library, and constitute the first peak of high coverage at

the two ends of assembled contigs (see Figure 5-3). The second peak at both ends probably comes from the paired-end read of some of the fragments that generated the first peaks. The dip in coverage between the peaks is probably due to a selection of fragment sizes for library construction that are 200 bp or longer. As a consequence, the paired-end 75 base reads from each end left a gap of approximately 50 bases of low coverage at the centre of the 200 bp fragments. We suggest two relatively simple solutions to minimize uneven coverage. First, select for fragments down to 150 bp in size to avoid the dip in coverage between peaks. Second, extend, if possible, the length of amplicons, thereby reducing the overall impact of high coverage of ends, and most likely get a longer internal region of relatively even read coverage

In summary, we have identified putative SNPs of interest, and if any of these SNPs can be confirmed, the laboratory will continue to scale up and improve this effort to develop a novel breeding tool for willow, a bioenergy crop of increasing importance in the northern hemisphere of the world.

5.6 Figures

Figure 5-1 Test of medium suitable for EMS mutagenesis of *S. viminalis* pollen.

Picture shows female inflorescences 2 weeks after application of pollen by three different methods: (A) Pollen applied with pipe cleaner (positive control). (B) Pollen re-suspended in PGM and thereafter applied by micropipettor to pistil stigmas. (C) Pollination suspended in paraffin oil and applied by micropipettor. Inflorescences in (A) and (B) show enlarged pistils indicative of successful pollination, whereas inflorescence in C show no enlarged pistils, indicative of failed pollination.

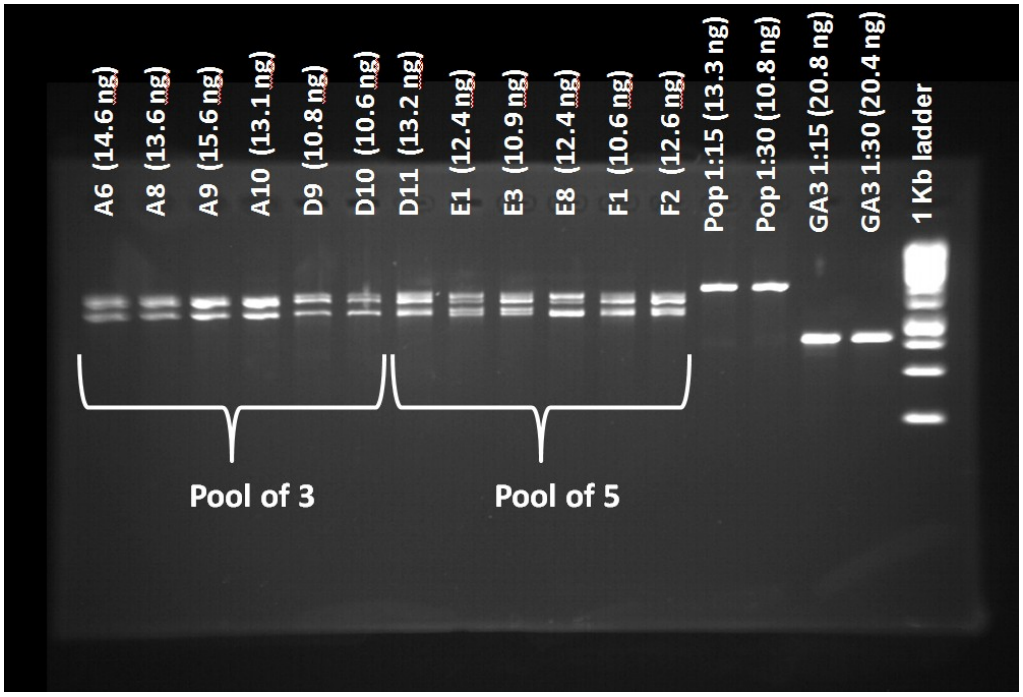


Figure 5-2 Sample of the gel-purified pools of PCR products from samples that were sent for sequencing using Illumina sequencing.

Twelve random samples from the 94 samples that were gel-purified were loaded into a 1.2% TBE agarose gel before sending them for sequencing. The pools were PCR products from either 3 genes or 5 genes groups. Pop 1:15 and Pop 1:30 are positive controls from poplar with known point mutation (G→A) mixed in 1:15 and 1:30 ratios. GA3 1:15 and GA3 1:30 are positive controls from *Arabidopsis* with known EMS-induced mutation (G→A) mixed in 1:15 and 1:30 ratios.

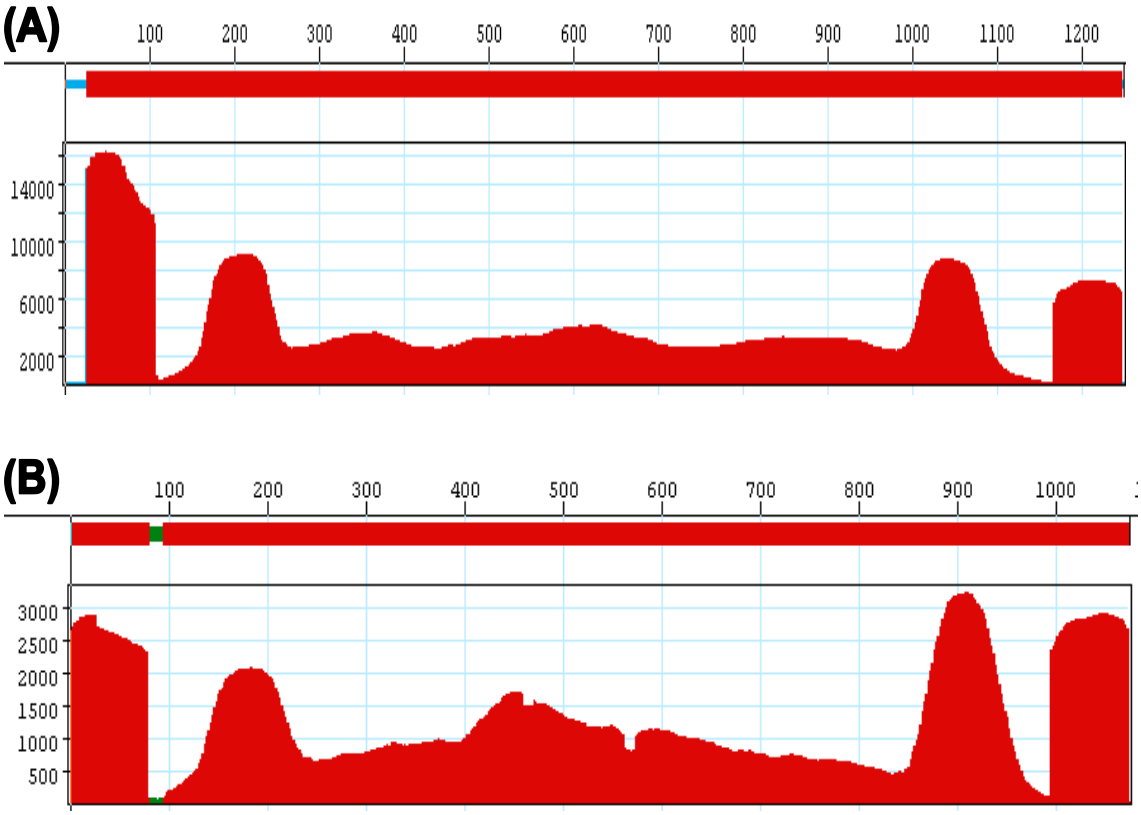


Figure 5-3 Examples of obtained coverage from Illumina sequencing of indexed pools of amplicons.

(A) Shows coverage of the 2CCOAMT2 amplicon from pool 2A and (B) shows coverage of the COMT-2 from pool 35B. Amplicon DNA sequence numbered from first to last base on X-axis, coverage on the Y-axis. The red areas illustrate coverage for each base along the amplicon sequence.

5.7 Tables

Table 5-1 The detection of positive controls.

Positive control	Pool	Dilution	SNP frequency
GA3	27A	1:15	6.5%
GA3	26A	1:30	3.3%
GA3	41B	1:15	7.1%
GA3	4B	1:30	4.1%
Pop	12A	1:30	1.9%
Pop	39B	1:30	1.3%

Table 5-2 Detected SNPs from the *Salix* population.
The SNPs listed in this table were detected in both libraries.

Gene	Pool	Position	SNP	SNP frequency	Depth	Effect on amino acid
2CCoAMT-2	2A	1187	G→A	1.1%	6686	Ala→Ala
2CCoAMT-2	30B	135	G→A	1.2%	1208	Ser→Asn
2CCoAMT-2	30B	145	G→A	1.6%	1772	Glu→Glu
4CL3-1	21B	771	C→T	1.2%	1228	Asp→Asp
4CL3-1	21B	789	C→T	1.1%	2276	Pro→Pro
COMT-2	46B	222	C→T	1.1%	1458	Cys→Cys
COMT-2	35B	264	C→T	1.3%	671	His→Tyr
CAD-2	20B	468	C→T	1.2%	1019	His→His
CAD-2	4B	507	G→A	3.8%	913	Gly→Arg
CAD-2	27B	521	G→A	1.3%	898	Ala→Ala
CAD-2	25B	647	G→A	1%	2734	Gly→Gly

5.8 References

Ball, J., Carle, J., & Del Lungo, A. (2005). Contribution of poplars and willows to sustainable forestry and rural development. *UNASYLVA-FAO-*, *56*(2), 3. UN FOOD AND AGRICULTURE ORGANIZATION. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Contribution+of+poplars+and+willows+to+sustainable+forestry+and+rural+development#0.

Berlin, S., Lagercrantz, U., Arnold, S. von, Ost, T., and Rönnberg-Wästljung, A. C. (2010). High-density linkage mapping and evolution of paralogs and orthologs in Salix and Populus. *BMC genomics*, *11*, 129.

Dickmann D., Isebrands J., Eckenwalder J. and Richardson J. (Eds.). (2001). *Poplar Culture in North America.* Ottawa, Canada: National Research Council (Canada) Research Press.

Gilchrist, E. and Haughn, G. (2005). TILLING without a plough: a new method with applications for reverse genetics. *Current opinion in plant biology, 8*(2), 211-215.

Hill, J., Nelson, E., Tilman, D., Polasky, S., & Tiffany, D. (2006). Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(30), 11206-10.

Keoleian, G., and Volk, T. (2005). Renewable Energy from Willow Biomass Crops: Life Cycle Energy, Environmental and Economic Performance. *Critical Reviews in Plant Sciences*, *24*(5-6), 385-406.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, *24*(3), 133–141.

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annual review of genomics and human genetics*, *9*, 387-402.

Morozova O and Marra MA. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 255-64.

Natural Resources Canada. (2010, April 26). Using willow for biomass production and wastewater disposal. Retrieved from http://canadaforests.nrcan.gc.ca/article/willow

Neuffer, M.G. (1994). Mutagenesis. In M. Freeling, and V. Walbot (Eds.), The Maize Handbook (pp. 212-219). New York: Springer-Verlag.

Ragauskas et al. (2006). The path forward for biofuels and biomaterials. *Science, 311*, 484-489.

CHAPTER 6: THESIS SUMMARY

A key approach to plant breeding has been to induce genetic variation by various means followed by self fertilization and screening for desired phenotypes in the F2 generation. Due to the long generation time and the great difficulties associated with controlled large-scale self fertilization in forest tree species, this approach has never been undertaken in forest tree species.

The research herein is aimed at developing procedures that allow for the identification of mutants in tree species based on genotype, thereby circumventing the almost insurmountable problem of large-scale selfing to uncover homozygous mutants based on phenotypes. We have taken two major approaches to both mutagenesis and screening. First, we mutagenized poplar calli with ethyl methane sulfonate (EMS), followed by regeneration of plants. We were able to successfully generate a *Populus* population, consisting of >6,000 plants, from EMS-treated calli of hybrid poplar using clonal micropropagation. We tested several DNA extraction protocols and found a suitable high-throughput protocol that is fast and that can generate good quality DNA that can be used for TILLING. We successfully extracted DNA from >6,000 plants using ChargeSwitch® gDNA plant kit. Treated population of hybrid *Populus* was generated and Targeting Induced Local Lesions IN Genomes (TILLING) technology was used to screen for mutants harbouring desired traits that can be used, and for the first time, in breeding. We also established TILLING in the

laboratory at SFU by developing a simplified TILLING protocol that is faster, less expensive and easier to analyze. Screening by the modified TILLING protocol identified the presence of single nucleotide polymorphisms (SNPs) or point mutations. Further analysis proved that they were pre-existing natural SNPs in the hybrid *Populus* lines used. We concluded that we were successful in generating EMS-mutagenized poplar trees, but, the frequency of point mutations or EMS-induced SNPs were possibly low and could not be identified using TILLING and that we probably needed a more sensitive technique for SNP detection.

We also successfully generated putative mutant population of willow trees (*Salix viminalis*), this time by treating the pollen with EMS, and we screened for mutants by high-throughput DNA sequencing. We were able to generate a non-chimeric EMS-mutagenized population of willow, consisting of ~700 trees. We extracted DNA using ChargeSwitch® gDNA plant kit and PCR amplified five genes of interest. The generated PCR products were gel purified and quantified, and then sent to BC Genome Science Centre in Vancouver for library construction and sequencing using Illumina sequencing. We were able to identify the positive controls that were included with the samples and generate stringent criteria that were then used for detecting EMS-induced SNPs. We have identified 11 SNP variants that are not present in either of the parents and that may therefore be due to EMS-induced mutation. We plan to assess individual plants in the pools that show novel SNPs to confirm the presence of novel SNPs indicative of mutation.

We also successfully generated a population of *Arabidopsis* plants from lines that are mutated in genes that are homologous to the genes selected in our *Populus* study, and that were tested for haploinsufficiency. Depending on the type of mutation, the generated *Arabidopsis* populations were genotyped using antibiotic and herbicide resistance, PCR, sequencing, or TILLING. In some cases, a combination of several screening techniques was used. After testing several lignin analysis techniques, we concluded that HR-TGA analysis was most suitable for our lignin analysis. Based on the results obtained from HR-TGA, we concluded that mutation in the CAD-1 gene had no effect on total lignin content and did not result in haploinsufficiency. We also analyzed the effect of mutations on the plants' morphology, including plants' total height and siliques length. Based on the effects on the plants' morphology, results showed that mutations in the GA3, CAD-1, *IRX4*, *TT5-1, GA5-1* and *FAH1-5* genes did not result in haploinsufficiency. Up to date, only mutants from CAD-1 had been analyzed with respect to lignin content. Further lignin analysis using HR-TGA analysis will be carried on the other mutant populations including mutants in IRX4 and FAH1-5 genes.

In summary, we have endeavoured on the development of novel, but also risky, breeding technology based on genomics technologies. So far, we have obtained encouraging, albeit preliminary, results using a combination of pollen mutagenesis and screening for mutants by the use of high-throughput DNA sequencing technology. We have also identified methods and materials

that will be valuable for further refinement of approaches and eventual success of the project.