AN ABSTRACT OF THE THESIS OF

Surbhi Nahata for the degree of Master of Science in Environmental Sciences presented on June 12, 2020

Title: Sequence Analysis and Vegetative Growth in Transgenic *Eucalyptus* with CRISPR-Cas9-induced Mutations in the *Eucalyptus* Homologs of the Floral Genes *EMBRYO DEVELOPMENT ARREST 33* (*EDA33)* and *TAPETAL DEVELOPMENT AND FUNCTION1 (TDF1)*

Abstract approved:

_____

Steven H. Strauss

An impediment to use of exotic and bioengineered trees in many places is their propensity for spread by pollen and/or seeds. Our laboratory has been using gene editing to induce mutations in floral genes as means to impart stable and reliable genetic containment when this is desirable from social (markets, regulation, public opinion) or ecological perspectives. Our studies focus on one of the most widely planted, productive, and sometimes invasive forest tree species, *Eucalyptus grandis x Eucalyptus urophylla*.

We used CRISPR-Cas9 to target two genes expected to be important to reproductive development in *Eucalyptus*. Based on studies in the model plant *Arabidopsis*, proteins encoded by the eucalypt homolog of *EDA33* (*EMBRYO DEVELOPMENT ARREST 33*) are expected to be essential for seed dispersal, and proteins encoded by the homolog of *TDF1* (*TAPETAL DEVELOPMENT AND FUNCTION1*) are expected to be essential for pollen development. Bioinformatic studies have shown that each has a single putatively orthologous, functional locus in

the *Eucalyptus* genome. This study focused on determination of the nature of mutations induced, with an emphasis on identifying biallelic loss-of-function mutations, and effects of loss-of-function mutation on vegetative growth and morphology in the greenhouse. Ongoing work in our laboratory is studying the effects of mutation of these genes on floral development and fertility.

We generated two CRISPR-Cas9 constructs, each with two guide RNAs that target each gene, and one Cas9-only (no guide RNAs) control construct. We then *Agrobacterium*-transformed hybrid eucalypt SP7 (*E. grandis x urophylla*), for which an efficient transformation system was available. We generated 17 and 18 gene insertion events for the *EDA33* and *TDF1* construct, respectively, and ten events for the Cas9 control. Taking advantage of the natural polymorphisms in the hybrid that we discovered after resequencing near to the target region, we generated and sequenced allele-specific PCR products for each locus. We identified ten insertion events in *EDA33* and six in *TDF1* events for which there were clear loss-of function mutations in both alleles. Overall, CRISPR-Cas9 was highly efficient in generating knockout mutations, with a rate (per transgenic regenerated shoot) of 83% in *EDA33* and 46% in *TDF1*. The knock-out events and five Cas9 control events were then micropropagated, acclimated, and transplanted into 942 cc^3 pots for a greenhouse trial to study variation in plant growth rate and leaf morphology.

The three month greenhouse trial used a randomized block design with an average of six ramets per genotype and 14 non-transgenic controls. We measured height and diameter at the start and end of the trial, and relative chlorophyll density, leaf area and weight, and leaf oil gland density at the end of the trial. In a linear

mixed effect model, we conducted an overall F-test where blocks and construct were fixed effects and events were considered random effects.

We found that blocks were highly statistically significant sources of variation ($P<<0.05$), which was not surprising as trees were sorted into blocks based on size at the start of the experiment. Presumably due to variation in propagation history, the wild type controls were larger at the outset of the study and continued to grow significantly differently from the transgenic trees, so comparisons between the knock-out and non-mutated events were predominantly based on the transgenic controls. At harvest, none of the traits were different at the 5% significance level among the transgenic genotype groups, though leaf area and old gland density were different at the 10% significance level. Equivalence testing of the differences in means of plant characteristics measured at harvest between the knock-out groups and transgenic controls showed that there was unlikely to be a 10% or greater reduction of any of the trait values in the knock-outs. Our results suggest that the functions of these genes are largely restricted to reproductive development in eucalypts, and imparting sterility through their loss-of-function is unlikely to affect productivity. Longer term studies, preferably in the field under normal forestry conditions, are needed to confirm these results.

Sequence Analysis and Vegetative Growth in Transgenic *Eucalyptus* with CRISPR-Cas9-induced Mutations in the *Eucalyptus* Homologs of the Floral Genes *EMBRYO DEVELOPMENT ARREST 33* (*EDA33)* and *TAPETAL DEVELOPMENT AND FUNCTION1* (*TDF1)*

by
Surbhi Nahata

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented 6.12.2020
Commencement June 2021

Master of Science thesis of Surbhi Nahata presented on June 12, 2020

APPROVED:

Major Professor, representing Environmental Science

Director of the Environmental Science Program

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Surbhi Nahata, Author

ACKNOWLEDGEMENTS

This journey of acquiring a master's degree has been exciting and enlightening. I am grateful for my mentors in the lab, academia, and my personal life. First, I would like to thank my major advisor Dr. Steve Strauss, for his constant support and understanding with my thesis. He truly is an inspirational teacher and a researcher; I have been fortunate to have been mentored and advised by him. His high standards for scientific research have pushed me to work hard and diligently. He also values science communication skills, preciseness, and high ethical standards in research, which has been a strong motivation for me. Moving forward I hope to incorporate such personality traits in my professional life. I also thank him for being very patient with me, despite the difficulties I faced with switching my major area and because I come from a different country with a very different education system and language. Second, I am deeply appreciative of the help provided by Dr. Estefania Elorriaga. I could not have imagined coming this far without her tutelage in the molecular biology lab and life in a scientific field. Again, she is a very inspiring and a strong woman. I value her knowledge and work ethics; I aspire to be like her one day. Third, I would like to thank Michael Nagel for helping me with understanding biotechnology techniques; I always ended up learning new things after conversing with him. Fourth, I would like to thank Cathleen Ma, her love for her profession is motivational. She has helped and mentored me in the tissue culture lab and greenhouse, and I cannot thank her enough for teaching me all the techniques with plant tissue culture. Fifth, I would like to thank Bahiya Zahl, for helping me with my research. Her articulate understanding of molecular biology is inspiring. Sixth, I would like to thank my committee members for helping me improve and being patient with me. Lastly, I cannot forget to thank my parents and the maternal side of my family, and especially my uncle, for providing me with their love, support, and encouragement. I would not have gotten this far without their care.

CONTRIBUTION OF AUTHORS

Michael Nagel designed the CRISPR-Cas9 sterility constructs and performed cloning and sequencing of SP7 wild type. Cathleen Ma performed the transformation aspects of plant tissue culture including propagation, cocultivation with *Agrobacterium*, regeneration/selection, and PCR confirmation. Dr. Estefania Elorriaga helped with primer designing and mutation analysis. Dr. Steve Strauss oversaw all the research performed in the studies.

This work was supported by USDA Biotechnology Risk Assessment Research, grant no. 2017-33522-27098 and the TBGRC industrial cooperative at Oregon State University. We thank Futuragene/Suzano for providing the SP7 hybrid eucalypt clone and transformation method.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

Figure Page

LIST OF TABLES

LIST OF APPENDIX FIGURES

Figure Page

LIST OF APPENDIX TABLES

Chapter 1: Scientific background

1.1 Scope and organization of thesis

1.1.1 General research goals

Field evaluation and commercialization of genetically engineered (GE) trees have been severely hindered due to a negative public reception and strict regulatory barriers. One kind of technology that may reduce negative perception and regulatory barriers is genetic containment, for which CRISPR-directed mutation of flowering-essential genes may be the best method. This research is part of a larger ongoing project in our laboratory directed toward this goal, with my focus on the rate of CRISPR-mutation, and whether floral mutations affect vegetative development (effects on flowering are under study by others in the laboratory).

The larger project can be categorized into the following major projects:

a) To examine the phenotypes (both vegetative and floral) of CRISPR-mutated, transgenic eucalypts through greenhouse trials in normal and rapid flowering genotypes;
b) To examine the extent of off-target mutagenesis from CRISPR expression; and
c) To develop and assess the effectiveness of an excision system for somatic removal of gene editing machinery.

This document contains information pertaining to the methods that were used for finding biallelic knockout mutants (i.e., where both copies of the target genes are rendered non-functional). Here we present the results of mutation assessments and a greenhouse trial to study vegetative growth analysis of different experimental groups, including transgenic vs. non-transgenic lines, and knock-out vs. non-mutated transgenic plants.

1.1.2 Organization of the document

This thesis contains three major sections each in general manuscript style, beginning with a broad literature review of areas important for my master's studies, methods and results of the research, and discussion and interpretation of results. My work will be part of a larger manuscript that includes results for flowering once that work is complete. Hence, the manuscript style is intended as a pedagogic exercise in scientific writing.

1.2 The *Eucalyptus* genus

The genus *Eucalyptus* includes over seven hundred species of flowering trees (Brooker and Kleinig, 1994). Within the family Myrtaceae, eucalypts are distinguished from other genera based on a lack of petals and the presence of opercula that covers the floral buds (Boland et al., 1984).

1.2.1 Morphology and reproduction of *Eucalyptus grandis*

For our research we used a *Eucalyptus grandis* x *urophylla* hybrid due to it ease of vegetative propagation and transformation *in vitro*. The genome size of *E. grandis* is around 640 megabase pairs (Myburg et. al, 2014). The chromosome number for the organism is 11-12 pairs and shows diploidy, which was suggested to help provide *Eucalyptus* with high genetic diversity, and thus adaptability to a range of environmental conditions (Rye and James, 1992). *E. grandis* is a tall, straight tree that could reach up to a height of 46m. Its bark, along with being smooth and deciduous, is white in color (Hill and Maiden, 1918). *Eucalyptus* displays opportunistic growth based on environmental conditions (Jacobs 1955; Pook 1984a). Its longevity can span from 100 to a maximum of 500 years. Leaves have an alternate arrangement on the plant and can be categorized into three stages of development: juvenile, intermediate, and mature. Leaf venation is moderately fine and oil glands for essential oils are present in the veins of the leaves. The characteristic aromatic odor of eucalypts comes from the volatilizing essential oils (Boland *et al*., 1991). Flowers are mostly bisexual, and a unit inflorescence consists of 1 to 11 flowers born on a common stock, the peduncle. *E. grandis* has an axillary umbel

with 3-10 flowers and floral buds are pyriform that are usually contracted in the middle, pedicellate and glaucous (Blakely, 1955). The color of the eucalypt flower is determined based on the stamens. Stamen are usually brightly and prominently colored. After fertilization, stamens fall off and the bud matures into a woody fruit, which dehisces longitudinally along the length of peduncle when the ovary roof splits into sectors (Boland *et al*., 1984). Pollination in *Eucalyptus* usually depends on the region, for example in temperate ecosystems wind pollination is observed whereas in tropical and sub-tropical regions animal pollination is more common (Regal, 1982). The *Eucalyptus* breeding system is of mixed mating type with preferential outcrossing (Griffin et al, 1987).

1.2.2 Global distribution

Eucalypts are indigenously found across the Australian continent and islands to its north (Hall et al. 1963, Williams and Potts 1996). However, owing to the economic and medicinal importance of eucalypts, they were introduced as an exotic species to other regions including South and North America, Africa, Europe, the Mediterranean Basin, the Middle East, China and the Indian subcontinent (Sellers, 1910).

1.2.3 Ecosystem services and uses

Eucalypts provides various ecosystem services and environmental benefits. They can be used reduce excess carbon dioxide in the atmosphere, provide stormwater benefits and purify air (UC San Diego Campus Forest, 2009). It also has various cosmetic and medicinal applications. The essential oil derived from eucalyptus is used as raw material in perfumery, food, beverages, aromatherapy and phytotherapy. Eucalyptol chemical derived from the extract contains antimicrobial and anti-inflammatory qualities (Minto et al. 2016). Additionally, eucalypts are used in plantation industry as a fast-growing rotation tree for timber, pulp and bioenergy. Primary producers of wood products derived from eucalypts are tropical countries like Brazil, Spain, Portugal and south Africa. Brazil is the largest exporter of eucalypts wood products. Plantations in Brazil are mostly owned by international and national companies like Suzano. It has been estimated that Brazil in

2005 had exported around US$3.5 billion amount worth of timber and paper products. It is believed that companies own around 1 million hectares of land, out of which 60% is dedicated for *Eucalyptus* plantations (global forest coalition, 2018).

Due to an increased interest in renewable energy resources, biofuel feedstock obtained from *Eucalyptus* holds potential as an alternative to petroleum-based products. Since quality requirements of biomass fuel are lower than for pulpwood, eucalypts are highly suitable due to their high wood density and shorter rotation times, thus maximizing yield (Spinelli et. al, 2016). *Eucalyptus* can be harvested for pulpwood and fiber within 6-8 years. Owing to a uniform and high fiber content, *Eucalyptus* pulp has been used for coated and uncoated free-sheet paper, bleach board, sanitary products and in cardboard boxes. However, there is also increased use of *Eucalyptus* for solid wood purposes, such as sawn wood, veneer and medium density fiberboard—as well as extenders in plastic and timber (Myburg, Potts et. al, 2007). Besides above mentioned uses of *Eucalyptus*, it has also been used in some countries like India as windbreak, and to control diseases such as malaria as part of wastewater management.

In *Eucalyptus*, growth, wood quality, disease resistance, and vegetative-propagation-related traits are considered as commercial traits of importance in breeding programs. According to Myburg, opportunities in *Eucalyptus* molecular breeding and genome research exists in community linkage maps, an integrated physical and genetic linkage map, comparative genome mapping, association genetics, integrative and comparative genomics.

Figure 1.1: *Eucalyptus grandis* flowers. The white filamentous structures are the anthers.URL: (https://www.flickr.com/photos/mercadanteweb/5778708890)

Figure 1.2: *Eucalyptus grandis* plantation in Hawaii (URL: https://www.flickr.com/photos/starr-environmental/24562059723)

1.3 Application of biotechnology in forestry and agriculture

1.3.1 A brief history of tree breeding, genetic modification and genetic improvement

I will write using the following definitions, which can vary among authors -

a) Tree breeding: Similar to natural selection, i.e. through natural mutation, recombination and selection pressure, tree breeding comprises artificial selection of individual or plus (superior phenotype) trees which is directional (i.e. focused on socio-economic needs and adaptive requirements) followed by controlled crosses and propagation of elite families or clonal varieties. Improved varieties of plantation trees are primarily used for industrial forestry but they can also be employed in agro-forestry or for enrichment of local forests (Pâques, 2013).
b) Forest genetics: It is the study of hereditary variation in forest trees (Wright, 1976).
c) Genetic modification: It involves unintentional or purposeful tweaking of genetic constitution of an individual tree or an entire population for human needs (Burdon and Libby, 2006). Unintentional changes can be both detrimental and beneficial.
d) Genetic engineering: It is the utilization of recombinant DNA methods for isolation, configuration, modification and transfer of genes in plants (James et al. 1998).
e) Tree improvement: “Application of forest genetics along with other fields such as tree biology, silviculture and economics, to develop genetically improved varieties of forest trees” (White, Adams, and Neale 2007).

Wood has been used throughout the human history from the bronze age to modern day as the major means for provision of shelter and energy for many societies. Human mobilization around the globe has been possible by employing wood and timber for basic purposes such as creation of fire, transportation, shelter and metal extraction. Wood has been used in construction of ships and as energy source in early locomotives. Fire generated through wood had provided early humans with heat necessary to survive in cold temperatures, bake bread from grain and derive salt from seawater (John Perlin,

1989). It is believed that hunter-gatherer humans dependence on wood for purposes such as cooking, shelter, tools and heating has influenced biodiversity of trees and genetic composition of selected species in a forest (Libby and Burdon, 2006).

There have been many episodes of deforestation in the past, but only recently have we understood the needed technique for sustainably maintaining forests and plantations for timber. Although, modern tree breeding began after the second world war with international and regional cooperation (du Cros, 2000), earlier efforts with the practice of tree breeding in the west was shown by John Evelyn. Upon a request from the British navy commissioners and The Royal Society, John Evelyn in his report "Sylva Or A Discourse of Forest-Trees and the Propagation of Timber in His Majesties Dominions" was presented on October 15th, 1662. It provided exhaustive information about the variety of trees that could grow in England along with their cultivation and uses (Evelyn, 1679). Moreover, there have been a few tree breeding and conservation acts that were supported by the British monarchy around 1600's, for instance, "Act for the Better Breeding, Increasing, and Preserving of Timber and Underwoods" (House of Commons Journal, 1610). However, the first documentation of scientific tree breeding is from 1749 in Sweden where a Marine command regarding the choice of seed source for *Quercus robur* (pedunculate oak) was put forward for ship construction (du Cros, 2000).

There is a vast literature on tree breeding, but I'll briefly touch on this subject as most of that literature is beyond the scope of this document. Du Cros, (2000) suggests the following prerequisites for tree breeding:

a) Presence of genetic variation of economic traits
b) Inheritance of individual traits
c) Knowledge of genotype environment interaction
d) A comprehension of breeding systems and genetic compatibility among species to be cultivated
e) An understanding of financial implications of tree breeding

Knowledge accumulated in the area of tree breeding and domestication is recent compared to domestication of agricultural crops. Forest genetics and tree improvement programs have helped in tree breeding. Biotechnology and genomics are usually used as modern tools to speed up the process of tree breeding and get improved wood products for biofuel, pulpwood and timber industry. Tree breeding starts with having genetic variations in the economic traits (du Cruss, 2000). Genetic markers are DNA phenotypes that reflect differences among population, species and individuals (Myburg et al., 2007). The variability in these markers have helped in developing genetic maps which can further help in various studies like Genome Wide Association study (GWAS) and Marker Assisted selection (MAS). GWAS and MAS help in selection process by selecting genotypes that might provide best phenotypes in given environmental conditions. Genomics can aid in understanding the influence of multiple genes [quantitative trait loci (QTL)] on the phenotype. Examples of some markers that can be used to study genetic variations in a population are SNPs (Single Nucleotide Polymorphism), RAPDs (Random Amplified Polymorphic DNA) and RFLPs (Restricted Fragment Length Polymorphisms).

The appearance of recombinant biotechnology further advanced tree breeding by providing researchers with the capacity to introduce novel genes from the same species (cisgenics) or a different species (transgenics). *Agrobacterium tumefaciens* mediated transformation of trees aids with introduction of desired genes in the plant genome (Nester, 2008). This ability to insert foreign DNA into a plant genome can help with cell wall modification, yield improvement, forest conservation and modification of flowering.

1.3.2 Tissue culture in forest biotechnology and clonal forestry

Deployment of tissue culture techniques in breeding program has been revolutionary. Tissue culture includes various techniques for *in vitro* cultivation of plants, seeds, organs, and embryos in a controlled and sterile environment. The primary techniques that are used in tree biotechnology entail micropropagation and organ regeneration. In micropropagation, a fully functional plant is generated using meristem tissues such as axillary primordia (Ikeuchi, 2016). Steps involved in micropropagation are: shoot multiplication within an aseptic media, shoot induction, and transfer of these plants into

soil (Gosal et al. 2010). Organ regeneration is the generation of a whole plant from an explant or mass of plant cells (Hill, 2013). These cells are usually somatic and totipotent, and they can specialize into any type of plant cell that these are derived from. Steps involved in organ regeneration are: callus induction, root formation and the transfer to soil media. Cytokinin to auxin ratio in the media are important in both kinds of propagation. Vegetative propagation is important since they help with asexual reproduction of plants and trees.

Molecular genetics, biotechnology and clonal forestry go hand in hand; the combination of these three fields help in maintaining and multiplying commercially important traits that might have been genetically engineered at an economical expense (Ahuja and Libby 1993). Clonal forestry can be defined as "large-scale deployment of relatively few, known superior clones that have proven their superiority in clonal tests." (environmentalpollution.in/forestry/clonal-forestry, S. Agarwal). In forestry, a clone can be defined in three ways:

a) Traditionally, a vegetative propagule of a plant;
b) In cell culture, a cell line that originated from different single cells of the parent organism;
c) In terms of molecular biology, bits of DNA or genes that have same nucleotide base pair sequence and are amplified to make many copies for study or use. (Ahuja and Libby 1993).

Clonal forestry plantations have been widely used with several plant species such as *Populus*, pine, and *Eucalyptus*. The ease of propagating rooted cuttings is a major consideration for commercial clonal forestry (Myburg et al., 2007). One of the disadvantages of clonal forestry is the susceptibility of tree population to a pest or insect due to lower biodiversity in the plantation. Clonal forestry has been widely used with *Eucalyptus* in countries like Brazil and South Africa. Many plantations use stecklings (rooted cuttings) instead of seedlings for producing new *Eucalyptus* trees. Clonal forestry captures full genetic potential of genotypes and gives a greater uniformity in *Eucalyptus* plantations.

1.3.3 Biotechnology for crop improvement

Biotechnology is broadly defined as "the use of living systems and organisms to develop or make products (Wikipedia 2020). More specifically, Biotechnology includes genetic engineering, genomics, genetic modification, application of genetic markers and vegetative reproduction methods (Sedjo, 2001). It can assist with traditional crop breeding by introducing or modifying genes that can help with pest resistance, increase nutritional value, and improve yield. Genomic tools and resources further facilitate crop breeding by helping researchers understand genotype-phenotype relationships (Castro et. al, 2012). Through utilizing next-generation sequencing and bioinformatics tools, we can discover new genes, regulatory sequences and develop molecular markers. Genomic selection, where entire genomes are indexed for making breeding selections, expedites the process of selection of traits that have economical value and provide resistance against diseases. Besides *in vitro* propagation of plants through tissue culture techniques described earlier, crops can be propagated both sexually and asexually. Asexual propagation includes techniques such as tissue (root, stem, cane and leaf bud) cutting, layering, separation and division, budding and grafting, and micropropagation. Sexual propagation involves seed germination, techniques to break dormancy, growing plants from seed and by transplanting seedlings (ncsu.edu/extension-gardener-handbook/13-propagation).

Crop domestication and improvement has been critical to the development of human civilization. Domestication of crop plants in the new world likely began around 9,000 years ago in the Balsas River Valley of tropical southwestern Mexico, where maize was obtained by genetic modification of Balsas teosinte (Hastorf, 2009). Advancement in food domestication over the years helped address people's demand for food. With the global human population projected to reach 9.7 billion by 2050 and to be mainly concentrated in developing countries (United Nations et al., 2019), problems of poverty, malnutrition, food security and degradation would worsen. To feed the growing population large masses of land will be needed for agricultural purposes causing land and forest degradation, soil erosion or salination and global warming. Along with the

concerns mentioned above, various kinds of pests and pathogens also lead to a loss in crop productivity due to their increased susceptibility as a result of climate change. Hence, application of biotechnological tools can be one of the approaches to tackle those issues (Miflin, 2000).

Agricultural biotechnology helps in improving crop productivity in many ways, including by reducing application of pesticides on crops and improving crop yield and nutrition (Persley, 1999). In 2018, around 70 countries had 191.7 million hectares of land planted with biotech crops. A ~113 fold increase in planting of biotech crop around the globe has been observed since 1996, with an accumulated area of 2.5 billion hectares. Top producers of biotech crops include five countries: the USA, Brazil, Argentina, Canada and India accounted for a total of 91% of global biotech crop area (ISAAA, Brief 54). According to the ISAAA, in the USA approximately 75 million hectares of land were planted with biotech crops such as soybeans (34 million hectares), maize (33 million hectares), cotton (5 million hectares), canola (900,000 hectares), sugar beets (491,000 hectares) and alfalfa (~1 million hectares).

Primary traits of importance present in biotech crops include insect and disease resistance, herbicide tolerance, altered nutritional profile and enhanced storage life. Few of the examples of genetically engineered crops include biotech papaya, golden rice, banana, apple and mushroom. Some prominent examples of biotech crop that were proven to be beneficial for consumers include:

a) Non browning apple – The artic apple was created using RNAi (RNA interference) suppression of genes that are associated with oxidation of polyphenol oxidase enzyme (PPO). When an apple's surface comes in contact with air or saliva of the consumer, browning in apple occur due to the activity of PPO. The non-browning trait introduced protects the flavor and nutrition of apple, it also reduces food wastage (Baker, 2018).

b) Golden Rice – Prof Ingo Potrykus and collaborators created Golden Rice in the 1990s by introduction of a multi-gene-coded biochemical pathway in rice's genome. This pathway produces beta-carotene, a precursor of Vitamin A, in rice which is naturally

absent in it. Populations in poor countries offer suffer due to Vitamin A deficiency, Golden Rice could meet the demands for Vitamin A in diets. Filipinos, in 2019, had become the first country to approve the utilization of GR2E, second genartion Golden Rice for its populace.

c) Rainbow papaya – *Papaya Ringspot Virus* (PRSV) causes serious disease in Papaya. Since 1992, this virus almost took over the entire hectares of Papaya farms in Hawaii. Transgenic papaya line 55-1 showed resistance to the virus and hence rainbow papaya was created, and its use had become widespread in early 2000s. Rainbow papaya helped farmers in earning money for their livelihoods. Similar to vaccination, transgenic papaya was introduced with Virus coat protein from PRSV strain HA 5-1. The protein provide papaya with resistance against the virus infestation (Ferreira and Pitz et al., 2002)

1.3.4 Genetic engineering in forestry and biotechnology for tree improvement

From the first genetically engineered (GE) poplar tree (Fillatti et al., 1987) to possibly the first GMO chestnut to be planted in the wild (Hill, AP news, 2019), genetic engineering has come a long way in the field of forestry. Application of GE is often accompanied with plant tissue culture and transformation techniques. GE holds promise as a tool for expediating tree breeding process by shortening generation times, introducing novel traits into proven genotypes and by reducing the cost of selection (Pena and Seguin, 2001). Through recombinant biotechnology, a gene of interest can be introduced into an organism using vectors such as plasmid and BAC (bacterial artificial chromosome). Utilizing methods such as site-directed mutagenesis, specific mutations can be made in the genome which can aid in studying the genes and obtaining beneficial products by creating a fairly accurate changes in the genome. This capability provides us with a tool for beneficial modification of tree to improve wood quality, alter fiber content, and for sustainable management of forest lands. Moreover, marker assisted selection (MAS) and genome wide association study (GWAS) can help us study the genes responsible for quantitative traits (Powell et al,.2018). Powerful genome sequencing tools such as next-generation sequencing (NGS) helps with comparative genomic approaches as well as in sequencing entire genomes of tree species (Kremer and

Neale, 2011). Genomic resources such as ESTs, reference genetic maps and SNPs are well developed in many tree genera and help us in deriving knowledge about the variations in genomes caused within species.

Although, GM trees have been grown experimentally in more than 700 field trials since 1987, its commercial utilization is minimal (Vidal and editor, 2012). Countries like China and Brazil have commercially approved the use of transgenic trees. In 2002, China had become the first country ever to approve commercial use of Bt Poplar (Lu and Jian-Jun, 2011). Followed by Brazil, when in 2015 FuturaGene's GM eucalyptus were approved for plantations, the company claimed that their GM eucalyptus could produce 20% more wood compared to the conventional variety (Ledford, 2014). Since, poplar is a model organism for studying trees, a great deal of biotechnological research has been performed on them. Following are some case studies for the application of genetic engineering in forestry.

Forest tree restoration and conservation

Biotechnology can play a crucial role in biodiversity restoration and conservation of genetic resources within and among species in a forest. Tools in molecular genetics can aid in distinguishing between native and non-native tree species, and for monitoring an endangered ecosystem. It can help us with knowledge in developing and interpreting information on the evolution and current state of an ecosystem (Gaston et al., 1995). For example, species specific probes can be used for analysis of natural hybrids and DNA fingerprints can be utilized for monitoring seed orchards for pollen contamination (Sutton and Grossnickle, 1999).

Due to increased global travel pests and insects can be easily introduced into a novel geographic location, where they can establish themselves as non-native species. Moreover, a warming climate expediates the process of colonization in both native and introduced insects. Natural biodiversity and plantation forestry suffer from these pest infestations. Biotechnology is a tool that could be utilized as a measure to control the pest

growth as well as to restore endangered plant species. Following are some examples to support the above assertation:

a) Chestnut (Castanea spp.) have been extensively studied using biotechnological tools and genomics; it is an apt example for the use of biotechnology in species conservation program (Nelson et al., 2014). Chestnut trees located on the eastern coast of the USA have a cultural, ecological and economic value associated with it. Unfortunately, in the late 1800s, introduction of the fungus *Cryphonectria parasitica* caused Chestnut blight, a cankerdiseases which wiped out most of the chestnut population. Researchers had taken several traditional tree breeding approaches to save this tree but to no avail. As a last resort, biotechnology was used to introduce wheat gene in the genome of chestnut tree. The enzyme produced by the gene breaks down the toxin called oxalic acid produced by fungus (Newhouse, The Washington post, 2018).
b) Other instances of species restoration using biotechnological tools include white ash (WA) tree restoration in the USA and European ash trees (*Fraxinus excelsior)* in the UK. EAB (Emerald ash borer) consumes tree vascular tissue and in doing so harms the tree. Researchers at Purdue University, USA tried to develop transgenic WA trees with Cry8Da protein that is toxic to EAB larvae and is naturally found in *Bacillus thuringiensis* SDS-502 (Pijut et al, 2014). European ash trees have been devasted by a fungal infection caused due to *Hymenoscyphus fraxineus.* Researchers at University of York, UK utilized associative transcriptomes to selectively breed for trees that were resistant to the fungus (Harper and Bancroft, 2018).

Improved growth rate and yield

"Growth improvements can be categorized into yield potential (greater growth under non-stressful conditions) and yield preservation (adequate growth or at least survival, under stressful conditions)" (Chang et al., 2018). Yield of wood can be improved by improving processes such as light interception, conversion of energy to biomass and allocation of biomass to the harvested portion of plant (Koester et al., 2014). Transgenes can be used to increase tree biomass by 1) Manipulation of phytohormone pathways 2) Improved uptake and utilization of water and nutrients 3) Modification of photosynthesis and carbon utilization methods. (Dubouzet et al., 2013)

Cell and wood properties

Cell walls are mainly composed of cellulose and lignin. The other less dominant components include hemicellulose and pectin. Research on biochemical components of cell wall have been important since economic value of woody biomass is to an extent determined by its cell wall. Improved cellulose content and its separation from lignin helps in supporting chemical, energy and byproducts derived from wood. Moreover, cell wall components are also used in biofuel production, providing an alternative to petroleum-based energy products (Chang et. al, 2018).

Cellulose, *Beta* 1-4 glucan, has been used as a raw material in fabric and paper industry. It is also used for production of cellulose derivatives such as cellulose esters and ethers. Recently, it has been utilized for fermenting compound and production of bioethanol (Koda and Uraki, 2015). Cellulose also provides strength for vertical growth in the trees. U.S. Department of Energy (DOE) first used hybrid poplar in their Bioenergy Feedstock Development Program for use as fuel in generating heat and electricity. Willows and *Eucalyptus* are the other trees that have been used since then for biofuel production. Altering cellulose content can help in production of biofuel. An increase in cellulose content causes cell wall to thicken, crystalize and enhance wood density. In some studies researchers had introduced *SuSy* (Sucrose Synthase) and *UGPase* (UDP-

glucose pyrophosphorylase) transgene which resulted in a higher soluble sugar content in leaf tissue, indicating altered sink strength in trees (Coleman et al., 2007, 2009).

Lignin is a complex and heterogenous cell-wall-bound phenolic polymer, it provides cell wall with strength, rigidity and impermeability to water (Saxena and Stotzky, 2001). Lignin is one of the reservoirs of stored carbon in plants (Tuskan et. al, 2006). Lignin content modification can alter a plant's defense mechanism against invading pathogens. A reduction in lignin content makes the plant cell wall thinner and susceptible to attacks by insects, herbivores and microbes (Halpin et al., 2007). However, it provides commercial benefits which includes increased pulping efficiency and reductions in mill contaminants during paper production (Rastogi and Dwivedi, 2008). The livestock industry also benefits from lignin reduction as fodder becomes more chewable by animals. Lignin quantity and quality can be modified by alteration of genes that encode enzymes for the lignin biosynthetic pathway. For example, down-regulation of expression of cinnamyl alcohol dehydrogenase (CAD), caffeoyl-COA O-methlytransferase (CCoAOMT), and caffeic acid O-methyltransferase (COMT) in monolignol biosynthesis pathway reduces the lignin content of wood (Chen et al., 2001).

Hemicelluloses are heterogenous group of branched polysaccharides that associate with both cellulose microfibrils and lignin (Chang et al., 2018). During the biochemical conversion of woody biomass, hemicellulose content and composition may limit saccharification by limiting the access of cellulase to cellulose (Himmel et al, 2007). RNAi suppression of glucuronoxylan glycosyl transferase genes in hybrid poplar demonstrated a reduction in glucuronoxylan content (hemicellulose biochemical component) and improved hydrolysis by cellulases (Lee et al., 2009).

Pectins are a group of heterogenous polysaccharides rich in galacturonic acid, it is mainly composed of two components i.e. homogalacturonan (HG) and rhamnogalacturonan I (RGI) which surround the cellulosic glycan network (Willats et al, 2001). Pectins are major components of the primary cell wall and a modification in pectin may alter saccharification yields in trees (Chang et. al, 2018).

Resistance to biotic and abiotic stressors

a) Drought and frost tolerance

Tree survival can be threatened in the periods of long-lasting drought. Drought tolerance in cells and tissues of trees can be enhanced by targeting molecular physiological factors responsible for osmotic adjustment, antioxidative defense and water use efficiency in trees. Biotechnological tools can help in increasing drought tolerance among plants by overexpressing the genes involved in stress sensing and signaling, such as the abscisic acid core pathway, and down-stream transcription factors (Polle et al., 2019). For example, an upregulation of genes such as *STOMAGEN*, *ERECTA*, and *STOMATA DENSITY AND DISTRIBUTION1 (SDD1)* can induce changes in stomatal development of poplar for survival in drought like conditions (Harfouche, Meilan, and Altman, 2014). Likewise, overexpression of *DREB* (dehydration responsive protein binding element) under the *RD29* promoter can enhance drought tolerance in plants by activating osmolytes (Zhou et al., 2012). Furthermore, understanding of the genetic basis for drought tolerance by utilizing techniques such as transcriptomeics, provenance testing, quantitative trait locus (QTL) mapping, and genotype association studies can aid in forest management by providing options for plus tree selection (Moran et al., 2017).

Trees have physiological mechanisms that help them survive in cold temperatures. Cold hardiness in perennial woody trees is a complex process that involves different mechanisms and stages, along with the genetic regulation of dormancy (Wisniewski, Nassuth, and Arora, 2018). The complex processes associated with cold hardiness in woody plants has been studied by consideration of both epigenetic and genetic regulation of cold hardiness, in combination with advanced genetic analysis tools such as GWAS. An example of genetically engineered freeze tolerant *Eucalyptus* is the clone AGEH427, it was developed through introducing plasmid pABCTE01 into the EH1 genotype of *E.grandis x E.urophylla* hybrid. The transgenic line survived in cold temperatures through production of CBF2 (C-Repeat Binding Factor) cDNA. The plasmid used contained a CBF2 expression cassette that comprised of cold-inducible promoter rd29A (Hinchee et al., 2011). However, it has yet to be approved for

commercial use by USDA, and its level of cold tolerance may not be adequate for commercial applications.

b) Insect and pest resistance

In the face of a changing climate, trees have developed susceptibility to a broad range of insects and pathogenic microbes. Economic losses incurred due to infected plantation trees reduce the productivity of intensive plantation forestry. Hence, biotechnology and genomic tools can be used for engineering resistance in plantation trees against these pests. Tools such as GWAS or Genomic Selection can be utilized for choosing candidate genes to achieve improved resistance against pests (Naidoo S. et al., 2019). Knowledge derived from proteomics, transcriptomics, metagenomics and metabolomics can further help with developing systems in trees for pest resistance. Insects that causes great loses to forest trees in USA include the mountain pine beetle (*Dendroctonus ponderosae*), gypsy moth (*Lymantria dispar*), southern pine beetle (*Dendroctonus frontalis*), and emerald ash borer (*Agrilus planipennis*) (Chang et al., 2018). Around 22 insect-resistant poplar varieties have been developed for pilot scale studies and small-scale field tests in China (Wang et al., 2018). Insecticidal genes that are employed for targeting insects can be both exogenous and endogenous, and have been obtained from plants, animals and microorganisms. Genes derived from plants usually encode for proteinase inhibitors, phytolectin, amylase inhibitors and chitinase. Genes from the bacterium *Bacillus thuringiensis* has been widely used in many tree systems for providing resistance against broad range of insects. *B. thuringiensis* often contains genes that encode for protein and toxins that could be lethal for insect larvae.

c) Salt tolerance

Salt accumulation in upper soil layers can be detrimental to the growth of plants. With increased instances of droughts and climate change, the occurrence of higher salinity in soil has become more frequent around the world (Chen and Polle, 2010). In such scenarios, one of the strategies could be to develop transgenic trees with higher tolerance to saline conditions. Researchers have developed a transgenic *Eucalyptus*

camaldulensis tree that contains the mangrin gene, an allene oxide cyclase homolog, which has a core protein domain that enhances salt tolerance in the host (Yu et al., 2013). Examples of other genes that have been used for enhancing salt tolerance in transgenic trees include glycine betaine biosynthetic *codA*, the DREB transcription factors and vacuolar membrane Na^+/H^+ (Khan et al. 2016). Glycine betaine is an osmoprotectant that provides protection for important cellular organelles during plant adaption to stressful conditions. DREB genes encode for transcription factors that regulates stress-tolerance responsive genes.

1.4 Biotechnologies for reproduction control

Plant reproduction could either be accelerated or suppressed based on the requirement of an organization. In this section, I'll be highlighting the techniques that can be used for genetic containment in plants.

Plant sterility in exotic species and transgenic trees can be important from the viewpoint of regulatory bodies, market needs and ecological reasons (Fritsche et al., 2018). Additionally, sterile trees may demonstrate increased wood production due to enhanced vegetative growth, reduce instances of pollen allergies and help in hybrid breeding (Strauss et al., 1995). Methods that have previously been used for the above purposes include organ ablation, transgene excision, RNA interference (RNAi), transgene-encoded protein interference and expression of genes that cause delays in onset of flowering. I discuss a few of the most advanced forms below.

1.4.1 Ablation approaches

Genetic ablation methodology is a tool which is often used for analysis of development processes and to determine role of a specific cell type in complex tissues (Thorsness et al., 1993). This method could also be used for attaining genetic sterility and insect resistance. It functions by employing a promoter in specific cells to control the expression of a deleterious gene encoding for a cytotoxin, this cytotoxin in response causes cell death (Burgess et al., 2002). Several cytotoxins, such as Diphtheria toxin A chain, Exotoxin A, Ricin toxin A chain, RNAse T1, can be used for the cell ablation

process (Day and Irish, 1997). However, the most commonly used cytotoxin for anther specific floral cell is a ribonuclease Barnase (Mariani et al., 1990). However, non-exclusivity of promoter expression outside the floral gene (Rottmann et al., 2000) may impair tree growth due to unintended effects of cytotoxin (Skinner et al., 2000). Few of the successful examples of ablation technique employed in trees and crops include: attainment of sterility in male and female reproductive organs in Arabidopsis thaliana and Nicotiana tabacum through microspore and megaspore mother cells ablation using SDS (SOLO DANCERS) and BARNASE fusion gene (Huang et al., 2016); researchers successfully used a male specific promoter PrMC2 to drive modified Barnase coding sequences for ablation in pine, eucalyptus and tobacco (Zhang et al., 2012).

1.4.2 Transgene excision

Site-specific recombinase system such as Cre-Lox from bacteriophage P1 can be employed for manipulation and removal of transgenes from transgenic plants (Gilbertson, 2003). However, this system is primarily used for two purposes that is to remove selectable marker from the chloroplast or nuclear genome and for targeting the transgenes to defined locations. A more recent application of it has been to remove transgene from the plant before seed and pollen germination. This system is efficient in transgene containment, however, its stability in field conditions and the activity of the recombinase system in large trees when they are flowering is difficult to assure. Moreover, this technique does not impair fertility and hence it would be wiser to use this system along with a sterility transgene (Brunner et al., 2007).

1.4.3 Gene suppression and targeted gene mutagenesis

RNA interference (RNAi)

Ribonucleic acid interference (RNAi) was first observed in the nematode Caenorhabditis elegans (Fire et al., 1998). RNAi involves suppression of gene expression or translation through specific degradation of mRNA molecules (Mansoor et al., 2006). RNAi has multiple uses in plant biotechnology, for example it could be used against plant pathogens or for overproduction of secondary metabolites which might have yield, health or environmental benefits. Moreover, it has also been used for conferring reproductive sterility (Lu et al. 2018).

Targeted mutagenesis

Genome editing is the primary means for targeted mutagenesis, and has been evolving rapidly in the past two decades. It involves creating double-stranded breaks (DSBs) at targeted sites in the genome that are incorrectly repaired, causing loss of function mutations at those sites. Some examples of site-directed mutagenesis methods include ZFNs (Zinc finger nucleases), TALENs (transcription activator-like effector nucleases) and CRISPR/Cas. The most advanced and efficient method for directed mutagenesis is based on CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) associated RNA guided endonucleases Cas9 (Type 2 CRISPR/Cas). CRISPR/Cas system was first discovered in archaea. It provides defense/adaptive immunity against bacteriophages and viruses in prokaryotes. This system was then developed for site-directed gene editing in the year 2013. It works by Cas9 endonuclease cutting at the desired DNA sequence within the endogenous genome. The endonuclease is guided by a short RNA search string. Cas9 contains two domains, i.e., HNH domain (cuts the complementary strand of crRNA) and RUcV-like domain (cuts opposite strand of double stranded DNA). For excision purposes, Cas9 identifies the PAM (protospacer adjacent motif) sequence in the genome. After an excision is made double stranded break can be repaired through non-homologous end joining or homology directed repair (Liu et al, 2017). CRISPR-Cas9 has revolutionized genetic modifications in plant biology. Since its

discovery it has been used in crops and plants such as cotton, tomato, potato, rice, wheat, eucalypts and poplar.

1.5 Roadblocks in commercial use of GE trees and crops

Major impediment to the wide commercialization and use of genetically engineered trees and crops stem from three areas: public acceptance, regulatory supervision, and ecological concerns regarding genetically modified organisms (GMOs).

1.5.1 Public acceptance

Genetically modified organisms are often viewed both in positive and negative lights among the public, based on factors such as educational levels, political standings, personal values, and perception of risks and benefits (Lucht, 2015). National public surveys in the US have shown that Americans are more likely to accept biotechnology for medical purposes or for human health improvement, but not as much for other uses like ameliorating meat quality (Hefferon and Funk, Pew Research Center, 2018). The skepticism regarding consumption and utilization of genetically modified organisms was strongly expressed in 1996 in Europe, with France developing a strong opposition against GM crops. Due to certain mishaps in the European Union with food safety and diseases, the public's trust of governmental organizations dropped. Various outlets played a part in fueling this mistrust, particularly non-governmental organizations (NGOs) and environmental organization's opposition against GM food (Bonny, 2003).

People have different views regarding genetically engineered food all around the world. For example, Chinese citizens mostly display a neutral to positive response to GM food and crops, however, some see them as a means to spread bioterrorism in China (Shoemaker, Cui, 2018). Citizens in the European Union view GMOs more harshly. Genetically modified food is either partially or absolutely banned in 19 out of 28 countries in the European Union. Countries such as Spain and Portugal have allowed Bt maize production (European green capital news). In many countries a high percentage of respondents to the polls accepted the use of both GM and CRISPR for food. For example, 56%, 47%,46%, 30% and 51%of those surveyed in the USA, Canada, Belgium, France,

and Australia viewed GM food in a positive light (Shew, Nalley et al., 2018). Commercialization of GM products is further complicated due to this lack of global consensus.

Acceptance of GMOs varies among groups of people, for example, farmers and plantation owners, are more likely to accept GM crops and trees if it helps them reduce the cost of producing crops. Consumers are usually unaware of the pros of GM crops and hence, they are more resistant to consuming food derived using biotechnological tools. Consumers are usually worried about the presence of "foreign DNA" in GM products and its implication for their health (Lucht, 2015). Another reason for lower public acceptance of biotech products is the knowledge gap between the final product and processes used in developing GE crops and trees. Studies have shown that people are more likely to purchase GM products if they are educated and aware about the process underlying the final GE product.

Moreover, an individual's values and personal beliefs also affect the use of GM crops and trees. Many people tend to avoid consuming biotech products due to religious and economical reasons. For example, some people show hesitance in purchasing GM products based on their worry about a corporate monopoly in agriculture and plantations. Big companies like Monsanto (now Bayer) may affect smaller farmers and potentially contribute to global agricultural income inequality (Lucht, 2015). Negative portrayal of genetically modified organisms by media outlets influences public reception of GMOs. Catchphrases such as "Frankenstein Forests", "Frakentrees", "Designer Trees", and "Terminator gene" inflict fear of deploying GM products for agricultural and forestry purposes among citizens.

Vandalism of field trials of GM products by GMO opponents adversely affects research in the plant biotechnology sector. In Europe, up until 2012, there had been 80 acts of vandalism against academic or governmental research on GMOS. These acts were mainly concentrated in 4 countries: France, Germany, the United Kingdom, and Switzerland (Kuntz, 2012). In 2001, few organizations on the west coast of the USA had become a target of "eco" terrorism. Millions of dollars were lost with successful acts of

vandalism in GM field trial sites located at Oregon State University in Corvallis, the University of Washington in Seattle, and GreenWood Resources (GWR) in Portland, Oregon (Strauss, 2016). In another case in Oregon, vandals had destroyed around 1,000 plants in a farm that had cultivated transgenic sugar beets (Biology fortified, 2013). Moreover, organizations such as Greenpeace and Earth Liberation Front have been historically involved in various GM crop field destruction acts.

However, GM crops and trees are not a panacea for world hunger problems, dwindling biodiversity, and sustainability. Factors such as societal values, cultural differences, economics, and population should be taken into consideration as well. And other genetic and management methods are often more efficient and reliable than GM methods, a difference that is greatly exacerbated by regulatory and market obstacles.

Studies have shown that people are open to the idea of using biotechnological solutions for problems that require immediate attention such as in addressing forest health threats by insects, pests, and pathogens; whereas not as much when compared to using genetic engineering for less tangible threats like climate change or improved tree growth (Needham, 2015).

1.5.2 Governmental and market regulations important to crop and forest biotechnology in the USA

The USA federal government designed a policy in 1986, the Coordinated Framework for Regulation of Biotechnology, which examines products developed using biotechnology. In USA, the primary agencies that are responsible for experimental testing, approval, and commercial release of transgenic organisms include the USDA's Animal and Plant Health Inspection Service (APHIS), the Environmental Protection Agency (EPA), and the Department of Health and Human Services' Food and Drug Administration (FDA) (Biotechnology topics, USDA).

APHIS deals with protecting agriculture from pests and diseases; and it does that through "regulated articles" (though a change to this system was just published and will take effect in coming months and years: USDA 2020). APHIS oversees the import,

handling, interstate movement, and release of regulated organisms into the environment. APHIS also holds the authority to deregulate a transgenic article based on scientific evidence. The EPA oversees the application of biopesticides to crops. Biopesticides are pesticides for pest control that are produced through plants and microbes which could be transgenic (www.epa.gov/pesticides/biopesticides). The FDA holds accountability for proper labeling of food and feed derived from transgenic organisms (www.fda.gov).

The USDA helps with market facilitation of biotech crops and food through programs such as Agricultural Marketing Service (AMS) and the Grain Inspection, Packers and Stockyards Administration (GIPSA) (usda.gov/biotechnology-frequently-asked-questions). Despite this facilitation, the testing, patenting, and commercialization of GM products take several years compared to non-GM products. Also, the high cost of testing and field trials lead to smaller biotech companies and public sector researchers to opt out of producing biotech products for the betterment of society.

Forest certification programs such as Forest Stewardship Council (FSC) and Sustainable Forestry Initiative (SFI), as well as others, have certified around 440 million hectares of forest around the globe for sustainable management of plantation land. One of the criteria for their certification is an absence of transgenic trees from plantation land. Such stringent requirements hinder progress in sustainable forestry management. Beneficial traits like disease resistance or increased yield could help plantation owners in tropical countries to manage their land sustainably and economically (AAAS, 2019).

1.5.3 Ecological concerns

With biotechnological research the primary concern is the confinement of biotech products to the product site, for example pharmaceutical products are mostly developed in laboratories. However, that is not the case with GM trees and crops; the main ecological concerns associated with genetically engineered trees and crops revolve around gene flow of transgenes from transgenic tree/crop to wild or feral populations.

1.6 Environmental impacts of engineering genetic sterility in trees and crops

Some cautions to use of GM trees and crops

As an evolutionary survival mechanism pollen from trees are dispersed at a long distance through wind pollination. Gene flow and horizontal gene transfer (escape of genes from genetically modified tree to wild or feral species) may pose ecological threats (Libby and Burdon, 2006). Furthermore, presence of GM crops and trees can cause plant weediness and higher survival rate of them in the natural ecosystem compared to non-GMO counterparts. An absence of pollen or seeds in plants can impact pollinators as well as nectar consuming organisms. It may also indirectly affect the business that depend on flower nectar such as honey farms/keepers. An increased usage of herbicide may lead to weediness of a non-native tree or crop species. Another problem with using genetically engineered trees involve intellectual property rights. Poor farmers may not be able to afford monetary fees that are imposed if pollen from a genetically engineered crop escapes into an adjacent non-GMO field.

Eucalyptus has been considered as an exotic species in many countries and regions such as India, Brazil, China, Indonesia and South Africa. The tree competes with natural forest tree biodiversity of a region by depleting nutrients and water from the soil. Eucalypt oil is highly flammable and it may cause forest fires. Moreover, monoculture in forest land can lead up to reduced biodiversity of birds and insects. Allelopathy due to eucalypts inhibits growth of other plant species (FAO report, 1993). Because of its impacts and exotic nature many *Eucalyptus* plantations are not accepted by environmental organizations. Matching of species to a region and provenance testing is

crucial; additionally, managing forests with a comprehension about nutrient cycling, growing eucalyptus with other trees and handling its plantation as a habitat for wildlife may help in solving acceptance issues (FAO, 1993).

Chapter 2: Sequence analysis and vegetative growth in transgenic *Eucalyptus* with CRISPR-Cas9-induced mutations in the *Eucalyptus* homologs of the floral genes *EMBRYO DEVELOPMENT ARREST 33* (*EDA33)* and *TAPETAL DEVELOPMENT AND FUNCTION1* (*TDF1)*

2.1 Introduction

Demand for wood has been predicted to rise over the next few decades (FAO et al., 2012), which trends until 2020 have shown to be true. The pressure put on the natural forest for logging often reduces biodiversity and causes other environmental problems (Pimm et al., 2014), particularly in tropical and subtropical environments. Globally, approximately one third of wood for timber is derived from plantations forestry, though it uses only 5% of the total forest land around the world (FAO, 2010).

Genetic engineering techniques are intended as tools to increase the health of forests, and to help improve the yield of wood at lower environmental costs. However, due to biological, economic and social constraints, including difficult and costly transformation in many species, stringent regulatory barriers, negative public acceptance, and ecological concerns, there is only modest research with genetically engineered trees outside of China. Commercialization and field trials of genetically engineered trees have been severely hindered. Our laboratory has focused on means for genetic containment of transgenic and exotic trees as a tool to increase public acceptance and reduce regulatory barriers.

There are many technologies for genetic containment in plants which have been discussed earlier in chapter one. In this research we focus on methods to mutagenize genes expected to be essential for sexual reproduction. We used CRISPR-Cas9 (Clustered Regulatory Interspaced Palindromic Repeat – CRISPR associated sequence 9) gene editing technology for generating site-directed mutations. CRISPR-Cas9 has been highly efficient as a tool for reverse genetics in plants (Elorriaga et al. 2018; Belhaj et. al, 2015; Montenegro, 2016; Quetier, 2016), and can in theory induce near-permanent sterility by mutations and/or deletions of essential floral development genes. CRISPR-Cas9 has been

a revolutionary technique for site-directed mutagenesis because of its high mutagenesis efficiency (Samanta et al., 2016), and its relatively low cost and ease of use compared to other nucleases such as ZFNs and TALENs (Elorriaga et al. 2018). The CRISPR-Cas9 system has been used to generate mutations in poplar homologs of the well-studied floral development genes *LFY* and *AG*, whose mutation is expected to give bisexual sterility (Elorriaga et al. 2018). We have extended this work by studying additional genes whose mutation in *Eucalyptus* might give male or female sterility specifically—with potential uses that include male-sterility to facilitate hybrid breeding or minimize pollen dispersal, and avoidance of seed dispersal to reduce invasiveness or facilitate seed collection from trees in seed orchards.

Genes selected for research and their roles in plant reproduction

We targeted the *Eucalyptus* homologs of two genes, *Embryo Sac Development Arrest 33 (EDA33)* and *Tapetum Development and Function 1 (TDF1),* for female and male sterility. These choices were based on three criteria. First, the genes had been shown, in a model plant species, to be critical for reproduction but not to have obvious effects on vegetative development. Second, the *Eucalyptus* homologs were actively expressed during floral development. Finally, bioinformatic studies of related genes in the *Eucalyptus* genome suggested that the chosen genes had a unique non-redundant function, thus that its mutation would be likely to cause sterility (further described in methods).

EDA33 – Embryo Sac Development Arrest 33

EDA33, also called *INDEHISCENT/IND,* encodes an atypical class of eukaryotic bHLH (basic helix- loop- helix) transcription factor that is necessary for normal development and rupture of the capsule valve margin and thus allows seed dispersal in *Arabidopsis thaliana.* The valve margin separates the valve (a wall that protects seedpod) from the replum (a structure that helps in attachment of the fruit to the plant). Seeds from the fruitpod are released by a spring-loaded mechanism, which enables the fruitpod to open and disperse the seed. EDA33 along with other key regulators in the floral

regulatory network such as SHP, ALC and FUL controls specification of the valve margin and allows differentiation of the lignified valve layer and dispersal of fruit (Liljegren et al.,2004). Through study in Arabidopsis, mutants of EDA33 were shown to have reduced fertility due to the inability of seed dispersal and valve margin development. In another study conducted in Arabidopsis, an EDA mutant with defects in embryo sac development were categorized into mutants having defects during the nuclear division phase of megagametogensis, mutants presenting abnormal nuclear numbers and positions, and mutants that became cellularized but failed in polar nuclei fusion (Pagnussat et al., 2005). Furthermore, in canola downregulation of IND gene produced a complete loss of fruit dehiscence, preventing any seed dispersal. We therefore hypothesized that a non-functioning *EDA33* would result in female sterility in *Eucalyptus*. However, because of the very distinctive capsule in *Eucalyptus* compared to that in the Brassicaceae, it is not clear exactly how fruit and seed release will be affected, or if there might be vegetative effects, such as on growth rate or leaf morphology as these were only superficially studied previously.

TDF1 – Tapetum Development and Function 1

In Arabidopsis, the *TDF1* gene codes for a putative R2R3 transcription factor which has high expression during anther development, particularly in the tapetum, meiocytes and microspores. The tapetum plays a critical role in pollen development by nourishing pollen grains during their maturation development, including to provide enzymes for callose dissolution. Because the tapetum breaks down in *TDF1* mutants, it is clear that TDF1 plays a critical role in tapetal differentiation and function. In Arabidopsis, the *TDF1* gene is expressed during the anther development stage, petal differentiation and expansion stage, and plant embryo cotyledonary stage (www. arabdiopsis.org). Loss-of-function *tdf1* mutants caused gross defects in anther development and tapetal function, particularly with respect to microspore maturation (Gu et al, 2008). We theorize that a similar effect will be observed in *Eucalyptus* for which both copies of its *TDF1* homolog were mutated, including complete male sterility.

2.2 Materials and Methods

After the *Eucalyptus* homologs of *EDA33 and TDF1* genes were identified by sequence alignment and comparison of tissue specific expression patterns (detailed below), *Agrobacterium*–mediated transformation was used for introduction of vector construct into the plant tissues. The DNA of regenerating shoots was extracted for checking the transgene presence in the plant genome via PCR (Mullis et al., 1986). We employed allele-specific-PCR to ensure there were knockout mutations in both alleles for plants whose phenotypes were to be studied. Plants with biallelic, loss-of-function mutations were then transplanted into rooting media first in the tissue culture room and then into the greenhouse in a randomized block design. To study whether vegetative growth differed between the knock-out lines and various controls, we measured several aspects of plant growth rate and leaf morphology. The general experimental flow is summarized in the figure 2.1.

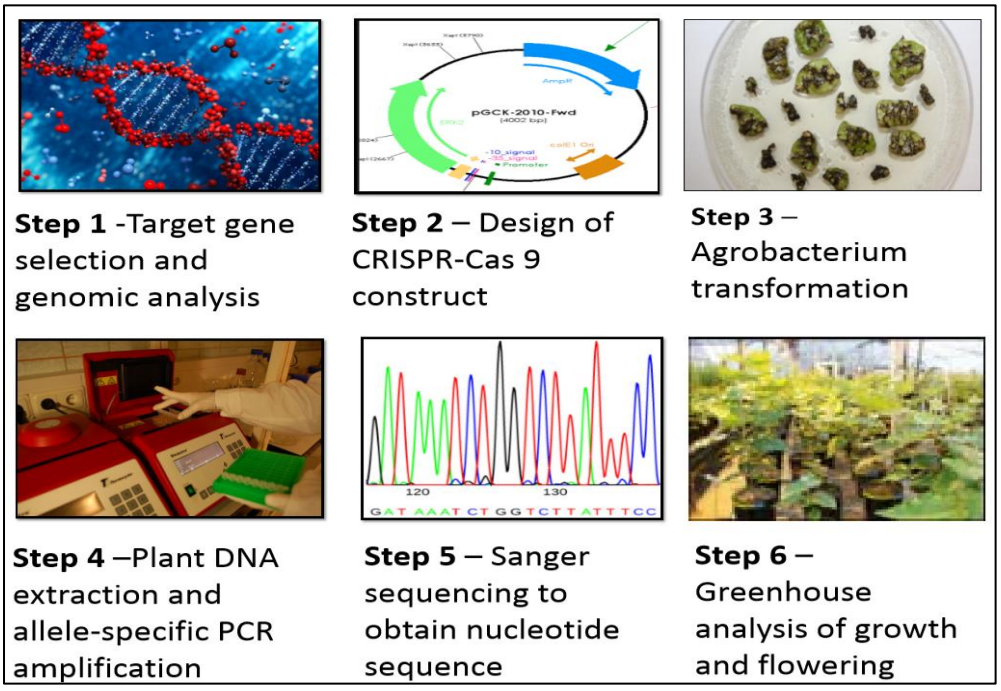


Figure 2.1: Overview of experimental design from target gene selection to greenhouse experiments. (URL: https://techonomy.com/wp-content/uploads/2014/05/DNA-

genomics-genetics.jpg; other images were taken from google images under the usage right labeled for Reuse)

2.2.1 Plant material and transformation

The *EDA33*, *TDF1* and Cas9-only (control) constructs were transformed into hybrid *Eucalyptus* clone SP7 (*Eucalyptus grandis x Eucalyptus urophylla*) generously provided by Futuragene (Rehovot, Israel) using an *Agrobacterium*-based, organogenic transformation method described earlier (Klocko et al. 2014). In brief, sterile leaves were wounded and cocultivated with *Agrobacterium* containing binary CRISPR-Cas9 plasmids then placed on callus-induction medium (CIM) for 48h in the dark followed by shoot-induction medium (SIM) for 6-8 weeks, with subculturing at 3 to 4- week intervals. After shoots became visible, explants were moved to shoot-elongation medium (helps the shoot to grow vertically upwards) for 2-3 weeks. All types of media contained the selection antibiotic hygromycin. Shoots from individual PCR-confirmed transformed events were propagated, leading to multiple identical ramets (trees) per independent transformation event.

2.2.2 CRISPR-Cas9 floral gene and target site selection

We assembled a list of floral-specific genes from the floral transcriptome of *Eucalyptus grandis* (Vining *et al*., 2014). Subsequently, we ran a BLASTn search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to find Arabidopsis homologs of late flowering genes in *Eucalyptus* that are expected to be critical for reproduction based on published literature, but did not have known effects on vegetative growth. We made sure that the floral genes that we chose were expressed only in floral tissues based on genomic databases available. The software a Simple Modular Architecture Research Tool (SMART) (Bork et al. 2015) was used to compare functional domains of Arabidopsis and *Eucalyptus* homologs to ensure that both homologs had key functional domains in common, and that no other genes in the genome shared high sequence similarity in these

domains. After selecting the floral target genes *EDA33* and *TDF1*, we resequenced those genes in the our SP7 hybrid to ensure our designed sgRNAs would have a perfect match to the chosen sequences.

For CRISPR mutation, we used the online tools sgRNA scorer (https://sgrnascorer.cancer.gov/; Chari et al,2017) and Cas-OFFinder (http://www.rgenome.net/cas-offinder/; Park et al, 2014) to pick dual guide RNA sites on the genes. sgRNA scorer was used to rank lists of possible sgRNA targets by predicted mutation rates, whereas Cas-OFFinder (Bae et al. 2014: http://www.rgenome.net/cas-offinder/) was used to search for potential off-target sites of the Cas9 RNA-guided endonucleases. While designing the CRISPR-Cas9 constructs we considered the following for selection of two different guide RNAs to target each gene:

a) They should match the gene at a position where a frame shift mutation or deletion would lead to a non-functional protein.
b) They should target both alleles of the target genes in our *E. grandis x E. urophylla* test hybrids.
c) They should not cause off target mutations at similar loci in the genome.
d) The distance between two guide RNAs was far enough to potentially create a large deletion when both sgRNAs were present.

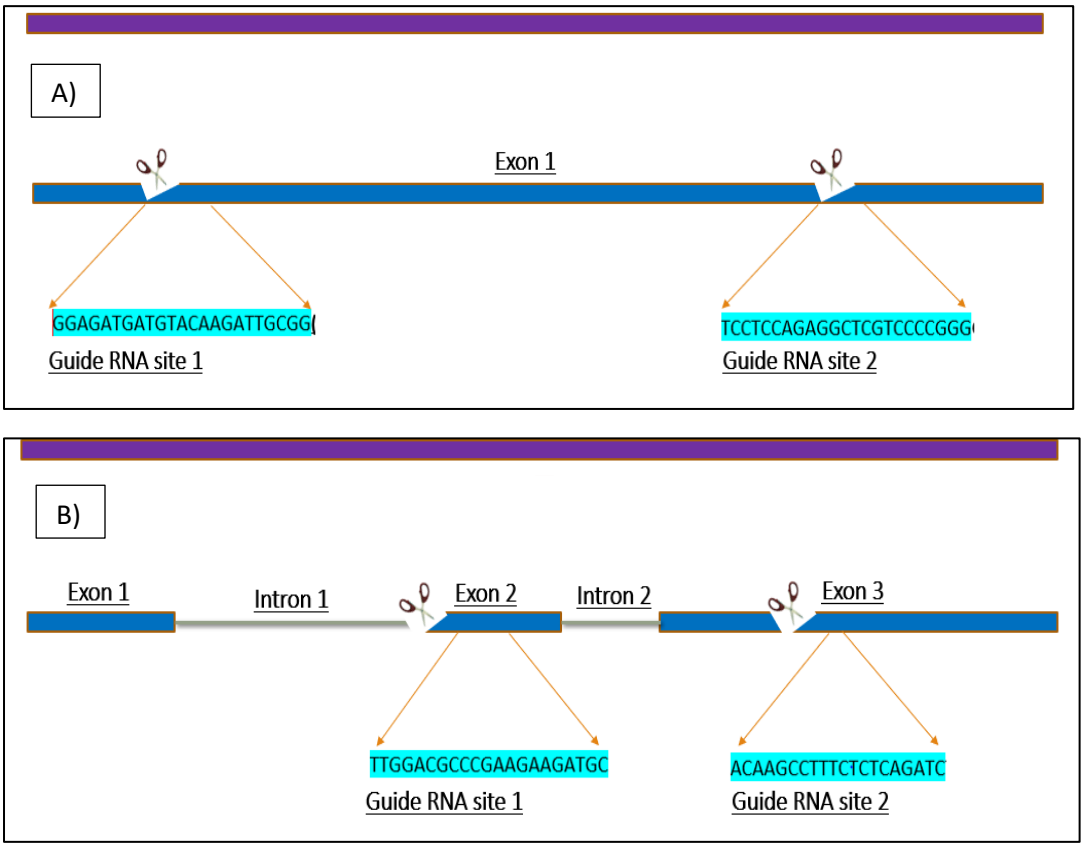


Figure 2.2: Location of CRISPR target sites: A) *EDA33*: Located on chromosome no 9, it contains only a single exon. The guide RNAs are 205 base pair apart on the exon. B) *TDF1*: Located on chromosome no 8, it contains three exons. The guide RNAs are located 887 base pairs apart on exon 2 and exon 3

2.2.3 CRISPR-Cas9 vector construction and transformation

We used Gibson assembly (Gibson *et. al*, 2009) for cloning fragments into plasmids. We assembled three constructs, two for targeting *EDA33* and *TDF1* (see fig), and an empty vector control for expression of Cas9 (see fig) in the absence of sgRNA. The fragments included a ubiquitin promoter U6 to initiate the transcription of sgRNA, a Scaffold to connect sgRNA to cas9 (only in sterility constructs), the endonuclease cas9 to cleave genes at sequences matching sgRNA and two sgRNA sequences to guide cas9 to the matching target sequence in the gene. Selectable marker hygromycin was also

integrated into the plasmid vector. A double 35S promoter was used to ensure high expression of Cas9 in the plant genome (Belhaj et al., 2013).

Addgene (http://www.addgene.org/empty-backbones/), an online tool, was used to order the plasmid backbone. However, the construct was assembled in the lab. Gibson Assembly enzyme mixes from New England Biolabs (Gibson et al, 2009) enabled us to recombine many fragments of DNA into one transformation vector. The details for the protocol and enzymes used for the mix can be found on NEB's website provided above. The primers used for cloning have been shown in the table 2.2.1.

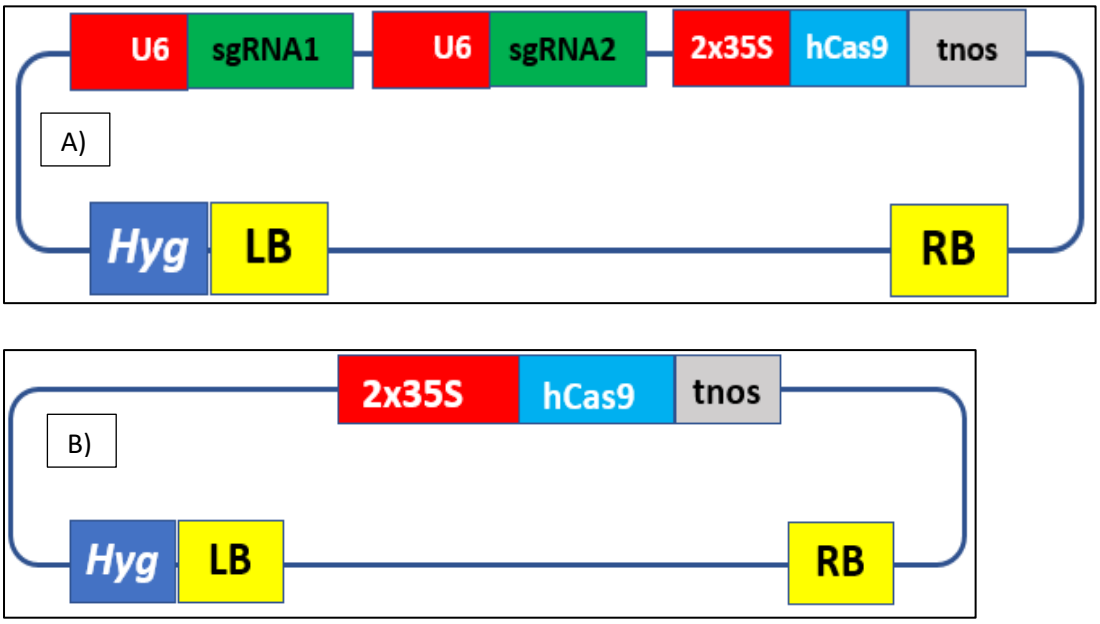


Figure 2.3: Construct design – A) At the top is the experimental plasmid vector that was used to target two small guide RNA (sgRNA) loci on *EDA33* and *TDF1*. B) At the bottom is the "control" plasmid for Cas9 that lacks sgRNAs. U6 promoter initiates transcription of sgRNAs; 2x35S, *double Cauliflower mosaic virus* (CAMV) 35S gene promoter, used for expression of Cas9; sgRNA1 is the guide RNA that matches target site 1 on the DNA and sgRNA2 is the guide RNA that matches target site 2 on DNA; hCas9 is the human codon-optimized Cas9 endonuclease gene sequence from *Streptococcus pyogenes*; tnos is termination region of the nopalene synthetase gene from *Agrobacterium tumefaciens*; *Hyg* is the gene that provides hygromycin resistance; LB and RB are left T-DNA and Right T-DNA borders, respectively.

2.2.4 Plant DNA extraction

We used the plant DNA extraction protocol suggested by Integrated DNA Technologies (IDT) (https://www.idtdna.com/pages/education/biotech-basics), which follows that of Keb-Llanes *et al*. (2002). In brief, after harvesting the leaf tissues from the plant, they were flash frozen in liquid nitrogen followed by storing them in -80 degree Celsius until the extraction of DNA. The leaf tissue for DNA extraction weighed approximately 0.30g and yields of DNA from extraction were typically 50-100 ng per gram of tissue. After quantification of all the DNA samples, they were diluted with 100 µl of nuclease free water. The list of buffer and reagents used can be obtained from the protocol cited above. The concentration and purity of genomic DNA were determined using a Nanodrop 2000 Spectrophotometer (www.nanodrop.com).

2.2.5 Transgene confirmation using end point PCR

We verified the samples for transgene presence by conducting PCR with Econotaq DNA polymerase (Lucigen, Middleton, Wisconsin, USA) and two sets of primers for the Cas9 (product size – 324 bp) and sterility constructs (product size – 648 bp). For Cas9, primers used were p201R (near right border) and StUbi3P218R (near left border); and for *EDA33* and *TDF1*, primers used were p201R (near right border) and ScaffoldF (near left border). Primer sequences and the number of PCR-positive events are shown in Tables 2.1 and 2.2, respectively.

Table 2.1: Primer sequences used for checking transgene presence in the genome.

Gene	Right border	Left border
Cas9 (p201R/StUbi3P218R)	5’ CGCGCCGAATTCTAGTGATCG	5’ ACATGCACCTAATTTCACTAGATGT
EDA33 and TDF1 (p201R/Scaffold)	5’ CGCGCCGAATTCTAGTGATCG	5’ GTTTTAGAGCTAGAAATAGCAAGTT

Table 2.2: Transformation efficiency of constructs with *Agrobacterium tumefaciens*. Shown are the number of positive events obtained after endpoint PCR using the primers shown in table 2.1.

Sterility construct (Hygromycin selection)	**Total number of explants transformed**	**Callus formation (%)**	**Shoot formation (%)**	**Positive transgenic events obtained**	**Transform-ation rates (%)**
Cas 9	316	42.7	16.2	10	3.2
EDA33	362	20.6	9.6	17	4.7
TDF1	428	39.8	18.2	18	4.2

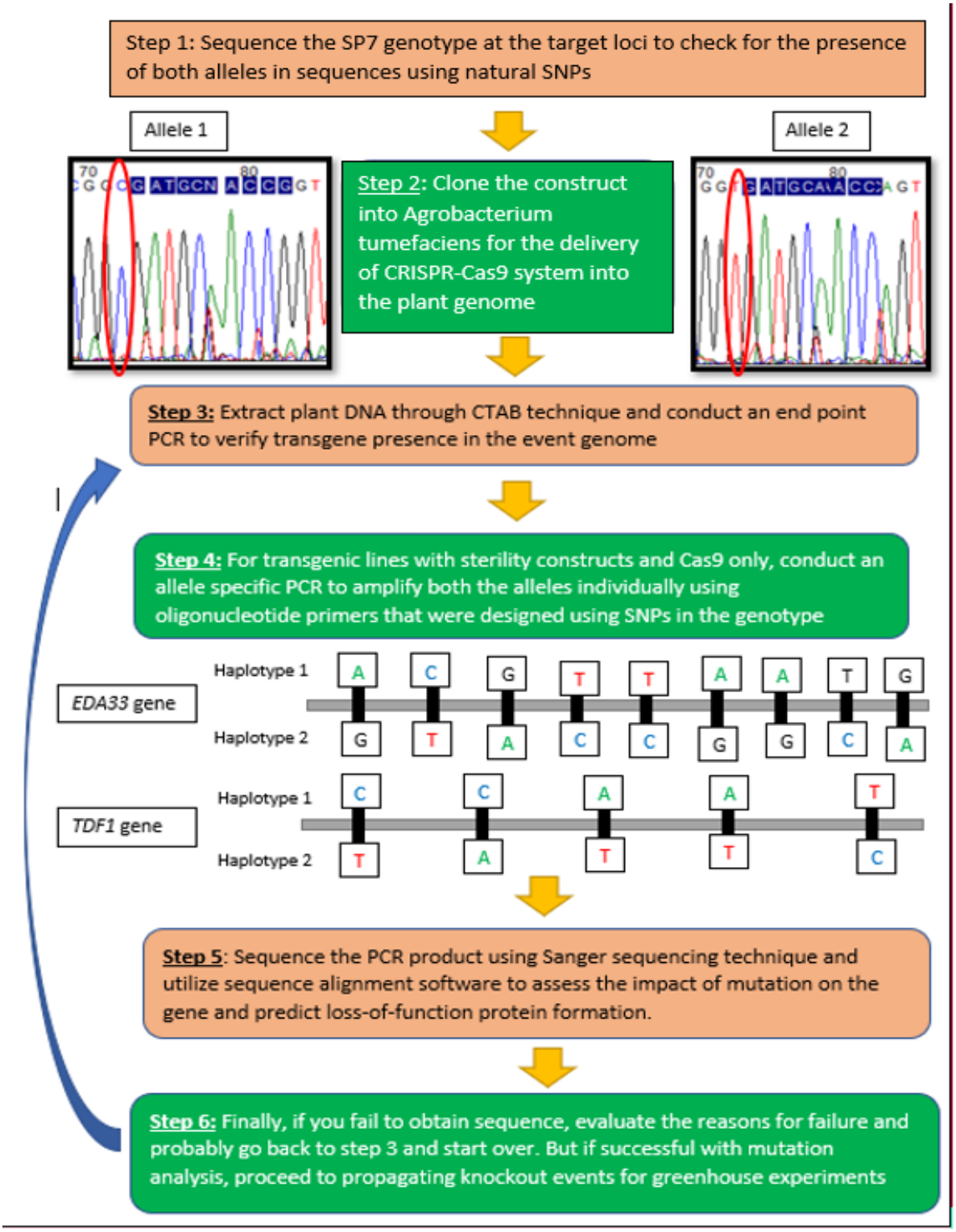


Figure 2.4: Pipeline for genetic investigation of mutations in the floral genes *EDA33 and TDF1*.

2.2.6 Allele-specific PCR for gene amplification

All the transgenic samples were amplified using allele specific polymerase chain reaction (ASPCR) (Wallace and Ugozzoli, 1991). Oligonucleotide primers for ASPCR were designed in a manner to selectively amplify a single allele based on presence of at least a single SNP (single nucleotide polymorphism) in the primer region of each allele.

We identified 5 natural SNP variants in *TDF1* gene and 9 in *EDA33.* The two haplotypes for *TDF1* were CCAAT (we named it as allele 1) and TATTC (we named it as allele 2). The haplotypes for *EDA33* were ACGTTAATG (allele 1) and GTACCGGCA (allele 2). We used these SNP variants in the haplotypes to ensure that both alleles had been amplified for each target gene. The whole gene sequence with SNPs can be found in appendix figure A.4

Oligonucleotide primers were designed through NCBI primer blast website (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Criteria considered for designing primers included:

a) Primer length was kept around 18-24 base pairs;
b) GC content for the primers were between 40-60%;
c) Melting temperatures (Tm) were between 50 to 65 C;
d) Primer pairs had similar Tm's with a maximum difference of 5 C;

A primer set contained of a forward primer which amplified DNA strands from 5' to 3' direction and a reverse primer which amplified DNA strand from 3' to 5' direction. Allele-specific ligonucleotide primers were designed so that a SNP was present at the 3' end. (Alleles which had the same SNP as that of primer were amplified whereas the other allele didn't get amplified due to the 3' mismatch on the SNP location.) In *EDA33*, the primers helped in amplifying genomic regions flanking both the target sites. However, a nested ASPCR was conducted for *TDF1* events due to a large distance between the two guide RNAs and an absence of SNPs outside the guide RNA locations.

Primers used for amplifying samples in the CRISPR-*EDA33* transformed samples were:a) For allele 1: EDA33_F1 and EDA33_R1 which gave a product size of 477 bp. B)

For allele 2: EDA33_F2 and EDA33_R2. Primers used for samples with *TDF1* amplified a single guide RNA with one PCR reaction. Primers for amplifying guide RNA 1 were TDF1_5prime F1 and TDF1_sg1_R1. Primers for amplifying guide RNA 2 were TDF1_sg2_F1 and TDF1_sg2_R2. Primer sequences can be found in the appendix, Table A.1

A thermocycler machine, BioRad C1000, was used for obtaining the PCR amplicons. PCR reactions were performed using Econotaq DNA polymerase (Lucigen, Middleton, Wisconsin, USA), 10xEconotaq DNA buffer, 10% Bovine Albumin serum (BSA), a 2.5mM ea mixture of dNTPs, oligonucleotide primer set and DNA. The concentration of each solution mixture was based off of the Table A.5 provided in appendix. Annealing temperature varied between 57 degree Celsius to 61 degree Celsius.

2.2.7 Agarose gel electrophoresis and Sanger sequencing

After amplification, electrophoresis was performed on all the samples at 130 V for 30 to 60 minutes depending on the agarose concentration in the gel (ranging from 0.8% to 1.5%). We used 1X tris-acetate-EDTA buffer for gels stained with red stain safedye (0.25 - 0.5 µg/µl). The amplified PCR products were visualized under UV light. PCR product bands were then excised from the gel using a clean razor and extracted using the Zymoclean Gel DNA Recovery Kit (Zymo Research) following the manufacturer's instructions.

The sequence of each purified PCR product were determined using the Sanger Sequencing service provided by Center for Genome Research and Biocomputing (CGRB) at Oregon State University (https://cgrb.oregonstate.edu/core/sanger-sequencing). Samples were prepared by following the guidelines provided by CGRB. A sample mixture contained 12 picomoles of primer, and templates at the concentrations provided by CGRB, dissolved in a final volume of 12µL in water. Template concentration were based on length of the PCR band.

After obtaining the sequences we used the software MEGA6 (Tamura et. al, 2013) to align sample DNA against the wild type. We first aligned the DNA nucleotide sequence and then we translated the sequence into an amino acid sequence. Protein sequences gave us an idea about possible frameshift mutations and changed amino acid sequences.

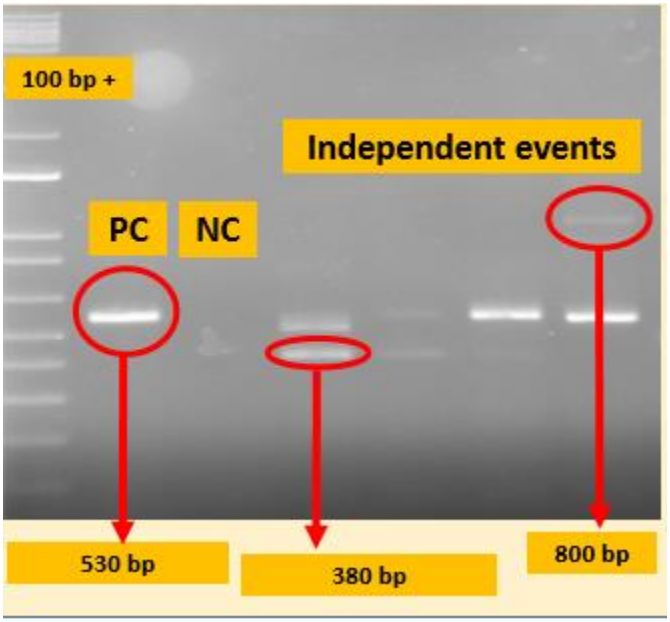


Figure 2.5: *eda33* gene mutations amplified by PCR. PC is positive control (plasmid DNA template), NC is negative control (water). The expected product size for PCR is 530 bp, but due to a large deletion we see bands of 380 bp in several events. In the last lane we can see a large insertion.

2.2.8 Greenhouse trial design and measurements

Plants were arranged in the greenhouse using a randomized block design with two blocks. Block designation was based on tree appearance and health at the outset of the planting; block one contained the larger trees whereas block 2 the smaller ones, both with all of the study genotypes represented. Within each block, a minimum of 1 ramet and a maximum of 5 ramets from each event were planted. An excel function, =RAND() was used for randomizing all the events and ramets into blocks. Plants were kept in the glasshouse and greenhouse for a total of 7 months. Over that period, we collected data at multiple points for measuring plant traits such as stem height and diameter, and leaf measurements that included chlorophyll content index, leaf area, leaf dry mass, and leaf oil gland density. Data collected were stored in an excel sheet for which a link has been attached and explained in the Appendix.

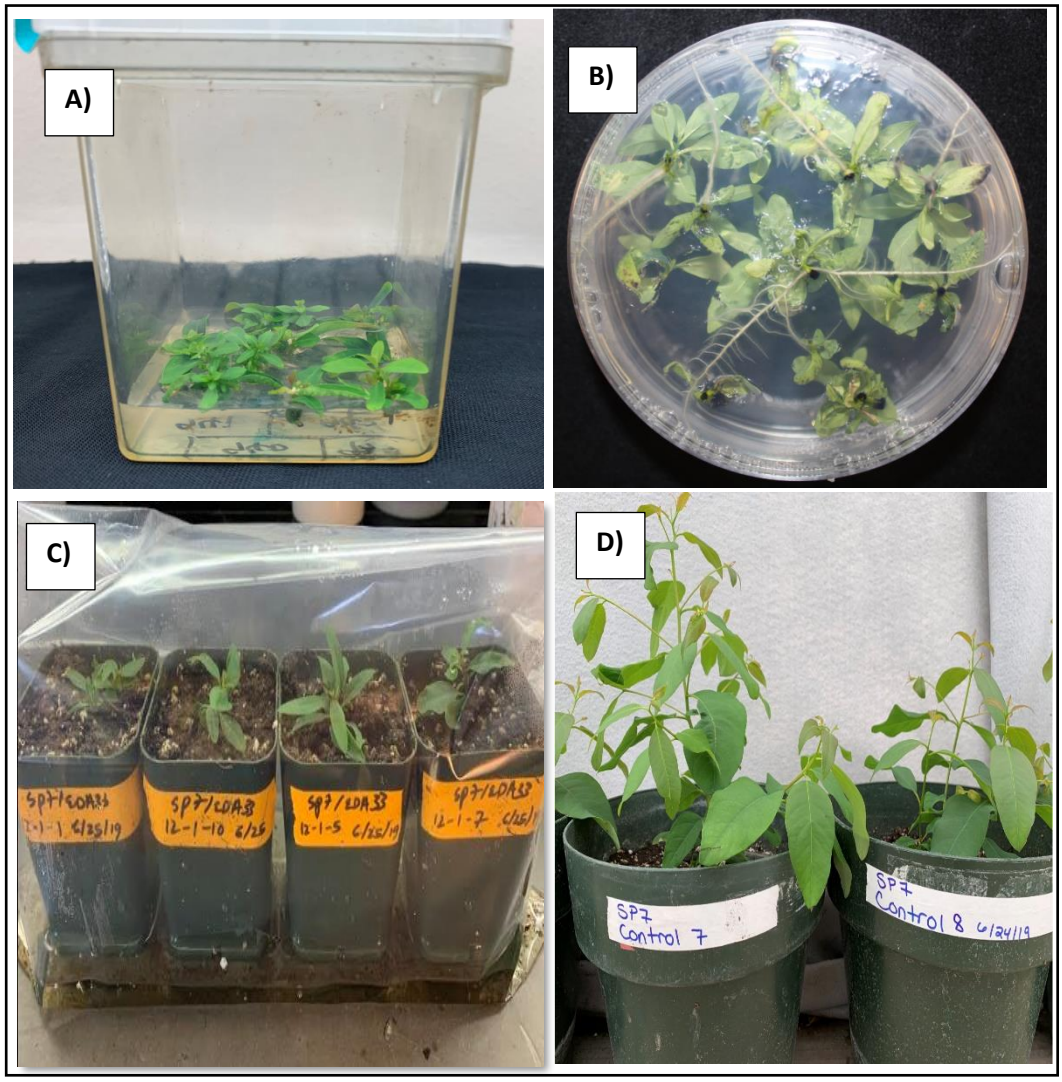


Figure 2.6: Plant phenotyping work flow – A) In this project it started with *in vitro* propagation of plants in Magenta box B) Plants were then transferred to rooting media in small dishes, C) Plants were then moved into small pots covered with a zip-lock plastic bag that are gradually opened to the air D) Plants were then moved into the greenhouse.

Figure 2.7: View of the greenhouse experiment shortly after transplanting (pots in middle).

Propagation and root induction

After confirming the knockout mutations, we propagated around 20 shoots per event in magenta boxes which summed up to 440 plants in total. The plants were allowed to grow in *Eucalyptus* induction media for a month in the boxes before transferring them into *Eucalyptus* rooting media (ERM).

Acclimating new trees in the greenhouse

Rooting trees were kept indoors in a lighted growth room/headhouse for a month before moving them to greenhouse. While indoors, they were placed in a covered zip-lock plastic bag to keep them moist. We opened the bag gradually after 21 days to acclimate plants to the ambient conditions. After the plants were moved to greenhouse, we transferred the plants into larger pots to let the plants root grow freely without space constraints. Conditions in the greenhouse were kept at 75-80-degree Celsius with 16 hours lighting.

Height and Diameter

We measured tree height and diameter twice, first at two weeks after plants were transferred into the big pots and the second observation was taken three months after the first measurement and towards the end of the experiment. Height was measured using a meter stick and diameter was measured at 2 inches above soil line using Meba IP54 Electronic Digital Calipers. We also took photographs of the plants at both points in time. In addition to analysis of the measure traits, for statistical analysis we also used several derived traits including volume index (height x diameter squared: VI), leaf density (area/weight: LD), and relative growth rate (RGR: which considers growth change over a period of time among the plants). RGR was calculated as natural log of the ratio of final and intial VI (ln VI2/VI1).

Leaf chlorophyll content

Relative chlorophyll content was determined based on three measurement per leaf on leaves collected from three locations on each plant (trunk at breast height, the middle and near the top of the crown). A SPAD 502 Plus Chlorophyll Meter was used for measurements.

Leaf area

Leaves were collected from three locations on each plant (from near the base of the plant, the middle of the plant, and the top of the plant). An HP Scanjet 8200 was then used for leaf scans. Each scan included the three leaves from the same tree and a ruler. Leaf areas were calculated using ImageJ software. In ImageJ, a scale was set based on the ruler from the scan photos. Then the image resolution was changed to 8-bit and the color threshold was adjusted to distinguish leaves from the background. The wand tool was used to outline all the 3 leaves; areas were calculated and recorded in an Excel sheet.

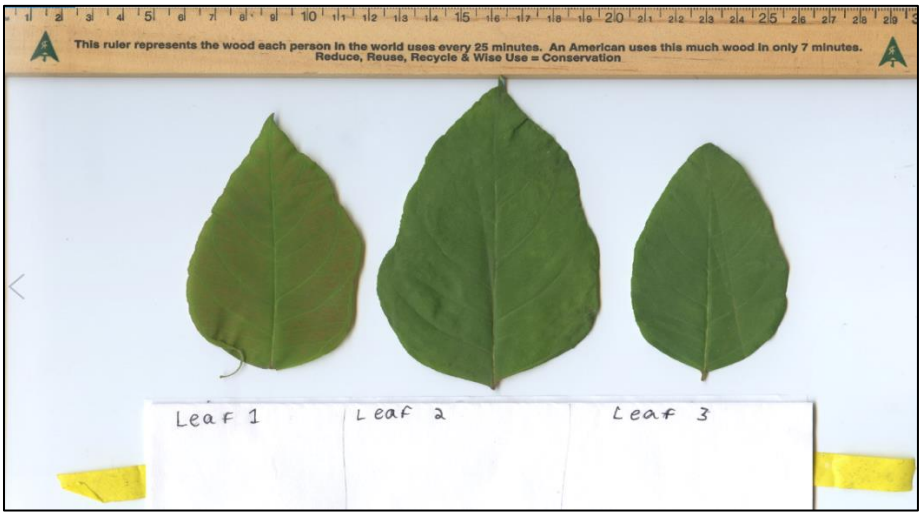


Figure 2.8: Example of leaf scans from a single plant viewed in ImageJ software

Leaf mass/dry weight

After measuring leaf area, leaves were placed in drying oven at 62 degree C for 10-12 days. All the three leaves from an individual tree were measured together on a Mettler AJ100 scale.

Leaf oil gland counts

Fresh leaves were used as to count oil gland density under the Keyence Digital microscope VHX-1000E. Leaves were placed on the microscope platform with the top surface of leaf facing upward towards the lens. In the software, two frames were used for calculating the number of oil glands per unit area.

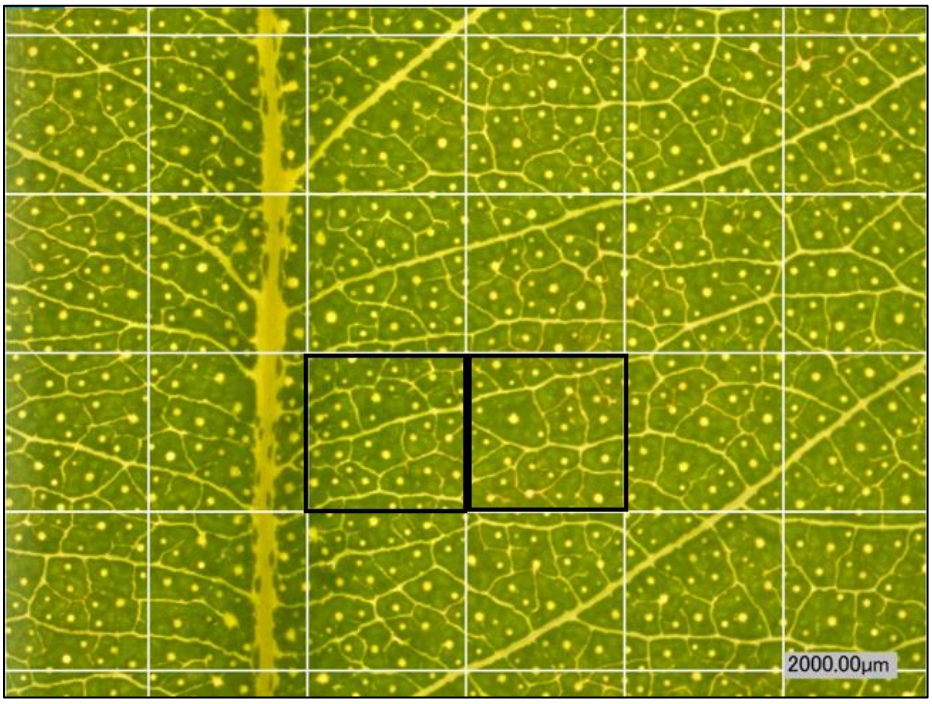


Figure 2.9: An example of oil gland seen in the Keyence microscope at 400x magnification. We used the two highlighted frames to estimate the density of oil glands in a leaf.

2.2.9 Statistical analysis

Statistical analysis was conducted using the R software version 4.0.0 (R Core Team, 2019). First, we graphically examined the distribution of each trait to look for the presence of outliers that might indicate incorrectly recorded data, or highly unusual plants as a result of damage or other factors. They were visualized using boxplots which were constructed using the package ggplot2 in R (Wickham, 2016). Differences among trait means for individual events were also visualized using bar plots.

We used a linear mixed effects model (LMM) (Zuur et. al, 2009) to test whether gene knockouts produced using either of the two CRISPR-Cas9 constructs (*EDA33* and *TDF1*) affected vegetative growth or leaf morphology compared to the wild type or transgenic (Cas9 only) controls. The model included the fixed effect of experimental group/construct (four types that were described above), the fixed effect of block (plants were arranged in blocks based on observed plant vigor), the random effect of event-within-construct, and residual error. Assumptions of homogenous variance and normality of errors were checked graphically with residual plots and histograms. When the assumption of homogenous variance was violated, we relaxed the assumption by allowing variances to differ by experimental group or by data transformations. Model fitting was performed using the R package "nlme" (Pinheiro et al., 2020). Estimated marginal means and 95% confidence limits from the fitted models were obtained using the package "emmeans" (Russell Lenth,, 2020).

The LMM was:

Model (LMM) = lme(Trait ~ Block + Construct, random = ~1|Events)

For each trait we report an overall F test for experimental groups along with planned comparisons among specific experimental groups from the LMM. The F tests indicate if there are any statistically significant (5%) differences in mean trait values between any of the knock-out events, including the transgenic or wild-type controls. Comparisons among the transgenic experimental group and control (*EDA33, TDF1 and*

Cas9only) were calculated using the "contrasts" function and the "pairs" function from the emmeans package. We used the Tukey HSD correction for a family of three comparisons. There were 3 comparison groups: 1) EDA33 vs. Cas 9; 2) TDF1 vs. Cas 9; 3) TDF1 vs. EDA33. We conducted statistical analyses both with and without SP7 wild type controls as inspection of growth data showed it was substantially larger even at the outset of the experiment, likely due to its different propagation history (see discussion).

In the equivalence test graphs, we designated an equivalence region with the goal of identifying a threshold below which the effects of the construct knockouts might be considered practically equivalent and thus acceptable for commercial uses (see discussion). This was based on a lower bound, set using negative 10% of the mean of the Cas9 control as a reference, as there was no reason to expect that knock-out of these genes would positively affect vegetative growth prior to reproduction (see discussion). To inform these comparisons, we used estimated confidence intervals (CI) as an indication of plausible values for the true difference in the difference seen. If the CI for a construct/knock-out to Cas9 difference overlapped the negative 10% difference line, the difference was assumed not to be different from 10%, and thus the construct/knock-out was equivalent to the Cas9 control for practical purposes.

We estimated the variance explained by the experimental group (fixed effect), transgenic events (random effect), and block/environment effects (residual effects) for each trait using the method for calculating the variance components of LMM's in Nakagawa and Schielzeth (2013). We fit new LMM's for each trait with the same structure as the models described above but without including block as a fixed effect. In these models the residual variance incorporates trait variation caused by both block and other environmental effects. The package insight (Ludecke et al. 2019) was used for extracting the variance or pseudo R^2 from the models.

2.3 Results

2.3.1 Genetic mutation analysis

High knockout rates in *EDA33*

We sequenced 16 independent events for investigating knock-out mutations among the two guide RNA sites within the *EDA33* gene. Out of those sequenced, only 12 events had both alleles at both sgRNA sites confidently defined (Table 2.3). Due to difficulties in obtaining chromatogram sequences for few alleles, 4 events had a single allele defined and we did not include them in the analysis. Out of the 12 defined events, 8 had frame-shift mutations or large deletions in both the alleles but at different locations in the guide RNA (heterozygous biallelic knockout), two had mutations at the same locations in both the alleles (homozygous biallelic knockout), and two had one mutated allele and one WT allele (heterozygous mutant). In summary, 10 of 12 independent events had both the alleles altered, making the putative total knockout rate 83%.

Moderate knockout rates in *TDF1*

We sequenced 16 independent events for investigating knock-out mutations among the two guide RNA sites within the *TDF1* gene. Out of those sequenced, only 13 events had both alleles defined (Table. 2.3). We did not include the remaining 3 events in the analysis as only one allele was defined for those events. Out of the 13 defined events, 6 had frame-shift mutations or large deletion mutations in both the alleles but at different locations in the guide RNA sites (heterozygous biallelic knockout), and 7 had one mutated allele and one WT allele (heterozygous mutant). In summary, 6 of 13 independent events had both the alleles altered to putatively make them non-functional, making the total knockout rate 46%.

No mutations detected in Cas9-Only transgenic controls

A total of 5 empty vector control events that had only the Cas9 gene sequence had no mutations in either of the alleles of *EDA33* or *TDF1*.

Mutation types correspond to activity in the guide RNAs

Mutations have been categorized into small deletions (deletions of < 5 base pair), small insertions (insertions of <5 base pair), large deletions (deletions of > 10 base pairs), large insertions (insertions of > 10 base pairs), inversions (when the entire nucleotide sequence between the guide RNAs have been reversed during the DNA repair mechanism), compound mutations (when both insertions and deletions were observed in the guide RNA locations simultaneously), and lastly, no mutation (no INDEL or inversion, and retained the same sequence as wild type). Number and percentages of the mutation types for allele 1 and allele 2 is shown in Tables 2.4 and 2.5, respectively.

In brief, *EDA33* mostly had large deletions of around 152 base pairs in both the alleles, with allele 1 having 33.3% and 41.7% mutations in guide RNA1 and guide RNA 2, respectively. Whereas, allele 2 had 50% large deletions extending from guide RNA1 to guide RNA2. Meanwhile, *TDF1* showed variability in mutations among the two guide RNA sites, with mostly no mutations or small deletions in both the alleles. Allele 1 had 53.9% wild type sequences in guide RNA1 and 46.2% wild type sequence in guide RNA 2. Allele 2 had 61% wild type sequence in guide RNA1 and 53.9% small deletions in guide RNA2. A detailed description of INDELS for both *EDA33* and *TDF1* gene is provided in Appendix Table A.2 and A.3.

Table 2.3 Mutation rate among *EDA33* and *TDF1;* for our greenhouse experiments we only used knockout events.

Gene	Total events studied	Heterozygous biallelic knockout	Homozygous biallelic knockout	Heterozygous mutant	No mutation-wild type
EDA33	12	8(66.7%)	2(16.67%)	2(16.67%)	0
TDF1	13	6(46.15%)	0	7(53.8%)	0

Table 2.4 Mutation types in allele 1 per guide RNA site.

Gene-sgRNA (Clone SP7)	Allele 1	Mutation type in each allele						
		Small deletion	Small insertion	Large deletion	Large insertion	Inversion	Compound mutation	No mutation
***EDA33*-sg1**	12	4(33.3%)	0	4(33.3%)	0	2(16.7%)	0	2(16.7%)
***EDA33*-sg2**	12	0	1(8.3%)	5(41.7%)	2(16.7%)	2(16.7%)	1(8.3%)	1(8.3%)
***TDF1*-sg1**	13	2(15.4%)	2(15.4%)	2(15.4%)	0	0	0	7(53.9%)
***TDF1*-sg2**	13	2(15.4%)	4(30.8%)	1(7.7%)	0	0	0	6(46.2%)

Table 2.5 Mutation types in allele 2 per guide RNA site.

Gene-sgRNA (Clone: SP7)	Allele 2	Mutation type in each allele					
		Small deletion	Small inser-tion	Large deletion	Large inser-tion	In-ver-sion	No mutation
***EDA33*-sg1**	12	3(25%)	0	6(50%)	0	0	3(25%)
***EDA33*-sg2**	12	3(25%)	1(8.3%)	6(50%)	0	0	2(16.7%)
***TDF1*-sg1**	13	2(15.4%)	0	3(23.1%)	0	0	8(61%)
***TDF1*-sg2**	13	7(53.9%)	1(7.7%)	3(23.1%)	0	0	2(15.4%)

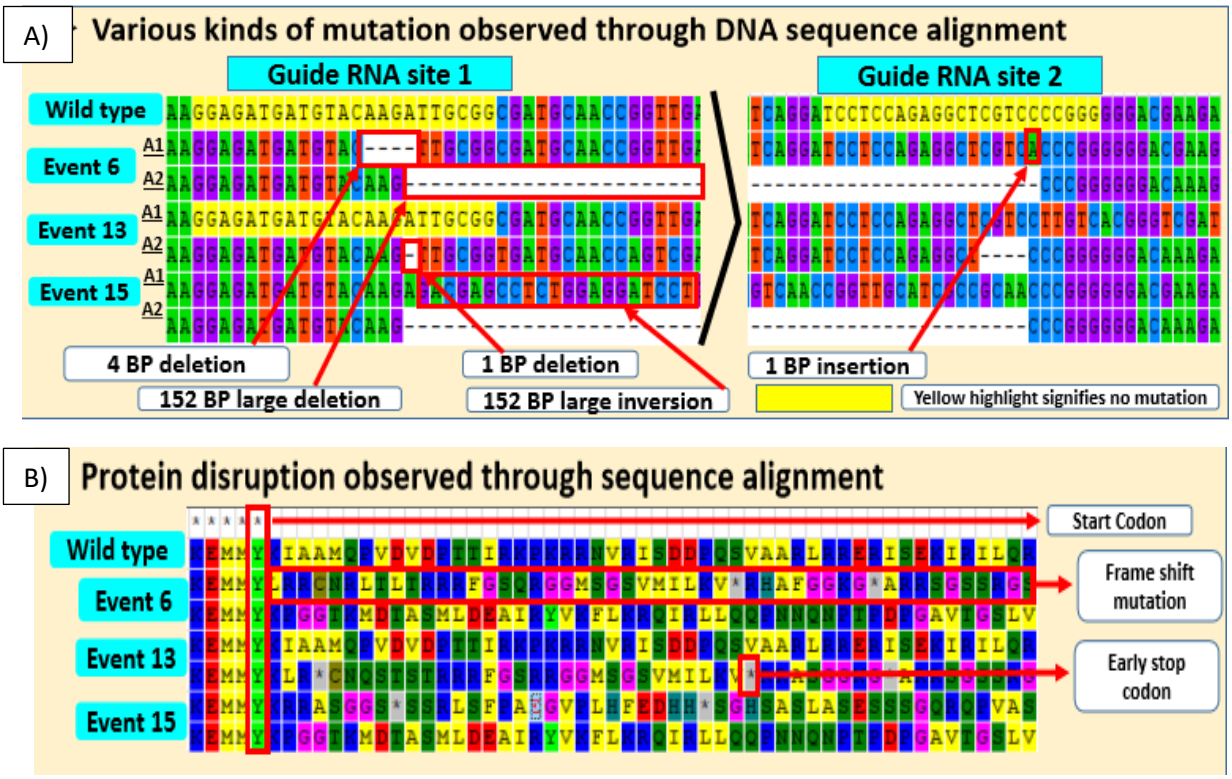


Figure 2.10: DNA and protein alignments of CRISPR mutants. A) DNA nucleotide sequence alignment; on the left are the event numbers with allele1 (A1) and allele2 (A2) labelled. Different types of mutations have been labeled at the bottom. All the events have been compared against the wild type. B) Protein alignment; frame shift mutations cause amino acid changes as shown in color coding, and early stop codons are indicated with asterisks.

2.3.2 Statistical analysis

The means and outliers for the experimental groups were visualized using boxplots. We could not find any clear outliers in our data, therefore we continued with our model fitting and analysis without removing any plants or events. The results of statistical analysis for each vegetative trait have primarily been divided into:

A) Report of overall F-test done on experimental groups with SP7 wild type control in consideration,
B) Report of overall F-test and comparisons done on experimental groups without SP7 wild type control,
C) Report of amount of variance in the response variable explained by experimental group, events, and environment.

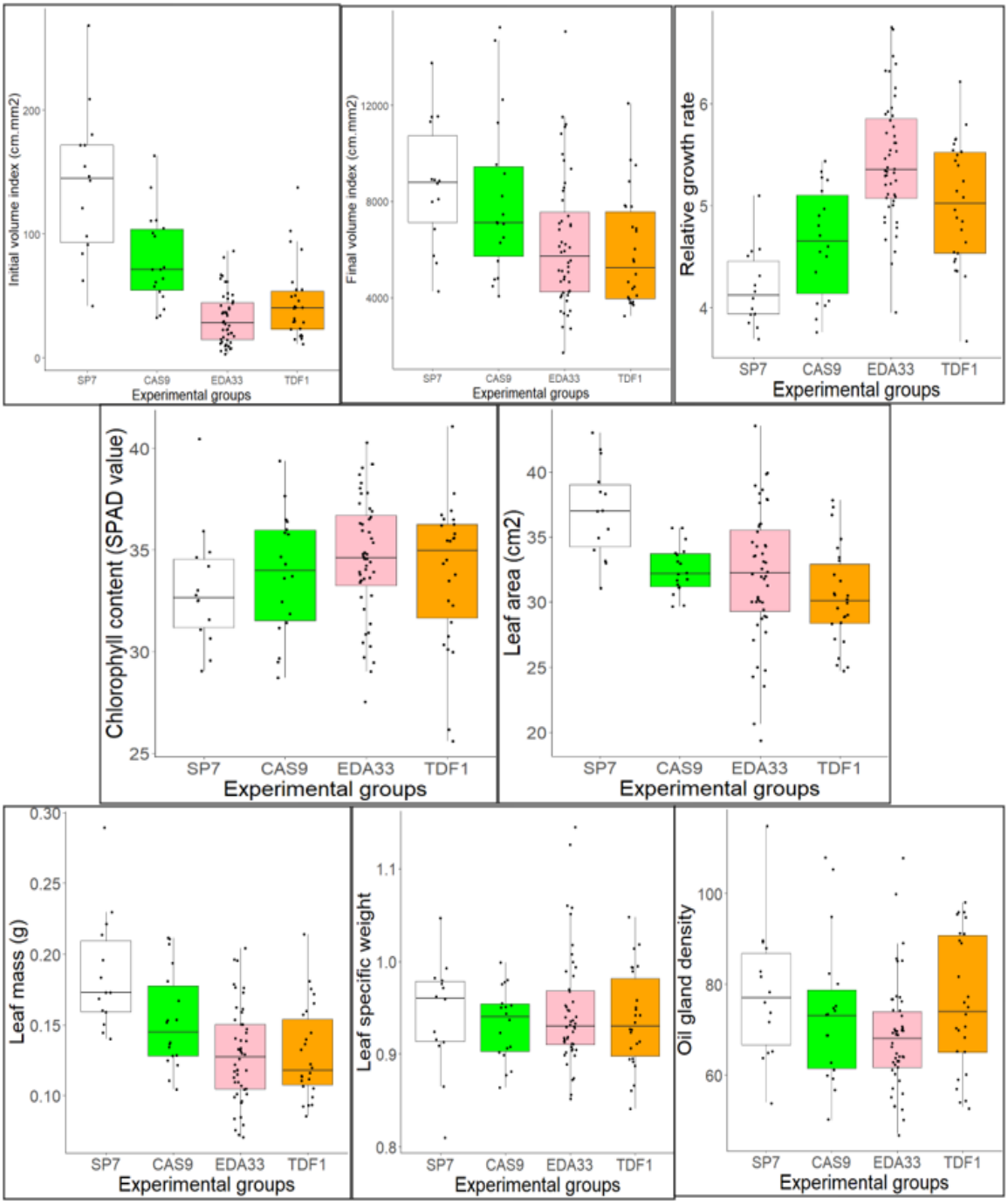
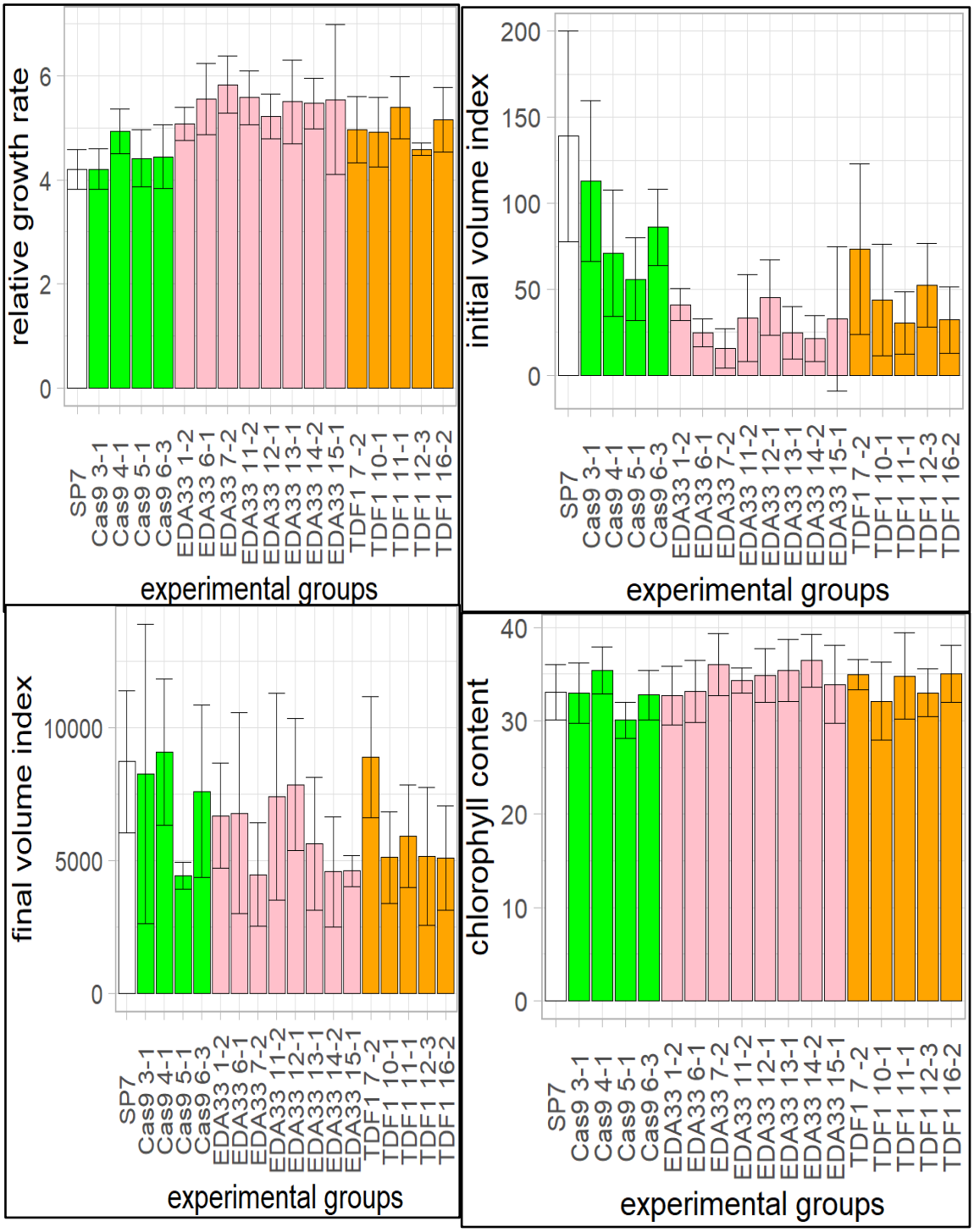


Figure 2.11: Boxplots displaying the mean of experimental groups and value of individual trees.The upper limit shows the biggest value for a trait and lower limit the smallest value



z **Figure 2.12:** Bar plots with standard deviation bars showing means of vegetative traits for individual events.

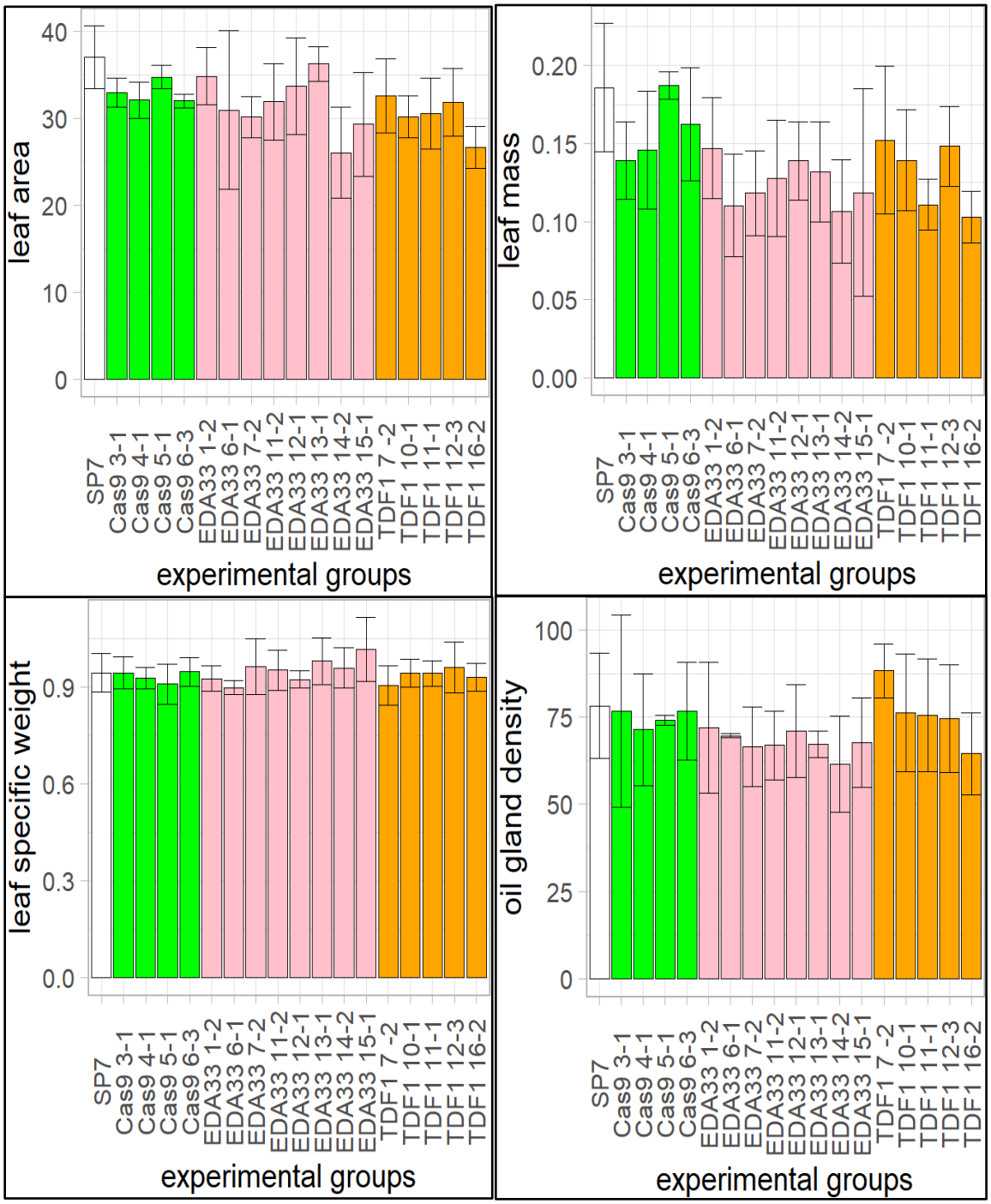


Figure 2.12 (continued): Bar plots with standard deviation error bars showing means of vegetative traits for individual event

Overall F-test

Results with wild type

Overall F- results are summarized in Table 2.6. Based on F-value and p-values, we found little evidence against the null hypothesis for vegetative traits such as final volume index, chlorophyll content, average leaf area, leaf specific weight and oil gland density. The estimated means among the experimental groups for the traits mentioned above were similar. However, we obtained strong evidence against the null hypothesis for traits such as relative growth, initial volume index and average leaf mass. The diagnostic plots to check the model fit can be found in the appendix Figure A.2.

Table 2.6: Overall F-test results with wild type

Vegetative trait	**Experimental group F-value $_{3,14}$**	**Experimental group p-value**	**Block p-value**
1. RGR	12.19	**3e-04**	**<0.0001**
2. Initial VI	18.95	**<0.0001**	**<0.0001**
3. Final VI	1.98	0.16	**0.0002**
4. Chlorophyll content	0.80	0.51	**<0.0001**
5. Average leaf area	2.60	0.09	0.53
6. Average leaf mass	7.59	**0.003**	0.20
7. Leaf specific weight	0.42	0.74	**0.03**
8. Oil gland numbers	2.7	0.08	0.87

Results without wild type

Overall F-test results (F-statistic value, degrees of freedom and P-value) demonstrating the degree of evidence against the null hypothesis that there were no differences in the estimated means of vegetative traits among any of the experimental groups in this study have been summarized in the Table 2.7. Based on F-value and p-values, we had little evidence against the null hypothesis for vegetative traits such as final volume index, chlorophyll content, average leaf area, average leaf mass, leaf specific weight and oil gland density. The estimated means among the experimental groups for the traits mentioned above were similar. However, we obtained strong evidence against the null hypothesis for the traits such as relative growth rate and initial volume index. Their values among the four experimental groups are significantly different.

Table 2.7 – Overall F-test results without wild type

Vegetative trait	**Experimental group F-value** $_{2,14}$	**Experimental group p-value**	**Block p-value**
1. RGR	11.87	**0.001**	**0.0001**
2. Initial VI	13.82	**5e-04**	**0.0001**
3. Final VI	1.34	0.29	**0.0005**
4. Chlorophyll content	0.70	0.51	**0.0001**
5. Average leaf area	0.95	0.14	0.23
6. Average leaf mass	3.27	0.07	0.18
7. Leaf specific weight	0.57	0.58	0.07
8. Oil gland numbers	2.77	0.097	0.57

Equivalence testing and estimated marginal means

Differences in the estimated means of vegetative traits between experimental groups and Tukey-adjusted 95% confidence intervals for the family of three comparisons around those estimated differences are summarized in Table 2.8 and Figure 2.13.

Table 2.8: Estimated confidence intervals for differences in construct type means.

Vegetative Trait	Comparison	Difference	Lower	Upper
1. RGR	EDA33 – Cas 9	0.90	0.41	1.38
	TDF1 – Cas 9	0.45	-0.08	0.98
	TDF1 – EDA33	-0.45	-0.91	0.01
2. Initial Volume index	EDA33 – Cas 9	-50.12	-75.25	-24.99
	TDF1 – Cas 9	-35.10	-63.60	-6.59
	TDF1 – EDA33	15.02	-5.91	35.96
3. Final Volume index	EDA33 – Cas 9	-1684.65	-4630.10	1260.79
	TDF1 – Cas 9	-1828.58	-4910.34	1253.16
	TDF1 – EDA33	-143.93	-2321.79	2033.92
4. Chlorophyll content	EDA33 – Cas 9	1.06	-1.50	3.63
	TDF1 – Cas 9	0.37	-2.43	3.18
	TDF1 – EDA33	-0.69	-2.94	1.56
5. Average leaf area	EDA33 – Cas 9	-0.84	-5.09	3.39
	TDF1 – Cas 9	-2.32	-6.96	2.31
	TDF1 – EDA33	-1.47	-5.24	2.28
6. Average leaf mass	EDA33 – Cas 9	-0.02	-0.05	0.001
	TDF1 – Cas 9	-0.02	-0.06	0.01
	TDF1 – EDA33	0.003	-0.02	0.03
7. Leaf Specific weight	EDA33 – Cas 9	0.015	-0.02	0.05
	TDF1 – Cas 9	0.06	-0.04	0.05
	TDF1 – EDA33	-0.01	-0.04	0.03
8. Oil gland number	EDA33 – Cas 9	-5.52	-15.56	4.50
	TDF1 – Cas 9	1.96	-9.17	13.09
	TDF1 – EDA33	7.48	-1.35	16.32

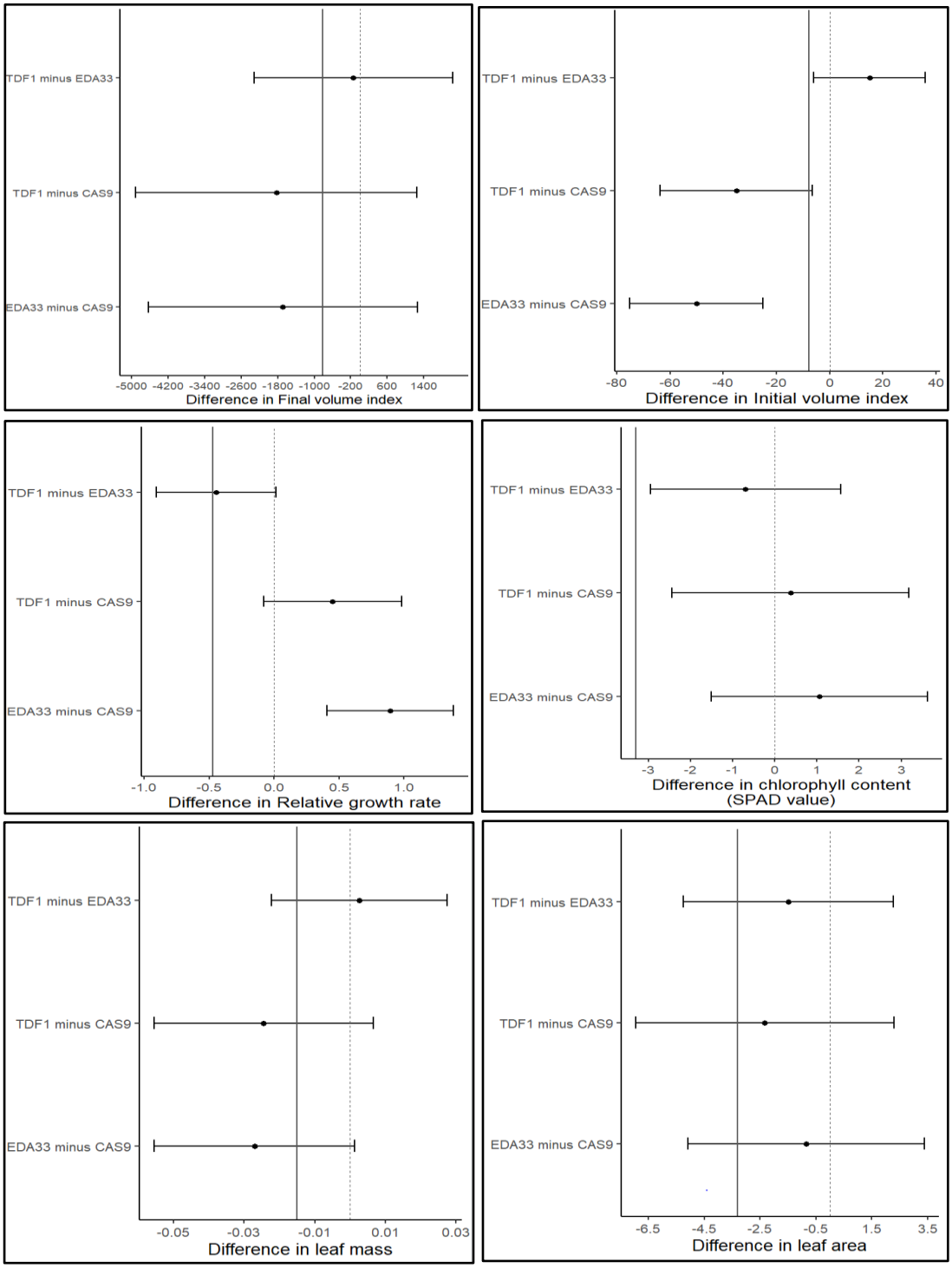


Figure 2.13: Equivalence testing at -10% reference of the mean (Appendix Table A.3) of Cas9 control. Confidence interval range can be seen for three comparisons

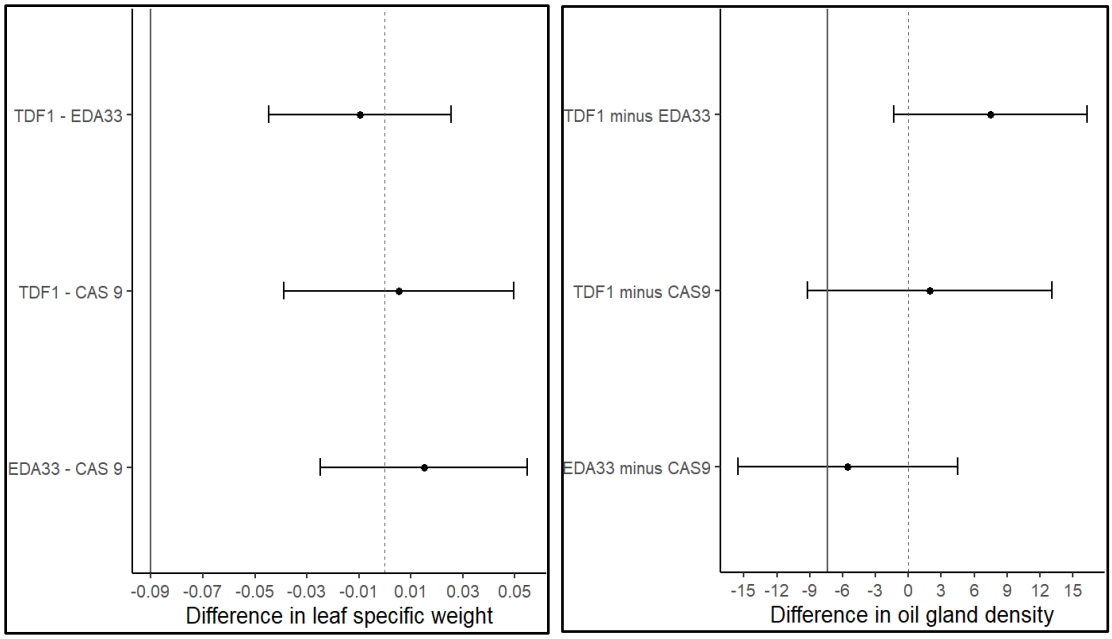


Figure 2.13 (continued): Equivalence testing at -10% reference of the mean of Cas9 control. Confidence interval range can be seen for three comparisons

The above graphs are helpful in visualizing the range of estimated true mean at a confidence interval of alpha 0.05. For leaf specific weight and chlorophyll content all the possible means are over the 10% boundary. Hence, it can be said that sterility constructs do not cause any negative effects in leaf characteristic mentioned. However, for other characteristics such as oil gland density, leaf area, leaf mass, relative growth rate and final volume index, the results are inconclusive. The mean below 10% could be of practical importance based on a forest manager's requirement. Although, the true means from model for all the growth characteristics except for initial volume index fall above the 10% mark, indicating that the growth is not getting affected negatively. The most likely reason for a lower value in initial volume index for EDA33 could be due to the transplanting shock. However, the growth normalizes in final volume index.

Variance Components

Variance components were estimated solely for quantifying the degree of importance of the various sources of variance. These included construct/control type, events, and residuals (environmental factors and blocks). The percentage value in the table represent the variance explained by each level. Value for the levels have been derived using proportion (variance divided by total model variance as a percentage).

Table 2.9: Variance in the data explained by experimental group, event and block/environmental factors

Vegetative trait	Experimental group	Event	Residual
RGR	38%	8%	54%
Initial VI	53%	0.5%	46%
Final VI	11%	10%	78%
Chlorophyll content	3%	5%	91%
Average leaf area	17%	15%	67%
Average leaf mass	25%	5%	70%
Leaf specific weight	0.09%	0.1%	99%
Oil gland number	7%	1%	92%
Mean	19.3%	5.6%	74.6%

2.4 Discussion

Our main results were that CRISPR-Cas9 was highly successful in generating site-directed, loss-of-function mutations in the reproductive genes *EDA33* and *TDF1* of *Eucalyptus,* and that greenhouse experiments support our expectation that the loss of function of these floral genes do not have large consequences for vegetative characteristics or growth rate in the greenhouse. We observed a high mutation rate in *EDA33* (83.3%) and *TDF1* (46.2%), with deletions, insertions, compound mutations and inversions. The small distance between the target guide RNAs as seen in *EDA33* (distance between guide RNAs was 162 bp), may have promoted its high rate of large mutations between the two guide RNAs; it showed a high rate of both large deletions (~74%) and inversions (~16%). Overall, our results that are similar to those observed in many other plant species, including *Arabidopsis thaliana, Brassica oleracea, Citrus*

sinensis, *Nicotiana benthamiana*, *Oryza sativa* and *Populus tomentosa* (Chen et al., 2017). A prior study conducted in our laboratory with CRIPSR-Cas9 directed against poplar homologs of the fertility genes *LEAFY* (*LFY*) and *AGAMOUS* (*AG*) also gave a high mutation rate (77.5%). It was based on 474 knock-out events, most of which were predominantly small indels, but also identified large deletions between the two guide RNAs employed similar to our results (Elorriaga et al., 2018).

With respect to vegetative effects of the knock-out of our target genes, to our knowledge there have been no prior randomized experiments, or even measurements, of the vegetative characteristics of mutants in *EDA33* and *TDF1*. Prior research conducted with *TDF1* (Zhu et al, 2008) and *EDA33* (Pagnussat et al, 2004) involved only the molecular characterization and functions of these genes with respect to reproductive development in Arabidopsis mutants.

We believe research similar to what we conducted would benefit from some revisions to our protocol. First, *in vivo* rooting among all the plants, including control and transgenic plants, should be carried out at the same time and use the same sizes and physiological conditions of explants. Despite our plans for this, due to an unexpected scheduling issue the control group (SP7 wild type) were transplanted in the rooting media two weeks *after* the transgenic trees already started rooting. This time gap, and their superior size and condition for reasons that are not clear to us, could be the possible reason for difference in the means of initial growth parameters (height and diameter) and relative growth rate among the four experimental groups. However, this transplanting effect was transient, and it became non-significant by the end of the experiment. The small initial size of the *EDA33* plants may also be the reason for its smaller volume at the start of the experiment compared to other genetic groups, and thus also for its statistically significantly larger RGR (a ratio of final to initial size).

Second, it is helpful for statistical analysis to have a balance in the number of data points among treatments (in this case transgenic and non-transgenic trees). The number of ramets (duplicates) for each event was different, resulting in a disproportionate number of trees between control and transgenic events (32 control trees vs 74 transgenic trees).

Since the transgenic group had many more trees than the control, this may be a reason for the very different levels of underlying variance, and thus statistical sensitivity to differences, that we observed (particularly for chlorophyll content).

Reliably sterile transgenic trees are expected from CRISPR technology, and may be helpful for relaxing public and regulatory concerns over escape of transgenic and exotic trees (Elorriaga et al. 2018). However, they also present some problems. For example, a completely sterile tree will be harder to breed and propagate, thus compromise the advancement of conventional breeding. A male-sterile, pollen-less tree such as we expect from biallelic mutation of *TDF1* would be expected to reduce the quality of honey as a forest plantation coproduct, which is common among eucalypt plantations in many parts of the world. It would also be less valuable to other kinds of fauna that feed on pollen (Strauss et al. 2017). A tree that does not release seeds, as we expect from biallelic mutation of *EDA33*, might be less valuable as a source of food for forest dwelling fauna. Finally, public sentiment against intentional inhibition of reproduction may render such technology unavailable in some places; for example, Brazil, a major grower of eucalypts, has a law against recombinant DNA modification of reproduction (Strauss et al. 2009).

After initial size and physiological differences were reduced by subsequent growth, statistical analysis of data through equivalence testing and F-tests suggested equivalent vegetative growth among the four experimental groups. Plant height and diameter differed initially when the plants were freshly transplanted into big pots in the greenhouse. However, growth among all the experimental groups stabilized in three months, resulting in similar final volume index among experimental groups. Research with lettuce transplantation provides evidence of similar recovery. Despite variation in initial size and transplant conditions, final yield is usually not impacted due to plants' quick and adaptive response to mechanical damage inflicted during *ex vitro* transplantation. Adaptive changes in root:shoot ratio appear to restore vigor most of the time (Struick et al, 2013).

Variation in transplant shock appears to have been a major cause of variation in early growth. Micropropagation is still poorly understood, but seems to cause high rates of damage and varied growth among plants transferred to *ex vitro* conditions (greenhouse or field). Plants *in vitro* are provided with special conditions in air-tight cultivation vessels, where they have a high dose of sugars and growth regulators. Along with decreased air turbulence and very high relative humidity, this results in abnormal plantlet morphology, anatomy and physiology (Kozai and Smith, 1995). When transferred into *ex vitro* conditions, plants tend to be under great stress, and act to correct these abnormalities in physiology and grow stably again after several weeks (Plzakova et al., 1999).

RGR was used in an attempt to correct for differences in initial size of plants. However, the standard approach to calculate RGR, using natural logarithm-transformed plant volume index change over time, is known to often yield a biased estimate of RGR (Poorter and Hoffmann, 2002). This bias usually arises from the fact that plants grow at faster pace in the starting compared to a later stage in the life cycle of tree. Larger trees show slowed growth rate compared to the smaller ones. There are different estimators for RGR that could be used, as well as different growth measurements; however, we choose not to harvest (destroy) trees for calculating their dry weight (Poorter and Hoffmann, 2002), and lacking other information could not choose among alternative RGR estimators in an unbiased manner.

To assess variation in vegetative morphology, we choose diverse aspect of leaf size, weight, shape, and physiology that could be assessed rapidly in large numbers of plants. These traits are also usually much more highly heritable and developmentally stable compared to height and diameter, thus might be more sensitive measures of genetic differences. Leaf morphology and chlorophyll density are important determinants of leaf photosynthetic potential. A defect in leaf morphology can affect eucalypt survival and productivity in the wild or in plantations. To our surprise, however, analysis of the sources of variance among traits did not show any clear trends when growth traits were compared to morphology traits, perhaps because of the large variability in transplant

conditions and timing discussed above. The very high statistical significance of block effects, however, is most prominent for the volume traits vs. the morphology traits, with the sole exception of chlorophyll density—which appeared to behave more like a growth than a morphology trait in having large block effects. The division of smaller and larger plants into blocks after transplant was clearly useful at reducing residual variance and thus improving statistical sensitivity.

Equivalence testing is usually used in toxicology and drug effect studies, where the main concern is not if there is simply a difference, but if that difference is large enough to have a practical consequence. The null hypothesis in equivalence testing effectively assumes group means to be different, unlike the traditional understanding of equal means among groups (Dixon et al., 2018). The Two-One-Sided-Tests (TOST) method is often employed to test equivalence. An upper and lower boundary (equivalence region) are set based on some reference value that expresses differences of consequence. Because our concern with floral gene knock-outs is mainly that growth might be impaired, we focused on whether there were practical *decreases* in vegetative growth traits. We therefore only set a lower boundary based on 10% of our reference (i.e., the Cas9 control). As shown through graphical analysis using confidence intervals around trait means, for none of our traits at harvest was there a 10% or greater reduction in trait value. We therefore conclude that we were unable to identify significant reductions in vegetative growth characteristics. However, given the limitations in our experimental procedures, including the variation in size, timing, and transplant conditions—which are represented in the large confidence intervals around all of our traits—it is also possible that our experiment lacked the precision needed to detect real differences that may exist.

The goals of this experiment were to produce CRISPR-Cas9 based mutations to disable selected floral genes, and then to see if the loss of gene function had detectable vegetative consequences. Despite a low rate of transformation in even this relatively easy to transform eucalypt genotype, we were easily able to produce the desired number of biallelic, loss of function genotypes and non-mutated controls. As in many prior studies in our laboratory and elsewhere, CRISPR-Cas9 is an extraordinarily efficient and precise

method for directed mutation. Our results, though limited in precision by experimental variations as discussed above, showed no evidence for perturbations to growth or leaf morphology once trees had stabilized in the greenhouse. All of them also had normal appearance, despite the continued presence and expected expression of the CRISPR-Cas9 locus. Thus, we consider the *EDF33* and *TDF1* genes promising candidates where female and male sterility, respectively, are desired in *Eucalyptus*, whether or not the CRISPR locus is excised or segregated away. Field studies, and analyses in other genotypes of this highly diverse genus, are logical next steps in research and development.

Chapter 3: Conclusion

CRISPR-Cas9 was efficient in generating knockout mutations among the sterility genes *EDA33* and *TDF1*. CRISPR against *EDA33* showed a high mutation rate (83.3%) with mostly large deletions, insertions and inversions observed. CRISPR against *TDF1* showed a moderate knockout mutation rate (46.2%), with mostly small deletions of one – two base pairs. Through statistical analysis we found that the presence of sterility constructs resulted in largely normal vegetative growth with no statistically significant abnormalities observed among transgenic knock-outs and transgenic controls. Additionally, equivalence testing further provides us with evidence of "practically" normal vegetative growth when compared with Cas9 control. True estimated means derived from the model were over the 10% lower boundary reference of the control.

Future research

1. Naturally, *Eucalyptus grandis* requires at least 2-3 years to initiate flowering. This long generation time can be a limiting factor for tree genetic research. Hence, we transformed *Eucalyptus grandis and Eucalyptus urophylla* hybrid (SP7) tree *in vivo* with a variety of plasmid vectors that enabled overexpression of *FLOWERING LOCUS T (FT)* resulting in rapid flowering of trees (Klocko et. al., 2015). The next step would be to hold a greenhouse experiment for characterizing the phenotype and morphology of flowers generated after knocking out *EDA33* and *TDF1* genes in *FT*-SP7 transgenic lines—work that is in fact underway in our laboratory.
2. Plant phenotype and growth is determined by genetic composition as well as environmental factors such as availability of sunlight, water, nutrition, humidity and temperature—and their interaction with genotype. Therefore, as a next step, it would be critical to conduct field trials for: A) Testing the effects of knockouts on flowering and vegetative growth; B) Testing stability of construct and allele conversion; C) Studying pleiotropic effects; and D) Inspecting for chimerism and off-target mutagenesis over time.

3. It would be highly desirable to remove the CRISPR components from the genome after mutagenesis. However, doing this by sexual segregation would take many years and disrupt the genotype of elite clones. Thus, it is important to develop means for somatic excision, such as by using induced recombinases after mutagenesis. This is an active area of work in our own and other laboratories.

Limitations with using transgenic trees not only revolve around strict regulatory barriers but also the infancy of scientific studies related to *Eucalyptus*. Along with biotechnological advances in *Eucalyptus*, *in vitro* transformation capacity should be improved as well. The very low regeneration ability of most types of *Eucalyptus* poses problems with transformation. The more time explants spend in tissue culture and transformation, the greater the chance of chimeras and somaclonal mutation. The strain of *Agrobacterium* is also highly important and needs to be matched to the eucalypt clone (Girijashankar, 2011). Therefore, more research needs to be undertaken with transformation.

CRISPR-Cas9 is a highly efficient gene editing technology in plant biology. However, there are a few problems with it. 1) The CRISPR/Cas9 vector can be quite large, making delivery of the system into plant genome difficult. 2) The mandatory presence of a PAM recognition site limits target site selection. 3) Transformation takes a long time to select and identify mutants; it took us 2 years to find positive transgenic events in *EDA33* (transformation rate of 4.7% transformation rate) and *TDF1* (4.2%). Therefore, a tissue culture-free GE system might be needed with CRISPR/Cas9. Since CRISPR-Cas9 is a relatively new gene editing system more research on stability of it in the genome is required (Manghwar et. al, 2019).

In summation, genes from genetically engineered trees are likely to migrate into feral or wild populations. But with correct and responsible measures taken with plant sterility the problems with this ecological risk, public concern and regulatory barriers may be mitigated. GE trees can help solve various environmental problems such as reduced biodiversity and sustainable silviculture. However, regulatory barriers and the high cost of GE crop production make commercialization and research challenging.

Hopefully, as research proceeds and is able to show benefits and technical reliability, these barriers will recede.

Bibliography

Aaron M. Shew, L. Lanier Nalley, Heather A. Snell, Rodolfo M. Nayga, Bruce L. Dixon, 2018, "CRISPR versus GMOs: Public Acceptance and Valuation - ScienceDirect." Volume 19, Pages 71-80, https://www.sciencedirect.com/science/article/pii/S2211912418300877.

"**Agroforestry.Org - Forestry Plantation**." Accessed February 1, 2020. https://agroforestry.org/2014-03-04-10-06-24/forestry-plantation.

Ahuja, Mulkh-Raj, and William J Libby. 1993. *Clonal Forestry I: Genetics and Biotechnology*. Berlin, Heidelberg: Springer Berlin Heidelberg. http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=3095919

Allona, Isabel, Matias Kirst, Wout Boerjan, Steven Strauss, and Ronald Sederoff, eds. 2019. *Forest Genomics and Biotechnology*. Frontiers Research Topics. Frontiers Media SA. https://doi.org/10.3389/978-2-88963-178-0

Andrew Newhouse. 2018 "Opinion | We Nearly Killed off These Trees. But Biotech Can Bring Them Back." Washington Post. https://www.washingtonpost.com/news/theworldpost/wp/2018/05/29/biotech/**.**

AO, WFP, and IFAD (2012). *The State of Food Insecurity in the World 2012: Economic Growth is Necessary but not Sufficient to Accelerate Reduction of Hunger and Malnutrition*. Rome: FAO Available online at: http://www.fao.org/docrep/016/i3027e/i3027e00.htm

Bae, Sangsu, Jeongbin Park, and Jin-Soo Kim. 2014. "Cas-OFFinder: A Fast and Versatile Algorithm That Searches for Potential off-Target Sites of Cas9 RNA-Guided Endonucleases." *Bioinformatics (Oxford, England)* 30 (10): 1473–75.

Baker, 2018 "Arctic Apples: A Fresh New Take on Genetic Engineering - Science in the News."http://sitn.hms.harvard.edu/flash/2018/arctic-apples-fresh-new-take-genetic-engineering/.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J., and Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* 32, 76–84. doi: 10.1016/j.copbio.2014.11.007

Belhaj, Khaoula, Angela Chaparro-Garcia, Sophien Kamoun, and Vladimir Nekrasov. 2013. "Plant Genome Editing Made Easy: Targeted Mutagenesis in Model and Crop Plants Using the CRISPR/Cas System." *Plant Methods* 9 (1): 39. https://doi.org/10.1186/1746-4811-9-39.

“Biotechnology.”, 2020. https://www.usda.gov/topics/biotechnology.

Blakely, W. F. 1955. A key to the eucalypts. 2d ed. Forest and Timber Bureau, Canberra, Australia.

Boland, D J & Boland, Douglas & CSIRO (1984). *Forest trees of Australia* (4th ed. rev. and enl). Nelson ; East Melbourne : CSIRO, Melbourne

Brooker, M. I. H., and Kleinig, 1994 D. A. *Field Guide to Eucalypts*. Melbourne

Brunner, Amy M., Jingyi Li, Stephen P. DiFazio, Olga Shevchenko, Brooke E. Montgomery, Rozi Mohamed, Hao Wei, et al. 2007. “Genetic Containment of Forest Plantations.” *Tree Genetics & Genomes* 3 (2): 75–100. https://doi.org/10.1007/s11295-006-0067-8.

Burdon, R.D. and W. J. Libby. 2006. Genetically Modified Forests: From Stone Age to Modern Biotechnology. Forest History Society, Durham, N.C. p 79

Burgess, Diane G., Edward J. Ralston, William G. Hanson, Matthew Heckert, Minh Ho, Tina Jenq, Joseph M. Palys, Keliang Tang, and Neal Gutterson. 2002. “A Novel, Two-Component System for Cell Lethality and Its Use in Engineering Nuclear Male-Sterility in Plants.” *The Plant Journal* 31 (1): 113–25. https://doi.org/10.1046/j.1365-313X.2002.01330.x.

Cary Funk and Meg Hefferon, 2018. “Most Americans Accept Genetic Engineering of Animals That Benefits Human Health, but Many Oppose Other Uses.” *Pew Research Center Science & Society* (blog). August 16, 2018. https://www.pewresearch.org/science/2018/08/16/most-americans-accept-genetic-engineering-of-animals-that-benefits-human-health-but-many-oppose-other-uses/

Chang, Shujun, Elizabeth L. Mahon, Heather A. MacKay, William H. Rottmann, Steven H. Strauss, Paula M. Pijut, William A. Powell, et al. 2018. “Genetic Engineering of Trees: Progress and New Horizons.” *In Vitro Cellular & Developmental Biology - Plant* 54 (4): 341–76. https://doi.org/10.1007/s11627-018-9914-1.

Chen, Baohui, Yuyu Niu, Haoyi Wang, Kejian Wang, Hui Yang, and Wei Li. 2020. “Recent Advances in CRISPR Research.” *Protein & Cell*, March. https://doi.org/10.1007/s13238-020-00704-y.

Chen, S., and A. Polle. 2010. “Salinity Tolerance of Populus.” *Plant Biology (Stuttgart, Germany)* 12 (2): 317–33. https://doi.org/10.1111/j.1438-8677.2009.00301.x

Christel Palmberg, 2002, Annotated bibliography on environmental social and

economic impacts of Eucalypts, Rome: FAO, http://www.fao.org/forestry/6172-01295154eb5459bc018e828076ff8a292.pdf

Coalition, Global Forest. 2018. "Fire and Plantations in Portugal: A Case Study on the Risks of Using Tree Plantations to Remove Carbon from the Atmosphere." Global Forest Coalition. September 28, 2018. https://globalforestcoalition.org/fire-and-plantations-in-portugal/.

Coleman, Heather D., Jimmy Yan, and Shawn D. Mansfield. 2009. "Sucrose Synthase Affects Carbon Partitioning to Increase Cellulose Production and Altered Cell Wall Ultrastructure." *Proceedings of the National Academy of Sciences* 106 (31): 13118–23. https://doi.org/10.1073/pnas.0900188106**.**

Cros, Eric Teissier du. 2000. "Situation of Forest Genetics: Research and Extension." In *Forest Genetics and Sustainability*, edited by Csaba Mátyás, 63:9–22. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-1576-8_3

Cui, Kai, and Sharon P. Shoemaker. 2018. "Public Perception of Genetically-Modified (GM) Food: A Nationwide Chinese Consumer Study." *NPJ Science of Food* 2 (June). https://doi.org/10.1038/s41538-018-0018-4.

Dean, Eric A, Paula M Pijut, and Micah E Stevens. 2014"Development of Transgenic North American White Ash (Fraxinus Americana) Expressing a Bacillus Thuringiensis Protein for Management of the Emerald Ash Borer," 1.

Dixon et al., 2012 "A Primer on the Use of Equivalence Testing for Evaluating Measurement Agreement - PubMed." Accessed May 31, 2020. https://pubmed.ncbi.nlm.nih.gov/29135817/.

Dubouzet, Joseph G., Timothy J. Strabala, and Armin Wagner. **2013**. "Potential Transgenic Routes to Increase Tree Biomass." *Plant Science* 212 (November): 72–101. https://doi.org/10.1016/j.plantsci.2013.08.006.

Elorriaga, Estefania, Amy L. Klocko, Cathleen Ma, and Steven H. Strauss. 2018. "Variation in Mutation Spectra Among CRISPR/Cas9 Mutagenized Poplars." *Frontiers in Plant Science* 9: 594. https://doi.org/10.3389/fpls.2018.00594.

"European Green Capital." 2019

https://ec.europa.eu/environment/europeangreencapital/countriesruleoutgmos/

Evelyn, John. 1679. *Sylva, or, A Discourse of Forest-Trees, and the Propagation of Timber...* London, England. http://search.proquest.com/docview/2240961998/citation/62E0E8B5DA47405FPQ/1

FAO (2010). *Forests and Genetically Modified Trees*. Rome: FAO

Fillatti, JoAnne J., James Sellmer, Brent McCown, Bruce Haissig, and Luca Comai. 1987. "Agrobacterium Mediated Transformation and Regeneration of Populus." *Molecular and General Genetics MGG* 206 (2): 192–99. https://doi.org/10.1007/BF00333574.

Fire, Andrew, SiQun Xu, Mary K. Montgomery, Steven A. Kostas, Samuel E. Driver, and Craig C. Mello. 1998. "Potent and Specific Genetic Interference by Double-Stranded RNA in Caenorhabditis Elegans." *Nature* 391 (6669): 806–11. https://doi.org/10.1038/35888.

"**Forestry | History, Types of Forests, Techniques, & Aims | Britannica**.". 2020 https://www.britannica.com/science/forestry.

Fritsche, Steffi, Amy L. Klocko, Agnieszka Boron, Amy M. Brunner, and Glenn Thorlby. 2018. "Strategies for Engineering Reproductive Sterility in Plantation Forests." *Frontiers in Plant Science* 9: 1671. https://doi.org/10.3389/fpls.2018.01671.

Gibson, Daniel G, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." *Nature Methods* 6 (5): 343–45. https://doi.org/10.1038/nmeth.1318.

Gilbertson, Larry. 2003. "Cre–Lox Recombination: Cre-Ative Tools for Plant Biotechnology." *Trends in Biotechnology* 21 (12): 550–55. https://doi.org/10.1016/j.tibtech.2003.09.011.

Girijashankar, V. 2011. "Genetic Transformation of *Eucalyptus*." *Physiology and Molecular Biology of Plants : An International Journal of Functional Plant Biology* 17 (1): 9–23. https://doi.org/10.1007/s12298-010-0048-0

Gonsalves, Dennis, Carol Gonsalves, Steve Ferreira, Karen Pitz, Maureen Fitch, Richard Manshardt, and Jerry Slightom. 2002 "Transgenic Virus Resistant Papaya: From Hope to Reality for Controlling Papaya Ringspot Virus in Hawaii," 12.

Gosal SK, Karlupia A, Gosal SS, Chhiba IM, Varma A. 2010 "Biotization with *piriformospora indica* and *Pseudomonas fluorescens* improves survival rate, nutrient

acquisition, field performance and saponine content of micropropagated *Chlorophytum* sp." Indian J Biotechnol. 9:289–297

Griffin AR , Moran GF Fripp YJ (1987) Preferential Outcrossing in *Eucalyptus regnans* F. Muell. *Australian Journal of Botany* **35**, 465-475. https://doi.org/10.1071/BT9870465

Grossnickle, S.C., Sutton, B.C. (1999) Applications of biotechnology for forest regeneration. *New Forests* 17, 213–226 https://doi.org/10.1023/A:1006501716404

Gu, Jing-Nan, Jun Zhu, Yu Yu, Xiao-Dong Teng, Yue Lou, Xiao-Feng Xu, Jia-Li Liu, and Zhong-Nan Yang. 2014. "DYT1 Directly Regulates the Expression of TDF1 for Tapetum Development and Pollen Wall Formation in Arabidopsis." *The Plant Journal* 80 (6): 1005–13. https://doi.org/10.1111/tpj.12694

Hakim MAnghwar, Keith Lindsey, Xianlong Zhang, Shuangxia Jin, 2019 "CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing: Trends in Plant Science." n.d. Accessed May 31, 2020. https://www.cell.com/trends/plant-science/fulltext/S1360-1385(19)30243-2.

Halpin, Claire, Simon C. Thain, Emma L. Tilston, Emma Guiney, Catherine Lapierre, and David W. Hopkins. 2007. "Ecological Impacts of Trees with Modified Lignin." *Tree Genetics & Genomes* 3 (2): 101–10. https://doi.org/10.1007/s11295-006-0060-2.

Harfouche, Antoine, Richard Meilan, and Arie Altman. 2014. "Molecular and Physiological Responses to Abiotic Stress in Forest Trees and Their Relevance to Tree Improvement." *Tree Physiology* 34 (11): 1181–98. https://doi.org/10.1093/treephys/tpu012.

Harper and Bancroft. 2018 "Developing a Genetic Test to Fight against Ash Dieback - Research, University of York." https://www.york.ac.uk/research/impact/fighting-ash-dieback/

Hastorf, Christine A. 2009. "Rio Balsas Most Likely Region for Maize Domestication." *Proceedings of the National Academy of Sciences* 106 (13): 4957–58. https://doi.org/10.1073/pnas.0900935106.

Hill, Kristine, and G Eric Schaller. 2013. "Enhancing Plant Regeneration in Tissue Culture." *Plant Signaling & Behavior* 8 (10). https://doi.org/10.4161/psb.25709.

Hill ex Maiden, 1918 [family MYRTACEAE], in Journ. Proc. Roy. Soc. N.S.W. 52: 501 (1918). F. White, F.F.N.R.: 296 (1952).—Blakely, Key Eucalypts,

Hill Michael, 2019 "High-Tech Chestnuts: US to Consider Genetically Altered Tree." https://apnews.com/e4dd339ee2ee408c8a40c9556aee318f.

Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315:804–807 https://doi.org/10.1093/bioinformatics/btu048.

Hinchee, Maud, Chunsheng Zhang, Shujun Chang, MIchael Cunningham, William Hammond, and Narender Nehra. 2011. "Biotech *Eucalyptus* Can Sustainably Address Society's Need for Wood: The Example of Freeze Tolerant *Eucalyptus* in the Southeastern U.S." *BMC Proceedings* 5 (Suppl 7): I24. https://doi.org/10.1186/1753-6561-5-S7-I24

HOFFMANN, WILLIAM A., and HENDRIK POORTER. 2002. "Avoiding Bias in Calculations of Relative Growth Rate." *Annals of Botany* 90 (1): 37–42. https://doi.org/10.1093/aob/mcf140

"**House of Commons Journal Volume 1: 15 February 1610** | British History Online."

https://www.british-history.ac.uk/commons-jrnl/vol1/pp393-394

https://en.wikipedia.org/wiki/Biotechnology#Definition

Huang, Jian, Ashley R. Smith, Tianyu Zhang, and Dazhong Zhao. 2016. "Creating Completely Both Male and Female Sterile Plants by Specifically Ablating Microspore and Megaspore Mother Cells." *Frontiers in Plant Science* 7. https://doi.org/10.3389/fpls.2016.00030

Ikeuchi, Momoko, Yoichi Ogawa, Akira Iwase, and Keiko Sugimoto. 2016. "Plant Regeneration: Cellular Origins and Molecular Mechanisms." *Development* 143 (9): 1442–51. https://doi.org/10.1242/dev.134668.

"ISAAA Brief 54-2018: Executive Summary | ISAAA.Org." https://www.isaaa.org/resources/publications/briefs/54/executivesummary/default.asp.

Jacobs, Maxwell R. 1955; Growth habitats of the *Eucalyptus*, Australia; Forestry and Timber Bureau

James, Rosalind R, Stephen P DiFazio, Amy M Brunner, and Steven H Strauss. 1998. "Environmental Effects of Genetically Engineered Woody Biomass Crops." *Biomass and Bioenergy* 14 (4): 403–14. https://doi.org/10.1016/S0961-9534(97)10077-0.

Keb-Llanes, Miguel, Gerardo González, Bartolomé Chi-Manzanero, and Diógenes Infante. 2002. "A Rapid and Simple Method for Small-Scale DNA Extraction InAgavaceae and Other Tropical Plants." *Plant Molecular Biology Reporter* 20 (3): 299–299. https://doi.org/10.1007/BF02782465.

Kelaine Vargas, 2009 "Ecosystem Services and Environmental Benefits" UC San Diego Campus Forest.

Kerbiriou, P. J., T. J. Stomph, E. T. Lammerts van Buere n, and P. C. Struik. 2013. "Influence of Transplant Size on the Above- and below-Ground Performance of Four Contrasting Field-Grown Lettuce Cultivars." *Frontiers in Plant Science* 4: 379. https://doi.org/10.3389/fpls.2013.00379.

Khan, Mohammad S., Muhammad A. Khan, and Dawood Ahmad. 2016. "Assessing Utilization and Environmental Risks of Important Genes in Plant Abiotic Stress Tolerance." *Frontiers in Plant Science* 7: 792. https://doi.org/10.3389/fpls.2016.00792.

Koester, R. P., Skoneczka, J. A., Cary, T. R., Diers, B. W., and Ainsworth, E. A. (2014). Historical gains in soybean (*Glycine max* Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. *J. Exp. Bot.* 65, 3311–3321. doi: 10.1093/jxb/eru187

Komor, Alexis C., Ahmed H. Badran, and David R. Liu. 2017. "CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes." *Cell* 168 (1–2): 20–36. https://doi.org/10.1016/j.cell.2016.10.044.

Kumar, Vikas, Sandeep Rout, and Manoj K Tak. 2015. "Application of Biotechnology in Forestry: Current Status and Future Perspective." *Nature Environment and Pollution Technology* 14 (3): 10.

Ledford, Heidi. 2014. "Brazil Considers Transgenic Trees." *Nature News* 512 (7515): 357. https://doi.org/10.1038/512357a.

Lee, Chanhui, Quincy Teng, Wenlin Huang, Ruiqin Zhong, and Zheng-Hua Ye. 2009. "Down-Regulation of PoGT47C Expression in Poplar Results in a Reduced Glucuronoxylan Content and an Increased Wood Digestibility by Cellulase." *Plant and Cell Physiology* 50 (6): 1075–89. https://doi.org/10.1093/pcp/pcp060

Letunic, Ivica, Tobias Doerks, and Peer Bork. 2015. "SMART: Recent Updates, New Developments and Status in 2015." *Nucleic Acids Research* 43 (Database issue): D257-260. https://doi.org/10.1093/nar/gku949.

Liljegren, Sarah J., Adrienne H. K. Roeder, Sherry A. Kempin, Kristina Gremski,

Lars Østergaard, Sonia Guimil, Daengnoy K. Reyes, and Martin F. Yanofsky. 2004. "Control of Fruit Patterning in Arabidopsis by INDEHISCENT." *Cell* 116 (6): 843–53. https://doi.org/10.1016/s0092-8674(04)00217-x.

Lucht, Jan M. 2015. "Public Acceptance of Plant Biotechnology and GM Crops." *Viruses* 7 (8): 4254–81. https://doi.org/10.3390/v7082819.

Lüdecke, Daniel, Philip Waggoner, and Dominique Makowski. 2019. "Insight: A Unified Interface to Access Information from Model Objects in R." *Journal of Open Source Software* 4 (38): 1412. https://doi.org/10.21105/joss.01412.

Lu, Haiwei, Amy L. Klocko, Amy M. Brunner, Cathleen Ma, Anna C. Magnuson, Glenn T. Howe, Xinmin An, and Steven H. Strauss. 2019. "RNA Interference Suppression of *AGAMOUS* and *SEEDSTICK* Alters Floral Organ Identity and Impairs Floral Organ Determinacy, Ovule Differentiation, and Seed-hair Development in *Populus*." *New Phytologist* 222 (2): 923–37. https://doi.org/10.1111/nph.15648.

Lu, Meng-Zhu, and Jian-Jun Hu. 2011. "A Brief Overview of Field Testing and Commercial Application of Transgenic Trees in China." *BMC Proceedings* 5 (S7): O63. https://doi.org/10.1186/1753-6561-5-S7-O63.

Mansoor, Shahid, Imran Amin, Mazhar Hussain, Yusuf Zafar, and Rob W. Briddon. 2006. "Engineering Novel Traits in Plants through RNA Interference." *Trends in Plant Science* 11 (11): 559–65. https://doi.org/10.1016/j.tplants.2006.09.010.

Marcel Kuntz (2012) Destruction of public and governmental experiments of GMO in Europe, GM Crops & Food, 3:4, 258-264, DOI: 10.4161/gmcr.21231

Mariani, Celestina, Marc De Beuckeleer, Jessie Truettner, Jan Leemans, and Robert B. Goldberg. 1990. "Induction of Male Sterility in Plants by a Chimaeric Ribonuclease Gene." *Nature* 347 (6295): 737–41. https://doi.org/10.1038/347737a0.

Mátyás, Csaba, and International Consultation on Forest Genetics and Tree Breeding, eds. 2011. *Forest Genetics and Sustainability*. Dordrecht; Springer.

McEwan, Andrew, Enrico Marchi, Raffaele Spinelli, and Michal Brink. 2020. "Past, Present and Future of Industrial Plantation Forestry and Implication on Future Timber Harvesting Technology." *Journal of Forestry Research* 31 (2): 339–51. https://doi.org/10.1007/s11676-019-01019-3.

Miflin, Ben. 2000. "Crop Improvement in the 21st Century." *Journal of Experimental*

Botany 51 (342): 1–8. https://doi.org/10.1093/jexbot/51.342.1.

Montenegro, M. (2016). *CRISPR is Coming to Agriculture — with Big Implications for Food, Farmers, Consumers and Nature. Ensia*. Available online at: http://ensia.com/voices/crispr-is-coming-to-agriculture-with-big-implications-for-food-farmers-consumers-and-nature/

Moran, Emily, Jeffrey Lauder, Cameron Musser, Angela Stathos, and Mengjun Shu. 2017. "The Genetics of Drought Tolerance in Conifers." *The New Phytologist* 216 (4): 1034–48. https://doi.org/10.1111/nph.14774.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich. 1986. "Specific Enzymatic Amplification of DNA in Vitro: The Polymerase Chain Reaction." *Cold Spring Harbor Symposia on Quantitative Biology* 51 Pt 1: 263–73. https://doi.org/10.1101/sqb.1986.051.01.032.

Myburg, Alexander A., Dario Grattapaglia, Gerald A. Tuskan, Uffe Hellsten, Richard D. Hayes, Jane Grimwood, Jerry Jenkins, et al. 2014. "The Genome of *Eucalyptus* Grandis." *Nature* 510 (7505): 356–62. https://doi.org/10.1038/nature13308.

Naidoo, Sanushka, Bernard Slippers, Jonathan M. Plett, Donovin Coles, and Caryn N. Oates. 2019. "The Road to Resistance in Forest Trees." *Frontiers in Plant Science* 10. https://doi.org/10.3389/fpls.2019.00273.

Nakagawa, Shinichi, and Holger Schielzeth. 2013. "A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models." *Methods in Ecology and Evolution* 4 (2): 133–42. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Neale, David B., and Antoine Kremer. 2011. "Forest Tree Genomics: Growing Resources and Applications." *Nature Reviews Genetics* 12 (2): 111–22. https://doi.org/10.1038/nrg2931.

Needham, Mark D, Glenn T Howe, and Joshua Petit. 2015 "Forest Health Biotechnologies: What Are the Drivers of Public Acceptance?," 37.

Nelson, C.D.; Powell, W.A.; Merkle, S.A.; Carlson, J.E.; Hebard, F.V.; Islam-Faridi, N; Staton, M.E.; Georgi, L. 2014. Biotechnology of trees: Chestnut. Biology and Biotechnology 35 p.

Nester, Eugene W. 2015. "Agrobacterium: Nature's Genetic Engineer." *Frontiers in Plant Science* 5 (January). https://doi.org/10.3389/fpls.2014.00730.

Pagnussat, Gabriela C., Hee-Ju Yu, Quy A. Ngo, Sarojam Rajani, Sevugan

Mayalagu, Cameron S. Johnson, Arnaud Capron, Li-Fen Xie, De Ye, and Venkatesan Sundaresan. 2005. "Genetic and Molecular Identification of Genes Required for Female Gametophyte Development and Function in Arabidopsis." *Development* 132 (3): 603–14. https://doi.org/10.1242/dev.01595.

Pâques, Luc E., ed. 2013. *Forest Tree Breeding in Europe: Current State-of-the-Art and Perspectives*. Managing Forest Ecosystems, volume 25. Dordrecht: Springer.

Peña, L., and A. Séguin. 2001. "Recent Advances in the Genetic Transformation of Trees." *Trends in Biotechnology* 19 (12): 500–506. https://doi.org/10.1016/s0167-7799(01)01815-7.

Pérez-de-Castro, A. M., S. Vilanova, J. Cañizares, L. Pascual, J. M. Blanca, M. J. Díez, J. Prohens, and B. Picó. 2012. "Application of Genomic Tools in Plant Breeding." *Current Genomics* 13 (3): 179–95. https://doi.org/10.2174/138920212800543084.

Perlin, J., 2005. *A forest journey : the story of wood and civilization* 1st Countryman Press paperback., Woodstock, VT: The Countryman Press.

Persley, Gabrielle, 1999 "APPLICATIONS OF BIOTECHNOLOGY TO CROPS: BENEFITS AND RISKS," 8.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2020). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-148, https://CRAN.R-project.org/package=nlme.

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., et al. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 344:1246752. doi: 10.1126/science.1246752

Polle, Andrea, Shao Liang Chen, Christian Eckert, and Antoine Harfouche. 2019. "Engineering Drought Resistance in Forest Trees." *Frontiers in Plant Science* 9. https://doi.org/10.3389/fpls.2018.01875.

Pospóšilová, J., I. Tichá, P. Kadleček, D. Haisel, and Š. Plzáková. 1999. "Acclimatization of Micropropagated Plants to Ex Vitro Conditions." *Biologia Plantarum* 42 (4): 481–97. https://doi.org/10.1023/A:1002688208758.

Quétier, F. (2016). The CRISPR-Cas9 technology: CLOSER to the ultimate toolkit for targeted genome editing. *Plant Sci.* 242, 65–76. doi: 10.1016/j.plantsci.2015.09.003

Rastogi, Smita, and Upendra Nath Dwivedi. 2008. "Manipulation of Lignin in Plants with Special Reference to O-Methyltransferase." *Plant Science* 174 (3): 264–77.

https://doi.org/10.1016/j.plantsci.2007.11.014.

Regal, 1982 Eucalypt Ecology: Individuals to Ecosystems, p.36

Rottmann, William H., Richard Meilan, Lorraine A. Sheppard, Amy M. Brunner, Jeffrey S. Skinner, Caiping Ma, Shuping Cheng, Lise Jouanin, Gilles Pilate, and Steven H. Strauss. 2000. "Diverse Effects of Overexpression of LEAFY and PTLF, a Poplar (Populus) Homolog of LEAFY/FLORICAULA, in Transgenic Poplar and Arabidopsis." *The Plant Journal* 22 (3): 235–45. https://doi.org/10.1046/j.1365-313x.2000.00734.x.

Royle, J. Andrew, Alain F. Zuur, Elena N. Ieno, Neil J. Walker, and Anatoly A. Saveliev. 2009 *Mixed Effects Models and Extensions in Ecology with R1*.

Russell Lenth, Henrik Singmann, Jonathon Love, Paul Buerkner, Maxime Herve, 2020 emmeans: Estimated Marginal Means, aka Least-Squares Means https://cran.r-project.org/web/packages/emmeans/index.html

Rye, B., & James, S. (1992). The Relationship Between Dysploidy and Reproductive Capacity in Myrtaceae. Australian Journal of Botany, 40(6), 829–848. https://doi.org/10.1071/BT9920829

Samanta, M. K., Dey, A., and Gayen, S. (2016). CRISPR/Cas9: an advanced tool for editing plant genomes. *Transgenic Res.* 25, 561–573. doi: 10.1007/s11248-016-9953-5

Saxena, Deepak, and G. Stotzky. 2001. "Bt Corn Has a Higher Lignin Content than Non-Bt Corn." *American Journal of Botany* 88 (9): 1704–6. https://doi.org/10.2307/3558416.

Sedjo, Roger, 2001 "Biotechnology in Forestry: Considering the Costs and Benefits,"

Sellers, C. H. 1910. *Eucalyptus: Its History, Growth, and Utilization*. A.J. Johnston.

Skinner, J. S., R. Meilan, A. M. Brunner, and S. H. Strauss. 2000. "Options for Genetic Engineering of Floral Sterility in Forest Trees." In *Molecular Biology of Woody Plants*, edited by S. Mohan Jain and Subhash C. Minocha, 64:135–53. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-2311-4_5.

Strauss, Steven H., Huimin Tan, Wout Boerjan, and Roger Sedjo. **2009**. "Strangled at Birth? Forest Biotech and the Convention on Biological Diversity." *Nature Biotechnology* 27 (6): 519–27. https://doi.org/10.1038/nbt0609-519.

Strauss, Steven H., Kristin N. Jones, Haiwei Lu, Joshua D. Petit, Amy L. Klocko, Matthew G. Betts, Berry J. Brosi, Robert J. Fletcher, and Mark D. Needham.

2017. "Reproductive Modification in Forest Plantations: Impacts on Biodiversity and Society." *New Phytologist* 213 (3): 1000–1021. https://doi.org/10.1111/nph.14374.

Strauss, Steve H., Mikaela Schmitt, and Roger Sedjo. 2009. "Forest Scientist Views of Regulatory Obstacles to Research and Development of Transgenic Forest Biotechnology." *Journal of Forestry* 107 (7): 350–57. https://doi.org/10.1093/jof/107.7.350.

Strauss, Steven H, William H Rottmann, Amy M Brunner, and Lorraine A Sheppard. 1995 "Genetic Engineering of Reproductive Sterility in Forest Trees," 22.

Sylvie Bonny,2003. Why Are Most Europeans Opposed to GMOs? - Factors Explaining Rejection in France and Europe. https://scielo.conicyt.cl/fbpe/img/ejb/v6n1/a04/bip/.

Tamura, Koichiro, Glen Stecher, Daniel Peterson, Alan Filipski, and Sudhir Kumar. 2013. "MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0." *Molecular Biology and Evolution* 30 (12): 2725–29. https://doi.org/10.1093/molbev/mst197.

Thorsness, M. K., M. K. Kandasamy, M. E. Nasrallah, and J. B. Nasrallah. 1993. "Genetic Ablation of Floral Cells in Arabidopsis." *The Plant Cell* 5 (3): 253–61. https://doi.org/10.1105/tpc.5.3.253.

Tuskan, G. A., S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, et al. 2006. "The Genome of Black Cottonwood, Populus Trichocarpa (Torr. & Gray)." *Science* 313 (5793): 1596–1604. https://doi.org/10.1126/science.1128691.

Ugozzoli, Luis, and R. Bruce Wallace. 1991. "Allele-Specific Polymerase Chain Reaction." *Methods* 2 (1): 42–48. https://doi.org/10.1016/S1046-2023(05)80124-0.

United Nations, Department of Economic and Social Affairs, and Population Division. 2019. *World Population Prospects Highlights, 2019 Revision Highlights, 2019 Revision.*

Uraki, Yasumitsu, and Keiichi Koda. 2015. "Utilization of Wood Cell Wall Components." *Journal of Wood Science* 61 (5): 447–54. https://doi.org/10.1007/s10086-015-1492-9.

"USDA APHIS | About the SECURE Rule." 2020. https://www.aphis.usda.gov/aphis/ourfocus/biotechnology/biotech-rule-revision.

Vecchio, Maria Gabriella, Claudia Loganes, and Clara Minto. 2016. "Beneficial and Healthy Properties of *Eucalyptus* Plants: A Great Potential Use." *The Open Agriculture Journal* 10 (1). https://doi.org/10.2174/1874331501610010052.

Vidal, John, and environment editor. 2012. "The GM Tree Plantations Bred to Satisfy the World's Energy Needs." *The Guardian*, November 15, 2012, sec. Environment. https://www.theguardian.com/environment/2012/nov/15/gm-trees-bred-world-energy.

Vining, Kelly J., Elisson Romanel, Rebecca C. Jones, Amy Klocko, Marcio Alves-Ferreira, Charles A. Hefer, Vindhya Amarasinghe, et al. 2015. "The Floral Transcriptome of *Eucalyptus* Grandis." *New Phytologist* 206 (4): 1406–22. https://doi.org/10.1111/nph.13077.

Wang, Guiying, Yan Dong, Xiaojie Liu, Guosheng Yao, Xiaoyue Yu, and Minsheng Yang. 2018. "The Current Status and Development of Insect-Resistant Genetically Engineered Poplar in China." *Frontiers in Plant Science* 9 (September). https://doi.org/10.3389/fpls.2018.01408.

Wickham, Hadley. 2016. "Introduction." In *Ggplot2: Elegant Graphics for Data Analysis*, edited by Hadley Wickham, 3–10. Use R! Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4_1.

White, T.L., W.T. Adams, and D.B. Neale. 2007. *Forest Genetics*. Wallingford, UNITED KINGDOM: CABI. http://ebookcentral.proquest.com/lib/osu/detail.action?docID=298973.

Willats WGT, McCartney L, Mackie W, Knox JP (2001) Pectin: cell biology and prospects for functional analysis. Plant Mol Biol 47:9–27

Williams, Kristen, and Bradley Potts. 1996. "The Natural Distribution of *Eucalyptus* Species in Tasmania."

Wisniewski, Michael, Annette Nassuth, and Rajeev Arora. 2018. "Cold Hardiness in Trees: A Mini-Review." *Frontiers in Plant Science* 9. https://doi.org/10.3389/fpls.2018.01394.

Wright, Jonathan W. 1976. "1 - Introduction." In *Introduction to Forest Genetics*" https://doi.org/10.1016/B978-0-12-765250-4.50005-8.

Yu, Liangliang, Hongpeng Chen, Jiayan Sun, and Laigeng Li. 2014. "PtrKOR1 Is Required for Secondary Cell Wall Cellulose Biosynthesis in Populus." *Tree Physiology* 34 (11): 1289–1300. https://doi.org/10.1093/treephys/tpu020.

Yu, Xiang, Akira Kikuchi, Takayoshi Shimazaki, Akiyo Yamada, Yoshihiro Ozeki, Etsuko Matsunaga, Hiroyasu Ebinuma, and Kazuo N. Watanabe. 2013. "Assessment of the Salt Tolerance and Environmental Biosafety of *Eucalyptus* Camaldulensis Harboring a Mangrin Transgene." *Journal of Plant Research* 126 (1): 141–50. https://doi.org/10.1007/s10265-012-0503-9.

Zhang, Chunsheng, Kim H. Norris-Caneda, William H. Rottmann, Jon E. Gulledge, Shujun Chang, Brian Yow-Hui Kwan, Anita M. Thomas, et al. 2012. "Control of Pollen-Mediated Gene Flow in Transgenic Trees." *Plant Physiology* 159 (4): 1319–34. https://doi.org/10.1104/pp.112.197228.

Zhou, Mei-Liang, Jiang-Tao Ma, Yang-Min Zhao, Ya-Hui Wei, Yi-Xiong Tang, and Yan-Min Wu. 2012. "Improvement of Drought and Salt Tolerance in Arabidopsis and Lotus Corniculatus by Overexpression of a Novel DREB Transcription Factor from Populus Euphratica." *Gene* 506 (1): 10–17. https://doi.org/10.1016/j.gene.2012.06.089.

.

APPENDICES

Appendix A: Supplementary figures for chapter 2

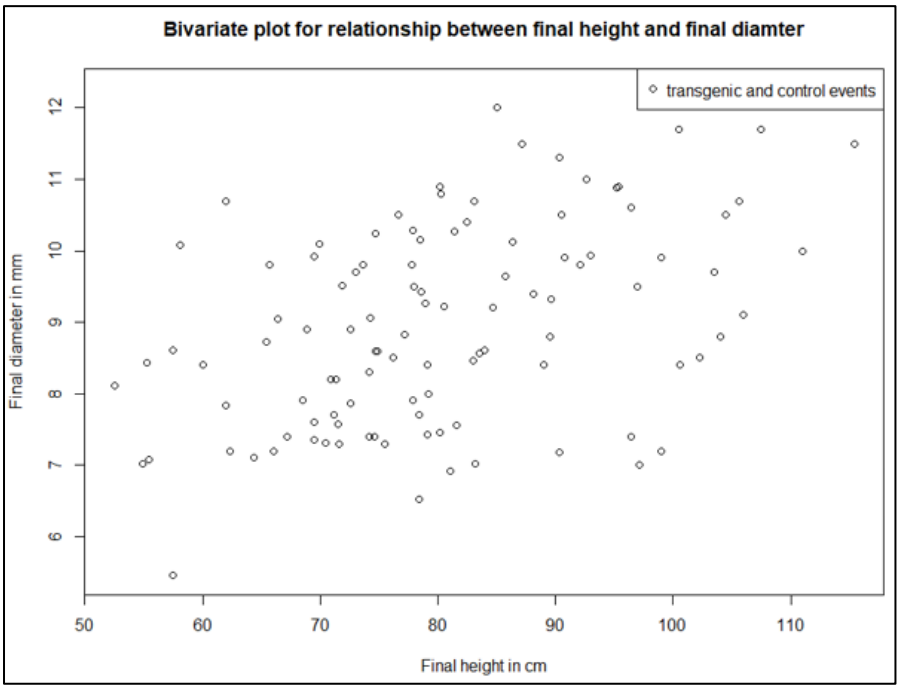


Figure A1: Bivariate plot showing the relationship between final height and final diameter of the trees.

Figure A2: Diagnostic plots from the greenhouse studies indicate variance is equal among the experimental groups, the assumptions of equal variance and normality are supported, and that there are no outliers in the data. Vegetative traits have a linear relationship with the experimental group type.

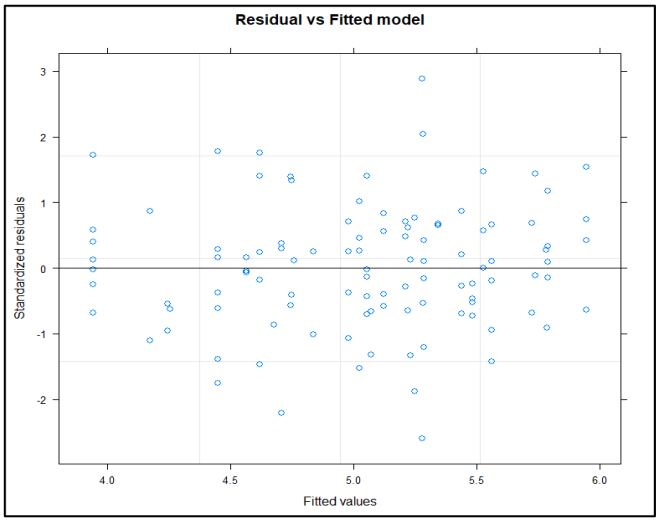
Diagnostic plots for linear mixed effect models

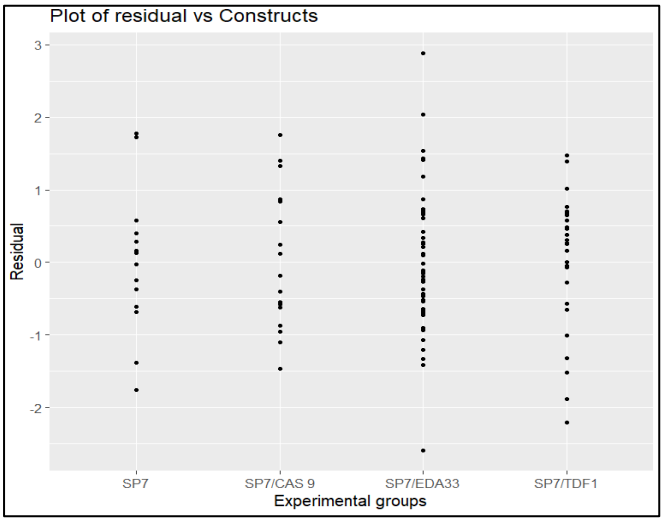
Residuals versus fitted plots show the spread of the standardized error terms. If the points are equally spread (homoscedastic) and bounce around the horizontal line with no clear pattern, we can tell that the assumption of equal variance is a good one. The residual is the vertical distance between a point and the line of fit, and the fitted value is the y-value corresponding to that point. If a pattern, for e.g. a cone shape, is observed in the spread of points, we should consider transforming the data and we can confirm that our data shows heteroscedasticity or unequal variance.

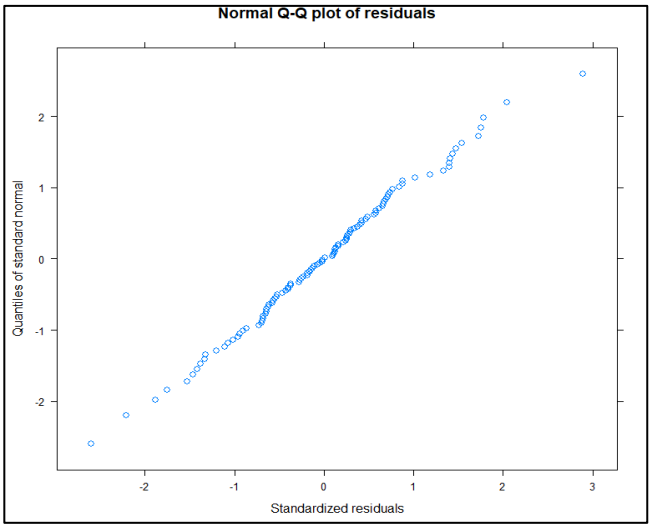
Residual versus explanatory variable graphs help us detect anything unusual occurring within the residual vs fitted plot due to the variance in explanatory variable categories.

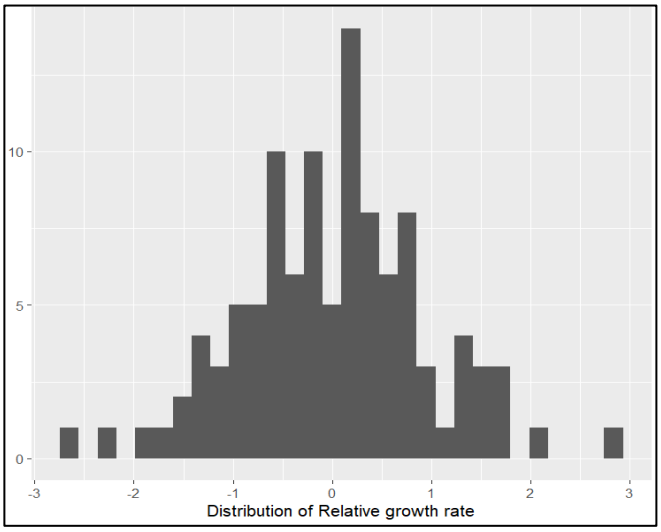
The normal quantile-quantile plot and histogram check for normality of the data. We look at the tail of the Q-Q plot to look for skewness in the data, although, it is normal to have some stray from the line along the ends. If points generally fall together, we can say that the assumption of normality of the data is a good one. Histogram has the same function as a normal Q-Q plot, but the skewness in the data is easier to detect.

1. **V**egetative traits – Relative Growth Rate

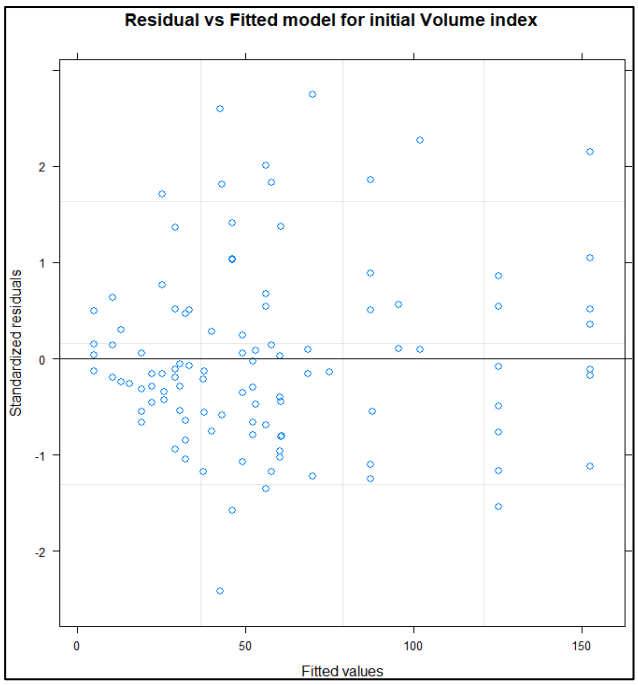


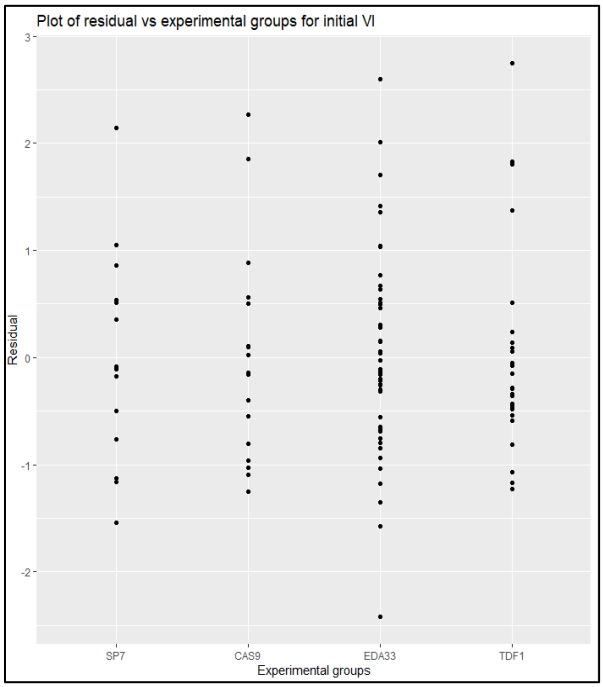


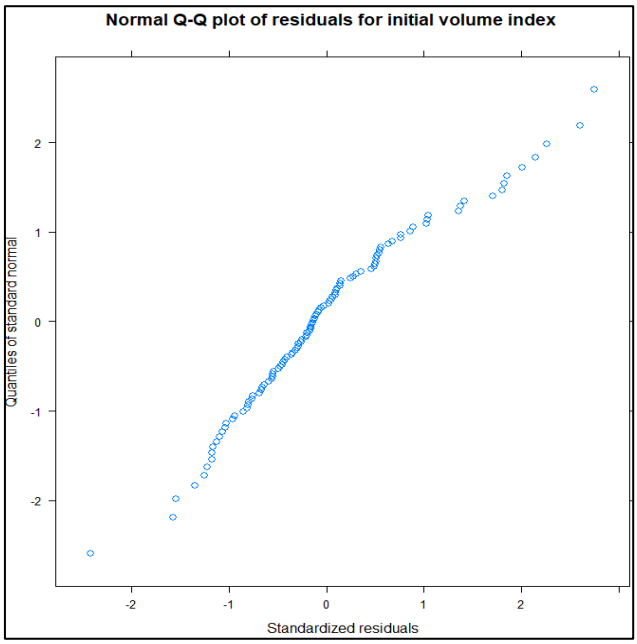


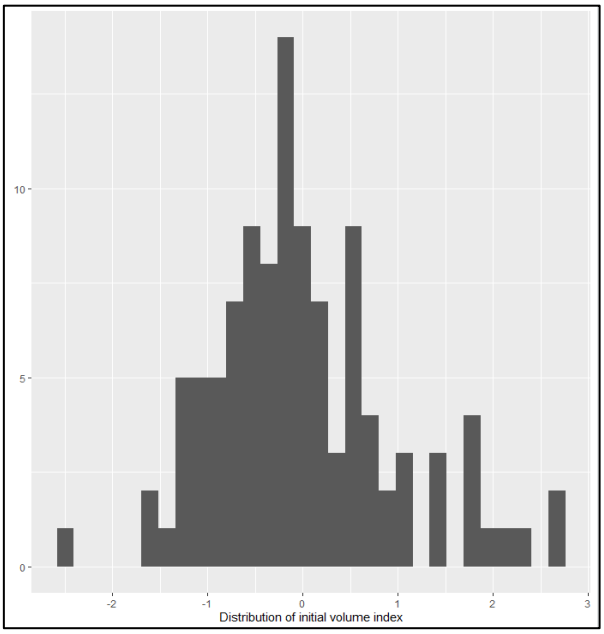


2. Vegetative trait – initial volume index

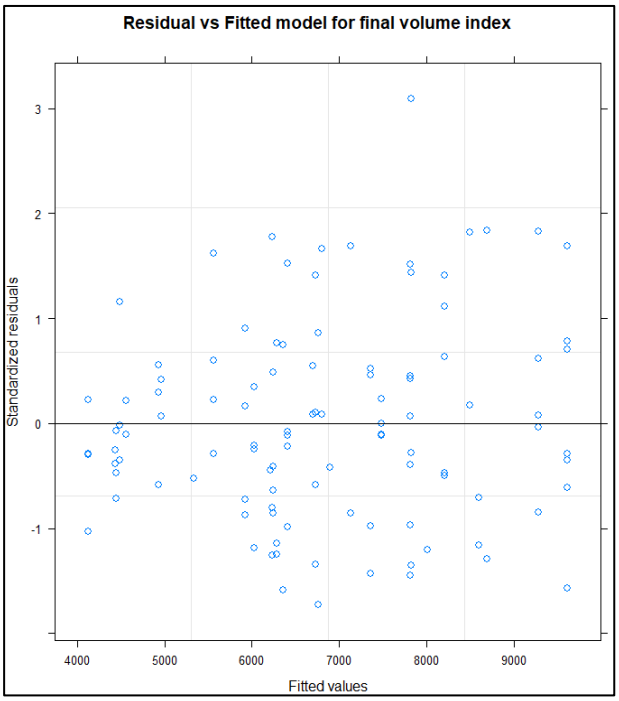


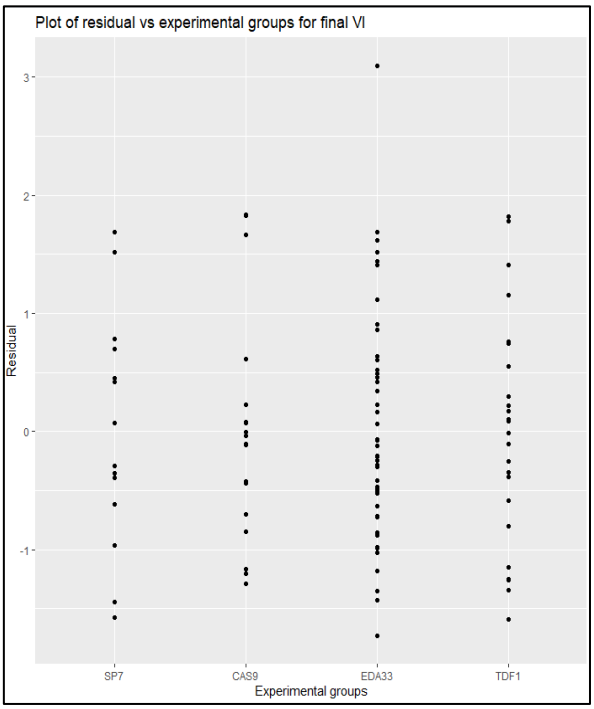


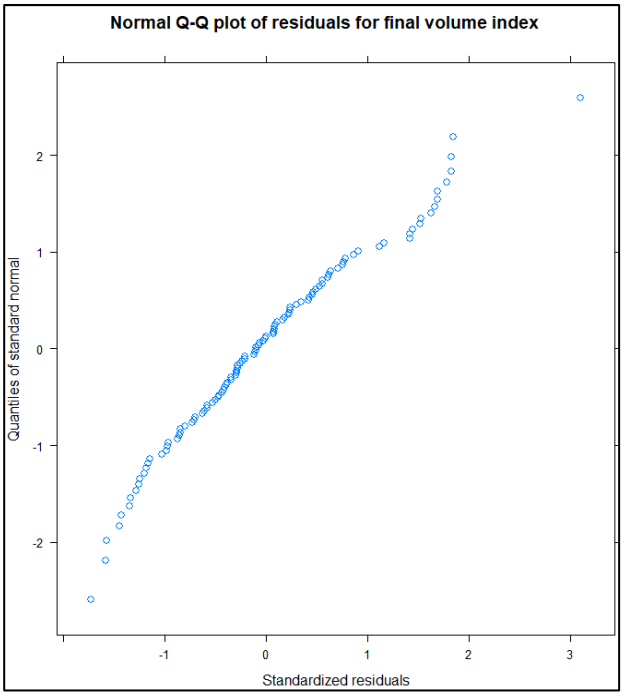


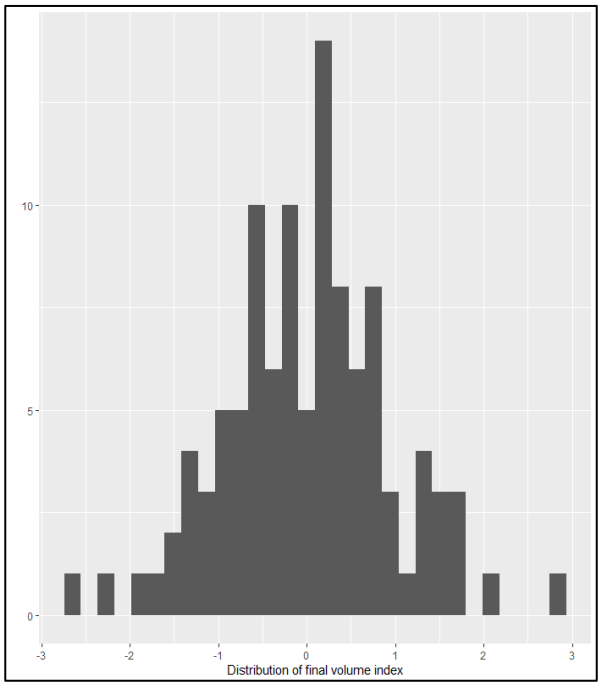


3. Vegetative trait – final volume index

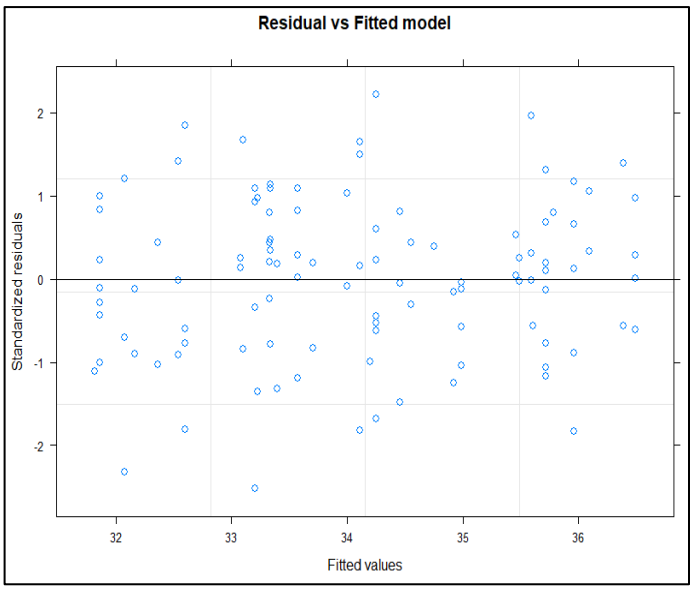


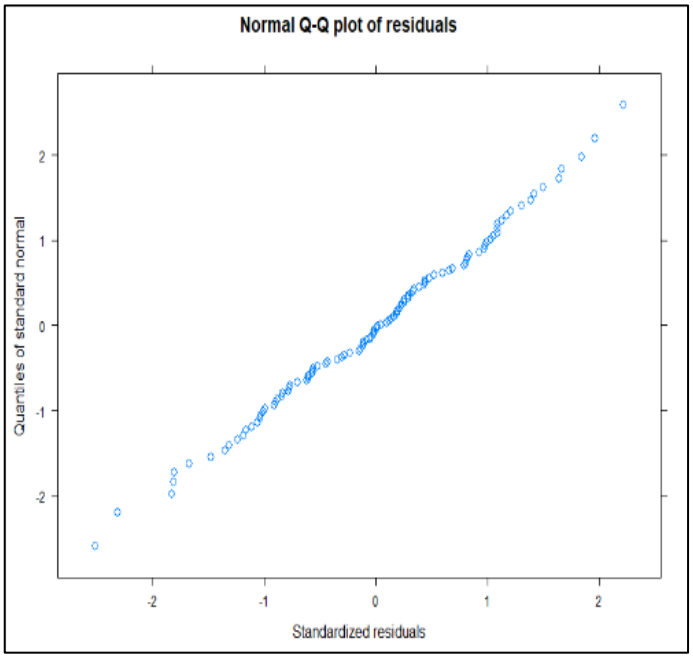


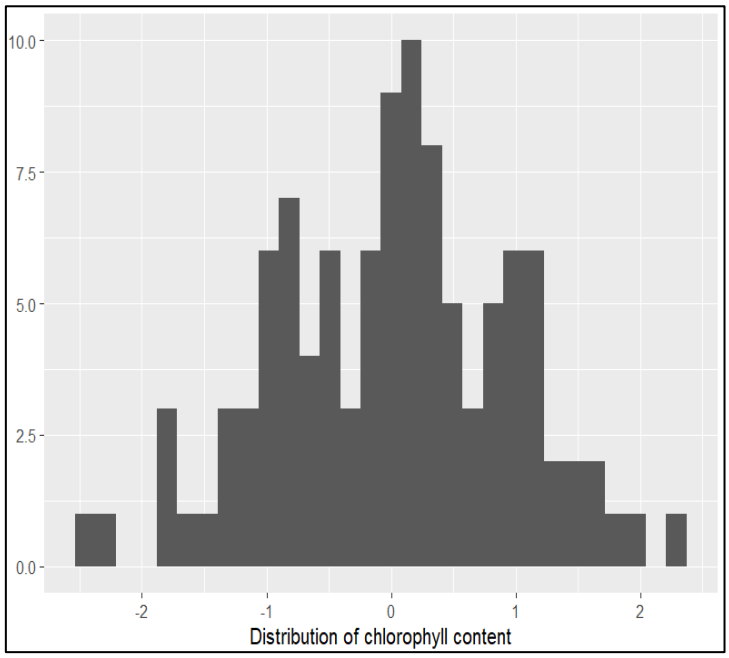




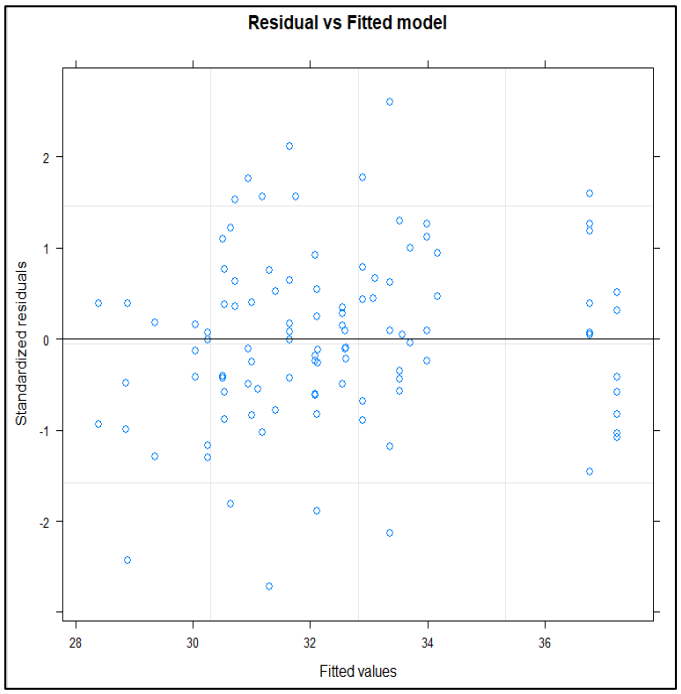
4. Vegetative trait – Chlorophyll content

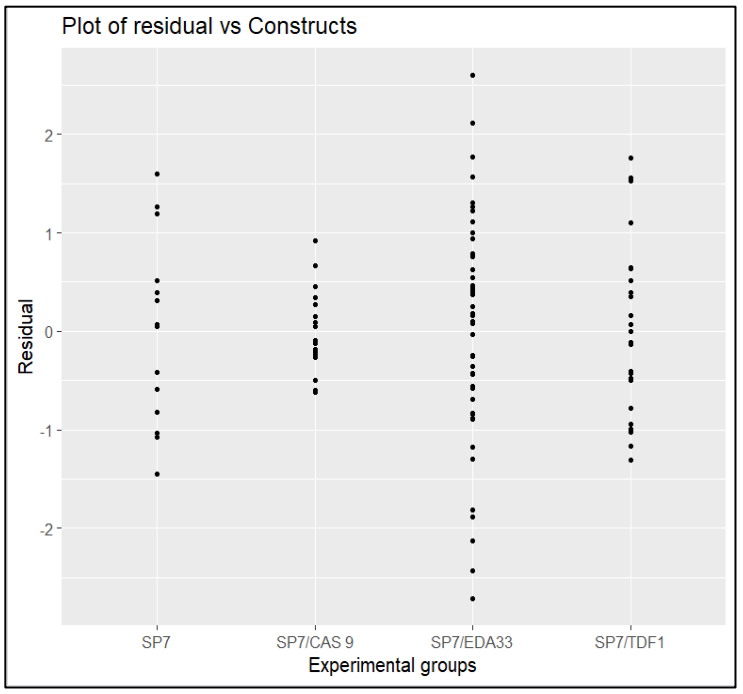


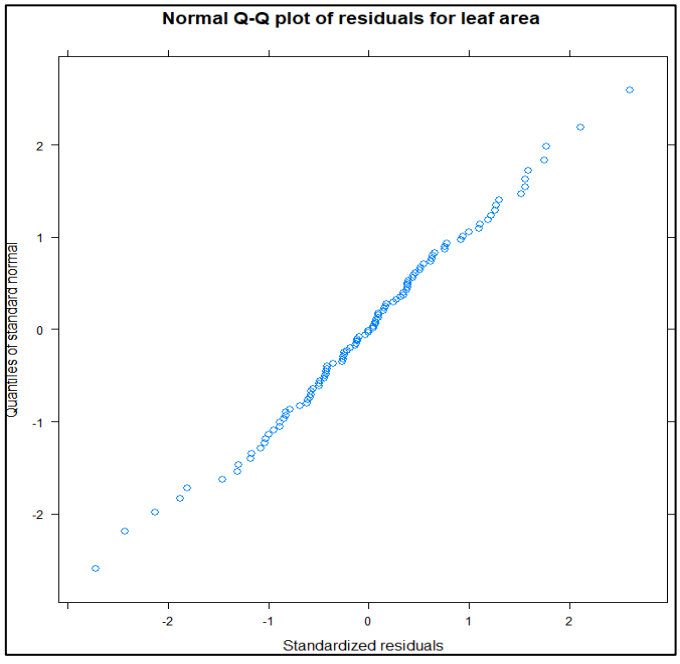


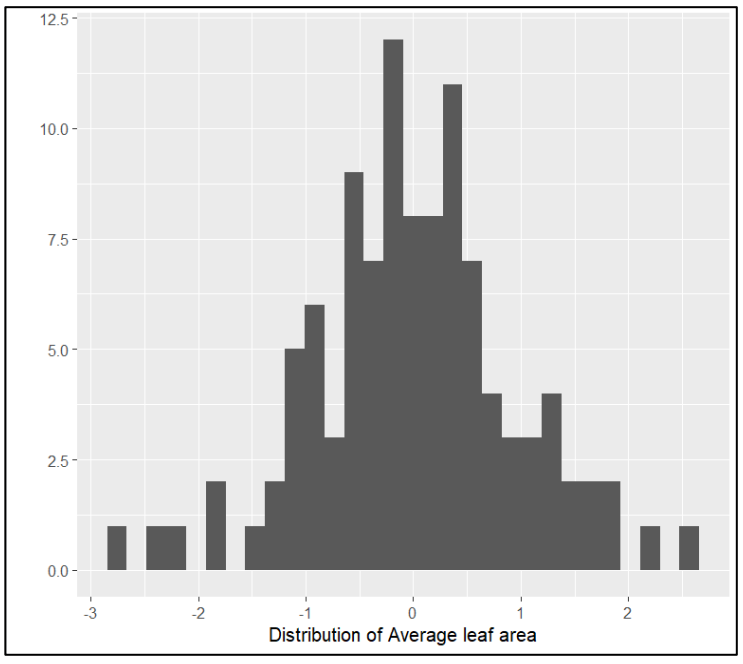


5. Vegetative trait – Average leaf area

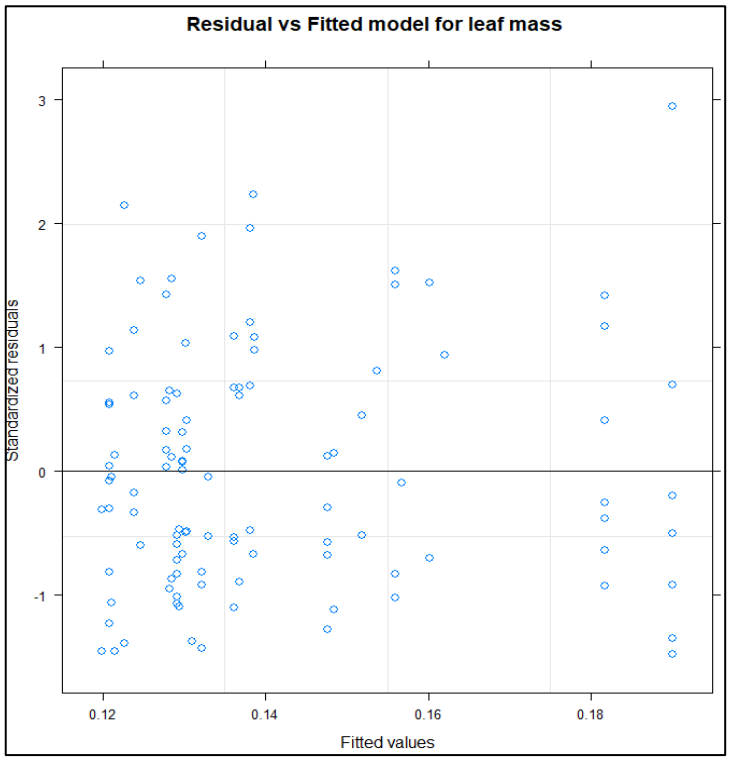


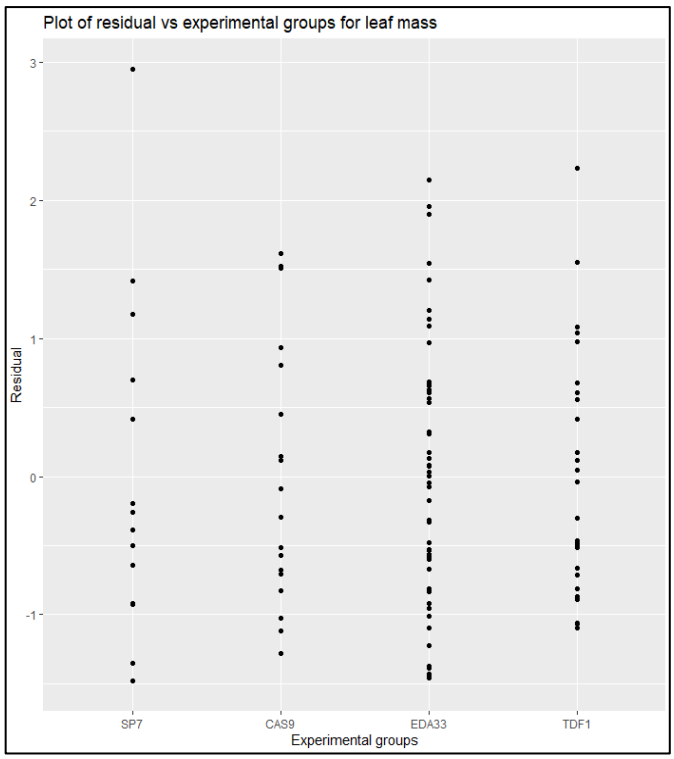


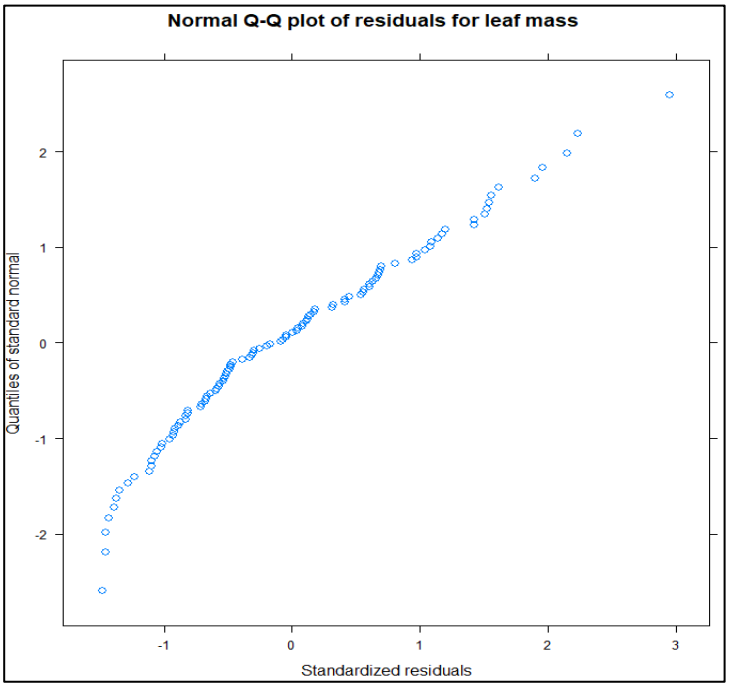


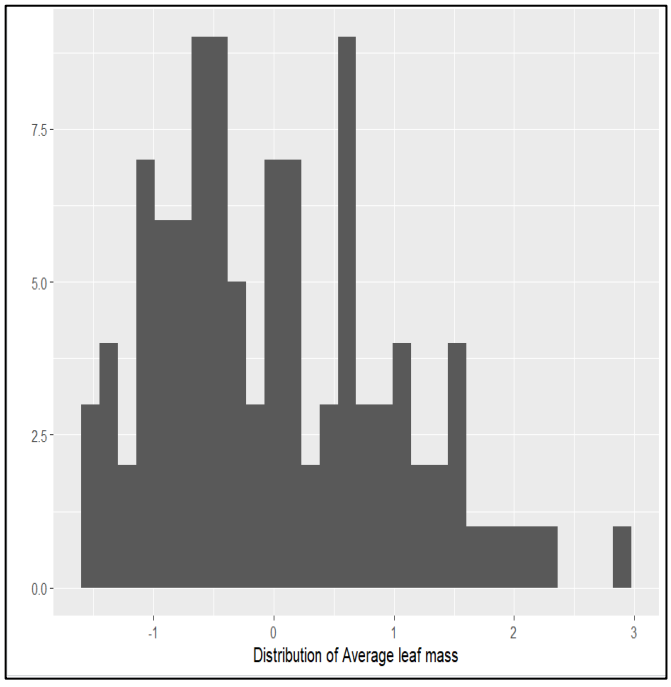


6. Trait of interest – Average leaf mass

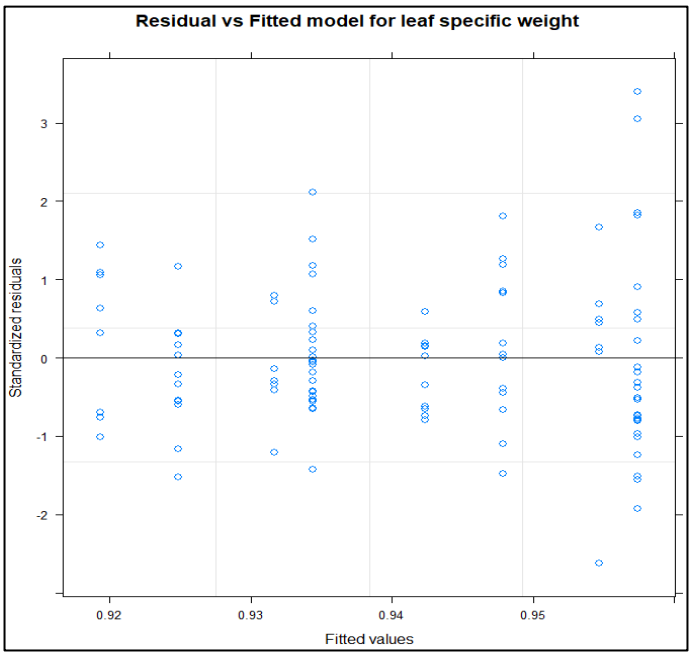


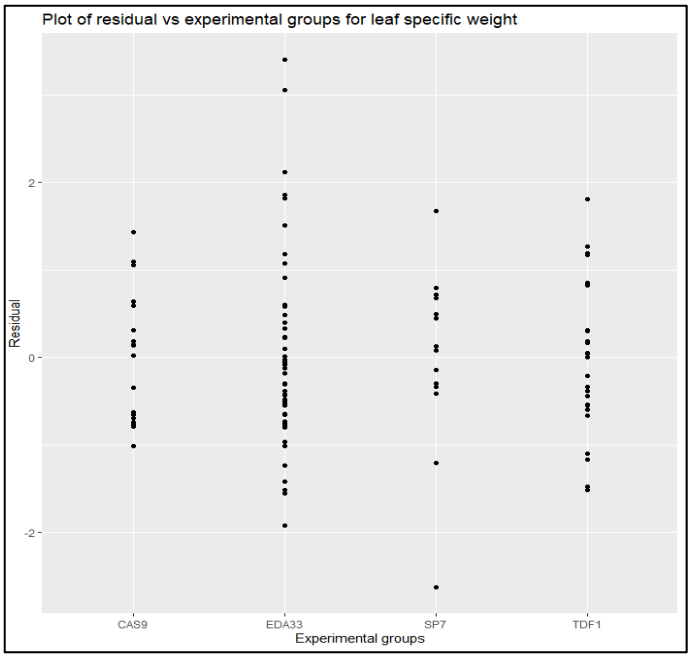


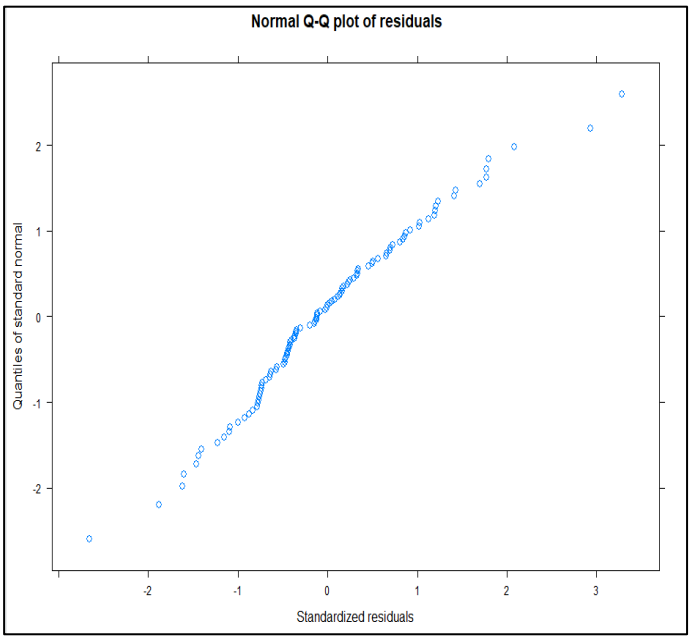


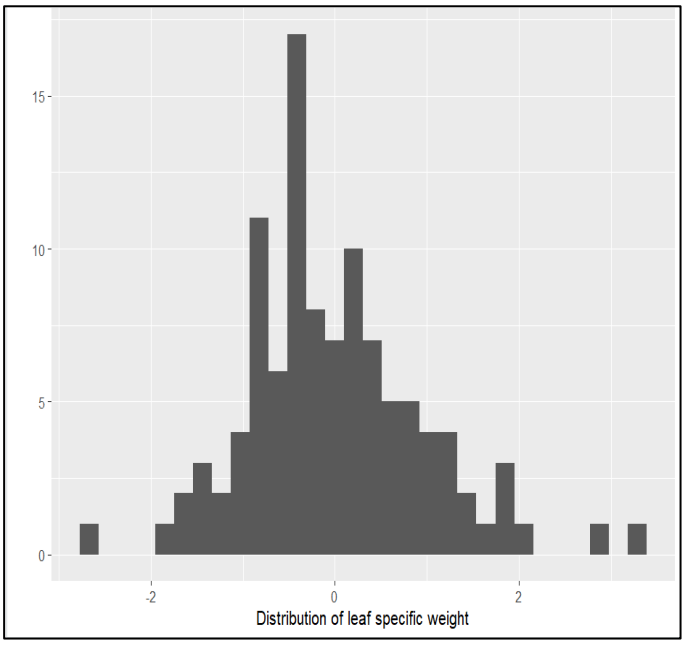


7. Trait of interest – Leaf specific weight

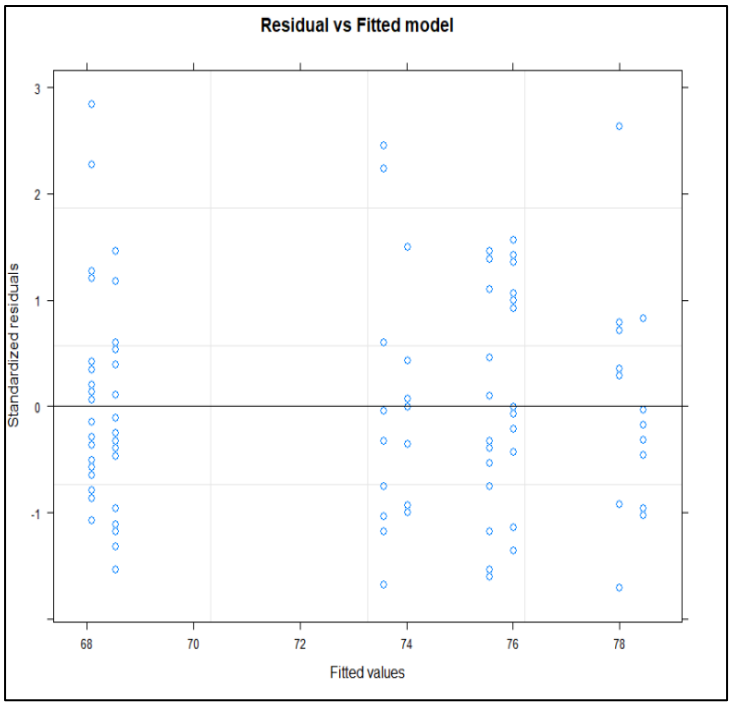


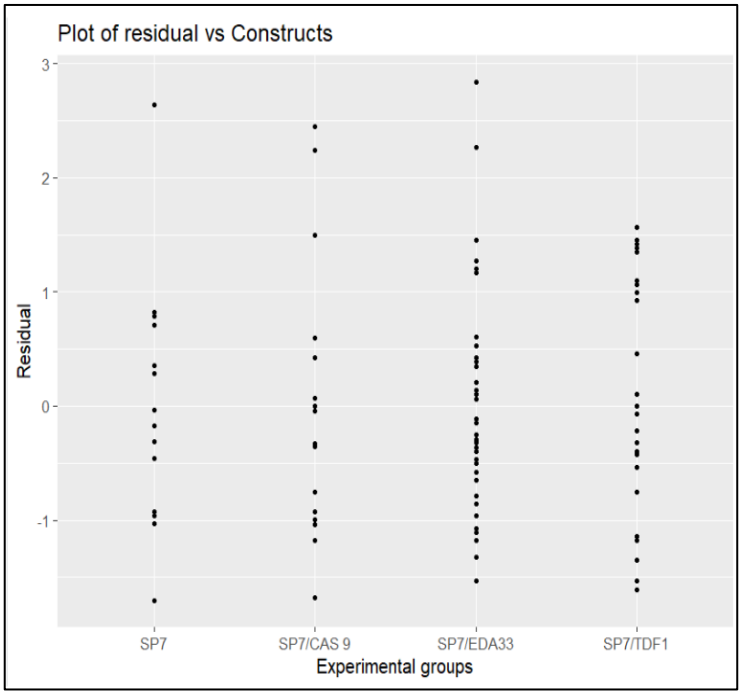


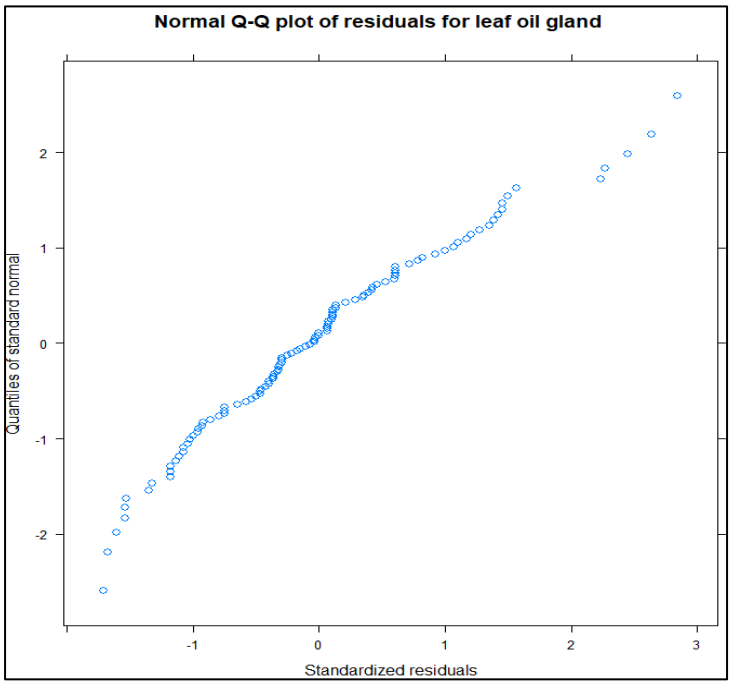


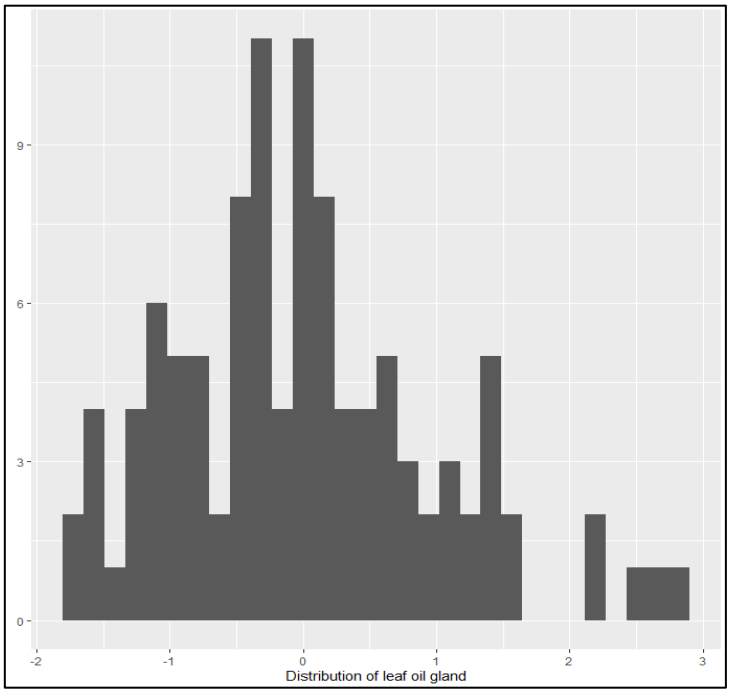


8. Trait of interest – oil gland density









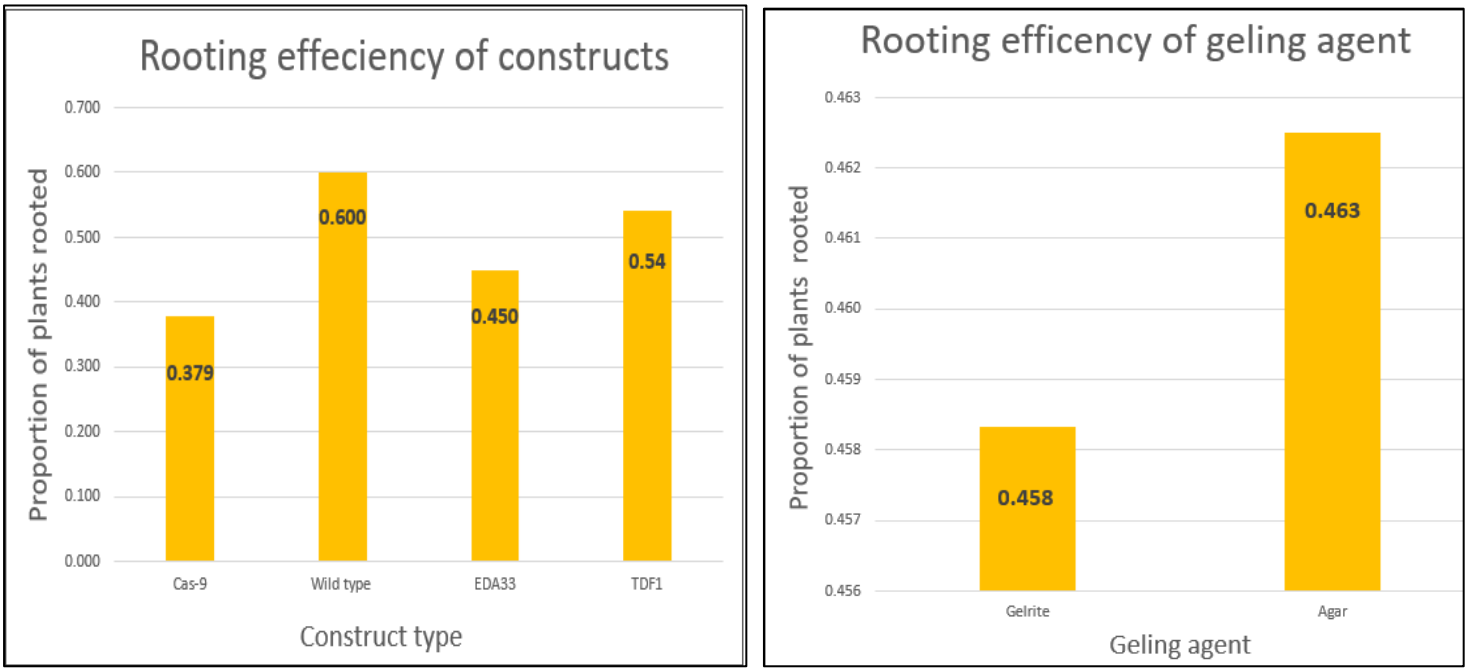


Figure A.3: Rooting efficiencies of plants *in vitro* based on the experimental group and kind of gelling agent (Agar vs Gelrite) before transferring them into the greenhouse.

>Eucgr.H04946 | Chr08:68856013..68856717 – *EDA33* Allele 1

5’ATGGACTTCAACCAAAGCAAGTTCTCGAACAACTTTTGGGATCTTGGCTTGGGCAT
GGAAGATCAAACCCTCCATTCCAATGACCAACATCATCATCAGCCGCCTTTCACTTCC
CTTTGGCCTAGCATCCATCTCCCATTAATGCACCAAACAACTCCAACTACTTCCCAGA
TTCCATCTTCCCACTTTGTGAATGATAGTAGCATTGGGGTTGTGGCAAATCAAATCGA
AGACAAGGATGAAGA**A**CCTGAAGAGGAGCTTGGAGCCATGAAGGAGATGATGTACA
AGATTGCGG**C**GATGCAACC**G**GT**T**GACGT**T**GACCCGACGACGATTCGGAAGCC**A**AAG
AGGCGGAATGTCCGGATCAGTGATGATCCTCAAAGTGTAGCGGC**A**CGCCT**T**CGGCGG
GAAAGGATAAGCGAGAAGATCAGGATCCTCCAGAGGCTCGTCCCCGGGGGGGAC**G**AA
GATGGACACGGCTTCGATGTTGGACGAGGCTATTCGCTATGTCAAGTTCTTGAAGCG
ACAAATCCGTTTGCTTCAACAGCCAAATAACCAAAACCCTACACCAGATCCGGGCGC
AGTCACGGGGAGCTTGGTGGGAGGGGATTGGCAAGGCGTGATTGCACCGAGCAAAC
CACCCTCCACTACCTCGTCATCGTCCTTCGAAGCACAGGTCGGATCAAGGTTTGGGTT
CTGGTCCAATGGTGGTGGTGGT**3’**

>Eucgr.H04946 | Chr08:68856013..68856717 – *EDA33* Allele 2

5’ATGGACTTCAACCAAAGCAAGTTCTCGAACAACTTTTGGGATCTTGGCTTGGGCAT
GGAAGATCAAACCCTCCATTCCAATGACCAACATCATCATCAGCCGCCTTTCACTTCC
CTTTGGCCTAGCATCCATCTCCCATTAATGCACCAAACAACTCCAACTACTTCCCAGA
TTCCATCTTCCCACTTTGTGAATGATAGTAGCATTGGGGTTGTGGCAAATCAAATCGA
AGACAAGGATGAAGA**G**CCTGAAGAGGAGCTTGGAGCCATGAAGGAGATGATGTACA
AGATTGCGG**T**GATGCAACC**A**T**C**GACGT**C**GACCCGACGACGATTCGGAAGCC**G**AAGA
GGCGGAATGTCCGGATCAGTGATGATCCTCAAAGTGTAGCGGC**G**CGCCT**C**CGGCGG
GAAAGGATAAGCGAGAAGATCAGGATCCTCCAGAGGCTCGTCCCCGGGGGGGAC**A**AA
GATGGACACGGCTTCGATGTTGGACGAGGCTATTCGCTATGTCAAGTTCTTGAAGCG
ACAAATCCGTTTGCTTCAACAGCCAAATAACCAAAACCCTACACCAGATCCGGGCGC
AGTCACGGGGAGCTTGGTGGGAGGGGATTGGCAAGGCGTGATTGCACCGAGCAAAC
CACCCTCCACTACCTCGTCATCGTCCTTCGAAGCACAGGTCGGATCAAGGTTTGGGTT
CTGGTCCAATGGTGGTGGTGGT**3’**

Figure A.4: *EDA33* gene sequence – Bold letter represent SNPs and blue highlighted letters are the guide RNAs on the gene. The strand shown is a template strand in the direction 5’-3’. The heading represents *EDA33’s* unique code in Phytozome.

>Eucgr.I02017 | Chr09:29814592..29816209 – *TDF1* Allele 1

CTTCCTCCTCCTCTTGTCTCAACGCATTCCTCATATTCATCTTCTCTGATCTCTCTCTTT
CTCTCGCTCTCTCTGTCAGCCGTCATGAAGAGACCGCCATGCTGCGACAAGTCGAAC
GTGAAGAGAGG**CCT**TTGGACGCCCGAAGAAGATGCCAAGATACTCGCATACGTATC
GACTTACGGGACTGGCAACTGGACTTTGGTCCCTGAGAAAGCTGGTCTTCTTTTATGT
TTCCAATCCTTTCATGAACAAGATCATTGTCCCTTCTCAGCTCGTTATTTCCATGTTTT
GATAGAGCTCAGTCTTACTTTAGGTGTGCCCGCGCGCGCACATGCAGGATTAAACAG
ATGTGGAAAGAGCTG**C**AGGCTTCGGTGGACCAATTACCTGAGGCCTGACCTTAAGCA
TGATGGCTTTTCCCCTGAAGAAGAAGATCTCATCAT**C**AACCTCCACAAAATCTTTGG
CAGCAGGTCTCACTAGTTTGCTCATCATTTCTCTGTCGAATGGCGTTATTCTTTTGAAT
GTATAAAGCCGATAGAATTTACAAGGAACGGAGTTTATCATGCATTTTCATTAGATT
GAATGAATGATTCAAATCGCTGCTCATTTTCCTGTCGAGTTCCCCTCCGAATTTCACG
GAGAAGACTCTTCCGCGGATGAGTGGATTAATGCAGACAGAGGGTTGTAGTTTTAAC
TTCGTAGATC**A**TGAGTTTGGAAGGAAACGTTAATGAACTCAACAGTTGGTGGGGTAA
AGTTCATTATCAACCCATCTTTCTGTCATGAAAGGTTTCGATGTGTGTAACTTAATTC
CCATGCATGCAGCATTTACGTATTTTCTAAGTACTCC**A**CTCACGCTTTTATTGGTTA**T**
ATGATATGGTCTTGTTGCAGATGGTCTCTAATTGCAAAACATCTGCCCGGAAGAACA
GACAACGATGTCAAGAACTACTGGAACACCAAGCTCAAGAAGAAGCTCCAAAAGAT
GGGAATCGATCCTTTAA**CCC**ACAAGCCTTTCTCTCAGATCTTCTCAGACTTTGAGAAC
ATGAGCGGCTGCCCAAACGCCAGACATCGCCAAATCTTGCCCGCGCCATCGTGTTTG
ACTCAGGTCCCTGCAGGCTCCAATTCTCACTTGGATGCGATCATGAAGCCCGTGATG
GAGCAAGTTCATGAGAATTTCACTGCCGAAAATCACCTCTCTTGGTCTCAGTACCAG
GTGGCGAACCAAGATGTTATGCAGCTGCAACCATATCAATGTGTATTGAGCGAGGTC
ACGTCTTCTTGTTCCTCGTCATCCTCTCCTACTCTGACACGATTTAACACGCAACAAT
CGGATGGTCCACCGCTTCCTCCTCCCTTTGTTGCCTGGACTGGCGATTCGATTTCGCA
TCAGCCATTTCCCACTGGCAATGTATTGCCGAAGCGCGAAGGAGACTTGCAGGATAT
AGTGTCTTCATCGTCCATCAACTGTGCGAGCGATATGGCAAAACAAGCGCTCCCCAA
CACCCCTGTCGGCACAATGGCTTGTAAAAACGAAGCCAAATGGGCCGAACCCGGAG
ATTCGTGCGAGGGAGCGCTAGAACTGGATTGCCATTTCGGAGACGGTTCGTACTCTT
GTTTCGGCTCTTTCACGGACGCCATCTTGGACAAGGACAGCAAGATGAGGTCGGAGT
TATTTCCAGAATTTATAGATGGACTTCTAGATTACTGA

Figure A.5: *TDF1* gene sequence – Bold letter represent SNP and blue highlighted letters are the guide RNAs on the gene. The strand shown is a template strand in the direction 5'-3'. Heading represent *TDF1* unique code in phytozome.

>Eucgr.I02017 | Chr09:29814592..29816209 – *TDF1* Allele 2

CTTCCTCCTCCTCTTGTCTCAACGCATTCCTCATATTCATCTTCTCTGATCTCTCTCTTT
CTCTCGCTCTCTCTGTCAGCCGTCATGAAGAGACCGCCATGCTGCGACAAGTCGAAC
GTGAAGAGAGG**CCT**TTGGACGCCCGAAGAAGATGCCAAGATACTCGCATACGTATC
GACTTACGGGACTGGCAACTGGACTTTGGTCCCTGAGAAAGCTGGTCTTCTTTTATGT
TTCCAATCCTTTCATGAACAAGATCATTGTCCCTTCTCAGCTCGTTATTTCCATGTTTT
GATAGAGCTCAGTCTTACTTTAGGTGTGCCCGCGCGCGCACATGCAGGATTAAACAG
ATGTGGAAAGAGCTG**T**AGGCTTCGGTGGACCAATTACCTGAGGCCTGACCTTAAGCA
TGATGGCTTTTCCCCTGAAGAAGAAGATCTCATCAT**A**AACCTCCACAAAATCTTTGG
CAGCAGGTCTCACTAGTTTGCTCATCATTTCTCTGTCGAATGGCGTTATTCTTTTGAAT
GTATAAAGCCGATAGAATTTACAAGGAACGGAGTTTATCATGCATTTTCATTAGATT
GAATGAATGATTCAAATCGCTGCTCATTTTCCTGTCGAGTTCCCCTCCGAATTTCACG
GAGAAGACTCTTCCGCGGATGAGTGGATTAATGCAGACAGAGGGTTGTAGTTTTAAC
TTCGTAGATC**T**TGAGTTTGGAAGGAAACGTTAATGAACTCAACAGTTGGTGGGGTAA
AGTTCATTATCAACCCATCTTTCTGTCATGAAAGGTTTCGATGTGTGTAACTTAATTC
CCATGCATGCAGCATTTACGTATTTTCTAAGTACTCC**T**CTCACGCTTTTATTGGTTA**C**
ATGATATGGTCTTGTTGCAGATGGTCTCTAATTGCAAAACATCTGCCCGGAAGAACA
GACAACGATGTCAAGAACTACTGGAACACCAAGCTCAAGAAGAAGCTCCAAAAGAT
GGGAATCGATCCTTTAA**CCC**ACAAGCCTTTCTCTCAGATCTTCTCAGACTTTGAGAAC
ATGAGCGGCTGCCCAAACGCCAGACATCGCCAAATCTTGCCCGCGCCATCGTGTTTG
ACTCAGGTCCCTGCAGGCTCCAATTCTCACTTGGATGCGATCATGAAGCCCGTGATG
GAGCAAGTTCATGAGAATTTCACTGCCGAAAATCACCTCTCTTGGTCTCAGTACCAG
GTGGCGAACCAAGATGTTATGCAGCTGCAACCATATCAATGTGTATTGAGCGAGGTC
ACGTCTTCTTGTTCCTCGTCATCCTCTCCTACTCTGACACGATTTAACACGCAACAAT
CGGATGGTCCACCGCTTCCTCCTCCCTTTGTTGCCTGGACTGGCGATTCGATTTCGCA
TCAGCCATTTCCCACTGGCAATGTATTGCCGAAGCGCGAAGGAGACTTGCAGGATAT
AGTGTCTTCATCGTCCATCAACTGTGCGAGCGATATGGCAAAACAAGCGCTCCCCAA
CACCCCTGTCGGCACAATGGCTTGTAAAAACGAAGCCAAATGGGCCGAACCCGGAG
ATTCGTGCGAGGGAGCGCTAGAACTGGATTGCCATTTCGGAGACGGTTCGTACTCTT
GTTTCGGCTCTTTCACGGACGCCATCTTGGACAAGGACAGCAAGATGAGGTCGGAGT
TATTTCCAGAATTTATAGATGGACTTCTAGATTACTGA

Figure A.5 (continued): *TDF1* gene sequence – Bold letter represent SNP and blue highlighted letters are the guide RNAs on the gene. The strand shown is a template strand in the direction 5'-3'. Heading represent *TDF1* unique code in phytozome.

Appendix B: Supplementary tables for chapter 2

Table A.1 Allele specific primers for mutation detection

Gene	Forward primer	Reverse primer
EDA33 Allele 1 EDA33_A1_F1/ EDA33_A1R1	5’ TCGAAGACAAGGATGAAGAG	5’ GAAGCCGTGTCCATCTTC
EDA33 Allele 2 EDA33_A2_F1/ EDA33_A2R1	5’ TCGAAGACAAGGATGAAGAA	5’ GAAGCCGTGTCCATCTTT
TDF1 sgRNA1 Allele 1 TDF1_5’F/TDF1_sg1A1R1	5’ TCAGCCGTCATGAAGAGACC	5’ CCATCTGCAACAAGACCATATCATG
TDF1sgRNA1 Allele 2 TDF1_5’F/TDF1_sg1A2R1	5’ TCAGCCGTCATGAAGAGACC	5’ CCATCTGCAACAAGACCATATCATA
TDF1sgRNA2 Allele 1 TDF1_A1F1/TDF1sg2R1	5’ AAACAGATGTGGAAAGAGCTGC	5’ TGCAGGGACCTGAGTCAAAC
TDF1sgRNA2 Allele 2 TDF1_A2F1/TDF1sg2R1	5’ AAACAGATGTGGAAAGAGCTGT	5’ TGCAGGGACCTGAGTCAAAC
EDA33 Universal primers	5’ GCCTAGCATCCATCTCCCAT	5’ CGAAGGACGATGACGAGGTAG
TDF1 Universal primers	5’ TCAGCCGTCATGAAGAGACC	5’ TGCAGGGACCTGAGTCAAAC

Table A.2 Mutation type in *EDA33*

	Allele 1		Allele 2	
	Site guide RNA 1	**Site guide RNA 2**	**Site guide RNA 1**	**Site guide RNA 2**
Event 1 -2	5 bp deletion	39 bp deletion	no mutation	1 bp insertion
Event 5 -1	152 bp deletion		152 bp deletion	
Event 6-1	4 bp deletion	1 bp insertion	152 bp deletion	
Event 7-2	152 bp deletion		152 bp deletion	
Event 8-3	2 bp deletion	compound mutation with large deletion of 62 bp	152 bp deletion	
Event 9-1	152 bp deletion		4 bp deletion	1 bp deletion
Event 10-1	1 bp deletion	wild type	wild type	
Event 11-2	wild type	498 bp insertion	1 bp deletion	4 bp deletion
Event 13-1	wild type	530 bp insertion	1 bp deletion	4 bp deletion
Event 14-2	152 bp inversion		152 bp deletion	
Event 15-1	152 bp inversion		152 bp deletion	
Event 16-1	45 bp deletion	51 bp deletion	wild type	

Table A.3 Mutation type in *TDF1*

	Allele 1		**Allele 2**	
	Site guide RNA 1	**Site guide RNA 2**	**Site guide RNA 1**	**Site guide RNA 2**
Event 4 -3	888 bp deletion		Wild type	Wild type
Event 5 -2	Wild type	Wild type	Wild type	1 bp deletion
Event 6-2	Wild type	Wild type	Wild type	1 bp deletion
Event 7-2	1 bp deletion	5 bp deletion	888 bp deletion	
Event 9-1	Wild type	1 bp insertion	Wild type	Wild type
Event 10-1	1 bp insertion	Wild type	1 bp deletion	2 bp deletion
Event 11-1	7 bp deletion	1 bp insertion	888 bp deletion	
Event 12-3	1 bp deletion	1 bp insertion	888 bp deletion	
Event 13-1	Wild type	1 bp insertion	Wild type	2 bp deletion
Event 14-3	Wild type	Wild type	Wild type	1 bp deletion
Event 15-1	Wild type	Wild type	Wild type	1 bp deletion
Event 16-2	1 bp insertion	4 bp deletion	1 bp deletion	1 bp insertion
Event 17-1	Wild type	Wild type	Wild type	1 bp deletion

Table A.4: Total number of plants survived after transferring them to greenhouse

TDF1 events	Total transplanted	Plants survived
7-2	8	7
10-1	8	7
11-1	10	10
12-3	7	7
16-2	9	8

EDA33 events	Total transplanted	Plants survived
1-2	9	8
6-1	5	4
7-2	10	10
8-3	4	2
11-2	10	9
12-1	12	10
13-1	5	5
14-2	4	7
15-1	5	3

Cas 9 only events	Total transplanted	Plants survived
1-1	4	1
3-1	4	3
4-1	10	9
5-1	3	2
6-3	6	6
Sp7 wild type	14	14

Table A.5: PCR cocktail mixture concentration

REAGENT	ul / Rx	Final Conc	4					
ddH20	16.12		64.5					
10 x Econotaq buffer	2.5	1 X	10.0					
10mg/ml BSA	0.25	0.1mg/ml	1.0					94C 1'
dNTP (2.5 mM ea)	1	0.2 mM	4.0					55C 1'
DNA	colony dab							72C 1'
Primer 1-F (10 uM)	0.75	0.6 uM	3.0					
Primer 1-R (10 uM)	0.75	0.6 uM	3.0					30cycles
ddH20	0.75	0.6 uM	3.0					
ddH20	0.75	0.6 uM	3.0					
Taq (5 U/ul)	0.125	1 U	0.5					
TOTAL	22.995		92.0					70.5
MINERAL OIL			0.0					
	22.995		23.0					

Appendix C: Rstudio code for statistical analysis and the associated Excel sheet are available upon request.