8-8-2023

Gene expression effects on productivity and stress tolerance in polyclonal plantings of Populus deltoides

Macy Gosselaar
Mississippi State University, OSHScheer@gmail.com

Gene expression effects on productivity and stress tolerance in polyclonal plantings of *Populus deltoides*

By

Macy Gosselaar

Approved by:

Austin Himes (Major Professor)
Courtney Siegert
Daniel Peterson (Minor Professor)
Heidi J. Renninger (Committee Member/Graduate Coordinator)
L. Wes Burger (Dean, College of Forest Resources)

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Forestry
in the Department of Forestry

Mississippi State, Mississippi

August 2023

Name: Macy Gosselaar

Date of Degree: August 8, 2023

Institution: Mississippi State University

Major Field: Forestry

Major Professor: Austin Himes

Title of Study: Gene expression effects on productivity and stress tolerance in polyclonal plantings of *Populus deltoides*

Pages in Study 138

Candidate for Degree of Master of Science

Polyclonal plantings of *Populus deltoides* are expected to display increased site resource use, productivity, and tolerance to stress through plasticity changes leading to niche differentiation (i.e changes to crown/canopy structures). In the present study, *P. deltoides* Clones S7C8, 110412, and polyclonal plots were tested for differentially expressed genes and enriched biological pathways between planting schemes. Transcriptomic analysis of leaves revealed upregulation of an active growth gene and gene family members that play important roles in plant stress and stress tolerance in polyclonal plantings. A gene associated with oxidative stress was upregulated in polyclonal plantings across all treatments. Secondary metabolic pathways including arginine and proline metabolism were upregulated in monoclonal plantings and downregulated in polyclonal plantings. Phenotypic results displayed greater aboveground biomass in polyclonal plantings. Results suggested a potential increased tolerance in polyclonal plantings to water and heat stress, including increased productivity and resource usage.

DEDICATION

Dedicated to my parents, Erma and Robert Gosselaar, and my brother Enrico Gosselaar. Thank you for your guidance and love. To my fiancé, Aidan Lenhart-Baker, thank you for your unconditional love and never-ending moral support. To God, thank you for the numerous blessings, opportunities and knowledge allowing me to complete this thesis.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the people that supported and guided me throughout this project. First, my major advisor, Dr. Austin Himes, for his creative, enthusiastic, and knowledgeable advice. Second, the graduate committee: Dr. Heidi Renninger, and Dr. Courtney Siegert, for providing their expertise and continuous encouragement and support. To my minor committee member, Dr. Daniel Peterson who highlighted my passion for genetics through teaching and for his insightful feedback and encouragement. Special thanks are also due to Dr. Chuan-Yu Hsu for assistance with biological samples and sequencing analysis, and Mark A. Arick for assistance with statistical analysis and questions. I would like to thank my fellow forestry graduate students for inspiring and creating life-long memories. In addition, I would like to thank the Department of Forestry, the Forestry and Wildlife Research Center of Mississippi State, and the Mississippi Agricultural and Forestry Experimentation Station for letting me utilize their resources and research stations. I would also like to thank the US Department of Energy (DOE) DE-EE0009280, US Department of Agriculture National Institute of Food and Agriculture, McIntire Stennis project 1023932 and Mississippi State's Institute for Genomics, Biocomputing and Biotechnology for funding this project. Thank you all for making this project possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Background: *Populus spp.* Genetic Response to Heat and Water Stress

The usage of well adapted and productive genotypes of *Populus spp.* is essential for the sustainability of forests as climate change is increasing heat and water stress globally (Niemczyk et al., 2019). As a model genomic organism, *Populus* is a favorable system for understanding diverse physiological and morphological processes, including environmental responses to abiotic and biotic stresses and serves as a basis for developing strategies for improving tolerance to stress in trees (Popko et al., 2010). Plants, such as *Populus spp.*, respond to water and heat stress through transcriptional changes, altering gene expression levels to maintain homeostasis within biological systems and improve stress tolerance (Ren et al., 2019; Yang et al., 2015). Previous transcriptomic studies have focused on water and heat stress in *Populus spp.,* with studies showing increased production of secondary metabolites, including the accumulation of proline to alleviate oxidative stress (Bita & Gerats, 2013). Similar studies have shown the regulation of secondary metabolism from plant defense transcriptional factors (TFs), such as "*WRKY"* TFs in response to plant stress signals (Meraj et al., 2020).

Water and heat stress have been previously shown to reduce vegetative growth and yield in *Populus* (Chen et al., 2014; Jia et al., 2017). Although, there is limited information available on how the combination of water and heat stress influences gene expression effects for *Populus* (Jia et al., 2017). High temperatures can increase evapotranspiration rates, making plant growth

directly limited by heat stress or indirectly from water shortage (Jia et al., 2017; Verlinden et al., 2013). Understanding *Populus'* genetic response to these stresses is essential for understanding tree response to future drought conditions and extreme high temperatures predicted by climate change (Ren et al., 2019).

1.2 Background: *Populus* as the Model Tree Species

Populus is an ideal model woody species for genomic studies of trees (Bradshaw et al., 2000; Jansson & Douglas, 2007; Taylor, 2002; Tuskan et al., 2006). *P. trichocarpa* was selected for genome sequencing because of its relatively small genome size (450-550 Megabasepair (Mbp)), but also because of its rapid growth and ease of experimental manipulation through the production of large amounts of clonal material (Taylor, 2002; Tuskan et al., 2006). The ability for *Populus* to adapt to diverse conditions as well as prominent genetic polymorphisms has provided researchers with a rich source of variation in *Populus* morphology and physiology (Bradshaw et al., 2000). Because of its fast juvenile growth, short-term physiological responses to environmental variables are rapid and pronounced, producing distinctive tree phenotypes in 1-3 years in field environments (Bradshaw et al., 2000; Brunner et al., 2004). Another attribute of the *Populus* genomic system is its ecological and intraspecific diversity. Levels of genetic diversity are high for molecular markers, such as simple sequence repeats (SSRs) for marker assisted selection, and for adaptive traits including vegetative phenology (Brunner et al., 2004). For example, assessing tissues, like leaves, at the genomic level may provide information on vegetative productivity and differentiation of photosynthetic activity between clonal varieties. In addition to fast growth and diversity, a wide range of *Populus* genomic resources are available publicly, such as The International *Populus* Genome Consortium based at Oak Ridge National Laboratory, USA (Tuskan et al., 2004). Likewise, the *Populus* nuclear and chloroplast genome

sequence are available for downloading and searching at the Joint Genome Institute website through the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute (Goodstein et al., 2012).

1.3 Background: Ecological benefits from *Populus spp.*

Populus spp. help solve a number of large-scale environmental problems, through the mitigation of pollution runoff, providing a source of carbon neutral energy, and supporting the green economy (Forrester et al., 2005; Richards et al., 2010). Phytoremediation, or directly using *Populus* to clean up contaminated groundwater is a common phytotechnology (Farraji et al., 2016), and includes gaining hydraulic control of sites by controlling the migration of pollutants from an area including riparian areas (Zalesny et al., 2012). The removal of excessive nitrogen (N) from groundwater and surface runoff is important for the southeastern United States as N fertilizers from agriculture in the region end up in the Gulf of Mexico causing excessive growth of phytoplankton and contributing to annual hypoxic zones with an estimated cost to the economy of $9.27-$31.97 per kg N (Compton et al., 2011). Compared to herbaceous plants, *Populus spp.* are excellent candidates for remediation as they quickly produce high biomass in stems and leaves, allowing them to store large amounts of pollutants, including agricultural N runoff (Shim et al., 2013). Similarly, planting *Populus spp.* in riparian zones, which are known to function as buffers, could reduce non-point source pollution from agricultural lands to streams, improving groundwater quality (Hefting et al., 2005, 2006).

Previous research has also demonstrated the production potential and sustainability of hybrid poplar biofuels (Stanton & Gustafson, 2019). *Populus spp.* are the only cellulosic feedstock candidates with a history of commercial use and a proven supply chain, and past paper, pulp and veneer industries of *Populus spp.* could accelerate the design of similarly efficient

biofuels production (Stanton & Gustafson, 2019). This makes *Populus spp.* a great candidate for genetic improvement as researchers can enhance its accumulation capacity of pollution runoff and growth rate (Shim et al., 2013). Through molecular genetic analysis, the identification of expressed genes involved in improved *Populus* characteristics could be used to pinpoint target genes for genetic engineering or future tree breeding programs (Bunn, 2004). Since the release of the black cottonwood (*P. trichocarpa)* genome in 2006, molecular approaches were quickly adopted to investigate underlying mechanisms of basic morphological and physiological processes (Wullschleger et al., 2013). The availability of high-throughput sequencing technology and a fully sequenced *P. trichocarpa* genome has accelerated comparative genomic and transcriptomic analysis, enabling new approaches for understanding mechanisms of growth, yield, and mitigation traits (Han et al., 2020; Luo et al., 2015). Genomic analysis has allowed researchers to compare changes in tree morphologies and phenolic chemistry at the molecular level, unveiling new information on interspecific and intraspecific variation of *Populus spp.* (Keim et al., 1989).

Although *Populus spp.* perform well in northern climates, it is important to capture the productivity ability of *Populus* adapted to a subtropical climate and the extended growing season of the southeastern United States (Langholtz et al., 2016). There is limited information on gene expression profiling for *P. deltoides* productivity, stress tolerance and N mitigation grown in different planting schemes in the Southeast United States (Han et al., 2020; Kuchma et al., 2022, 2022; Luo et al., 2015; Richards et al., 2010) The research presented here contributes to the genomic knowledge of *P. deltoides* plantations in the southeastern United States by analyzing the gene expression of two *P. deltoides* clonal varieties grown in monocultures and grown in intimate polycultures.

1.4 Transcriptome Analysis for Differential Gene Expression

Transcriptome analysis plays a necessary role in deciphering *Populus* genetic networks, and establishing molecular biomarkers that respond to environmental challenges, including pollution and abiotic stress (Jiang et al., 2015). The transcriptome serves as a starting point for analyzing underlying changes in the physiological and morphological systems of *Populus spp.* across a wide range of study designs (Luo et al., 2015; Luo & Zhou, 2019). The transcriptome consists of all RNA transcripts produced in a particular tissue type. This includes coding transcripts, like messenger RNA (mRNA), and non-coding RNA (ncRNA). Although ncRNAs are transcribed RNAs involved in transcript regulation, ncRNAs are not transcripts (de Klerk & ‘t Hoen, 2015). Alternatively, mRNAs are transcripts containing information for protein production. While genes are made up of DNA, proteins are made of amino acids coded by mRNAs. Genes carry genotype information, while proteins express the phenotypes through the regulation of gene expression (de Klerk & ‘t Hoen, 2015) (Fig. 1.1). Phenotypic plasticity of organisms like *Populus spp.*, including physiological and developmental plasticity may arise from alternative transcription initiation, including co-transcriptional regulatory mechanisms, which modify protein function and gene expression levels, resulting in the expression of different transcripts and proteins from the same gene (de Klerk & ‘t Hoen, 2015).

A multitude of technologies, including hybridization or sequence-based approaches, have been developed over the years to quantify the transcriptome (Wang et al., 2009). One hybridization approach, microarray analysis, converts mRNA samples into complementary DNA (cDNA), and each sample is labeled with a fluorescent probe (Lowe et al., 2017; Stark et al., 2019; Wang et al., 2009). Hybridization occurs when the cDNA molecules bind to the target DNA molecules on a microscope slide, and the microscope slide is scanned to determine the

color probe output (green, red or yellow), measuring the expression of each gene printed on the slide (Lowe et al., 2017). Although hybridization-based approaches are high throughput and relatively inexpensive (Wang et al., 2009), limitations include reliance upon existing genomic knowledge and complicated normalization methods when comparing expression levels across different experiments (Lowe et al., 2017; Wang et al., 2009). Sequence-based approaches directly determine the cDNA sequences but technologies such as the first-generation Sanger sequencing are relatively low throughput and expensive (Wang et al., 2009). Next generation sequencing, such as RNA-Seq, is less expensive than Sanger sequencing and is a deep sequencing technology that is more sensitive in detecting genes with very low expression levels, and more accurate in detecting the expression levels of abundant genes compared to other hybridization and sequence-based approaches (Fu et al., 2009; *Sequencing Platforms | Illumina NGS Platforms,* n.d.; Zhao et al., 2014). The sensitivity of RNA-Seq can be further enhanced by enriching classes of RNA of interest, such as mRNAs (Lowe et al., 2017). mRNAs can be separated by using immobilized oligonucleotides probes, specifically targeting poly-A tails of mRNA and removing non-coding RNAs, such as uninformative ribosomal RNAs (rRNAs). Similarly, RNA-Seq provides unprecedented detail about transcriptional features, such as novel transcribed regions, allele-specific variation, and alternative splicing (Lowe et al., 2017; Stark et al., 2019; Wang et al., 2009), and is not limited by prior genomic knowledge unlike most hybridization approaches (Wang et al., 2009).

RNA-Seq is carried out on numerous sequencing platforms and instruments. One of these instruments includes Illumina's NovaSeq 6000 next generation sequencing system (*Sequencing Platforms | Illumina NGS Platforms*, n.d.). The *Illumina* sequencing workflow consists of four steps: sample preparation, cluster generation, sequencing, and data analysis. Sample preparation

includes library preparation, allowing compatibility of RNA samples with the sequencing instrument (*NGS Workflow Steps | Illumina Sequencing Workflow*, n.d.). Sequencing libraries are created through fragmented DNA, attaching special adapters containing complementary sequences allowing the fragments to bind to the flow cell (*NGS Workflow Steps | Illumina Sequencing Workflow*, n.d.). Libraries are loaded onto the flow cell and placed on the NovaSeq 6000 next generation sequencing system. Clusters of DNA fragments are amplified through repeated denaturing and extension cycle from a single template model, resulting in amplification of millions of unique clonal clusters across *Illumina's* flow cell. This process is known as cluster generation or bridge amplification (*NGS Workflow Steps | Illumina Sequencing Workflow*, n.d.). The clusters are then used as templates for sequencing. Chemically modified nucleotides bind to the DNA template strand through natural complementarity, with each nucleotide containing a fluorescent tag and a reversible terminator that blocks incorporation of the following base (*NGS Workflow Steps | Illumina Sequencing Workflow*, n.d.). Fluorescent tags indicate which nucleotide has been added to the template and the terminator is cleaved so the next base can bind. This results in true base-by-base calling, eliminating nucleotide error and enabling accurate data (*NGS Workflow Steps | Illumina Sequencing Workflow*, n.d.). After the forward DNA strand is read, the reads are washed away, and the process repeats for the reverse strands. This method is known as paired-end sequencing. Following sequencing, data analysis includes identifying nucleotides ("sequences"), and the predicted accuracy of the nucleotides being called (base-calling) (*NGS Workflow Steps | Illumina Sequencing Workflow*, n.d.).

1.5 RNA-Seq for Differential Gene Expression Analysis

1.5.1 Eukaryotic Directional Messenger RNA-Seq

RNA-Seq is a high-throughput sequencing tool for measurement of RNA expression levels from the transcriptome. (Mortazavi et al., 2008). Reduced costs and enabled performance detail for transcriptional features, such as allelic specific variation and for profiling gene expression levels has made it a ubiquitous tool for gene expression analysis (Mortazavi et al., 2008; Stark et al., 2019). Up to 90% of the eukaryotic genome is transcribed into RNA, however only 2% give rise to protein products (Pauli et al., 2011; Yu et al., 2019). The main goal for this project was to sequence gene-coded transcripts or mRNAs, which was done using the *Illumina* RNA-Seq approach. Rather than directly sequencing mRNA products, the Illumina RNA-Seq approach sequences complementary DNA (cDNA) products reverse transcribed from RNAs (‘MRNA Sequencing (MRNA-Seq)’, 2023). cDNA is the result of reverse transcribing RNA into DNA. cDNA reads in the same direction as the original mRNA is synthesized (in the 5’ → 3’ direction) and contains a visible poly(A) tail. A single conventional reading direction simplifies data handling (‘MRNA Sequencing (MRNA-Seq)’, 2023). Eukaryotic mRNA sequencing through Novogene selectively enriches and captures single-stranded mRNAs via poly (A) capture and converts it to cDNA for RNA library preparation, revealing gene expression profiles or continuous variation within the transcriptome (‘MRNA Sequencing (MRNA-Seq)’, 2023) (Fig. 1.2).

1.5.2 Overcoming Bias in Analysis Methods for RNA-Seq Data

Transcript length bias or gene length bias for detecting differential gene expression is a feature of current protocols for RNA-Seq technology (Chen et al., 2023). One effect of this bias reduces the ability to unveil differential gene expression for shorter genes (Aban et al., 2008;

Gail, 1974; Young et al., 2010). Shorter genes do not contain the same amount of coverage or detectability due to a decrease in count data, or number of sequences reads, which decreases the power of statistical tests (Aban et al., 2008; Gail, 1974). Likewise, with RNA-Seq, the expression level of transcript is limited to sequencing depth, or the number of unique reads of each region of a sequence. This can be overcome by increasing the number of replicate samples, significantly increasing the ability for statistical tests to detect differences in gene expression (Rapaport et al., 2013). The number of determined biological replicates utilized for this study will be further discussed throughout this chapter.

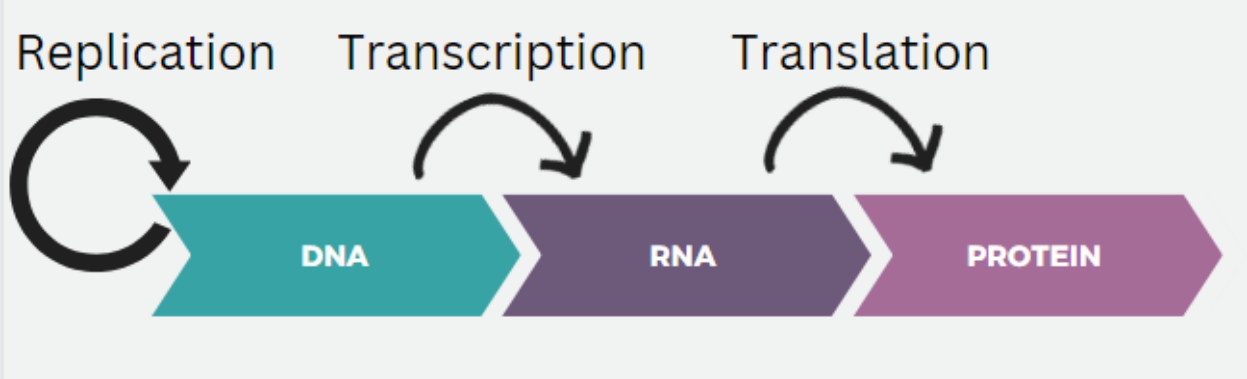


Figure 1.1 Central Dogma of Molecular Biology

Central Dogma displays the process of genetic information flowing from DNA to RNA to protein. Following DNA replication, DNA sequences provide instructions for functional products in the form of functional units called genes. Each gene provides instructions for a functional product, also known as a protein. Genes specify the functions of proteins through transcription and translation. This study is interested in the process of transcription. The DNA sequence of a gene is copied to make an RNA molecule but has not yet been fully translated into proteins, allowing us to analyze the first step in gene expression.

1.5.3 Transcriptome Data Analysis

Typically, millions of sequences or reads are produced from experimental analysis across several samples. The primary goal of transcriptome data analysis for this study is to identify genes that change in abundance between conditions, or between monoclonal and polyclonal plantings for this study (Anders et al., 2013). *Salmon* is a fast program, producing highly accurate transcript-level quantification estimates from RNA-Seq data (Patro et al., 2017). Indexing in *Salmon,* or aligning a reference transcriptome to sequencing data, allows one to quantify RNA-Seq data quickly, potentially identifying the genomic locus of transcripts (Patro et al., 2017). The *edgeR* software package (empirical analysis of differential gene expression (DGE) in R), from the Bioconductor project, was designed for the analysis of replicated count-based expression data, or the number of sequence reads that originate from a particular gene (Chen et al., 2016; McCarthy et al., 2012a; Robinson et al., 2010; Robinson & Smyth, 2007, 2008). The higher the number of counts for a particular gene, the more reads associated with that gene and the assumption that there is a higher-level expression of that gene (Chen et al., 2023). The *edgeR* software package requires an input of the table of counts, the total number of reads and a factor specifying the experimental group for each sample (Chen et al., 2023; Robinson et al., 2010). Two levels of variation can be distinguished from RNA-Seq data (Chen et al., 2022). First, between RNA samples, the relative abundance will vary for each gene due to biological causes and second, the measurement error for the abundance of each gene estimated by the sequencing technology (Chen et al., 2022). The Biological Coefficient of Variation (BCV) in *edgeR* allows one to visualize and estimate gene abundance and dispersion between RNA samples. The BCV represents the coefficient of variation that would remain between biological replicates if sequencing depth could be increased indefinitely (Chen et al., 2022). The software

models count data using an over dispersed Poisson model, and the Empirical Bayes procedure to moderate the degree of overdispersion across genes (Robinson et al., 2010). Through the sharing of information between genes, the Empirical Bayes procedure reduces dispersions towards one value (Robinson & Smyth, 2007). Once dispersion accounts are estimated, DGE data can be visualized through log ratio versus abundance plots, with significant genes determined through thresholds set by log fold changes.

Once significant differentially expressed genes are identified between polyclonal and monoclonal plantings, the function of those genes can be explored by conducting a Gene Ontology (GO) analysis. The ontology is grouped into sets of classes describing biological domains within three terms: molecular function (MF), cellular component (CC), and biological process (BP). The molecular function describes the molecular-level activities performed by gene products, (e.g. *catalyst*, or *transport*) and does not specify where, when or in what context the activity takes place (*Gene Ontology Overview*, n.d.). Cellular component describes the locations in which a gene product performs a function (e.g. *mitochondrion*) (*Gene Ontology Overview*, n.d.). Biological process describes the biological programs accomplished by various molecular activities (e.g. *DNA repair*) (*Gene Ontology Overview*, n.d.). First, the GO Consortium creates genes sets, or groups differentially expressed genes that are annotated by the same GO term (MF, CC, and BP) (Ashburner et al., 2000). GO terms can be structured into a loosely hierarchical graph and the three GO terms (MF, CC, and BP) are each represented by a separate root ontology term (Ashburner et al., 2000). All terms can be traced to a root term, although there may be numerous different paths of intermediary terms to an ontology root, providing information on the relationships between GO terms, and gene sets (*Gene Ontology Overview*, n.d.).

In conjunction with the GO analysis, a pathway enrichment will help us gain mechanistic insight into gene sets generated from a GO analysis (Kanehisa et al., 2023a) (Fig. 1.2). At present, the GO does not represent the dynamics required to fully describe a biological pathway (*Gene Ontology Overview*, n.d.). Pathway enrichment analysis identifies biological pathways that are enriched in a gene set more than would be expected by chance (Reimand et al., 2019). The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database for understanding high-level functions of biological systems from molecular-level information (Kanehisa, 2019; Kanehisa et al., 2023a; Kanehisa & Goto, 2000). KEGG databases are categorized into four different aspects: systems information, genomic information, chemical information, and health information. KEGG pathway, in the systems information category, contains a collection of manually drawn pathway maps representing experimental knowledge on metabolism and molecular functions of the cell for many organisms including *P. trichocarpa* (Kanehisa et al., 2023a). KEGG orthology, in the genomic information category, provides annotated ortholog groups of genes in *P. trichocarpa* and other genomes (Tuskan et al., 2006). Orthologous genes are genes in different species that evolved from a common ancestral gene and retained the same function during evolution (Kanehisa et al., 2023a). Orthologs provide a reliable prediction of gene function for KEGG genomes including newly sequenced genomes (Kanehisa et al., 2023b).

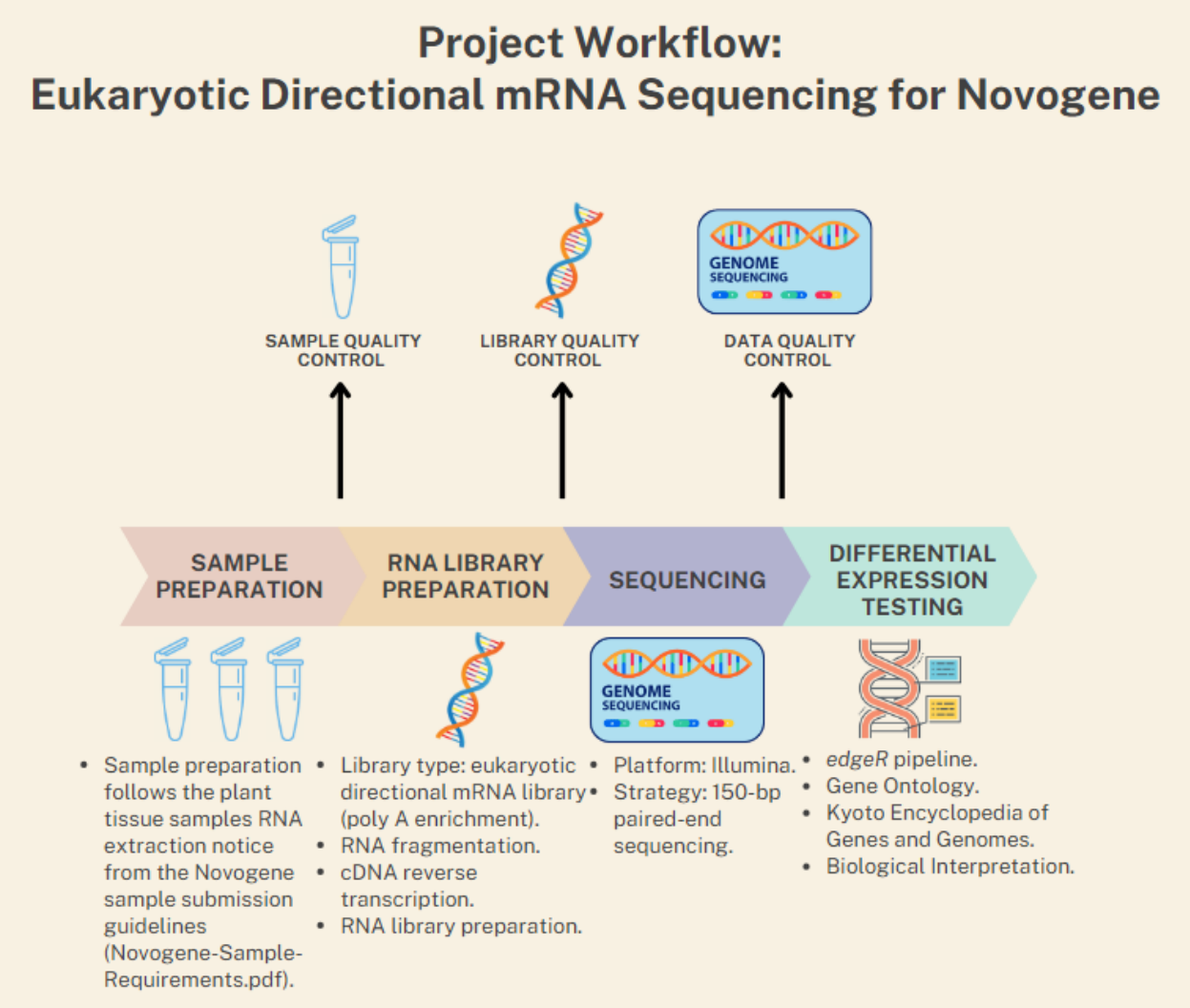


Figure 1.2 Project Workflow: Eukaryotic Directional mRNA Sequencing for Novogene

Project workflow for eukaryotic directional mRNA sequencing through Novogene company. Sample preparation follows Novogene's sample submission guidelines. Sample quality control is performed to ensure samples meet criteria for RNA sequencing. Library preparation followed the eukaryotic, directional mRNA library (poly A enrichment). Library quantification and quality control guarantees high quality data output. 150-bp paired-end sequencing through the Illumina platform was utilized for sequencing and the resulting data are checked for quality. From sequenced data, analysis tools including the *edgeR* pipeline, Gene Ontology and Kyoto Encyclopedia of Genes and Genomes are utilized for biological interpretation.

1.6 The Importance of Field Trials of *Populus spp.*

The advancement of tree molecular biology and genetics for *Populus spp.* is dependent on extensive field research (Strauss et al., 2016; Strauss & Irwin, 2004, 2004; Valenzuela & Strauss, 2005). *Populus spp.* offer researchers a substantial source of variation in tree morphology, anatomy, physiology, phenology, and response to biotic and abiotic stress through adaptations to diverse conditions and prominent genetic polymorphism in local populations (Bradshaw et al., 2000). Although genomic studies with *Populus spp.* can be conducted in greenhouse conditions, limited populations, tree sizes and physiological characteristics as well as cost are limiting factors (Brunner et al., 2004). In addition, field trials allow researchers to evaluate productivity and identify potential ecosystem impacts of *Populus spp.* (Zalesny et al., 2019). Functional genomics can analyze direct changes in the expression of single genes, providing an increase in the ability to link single genes to various phenotypes found in field trials of *Populus spp.* (Brunner et al., 2004).

1.7 Polyclonal Plantings of *Populus spp.*

Gene expression changes resulting from interactions between *P. deltoides* polyclonal plantings could help us target previously identified interpretable gene-biomarker associations or candidate genes related to increased tolerance to stress, nitrogen uptake and enhanced growth for future testing in *P. deltoides* (Richards et al., 2010). *Populus spp.* display high natural phenotypical variation related to its geographical distribution and high intraspecific variability in traits (McKown, Klápště, et al., 2014). Specific traits, such as growth, are genetically complex (McKown, Klápště, et al., 2014). For example, multiple traits, including growth, are genetically correlated and may be polygenic or involve two or more nonallelic genes (McKown, Guy, et al., 2014). Previous genome wide association studies for *P. trichocarpa* have identified single

nucleotide polymorphisms (SNP) associated with traits such as height and volume gain for *P. tricocharpa* (Chhetri et al., 2019; Fahrenkrog et al., 2017; McKown, Klápště, et al., 2014). For this study, it is important to determine if previously identified gene-trait associations related to growth and resource usage exhibit the same function in *P. deltoides.* Similarly, it is important to determine if these previously identified gene-trait associations or other genes that have been previously functionally identified and are not associated with specific traits in *P. deltoides* are differentially expressed in polyclonal plantings where increased growth is observed.

Previous studies have identified that interspecific or intraspecific competition could lead to substantial changes in productivity, including photosynthetic capacity (Duan et al., 2014), changes in root structures (Richards et al., 2010), and nitrogen uptake (Duan et al., 2014; Miller et al., 2007). Several studies have shown that species richness or species diversity increases productivity (Erskine et al., 2006; Forrester et al., 2006; Kelty, 2006; Piotto, 2008; Pretzsch, 2005; Richards et al., 2010). Because of the high level of phenotypic variability from *Populus spp.* (Bradshaw et al., 2000), polyclonal plantings of *P. deltoides* may be a reasonable analog for mixed species plantings (Bradshaw et al., 2000; Brunner et al., 2004). Differences in physiological (e.g. contrasting nitrogen use efficiency) and morphological plant traits for inter-specific competition (competition of different species) have been the basis for species selection for reducing competition between species, leading to greater productivity, growth and nutrient uptake (Richards et al., 2010). Differences in physiological and morphological plant traits for intraspecific competition (competition between the same species, or clonal varieties) have also been shown to drive productivity (Jose et al., 2006; Richards et al., 2010; Rowe et al., 2005; Schmid & Kazda, 2002). Individual trees within polyclonal plantings of the same species may modify soil characteristics through the adjustment of fine root architecture increasing nutrient

uptake within their root zone resulting in greater soil nutrient availability and productivity (Richards et al., 2010). Greater productivity may be due to a reduction in competition of *Populus* polycultures, likely resulting from niche differentiation (Fig. 1.3). Reduced competition has been previously hypothesized to be the result of clones occupying different ecological niches through evidentiary morphological, phenological and metabolic variation between *Populus* varietals (Richards et al., 2010).

Changes in gene expression are the basis for phenotypic variability and niche differentiation of *P. deltoides*. Biogeographical distributions and intraspecific interactions greatly influence phenotypic variability and *Populus* plasticity (McKown, Guy, et al., 2014). Several studies have reported morphological and physiological responses as well as transcriptomic responses to varying nitrogen levels in *Populus* roots (Wei, Yordanov, Georgieva, et al., 2013; Wei, Yordanov, Kumari, et al., 2013). Comparative transcriptomic analysis has investigated leaf size and development for *P. deltoides*,and *P. simonii*, identifying candidate genes involved in molecular mechanisms for leaf development and photosynthetic capacity (Zhang et al., 2021). In addition, a recent study analyzed the gene expression effects of *Populus* clones in monoclonal and polyclonal stands with black locust (*Robina pseudoacacia* L.) stands, identifying greater expression levels of stress and defense response genes for monoclonal plantings, and no differentially expressed genes in the total comparison of monoclonal and polyclonal plantings as a result of high genetic clone specific variability (Kuchma et al., 2022). However, information regarding gene expression and intraspecific variation of *P. deltoides* field clonal varieties has not been extensively studied. (Luo et al., 2015), and further research is required to fully understand gene expression effects of *P. deltoides* polyclonal plantings.

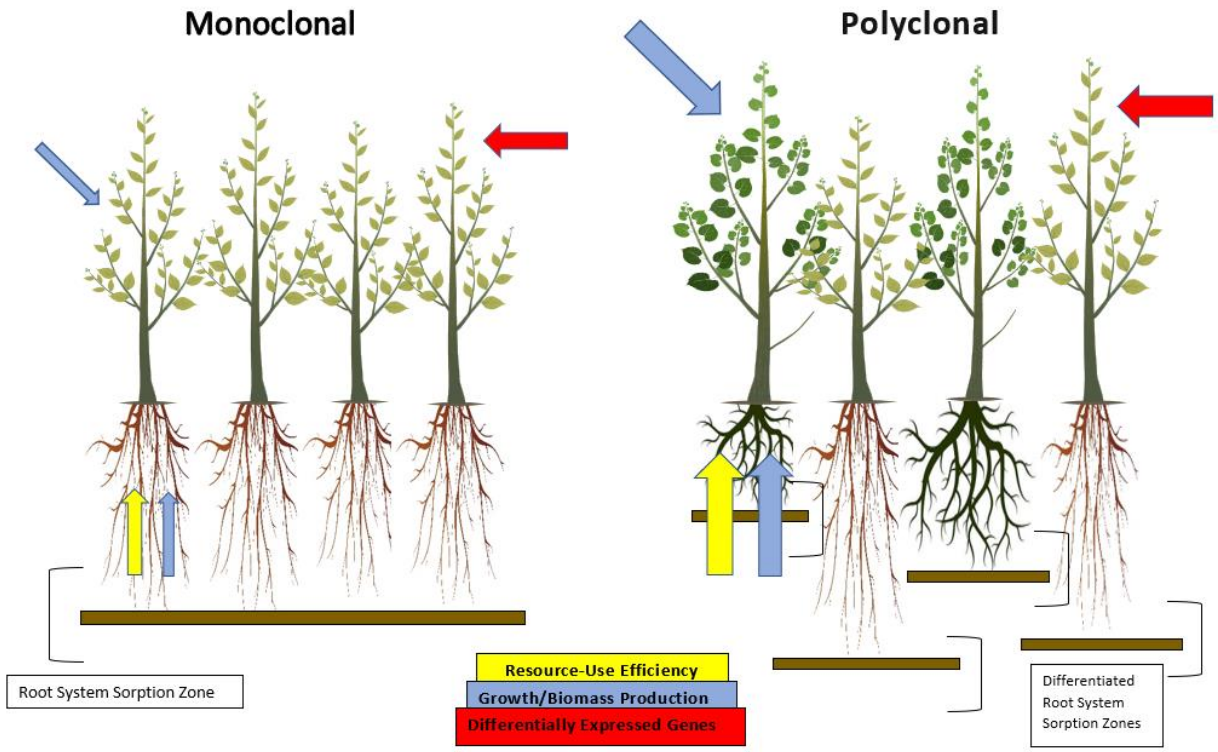


Figure 1.3 Example of Potential Niche Differentiation between Monoclonal and Polyclonal Plantings of *P. deltoides*

Monoclonal and polyclonal plantings and their hypothetical canopy and root structures. It is expected that changes in canopy and root structures, two mechanisms in niche differentiation, will be greater between polyclonal plantings. A process of niche differentiation includes clones working together through adaptations of differentiated root system sorption zones (yellow and blue arrows) and differentiated canopy structures (blue arrow). This will allow *P. deltoides* polyclonal plantings to expand their fibrous roots occupying greater and different portions of the soil compared to monoclonal plantings. Changes in canopy structure (e.g. increased leaf area index) may lead to greater photosynthetic capacity, increasing growth and nitrogen uptake. It is expected that differentially expressed genes to be present in all treatments (red arrows), but polyclonal plantings will display greater gene regulation in genes associated with growth and nutrient uptake.

1.8 Ecological Benefits from Genomic Research

Increased above ground biomass and carbon sequestration are a few ways polyclonal plantings could regulate ecosystem functions, promoting ecosystem services (Gamfeldt et al., 2013; Pretzsch, 2005; Pretzsch & Schütze, 2009). Polyclonal plantings could be an alternative to

monoclonal plantings in achieving positive effects for productivity (Kuchma et al., 2022; Richards et al., 2010), meeting economic, silvicultural, and environmental objectives (Forrester et al., 2005). More diverse plantations, such as polyclonal *Populus* plantings, have a higher chance of containing species that are highly efficient in their use of limiting resources and are responsible for an increase in productivity (Erskine et al., 2006; Loreau & Hector, 2001). Polyclonal plantings of *P. deltoides* may adjust their fine root architecture, may take up nutrients at different times of the year or may alter their preference for different forms of nutrients compared to monocultures, increasing resource use and productivity (Richards et al., 2010). To optimize the positive effects from polyclonal plantings, field trials and genetic analysis are needed to monitor such effects, and to select clones that will complement each other (Kuchma et al., 2022). While transcriptomic field studies are more challenging, they are essential to reveal and confirm complex gene expression responses to different environmental conditions (Izawa, 2015; Kuchma et al., 2022). Explanatory models, such as average soil carbon storage and average tree biomass production phenotypic results have shown that productivity increases with tree species richness (Gamfeldt et al., 2013), and similar responses may be possible by increasing genetic diversity of *Populus* plantations through polyclonal plantings. The positive effects on polyclonal plantings should and can be further confirmed with transcriptomic analysis. Researchers can pinpoint the exact location where genomic changes are occurring on the genome and select them for future testing for polyclonal plantings (Izawa, 2015; Kuchma et al., 2022; Stark et al., 2019).

1.9 Research Questions

This research fills gaps by assessing how the genetic expression of *P. deltoides* clones change when planted in monoclonal plantings or polyclonal plantings in the southeastern United

States. By genetically examining different planting schemes of *P. deltoides* clonal varieties, this research aims to test the hypothesis that differential gene expression (DGE) is an underlying molecular mechanism contributing to overyielding (polycultures are more productive than the average of their monocultures), increased resource utilization, and increased stress tolerance in polycultures through regulation of morphology and physiology. Furthermore, this research aims to identify differentially expressed genes and biological pathways associated with increased stress tolerance, increased productivity, and increased nitrate content in leaves of polyclonal plantings. This research will contribute to an increased understanding of gene expression between different planting schemes of *P. deltoides* clonal varieties. Genomic analysis of *P. deltoides* will fill gaps left by similar studies in field trial settings that lack information on *P. deltoides* clonal varieties, specifically, and when adapted to a subtropical climate with longer growing seasons.

The overall objective for this study is to determine if differentially expressed genes are an underlying molecular mechanism that may explain differences in stress tolerance, productivity and resource usage between monoclonal and polyclonal plantings, via inducing morphological and physiological plasticity.

I plan to attain the overall objective by pursuing the following specific objectives and associated hypotheses:

1. Determine if there are differences in nitrogen uptake and biomass production between *P. deltoides* monoclonal plantings and polyclonal plantings.

   Hypothesis: Polyclonal plantings of *P. deltoides* will display higher resource use and productivity than monoclonal plantings.

2. Determine the regulation of differentially expressed genes (upregulated or downregulated) for both *P. deltoides* varieties in monoclonal and polyclonal plantings.

   Hypothesis: Polyclonal plantings of *P.deltoides* will display regulation of genes associated with increased productivity and resource usage.

3. Identify the function of genes associated with stress, productivity and resource usage and how those genes function in biological processes, such as nitrogen metabolism and growth.

   Hypothesis: Polyclonal plantings of *P. deltoides* will display gene function in biological processes associated with increased stress tolerance, productivity and resource usage.

These objectives will promote:

1. Better understanding of the role gene expression plays in intraspecific variation for *P. deltoides*, providing information on already studied gene-trait associations related to stress, growth and resource usage or new candidate genes that have not been associated with specific traits but should be further analyzed in future studies, supporting future clonal selection for larger field varietal studies and tree improvement programs.
2. Better understanding of the implications in polyclonal plantations for increasing stress tolerance, productivity and resource usage.

CHAPTER II

MATERIALS AND METHODS

2.1 Experimental Design and Sampling Methods

2.1.1 Study Area and Planting Schemes

This research is a part of the *Populus* in the Southeast for Integrated Ecosystem Services (PoSIES) project and utilizes a new field trial as part of a larger Department of Energy grant (Department of Energy (DOE) DE-EE0009280) for evaluating *Populus spp.* clonal varieties' productivity and ecosystem services. The site is located at the Pontotoc Ridge-Flatwoods Experimentation Station near Pontotoc, Mississippi (Fig. 2.1). The research site is a part of Mississippi State University's Agricultural and Forestry Experimentation Station. One replicate of the larger project was utilized for this study (Fig. 2.2). This project focused on pure *P. deltoides* (DxD: *P. deltoides*) with contrasting nitrogen use. Clonal selection for contrasting nitrogen use for the PoSIES project was based on data collected from a previous field trial, with Clone 110412 displaying lower nitrogen use efficiency and lower nitrogen percentage in leaf tissues compared to Clone S7C8 (Renninger et al., 2022). Clone 110412 originated from Bolivar County in Mississippi, near the Mississippi River, and Clone S7C8 originated from Brazos County in Texas, near the Brazos River (Jeffreys, 2005). Information on the geographic origins of both clones came from former trials established by the USDA Forest Service and Oklahoma State University (Jeffreys, 2005). Dormant, 38 cm unrooted cuttings of selected eastern cottonwood were obtained from Big River Cottonwood Nursery in Winnsburro, LA. Prior to

planting and to provide initial protection against cottonwood leaf beetles, cuttings were soaked in water and Admire Pro® systemic insecticide following the same rate of 0.14 fluid oz. per 1 gallon (1.09mL/L) of water utilized in a previous study with *P. deltoides* (Dahal et al., 2022). Plantings were established on April 21st, 2021 using a split plot variation of a complete randomized block design. The whole plot factor consists of inoculation with a mixture of endophytic bacteria obtained from Intrinsyx Bio in Sunnyvale, CA and no inoculation (i.e. control). The split-plot factor consists of clonal composition of 12 different planting combinations of varietals. All planting treatments consist of 30 tree blocks planted in a 6 rows × 5 columns arrangement on a 1.8m × 1.8m spacing. The single row of trees around the perimeter of each plot served as buffer trees with only the inner 12 trees (4 rows x 3 columns) serving as trial trees. Prior to planting establishment, flumioxosin was applied to the ground as preemergent weed control. Cuttings were planted near riparian areas adjacent to agricultural fields to intercept polluted surface runoff and shallow ground water before it enters the stream. Vegetative competition was treated as needed through the first two growing seasons with a combination of glyphosate herbicide, mowing, discing, and hand weeding to keep trees in free to grow conditions. For this study, I chose to focus on two clones and two subplots in one complete replicate of the main study to minimize variability in site factors. Preliminary results did not show significant effects of the whole plot factors on gene expression, but any potential differences observed were captured by plot effects. Planting schemes for sequencing analysis consisted of two replicate plots for each planting scheme for a total of four monoclonal plots (two of each clone) and two polyclonal plots (Fig. 2.3).

Figure 2.1 Populus Clonal Varieties at the Pontotoc-Ridge Flatwoods Experimentation Station

(Left) Established clone trials at the Pontotoc-Ridge Flatwoods Experimentation Station near Pontotoc, Ms. Picture taken on July 21st, 2021. (Right) Hydraulic lift utilized for leaf tissue sample collection. Picture taken on July 28th, 2022.

Figure 2.2 Map of Pontotoc County, Mississippi, and Google Earth view of Experimentation Site

(Left) Map of the state of Mississippi. Green Star designates location of Pontotoc, Mississippi. (Right) Site of whole plot utilized for experimental analysis at the Pontotoc-Ridge Flatwoods Experimentation Station near Pontotoc, Mississippi. The whole plot is outlined in yellow. Plots used for leaf sample collection are highlighted in red.

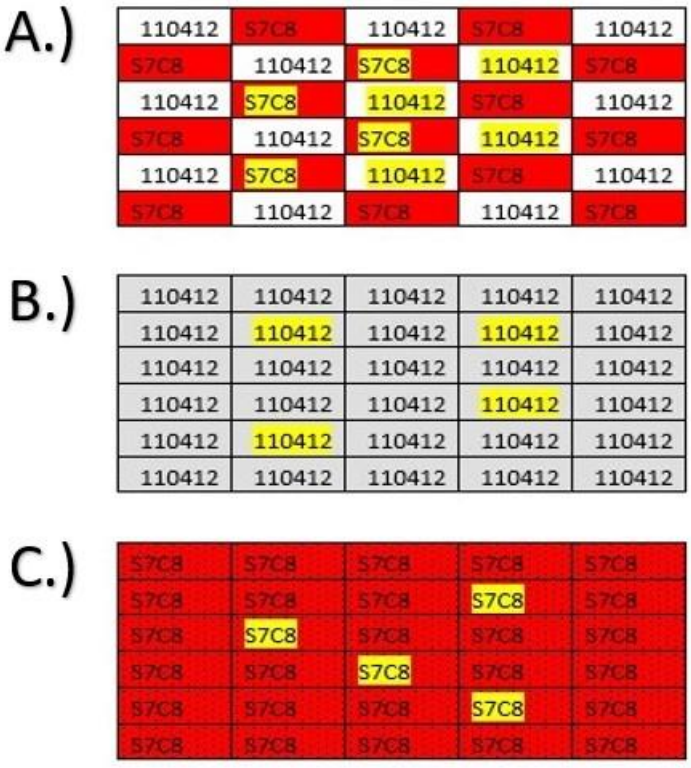


Figure 2.3 Planting Schemes of P. deltoides for Experimental Analysis

Polyclonal and Monoclonal planting design for one replicate for experimental analysis. A.) An example of eight individual biological replicates (highlighted) chosen for sampling analysis for polyclonal plantings. B.) and C.) An example of four individual biological replicates (highlighted) chosen for sampling analysis for monoclonal plantings. Leaf samples were chosen based on leaf quality and integrity.

2.1.2 Groundwater Collection

Collaborators on the PoSIES project installed shallow ground water wells outside the plots, on the agricultural field side and inside the plots. Groundwater wells were established at half of the whole plot factor, or plots containing no endophyte treatments. The wells consist of 10-cm diameter PVC pipe installed to a depth of 2 m, backfilled with fine sand, and sealed with bentonite clay. The wells were slotted and screened belowground. Samples were collected monthly to bimonthly during the growing season. Prior to sample collection, wells were

manually evacuated and allowed to refill to ensure fresh groundwater collection. Water samples were placed on ice, returned to the laboratory, and stored at 4℃ until processing. Inorganic nitrogen (Nitrate: NO_3^- and Ammonium: NH_4^+) was determined calorimetrically (AQ300 Discrete Analyzer, Seal Analytical) on samples that have been filtered to remove particulates < 0.45um. (NO_3^- and NH_4^+) were measured from water samples taken at each well to estimate the amount of shallow ground water nitrogen taken up by the *Populus* trees from the agricultural field before the shallow ground water enters the stream system. Nitrate is negatively charged and does not bind to soils, allowing excessive amounts to leach into water sources. Total nitrate (NO_3^-) concentration (mg N/L) averages, and standard errors were determined for Clone S7C8 and 110412 monoclonal plots, polyclonal plots and control wells. The collection time period analyzed for this project ran from June 3rd, 2021 through July 25th 2022. Since groundwater quality was measured in accordance with the larger PoSIES project and is not time specific to the collection of leaf tissue samples for gene expression analysis, only averages and standard errors were reported.

2.1.3 Tree Measurements and Leaf Area Index

At the end of the first growing season, all trees total height and diameter at stem base were measured. At the end of the second growing season, tree total height and diameter at breast (DBH) height (1.3m) were measured. A subset of trees was destructively sampled at the end of the second growing season throughout the larger study. Dry weights of sampled trees were utilized to develop allometric estimates of aboveground biomass based on DBH and height.

Leaf area index (LAI) was measured nondestructively to assess phenology of the *P. deltoides.* Collaborators on the project used the LAI 2200 from LiCOR Biosciences Inc. located in Lincoln, NE to collect measurements. Direct sun conditions for LAI collection could lead to

reflected and transmitted light off leaves that must be accounted for. Scatter corrections were made for measurements collected in direct sun conditions. For this project, the collection period analyzed for LAI was July 27th, 2022. Average LAI and standard errors were determined for the leaf tissue samples for Clone S7C8 and 110412 monoclonal plots and polyclonal plots. Each clonal monoclonal plot was analyzed separately from polyclonal plots (df=2). A one-way ANOVA was performed to compare the effect of planting schemes on average LAI. Following an ANOVA, a Tukey's HSD Test was performed to determine which planting schemes were significantly different in average LAI.

2.1.4 Leaf Material

At the end of July, or the peak of the second growing season, the leaves from two *P. deltoides* varieties were collected from the six monoculture and polyculture plots for RNA-Seq and combustion elemental analysis. On July 28th, 2022 a hydraulic lift device was used to collect leaf samples from the terminal shoot. Between the fifth to twelfth leaf from the terminal, three fully expanded and undamaged leaf samples were cut from the tree for each individual biological replicate. Samples were not selected from the terminal bud as this is where new growth is coming from. Using a leaf cutter tool, leaf samples from four individual biological replicates (i.e. four trees) were taken from the planted monoclonal plots and leaf samples from eight individual biological replicates, four from each clone, were taken from planted polyclonal plots, totaling to 32 sampled trees and biological replicates for this project (Fig. 2.3). All samples were collected in a three-hour time frame from 8am-11am under sunny conditions to minimize temporally related gene expression changes between plots. The standard workflow for differential gene expression (DGE) analysis using *edgeR* involves RNA-Seq with a sequencing depth of 10-30 M reads per library and at least three biological replicates per sample (Chen et al. 2023). Biological

replicates were chosen based on leaf quality and integrity. During sampling, a pair of scissors was used to cut each leaf sample in half down the central vein. One half of the leaf was placed in a 50ml labeled sample collection vial and flash frozen on dry ice for RNA extraction and sequencing analysis, and the other half was placed in a labeled Whirl-Pak® write-on bag and stored on ice for elemental analysis. In between cuttings, the leaf cutter tool and pair of scissors was sterilized with a 30% mixture of pure microbiology grade ethanol alcohol. The samples were immediately transported back to the laboratory. Samples for the combustion elemental analyzer were placed in brown paper bags in an oven dehydrator at 60℃ for 48 hours to remove excess moisture. Flash frozen samples for RNA extraction were placed into a freezer at -80℃ until extractions occurred.

2.1.5 Elemental Analyzer Analysis

Sample preparation and analysis followed the protocol for organic elemental analysis on the elemental analyzer ECS 4010 from Costech Analytical Technologies, Inc. located in Valencia, CA. Oven dehydrated leaf samples were removed from the oven and ground to a fine powder before placing them back in the dehydrator oven to dry overnight at 60℃. After drying, samples were placed in desiccators, allowing them to cool to room temperature for 1 hour. Between 2 to 4 mg of the homogenized leaf sample were weighed and used for subsequent analysis. Samples were placed into 3.5 x 5mm pressed tin capsules. Atropine, a chemical with a known concentration of nitrogen and carbon, was used to create a calibration curve for the leaf samples. The elemental analyzer ECS 4010 output nitrogen and carbon amounts in milligrams (mg), and percentages were calculated from nitrogen and carbon amounts (mg) vs. the weight of the grounded sample powder. Average total leaf carbon and total leaf nitrogen percentages and standard errors were determined for the leaf tissue samples. Since we were interested in

examining changes in nitrogen uptake, a one-way ANOVA was performed to compare the effect of planting schemes on total leaf nitrogen percentages. For this project, we were also interested in changes in carbon to nitrogen ratios (C:N) ratios, and a one-way ANOVA was performed to compare the effect of planting schemes on C:N ratios. Each clonal monoclonal plot was analyzed separately from polyclonal plots (df=2).

2.1.6 Total RNA Extraction, Purification and Quality Controls

Prior to extraction, mortar and pestles were sterilized in an autoclave machine with deionized water for 15 minutes and set out to completely cool. Frozen tissue samples of 100mg were weighed and ground to a fine powder in liquid nitrogen using the sterilized mortar and pestle. Qiagen's RNeasy plant mini kit was used for purification of total RNA from plants (*RNeasy Plant Mini Kit*, n.d.). Total RNA was extracted following the protocol for purification of total RNA from plant cells and tissues from the RNeasy Mini Handbook (*RNeasy Mini Handbook - (EN) - QIAGEN*, 2019). Before elution of each RNA sample occurred, Qiagen's RNase-Free DNase Set was used for efficient on-column digestion of DNA during the RNA purification process (*RNeasy Mini Handbook - (EN) - QIAGEN*, 2019). Purified RNA was stored at -80°C in RNAse-free water. 1µl of each extraction was analyzed spectrophotometrically using a Nanodrop™ One Spectrophotometer (Thermo Fisher Scientific, 2023). At least 200ng/µl of each sample was loaded onto a 1% (w/v) agarose gel stained with ethidium bromide. All samples displayed clear bands corresponding to ribosomal RNA (rRNA), absence of DNA contamination and no degradation (Fig. 2.4). Additionally, RNA integrity numbers (RIN) for quality controls (QC) ran by Novogene biotech company, discussed in the following section, showed appropriate rRNA peaks with RIN values, verifying quality RNA results for downstream applications (Babu & Gassmann, 2016).

Figure 2.4 Agarose Gel Electrophoresis to Assess RNA Quality

Example of agarose gel electrophoresis for eight biological replicates from the experimental analysis. Non-degraded RNA displays two consecutive and sharp 25S and 18S rRNA bands in a 2:1 ratio. 25S and 18S rRNA bands are circled in red.

2.2 Data Analysis

2.2.1 Messenger RNA-Seq and Quantifying Samples

RNA samples were sent to Novogene for eukaryotic mRNA sequencing and transcriptome profiling. Samples were prepared and shipped according to the manufacturer's instructions (‘MRNA Sequencing (MRNA-Seq)’, 2023). Before sequencing began, Novogene measured the RIN for each sample through the Agilent 2100 Bioanalyzer system (Agilent Technologies, Inc., 2023). Samples “passed” the QC if rRNA peaks were well resolved and automatically identified by the software (Babu & Gassmann, 2016). 31 out of 32 samples “passed” the RNA QC, with one sample displaying abnormal rRNA peaks. This sample was extracted again and resubmitted to Novogene, passing with well resolved rRNA peaks.

Following sample RNA QC, paired-end sequencing (PE150) with a total of 300 bp sequences for each RNA molecule was run on the Illumina NovaSeq 6000 ('MRNA Sequencing (MRNA-Seq)', 2023). Sample test, library preparation and sequencing directly impact the analysis results (*Illumina | Sequencing and Array-Based Solutions for Genetic Research*, n.d.). To guarantee reliable data, QC was performed at each step of the workflow. Following sequencing, raw data files were received from Novovgene and stored in the form of two separate compressed FASTQ files for each sample, each containing sequencing reads and corresponding sequencing quality. The *P. trichocarpa* v4.1 reference transcriptome from the Joint Genome Institute's Phytozome v13 was used in the *Salmon* software tool in Linux to build an index, map sequences, or reads, and quantify transcripts (Goodstein et al., 2012; Patro et al., 2017). The reference transcriptome (in FASTA format) and raw sequences reads (in FASTQ format) were utilized as input for "quasi-mapping" and quantification. Quasi-mapping requires a reference index to determine the position and orientation for where fragments best map prior to quantification, providing the transcriptome in a format that is easily and quickly searchable (Srivastava et al., 2016). In *Salmon*, length normalization occurs during quantification before transcripts are converted back to read counts. *Salmon*'s output file was in the form of a quantification file (quant.sf). Quantification files were inputted into R statistical software via the tximport Bioconductor package (R Core Team, 2022; Soneson et al., 2015).

2.2.2 Testing for Differential Gene Expression

The *edgeR* Bioconductor package in R statistical software was utilized for differential gene expression (DGE) analysis (Chen et al., 2016; McCarthy et al., 2012a; R Core Team, 2022; Robinson et al., 2010). Count data for DGE analysis represents the number of sequences reads that originate from a particular gene. Genes with very low counts across all libraries provide little

evidence for differential expression (Chen et al., 2023). 7844 genes that were not expressed in any samples or planting treatments and/or with very low counts were filtered out prior to further analysis. The surviving gene summary consisted of 26783 genes for subsequent analysis. *EdgeR* observed relative changes in expression levels between polyclonal and monoclonal plots. Since fewer than 12 biological replicates are used for each planting treatment, a superior combination of true positive and false positive performances makes *edgeR* the best tool for RNA-Seq analysis in this instance (Schurch et al., 2016). Similarly, because this study was not examining sample-specific effects, normalization was not conducted. Normalization issues only arise to the extent that technical factors display sample-specific effects (Chen et al., 2023). Although, sequencing depth, the most important technical factor affecting read counts, is automatically adjusted as represented by different library sizes and the function calcNormFactors (calculation Normalization (scaling) factors) for computed effective library sizes from scaled counts to account for composition biases between samples (Chen et al., 2023). *EdgeR* was utilized to fit a Poisson negative binomial distribution to model read counts for each gene in each sample, and account for variability between biological replicates (Chen et al., 2023; McCarthy et al., 2012b). The *P. trichocarpa* v4.1 reference genome from the Joint Genome Institute's Phytozome v13 was used to detect DGE (Goodstein et al., 2012). Dispersion estimates for relative gene abundance were visualized with a scatterplot of the biological coefficient of variation (BCV). Once dispersion estimates were attained, a general linearized model likelihood ratio test was utilized for DGE analysis. Three appropriate models via contrast matrices were created for DGE (Table 2.1). The functions glmFit() and glmLRT() from the *edgeR* package (Chen et al., 2022) were used for testing to compare polyclonal plots to baseline monoclonal plots (Clone S7C8 + Clone 110412 Polyclonal vs. Monoclonal) and to compare clones in their respective planting

schemes (Clone 110412 Polyclonal vs. Monoclonal, and Clone S7C8 Polyclonal vs. Monoclonal). The experimental analysis observes gene expression between polyclonal and monoclonal plantings. The fold change is usually given as a calculated log2 experimental group/control ratio. For this study, the ratio is defined as polyclonal/monoclonal. For example, if gene A has an average expression of 100 mapped reads in polyclonal plantings and 50 reads in monoclonal plantings, the logFC is 2 for polyclonal plantings compared to monoclonal plantings or a 2-fold upregulation for polyclonal plantings. For each model, the *P. trichocarpa* v4.1 genome was utilized as a reference-based genome and genes were established and grouped by expression level into gene sets, or biologically relevant groups (Goodstein et al., 2012). Overexpressed and underexpressed genes in each model had a positive or negative base 2 logarithmic fold change (logFC) values (Jabato et al., 2012). Expression levels thresholds for each model consisted of false discovery rates (FDR) less than or equal to 0.05, p-values less than or equal to 0.05, and a logFC greater than 1 or less than -1 (McDermaid et al., 2019).

Table 2.1 Design Matrices for Differential Expression Analysis

Model Names	Design Matrix	Contrast	Output
1-Polyclonal vs. Monoclonal	~0+Planting+Clone+Plot	Polyclonal.vs. Monoclonal	Clone S7C8+Clone 110412 Polyclonal vs Monoclonal
2-Clone 110412 Polyclonal vs. Clone 110412 Monoclonal	~0+Planting:Clone+Plot	White.Poly.vs.Mono	Clone 110412 White Polyclonal vs White Monoclonal
3-Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal	~0+Planting:Clone+Plot	Red.Poly.vs.Mono	Clone S7C8 Red Polyclonal vs Red Monoclonal

Specified design matrices for differential expression analysis in *edgeR*. The design matrix records which planting schemes were applied to each samples and defines how the experimental effects are parametrized in the linear models. Contrasts observe genes more highly expressed in polyclonal plantings. Each clone is designated with a specific color (red for Clone S7C8 and white for Clone 110412)

2.2.3 Gene Ontology (GO) and Pathway Analysis

To interpret the DGE analysis output in a biological context, a Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Enrichment Pathway Analysis were performed in R statistical software (R Core Team, 2022). The GO's knowledgebase is the world's largest source of information on the function of genes (Ashburner et al., 2000; Carbon et al., 2009; Day-Richter et al., 2007; Mi et al., 2019). Annotation maps for the entire GO were obtained from the Go.db Bioconductor package (Carlson et al., 2019). Additionally, Bioconductor's AnnotationDbi package was utilized jointly with the GO database (Pagès et al., 2023). The *P. trichocarpa* v4.1 annotation file contains GO terms from the GO database, specific to the *P. trichocarpa* genome and was used to annotate gene sets for each contrast matrix (Goodstein et al., 2012). A gene set consists of genes assigned to a specific GO term (Chen et al., 2016). Under each model, a given set of genes is either upregulated or downregulated. The GO enrichment analysis found GO terms that were over-represented or under-represented using

annotations for the genes in each gene set (Chen et al., 2023). In conjunction with the GO analysis, the fry() function in the Limma package in R statistical software was used for gene set testing (Chen et al., 2023; R Core Team, 2022). The Fry function tests whether any genes in each gene set are differentially expressed using random and infinite rotations of the residual space of each contrast matrix (Chen et al., 2016). For each model, the output of the analysis contained the GO ID, the number of genes associated with that ID, the net direction of change, p-values, FDRs, annotation terms, and ontology (Table C.1). Restrictions for each model consisted of FDRs less than or equal to 0.05, and p-values less than or equal to 0.05. These restrictions output no statistically significant GO IDs with FDR values less than or equal to 0.05. To better interpret GO terms with non-significant FDR values, the p-value cutoff was reduced to less than or equal to 0.01.

Following the GO analysis, a Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment pathway analysis was applied to better understand the related functions and pathways of the differentially expressed genes (CD Genomics, 2023). KEGG's database contains a collection of genomes, biological pathways, creating a computational representation of biological systems (Kanehisa, 2019; Kanehisa et al., 2023b; Kanehisa & Goto, 2000). KEGG Orthology (KO) database provides genomic information in terms of functional orthologs defined from experimentally characterized genes and proteins in specific organisms, and proteins and genes are used to assign orthologous genes in other organisms based on sequence similarity (Kanehisa et al., 1995). Since the Phytozome reference did not contain KEGG annotations, or KEGG ortholog identifiers (K numbers), Phytozome annotations were mapped to Entrez IDs from the National Center for Biotechnology Information (NCBI) database for KEGG analysis (Sayers et al., 2021). A reciprocal best hits blast (RBHB) was conducted for ortholog determination

(Goodstein et al., 2012). A RBHB is found when transcripts encoded by two genes, each in different genomes (i.e. the two *P. deltoides* clones), find each other as the best scoring match in the other genome (Ward & Moreno-Hagelsieb, 2014). The transcripts from Genome A (Clone S7C8) are aligned to the transcripts from Genome B (Clone 110412). If the top alignment is the same in both genomes, the transcripts are considered orthologs (Moreno-Hagelsieb & Latimer, 2008). An RBHB is a common method used for rough assessment of orthologs, or homologous sequences, between transcript files, or genomes. For KEGG, I imported transcript-level estimates using tximport, and the tx2gene database was used to link transcript ID to gene ID for the reciprocal blast orthologs (Soneson et al., 2015). The Fry function in the Limma package was utilized to filter genes based on expression levels, contrast matrix information and whether pathways contained genes or gene groups. 4207 genes were not associated with KEGG Pathway IDs. The surviving gene summary consisted of 23316 genes for pathway analysis. Pathway analysis output for each contrast matrix contained the pathway ID, the number of genes associated with that pathway, net change of direction and description of each pathway. Mixed analysis of p-values and FDRs were also utilized to analyze the magnitude of the expression change only, disregarding gene direction (upregulated or downregulated). Since pathways tend to have positively correlated expression levels (Hong et al., 2014), this is helpful in finding pathways with inversely correlated genes.

CHAPTER III

RESULTS

3.1 Groundwater Data

Total nitrate concentration averages were determined for the first and second growing seasons for monoclonal, and polyclonal plots and control wells (Fig. 3.1). The timeline for sample collection for this study ran from June 3rd, 2021 through July 25th 2022. Dates with no sampled control wells: August 17th, 2021, September 13th, 2021, September 29th, 2021, October 13th, 2021, June 15th, 2022, and July 25th, 2022. Wells sampled on September 29th, 2021 and July 25th, 2022 did not have all planting schemes sampled. Since groundwater quality was measured in accordance with the larger PoSIES project and is not time specific to the collection of leaf tissue samples for gene expression analysis, only averages and standard errors were reported.

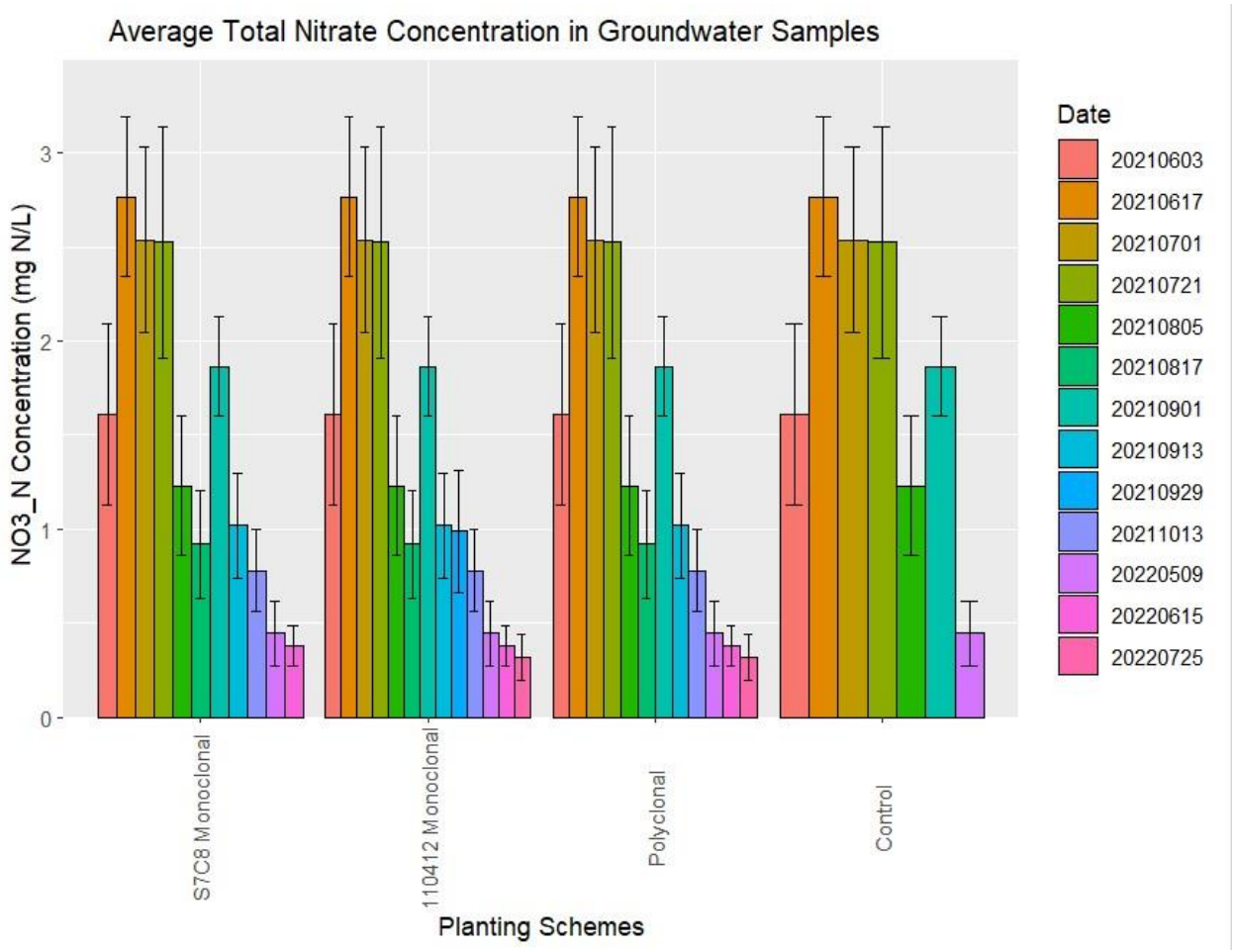


Figure 3.1 Average Nitrate Concentrations for First and Second Growing Seasons

Average nitrate concentrations for first and second growing seasons in groundwater samples for monoclonal and polyclonal plantings schemes with controls. Graph displays date of collection and the average nitrate (NO3_N) concentration in (mg N/L). The timeline for sample collection for this study ran from June 3rd, 2021 through July 25th 2022. Dates with no sampled control wells included August 17th, 2021, September 13th, 2021, September 29th, 2021, October 13th, 2021, June 15th, 2022, and July 25th, 2022. Wells sampled on September 29th, 2021 and July 25th, 2022 did not have all planting schemes sampled.

3.2 Survival, Biomass and Leaf Area Index Results

At the end of the second growing season, all inner sample trees for the two polyclonal plots survived. 11 out of the 12 sample trees survived the two S7C8 monoclonal plots. 110412 monoclonal plots had 8 and 10 out of 12 trees survive. After the first two years of growth, above

ground biomass in megagrams of biomass per hectare (Mg/ha) was estimated (Table B.2). The average above ground biomass for monoclonal plots was 6.37 Mg/ha. Both polyclonal plots had greater above ground biomass than any of the monoclonal plots and had an average of 11.463 Mg/ha. LAI averages and standard errors were determined for Clone S7C8 and 110412 monoclonal plots and polyclonal plots at the end of July (Fig. 3.2). Each monoclonal plot was analyzed separately from polyclonal plots (df=2). A one-way ANOVA revealed that there was a statistically significant difference in average LAI between monoclonal and polyclonal plantings (F=3.99, $p < 0.05$). Following an ANOVA, a Tukey's HSD Test was performed to determine which planting schemes were significantly different in average LAI. Tukey's HSD Test for multiple comparisons found that the average LAI was significantly different between Clone 110412 monoclonal plots and polyclonal plots (p=<0.05, 95% C.I.= [0.1096330, 2.4031320]). There was no statistically significant difference between Clone S7C8 monoclonal plots and polyclonal plots (p=0.098), and Clone S7C8 and Clone 110412 monoclonal plots (p=0.894).

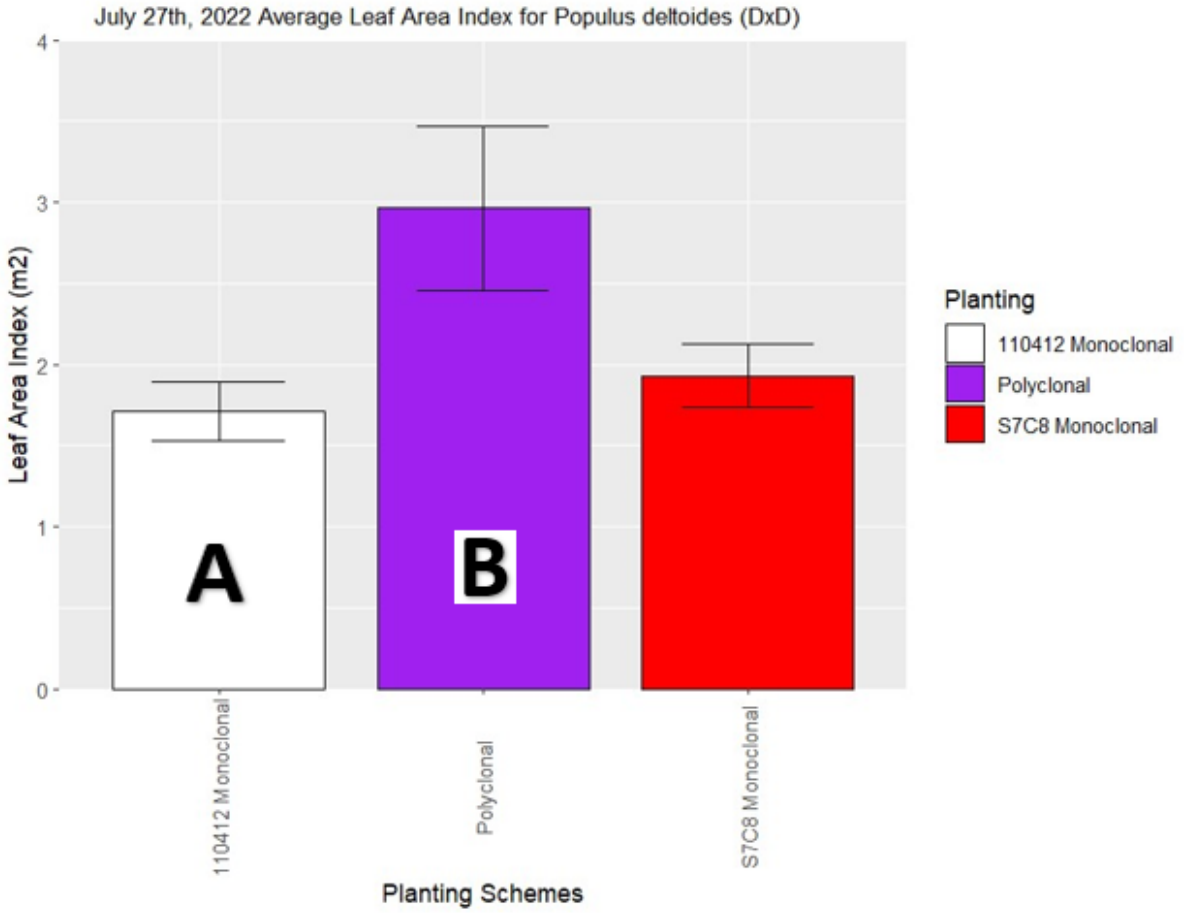


Figure 3.2 Average Leaf Area Index (LAI) for Planting Schemes of *P. deltoides* for July 27th, 2022

Average leaf area index (LAI) in m^2 for planting schemes from experimental analysis. Standard error bars are denoted by black lines. Tukey's HSD Test for multiple comparisons found that the average LAI was significantly different between Clone 110412 monoclonal plots (A.) and polyclonal plots (B.).

3.3 Results from Elemental Analysis

The average total leaf nitrogen percentage for polyclonal plots was 2.07% (Fig. 3.3). Clone S7C8 monoclonal plots had an average total leaf nitrogen percentage of 2.16%, and Clone 110412 monoclonal plots had an average of 2.03%. A one-way ANOVA revealed there was no statistically significant difference in average total leaf nitrogen percentages between monoclonal and polyclonal plantings ($p=0.507$). The average total leaf carbon percentage for polyclonal plots was 45.36%. Clone S7C8 monoclonal plots had an average total leaf carbon percentage of 44.41%, and Clone 110412 monoclonal plots had an average of 46.93%.

C:N ratios and average values of C:N ratios were calculated by dividing the carbon percentage by the nitrogen percentage (Fig. 3.4). The ratio signifies the amount of carbon in mg per 1 mg nitrogen in leaf tissue samples. Polyclonal plots had an average C:N ratio of 21.91:1. S7C8 and 110412 monoclonal plots had an average C:N ratio of 20.70:1 and 23.30:1. A one-way ANOVA revealed that there was no statistically significant difference in C:N ratios between monoclonal and polyclonal plantings ($p=0.111$).

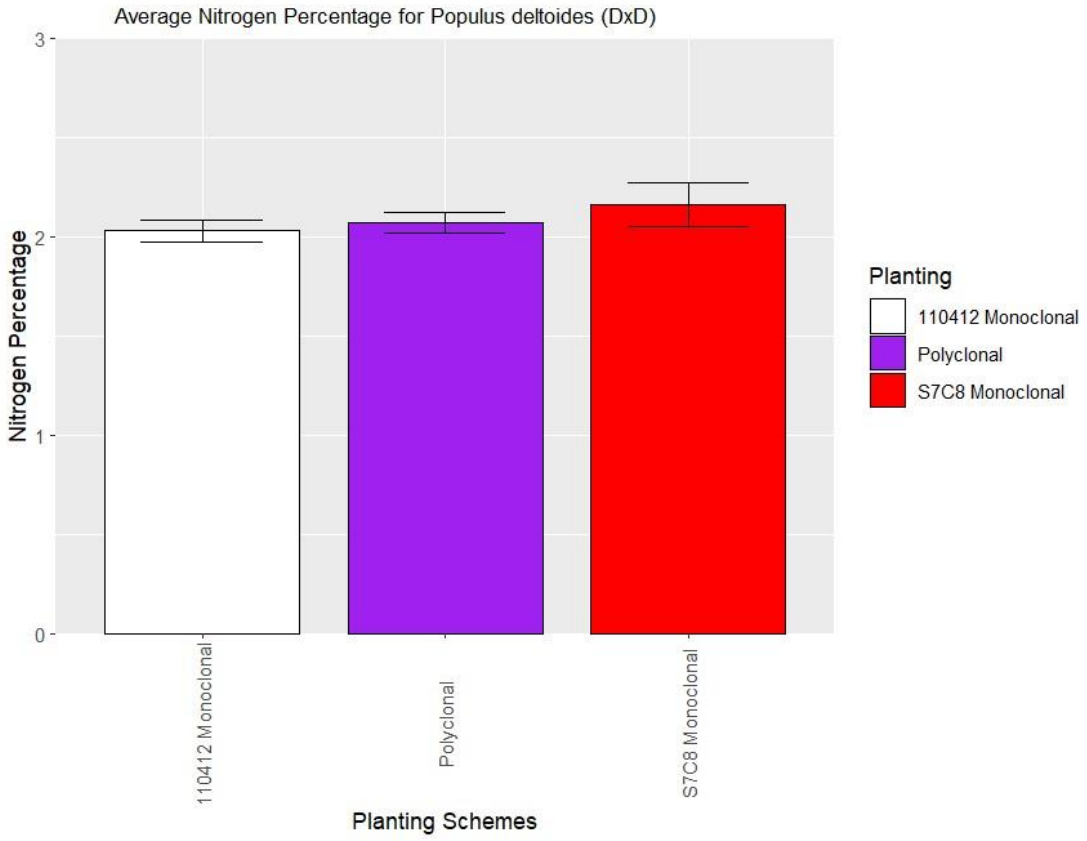


Figure 3.3 Average Nitrogen Percentage for Planting Schemes of *P. deltoides*

Average total leaf nitrogen percentage for planting schemes from experimental analysis. Standard error bars are denoted by black lines.

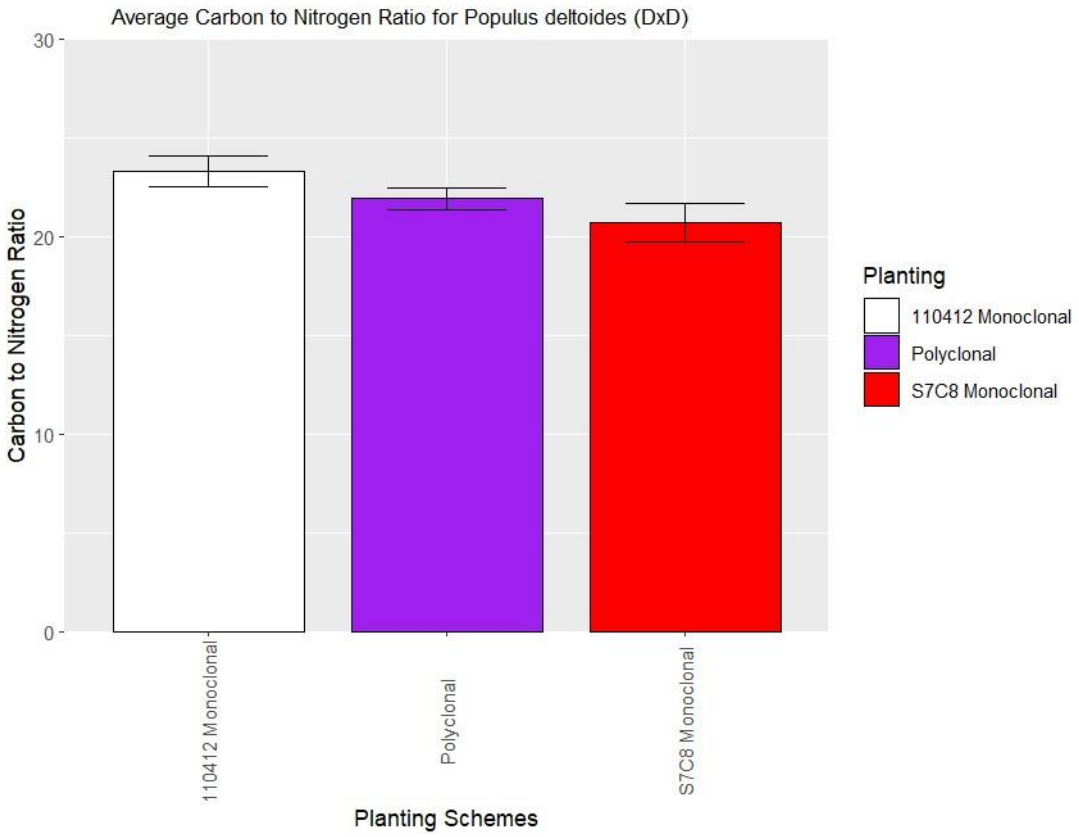


Figure 3.4 Average Carbon to Nitrogen Ratio for Planting Schemes of *P. deltoides*

Average leaf carbon to nitrogen ratios for planting schemes from experimental analysis. Standard error bars are denoted by black lines.

3.4 Quality of RNA Samples and Quantifying Transcript Expression

All 32 total RNA samples, including the resubmitted sample, were used for subsequent analysis. All samples displayed a ratio of absorbance at 260 and 280nm between 1.9 and 2.1 (Table A.1). 260/230 ratios, a secondary ratio to assess RNA purification and contamination, ranged from ~1.4 to 2.3. While some absorbance 260/230 ratios (4NEMS_3: 1.422, and 4ES_4:1.498), were lower than suggested in the protocol (above 1.5) (*RNeasy Mini Handbook - (EN) - QIAGEN*, 2019), RNA gels showed intact RNA bands (25S and 18S in a 2:1 ratio). RNA Integrity Numbers (RIN) ran by Novogene for QC, displayed well resolved rRNA peaks,

automatically detected by the Agilent 2100 Bioanalyzer system (Agilent Technologies, Inc., 2023). RIN for leaf tissue samples ranged from 5.9-8.3 (Figure A.1). One sample failed the RIN QC test with an RIN of 2.5 and small, fragmented rRNA peaks. This sample was resubmitted to Novogene, passing with well resolved rRNA peaks and an RIN sample of 8.1.

For this study, I generated roughly 4.8 billion high quality paired-end reads covering 354 GB of sequence data with a sequence of 150 bp (Table 3.1). The dataset of raw reads was filtered by Novogene to deliver clean data. Paired reads are discarded for the following situations: when one read contains adapter contamination; when one read contains more than 10 percent of uncertain nucleotides; and when one read contains more than 50 percent low quality nucleotides (‘MRNA Sequencing (MRNA-Seq)’, 2023; Yan et al., 2013). After cleaning, less than 1 percent of reads were removed, obtaining almost 4.8 billion reads, covering 351 GB of sequence data. Abundance estimates, or transcript expression estimates were obtained using the *Salmon* alignment tool (Patro et al., 2017). Mapping rates averaged 87.6% for all samples, with quantification files covering 417 Mb of output data.

Table 3.1 Overview of Sequencing for Raw Data Without and With Cleaning

	Raw Data (Without Cleaning)	Raw Data (With Cleaning)
Total Number of Raw Reads	4,806,441,796	4,767,971,544
Size (GB)	354.53	351.69
Sequencing Length	150	150
% of Erased Reads	0%	<1%

Paired reads are discarded for the following situations: when one read contains adapter contamination; when one read contains more than 10 percent of uncertain nucleotides; and when one read contains more than 50 percent low quality nucleotides (‘MRNA Sequencing (MRNA-Seq)’, 2023; Yan et al., 2013).

3.5 Differentially Expressed Genes and Functional Annotations for Leaf Tissue Samples

The dispersion plot allowed us to visualize gene abundance and dispersion from the experimental analysis (Fig. 3.5). Differentially expressed genes were determined for all three models at a 5% FDR (Fig. 3.6). In *edgeR*, I obtained 91 differentially expressed genes for Polyclonal vs. Monoclonal, one differentially expressed gene for Clone 110412 Polyclonal vs. Clone 110412 Monoclonal, and 47 differentially expressed genes for Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal (Fig. 3.7). Of the 91, 1, and 47 obtained differentially expressed genes, 69 (76%), 0 (0%), and 31 (66%) gene annotations were available for each contrast matrix through the Joint Genome Institute's Phytozome (Goodstein et al., 2012). Compared to monoclonal plantings, the most highly expressed genes in polyclonal plantings were Potri.015G136400 *DEHYDRATION RESPONSE ELEMENT B1A* (logFC 7.2392, FDR 0.02508, and p-value 7.58e-05), and Potri.002G014000 *PROLYL OLIGOPEPTIDASE FAMILY PROTEIN* (logFC 5.7673, FDR .00683, and p-value <0.001) (Table B.3). In Clone 110412 polyclonal plantings, the only significant differentially expressed gene was Potri.06G219800 (logFC 1.3487, FDR 0.0166, p-value <0.001). No annotations were available for this gene through the Joint Genome Institute database. Because this gene was significantly expressed across all three treatments (Clone S7C8 Polyclonal vs Clone S7C8 Monoclonal logFC 1.3342, FDR 0.003318, p-value <0.001, and Polyclonal vs. Monoclonal logFC 2.683, FDR 1.27e-07, p-value <0.001), both transcript FASTA files were blasted against nucleotide collection (Nucleotide Blast;nt) from the National Center for Biotechnology Information (NCBI) database (Altschul et al., 1990). Sequences producing significant alignments predicted an OXIDATIVE STRESS 3 LIKE 1 (LOC7491986) for *P. trichocarpa* with 100% identity (Altschul et la., 1990) (Fig. 3.8). Compared to Clone S7C8 monoclonal plantings, the most highly expressed gene in Clone S7C8

polyclonal plantings was Potri.002G014000 *PROLYL OLIGOPEPTIDASE FAMILY PROTEIN* (logFC 4.9181, FDR .004802, p-value <0.001) (Fig. 3.9) (Table B.4).

Polyclonal vs. Monoclonal and Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal expressed the same *EXOCYST SUBUNIT EXO70 FAMILY PROTEIN H7* (Polyclonal vs. Monoclonal, Potri.001G234600, logFC: 3.0696, FDR:0.002546, p-value: <0.001). Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal, Potri.001G234600, logFC: 1.7297, FDR: 0.04905, p-value: <0.001), with polyclonal plantings displaying upregulation for both models. Similarly, a multitude of *WRKY DNA-BINDING PROTEINS* were upregulated for Polyclonal vs. Monoclonal and Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal. Polyclonal plantings displayed upregulation of three proteins (Polyclonal vs. Monoclonal, *WRKY DNA-BINDING PROTEIN 60,* Potri.018G019800, logFC: 2.3660, FDR:0.0043, p-value: <0.001, *WRKY DNA-BINDING PROTEIN 33,* Potri.016G128300, logFC:2.6660, FDR: 0.01047, p-value: <0.001, and *WRKY DNA-BINDING PROTEIN 40,* Potri.003G182200, logFC: 1.9186, FDR:0.01719, p-value: <0.001). I also identified down regulation of the *ROOT HAIR SPECIFIC 2* gene for polyclonal plantings (Polyclonal vs. Monoclonal, Potri.019G063500, logFC: -1.6353, FDR: 0.006197, p-value: <0.001). In conjunction with *WRKY DNA-BINDING PROTEINS*, I identified *WRKY FAMILY TRANSCRIPTION FACTORS* (TFs) for Polyclonal vs. Monoclonal and Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal (Polyclonal vs. Monoclonal, *WRKY FAMILY TRANSCRIPTION FACTOR*, Potri.003G138600, logFC:4.4220, FDR:8.846e-05, p-value: <0.001, *WRKY FAMILY TRANSCRIPTION FACTOR*, Potri.014G096200, logFC:2.5211, FDR:0.0009037, p-value: <0.001, *WRKY FAMILY TRANSCRIPTION FACTOR*, Potri.001G092900, logFC:2.2230, FDR:0.01781, p-value: <0.001, Clone S7C8 Polyclonal vs.

Clone S7C8 Monoclonal, *WRKY FAMILY TRANSCRIPTION FACTOR*, Potri.003G138600, logFC:2.2159, FDR:0.02748, p-value: <0.001).

Gene ontology (GO) classified genes into one subcategory of cellular components for Clone S7C8 polyclonal plantings. The GO term was identified as a *PROTON-TRANSPORTING V-TYPE ATPASE, V1 DOMAIN* (GO:0033180, net direction: down, FDR 0.035, p-value 1.8e-05). No additional models with significant FDRs were identified for this study (Table C.1).

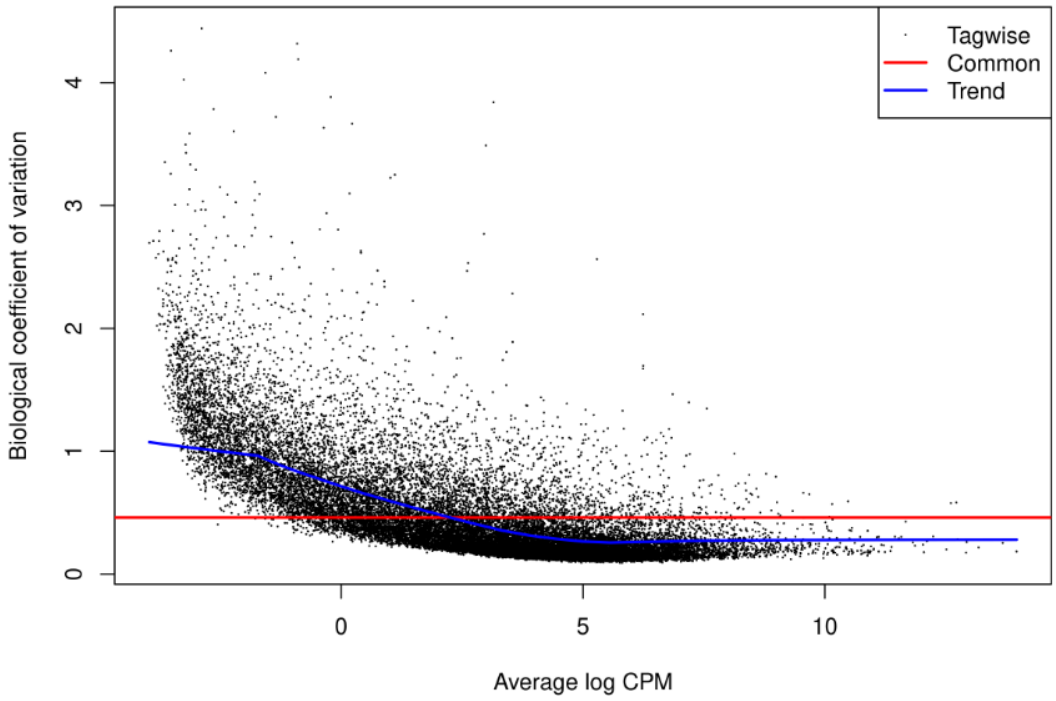


Figure 3.5 Scatterplot of the Biological Coefficient of Variation (BCV) vs. The Average Gene Abundance for Leaf Tissue Samples

Scatterplot of the biological coefficient of variation (BCV) against the average abundance of each gene (log CPM or logarithm of counts per million reads). The tagwise (black dots) represent the BCV if it were calculated for each individual gene. The relative abundance for each gene will vary due to biological causes (Chen et al., 2022). The blue line represents the trend of this dataset. The red line represents the BCV of all samples if a common dispersion value over all genes were used. Negative binomial dispersions tend to be higher for genes with very low read counts, and dispersion trend tends to decrease smoothly to a constant value for genes with larger counts (Chen et al. 2022), allowing the observation for genes that are consistent between biological replicates. A general linearized model was used to determine the significant difference of counts for a transcript across planting schemes.

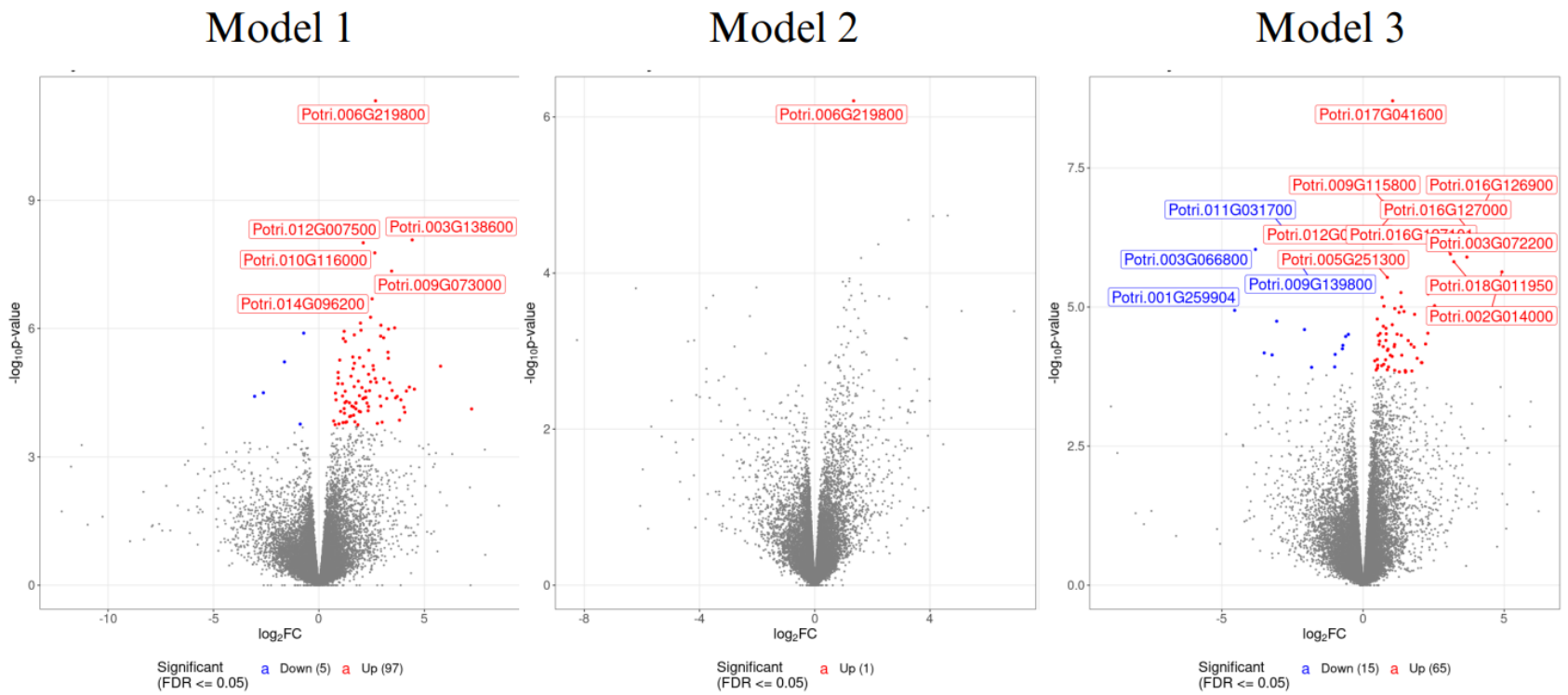


Figure 3.6 P-value vs. log2 Ratio (fold change) for All Three Models from Experimental Analysis

Volcano scatterplots for Model 1: Polyclonal vs. Monoclonal, Model 2: Clone 110412 Polyclonal vs. Clone 110412 Monoclonal and Model 3: Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal displaying statistical significance (P-value) versus the log2 ratio (fold change) for each model at a 5% FDR (graphing scales are not consistent). For each model, the most upregulated genes are highlighted in red towards the right, and the most downregulated genes are highlighted in blue towards the left. The most statistically significant genes are towards the top of the scatterplot. Genes were later filtered further based on logFC restrictions (greater than 1 or less than -1).

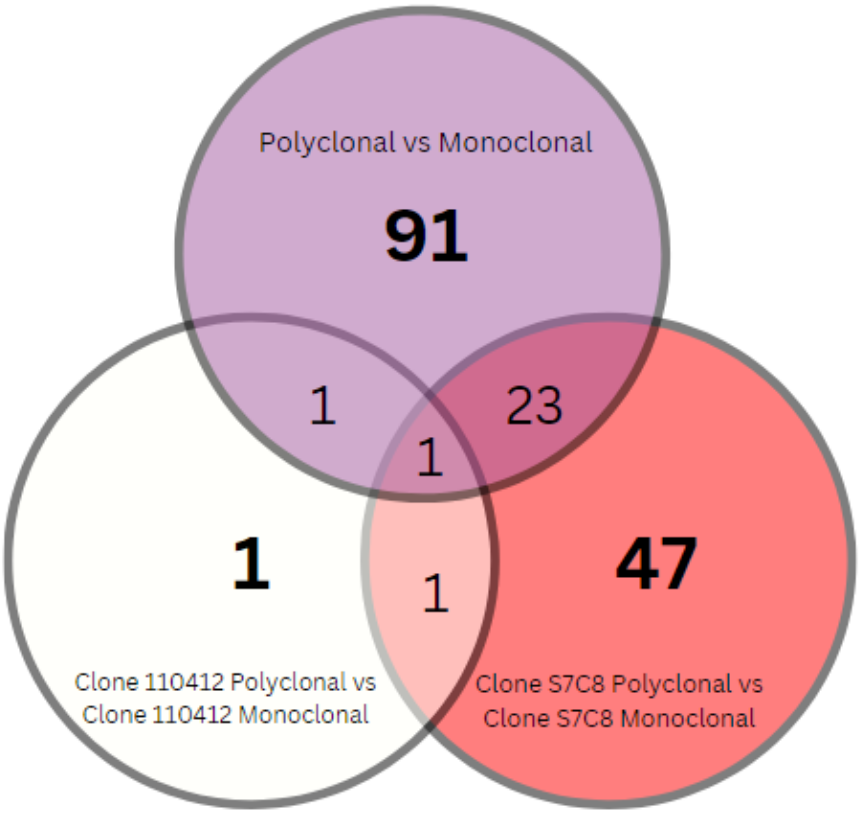


Figure 3.7 Venn Diagram for Differentially Expressed Genes for All Three Models

Venn Diagram displaying the number of differentially expressed genes in leaf tissue samples between planting schemes of *P. deltoides* clonal varieties. Large, bolded numbers indicate the total differentially expressed genes for each model, and smaller numbers indicate common differentially expressed genes between models.

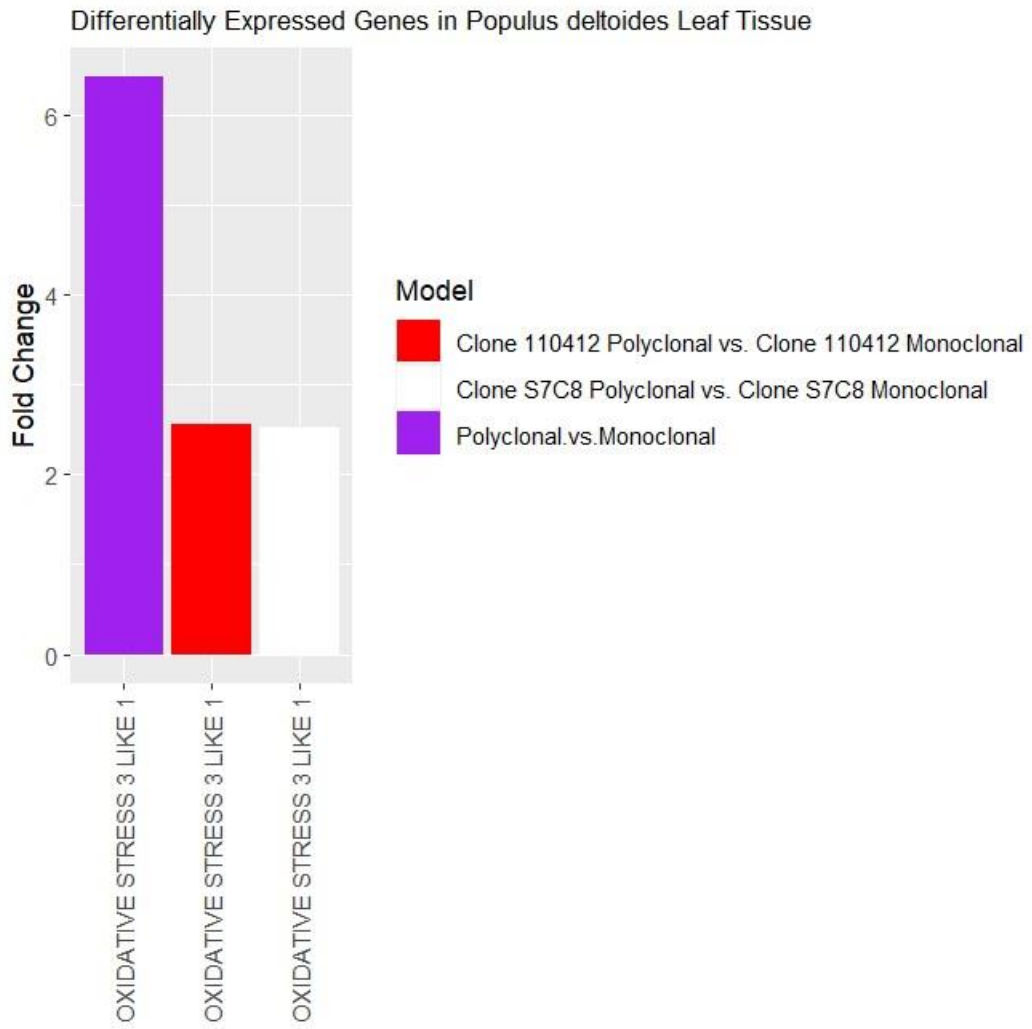


Figure 3.8 Identified OXIDATIVE STRESS 3 LIKE 1 Gene Fold Change Across All Three Models

Graph displays Potri.006G219800 fold change for all three models and planting schemes. Potri.006G219800 was blasted in the National Center for Biotechnology Information database. Sequences producing significant alignments predicted an OXIDATIVE STRESS 3 LIKE 1 gene. All three treatments contained the same locus name and locus name was graphed for each model. Fold change was roughly 2.5 for individual clones and tripled to 6.4 for all polyclonal plantings.

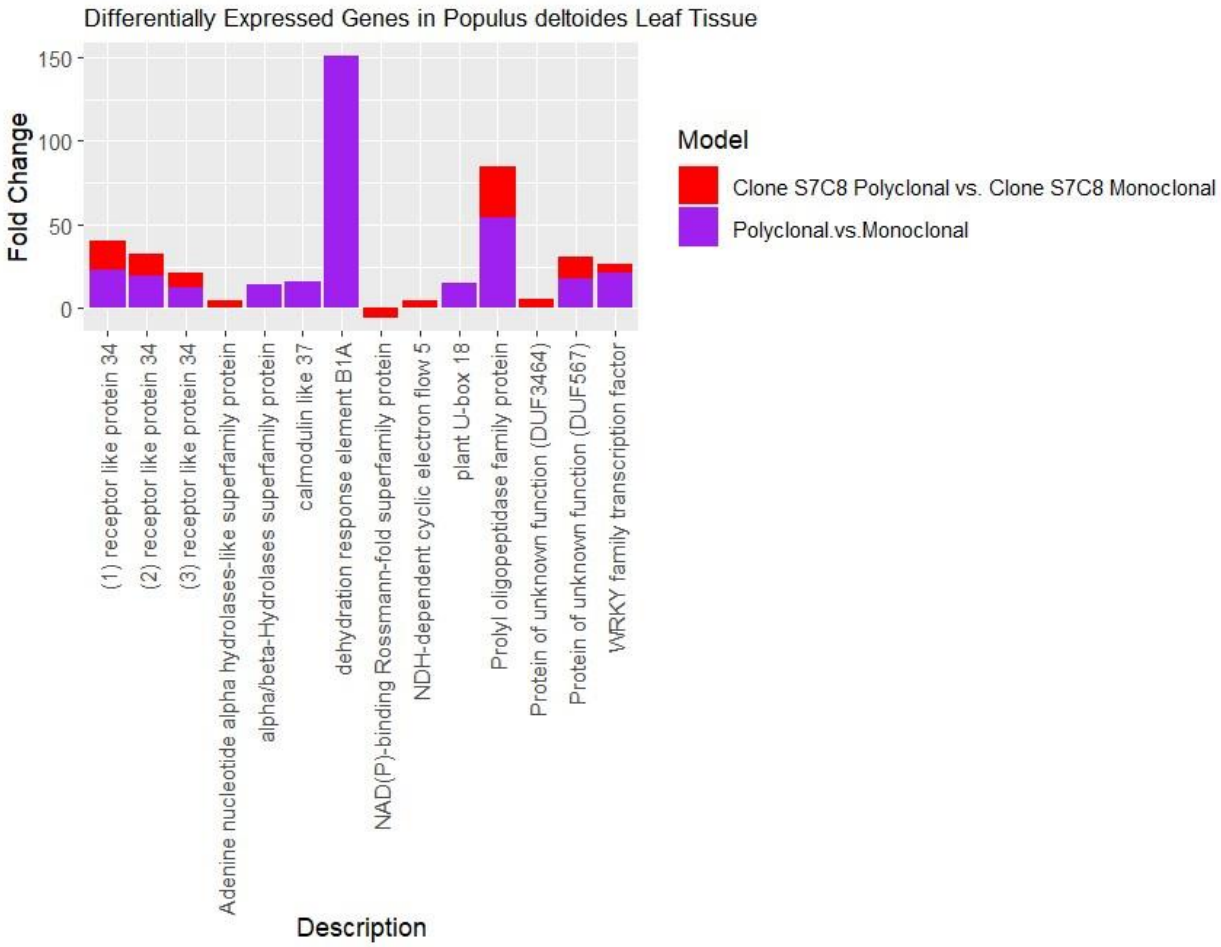


Figure 3.9 Top Ten Differentially Expressed Genes for Model 1: Polyclonal vs. Monoclonal and Model 3: Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal

Graph displays the top ten largest fold changes (from logarithmic scale base 2 (log2)) of differentially expressed genes for Polyclonal vs Monoclonal and Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal. All three receptors like protein 34 had different positions on the chromosome and different locus names ((1) receptor like protein 34-Potri.016G126900, (2) receptor like protein 34-Potri.016G127000, (3) receptor like protein 34-Potri.016G127101).

3.6 Pathway Analysis of Differentially Expressed Genes in Leaf Tissue Samples

The analysis revealed nine enriched KEGG pathways for Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal and no significant pathways for Polyclonal vs. Monoclonal and Clone 110412 Polyclonal vs. Clone 110412 Monoclonal at a 5% FDR. Multiple metabolic pathways were downregulated in Clone S7C8 polyclonal plantings. The most significant downregulated

pathway was arginine and proline metabolism (p-value 0.00033, FDR 0.026). Additional pathways included: tropane, piperidine and pyridine alkaloid biosynthesis, other types of o-glycan biosynthesis, isoquinoline alkaloid biosynthesis, phenylalanine, tyrosine and tryptophan biosynthesis, galactose metabolism, pyruvate metabolism, phenylalanine metabolism, and tyrosine metabolism (Table C.2). Additionally, a mixed KEGG analysis was included to observe the magnitude of the expression change with a pathway, disregarding the direction of genes expressed. No pathways with significant FDRs were identified for all three models for the mixed KEGG analysis (Table C.3).

CHAPTER IV

DISCUSSION

4.1 Planting Schemes Induce Transcriptomic Regulation Processes That Underlie Morphological and Physiological Acclimation in the *P. deltoides* Leaf Transcriptome

Previous global transcriptome analysis studies have identified NIA*2, GDH2 and ASN1* genes linked to metabolism of nitrogen and amino acids, with *NAI2* being previously observed in the roots of other *Populus spp.* (Li et al., 2012; Luo et al., 2015). Although I did not identify specific genes for regulation of nitrate transport and nitrogen metabolism, I did identify genes that influence the growth of the *P. deltoides* and could affect nitrogen metabolism within plants. The LAI results showed greater leaf area for polyclonal plantings. Although, groundwater results displayed similar nitrate concentration averages between planting schemes throughout the growing seasons, results from LAI support that more N was likely taken up by polyclonal plots as greater leaf area likely results in more leaf mass containing more total N in polyclonal plantings even with similar nitrogen percentages.

For both Polyclonal vs. Monoclonal and Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal, I identified upregulation of the *EXOCYST SUBUNIT EXO70 FAMILY PROTEIN H7,* with the gene displaying a larger fold change for polyclonal plantings in the Polyclonal vs. Monoclonal model. The *EXO70* isoform *H7* has been found to be most abundantly expressed in the root maturation zone and in root hairs (Pečenková et al., 2020). Additionally, this specific isoform has been identified in the leaf mesophyll, and where the primary photosynthetic cells in the leaves are located (Pečenková et al., 2020). From previous genome-wide association studies

on *P. trichocarpa*, different proteins within the same family such as Potri.010G250500 *SUBUNIT EXO70 FAMILY PROTEIN G1* have been associated with active growth rate, height and bole and whole-tree mass, and effects of the single nucleotide polymorphism (SNP) linked the allele with substantially greater biomass overall (McKown, Klápště, et al., 2014). Additionally, related *EXOCYST* components, such as, (*EXO70A1)* in *Arabidopsis* mutants showed altered cellular development and organ morphogenesis (Synek et al., 2006). While *Exo70* genes have been previously studied in plants and specific genes (*EXO70A1)* have been identified to play an important role in plant and pollen development (Synek et al., 2006), these associations could possibly link the observed *EXO70 H7* isoform to novel functionality in the *P. deltoides* clonal varieties in regard to increased growing capacity and plasticity within polyclonal plantings.

Although, the *ROOT HAIR SPECIFIC 2* GENE, which controls localized cell growth in plants (Gilroy & Jones, 2000), was downregulated with the gene displaying a larger fold change for polyclonal plantings in the Polyclonal vs. Monoclonal model. Roots hairs are essential for water absorption and nutrient uptake and the highly plastic *Populus* plant root system is able to respond developmentally to nitrogen signaling, enabling colonization into nitrogen-rich patches of the soil, increasing resource usage and nitrate uptake (Liu et al., 2020; Liu et al., 2021). Root hairs have also been shown to play vital roles in plants coping with biotic and abiotic stress (Liu et al., 2021). The contrast in regulation with these two specific genes likely indicates active growth and a potential source of stress in polyclonal plantings. Stress in polyclonal plantings can be further supported by upregulation of the *DEHYDRATION RESPONSE ELEMENT B1A*, which has been linked to stress in *Arabidopsis* (Su et al., 2013), and upregulation of the *OXIDATIVE STRESS 3 LIKE 1* gene across all models.

Although polyclonal plantings most likely were able to cope with this stress due to greater productivity compared to its monoclonal counterparts, the upregulation of a potential active growth gene (*EXO70 H7)*, further supports active plant growth in polyclonal plantings. Interestingly, a previous genome-wide association study in maize found that root hair length dimension was negatively correlated with several morphological and metabolic traits, including plant growth (Liu et al., 2021). This is important as *P. deltoides* polyclonal plantings could be utilizing other features of their root system allowing them to respond developmentally, supporting active plant growth. More significant root hair formation could be occurring within monoclonal plantings as individual clones, such as Clone 110412 with low nitrogen use efficiency, may not be meeting their nutrient needs, forcing colonization into different portions of the soil.

Total leaf area is closely correlated with total biomass and is a determining factor of productivity for *Populus* (Ceulemans, 1990). July is the peak of the growing season for *Populus* leaf formation before leaf senescence occurs (Braatne et al., 1996). As the growing season continues and younger leaves mature, increased photosynthetic capacity should occur due to a greater number of chloroplasts per cell (Bauer & Thöni, 1988). Although I observed greater average LAI between planting schemes for the month of July, further statistical analysis, such as comparing LAI with leaf gas exchange data may better elucidate changes in LAI. However, polyclonal plots had more aboveground biomass than either monoculture, suggesting transgressive overyielding. This, along with the observed greater average LAI between planting schemes could support increased plasticity within polyclonal plantings. Specific genes identified, including a potential novel functioning *EXO70 H7* gene identified in polyclonal plantings, need to be further tested to determine their functionality in *P. deltoides* clones. In addition to the leaf

transcriptome, this idea needs to be further supported by analyzing below ground biomass and gene expression and regulation from the root transcriptome of *P. deltoides*.

Additional genes have been identified to influence the growth of *P. deltoides.* A previous genome-wide association study identified Potri.016G128300 *WRKY33 (WRKY DNA-BINDING PROTEIN 33)* to contain single nucleotide polymorphism (SNP) markers associated with log volume and growth rate within *P. trichocarpa* (McKown, Klápště, et al., 2014). For this study, I identified the same *WRKY DNA-BINDING PROTEIN 33* upregulated for polyclonal plantings for Polyclonal vs. Monoclonal, as well as upregulation of *WRKY DNA-BINDING PROTEIN 60 (WRKY60),* and *WRKY DNA-BINDING PROTEIN 40 (WRKY40).* While *WRKY33* has previously been identified with SNP markers associated with biomass (McKown, Klápště, et al., 2014), upregulation of *WRKY33* has also been previously associated with oxidative stress in *A. thaliana*, as have *WRKY 60,* and *WRKY40* (Jiang et al., 2017). In plants, responses to stress need several signaling molecules, including salicylic acid and abscisic acid (Danquah et al., 2014; Jiang et al., 2014). A genome-wide association study of *Populus WRKY TRANSCRIPTION FACTORS* found *WRKY60* to be significantly affected by salicylic acid, indicating this gene plays an essential role in regulating this specific signaling pathway (Jiang et al., 2014). In addition, the abscisic acid signaling pathway has been identified as a central regulator of abiotic stress in plants, imposing adaptive physiological responses related to defense mechanisms in plants and changes in gene expression (Danquah et al., 2014). A previous study in *Arabidopsis* found that *WRKY40* negatively regulated plant abscisic acid, increasing sensitivity for inhibition of root growth, while *WRKY60* reduced sensitivity to abscisic acid (Chen et al., 2010). *WRKY* 60 has been identified as a “weak” transcriptional activator and *WRKY40* a “weak” transcriptional repressor in *Arabidopsis*, meaning that they have less of an influence on boosting gene transcription or

decreasing transcription compared to other transcriptional factors (Chen et al., 2010). Interestingly, polyclonal plantings were simultaneously upregulated for both *WRKY* activators and suppressors, and other identified differentially expressed genes, such as upregulation of stress-like genes in polyclonal plantings support the idea that both *WRKY60* and *WRKY40* may have a stronger effect on regulating gene expression in *P. deltoides,* in regard to coping with abiotic stress. The downregulation of *ROOT HAIR SPECIFIC 2* GENE supports the idea of *WRKY40* inhibiting root growth in *P. deltoides* polyclonal plantings. Inhibition of root growth belowground does not necessarily negatively impact biomass production, as polyclonal plantings in this study were overyielding. For this study, only aboveground biomass was estimated. An important consideration is that polyclonal plantings may be allocating more biomass aboveground and less belowground. In regard to *WRKY60,* there has been growing evidence that salicylic acid is an important growth regulator in plants, and salicylic acid accumulation in plants provides protection in plants against abiotic stresses (Khan et al., 2015). *P. deltoides* plantings may display an increase in tolerance to stress, specifically abiotic stress in polyclonal plantings from the upregulation of *WRKY60*, and potential changes to salicylic acid accumulation. As previously mentioned, the upregulation of *EXOCYST SUBUNIT EXO70 FAMILY PROTEIN H7,* which has previously been abundantly expressed in the root maturation zone and in root hairs, identified in photosynthetic cells of leaf tissue (Pečenková et al., 2020), and has been previously associated with active growth rate in *P. trichocarpa,*(McKown, Guy, et al., 2014), would support this idea as polyclonal plantings continue to develop, and cope with stress. *WRKY 60* and *WRKY40* may strongly regulate gene expression, specifically genes related to stress and growth within *P.deltoides* clonal varieties. However, further gene testing is needed to better understand their individual functions in *P. deltoides*.

Other *WRKY FAMILY TRANSCRIPTION FACTORS* (TF*s)* were identified to be upregulated for polyclonal plantings in Polyclonal vs. Monoclonal and Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal. I identified three different TFs for Polyclonal vs. Monoclonal (Potri.003G138600, Potri.014G096200, and Potri.001G092900) and one for Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal (Potri.003G138600). As one of the largest families of TFs in higher plants, *WRKY* plays an essential role in modulating plant growth, development and responses to stress (Lv et al., 2020). Oxidative stress, including the excessive production of hydrogen peroxide (H_2O_2), occurs when plants are exposed to light stress, high light intensities, and high temperatures (Inzé & Montagu, 1995; Jiang et al., 2017). From the results, oxidative stress, such as water stress in polyclonal plantings, could be occurring due to belowground moisture stress from high competition from water. Additionally, I observed upregulation of the *DEHYDRATION RESPONSE ELEMENT B1A* for polyclonal plantings in Polyclonal vs. Monoclonal. The abscisic acid (ABA) hormone and independent pathway plays a key role in acclimation to stresses during vegetative development in plants and includes members of the *DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN (DREB)* family (Su et al., 2013). *DEHYDRATION RESPONSE ELEMENT B1A* has been identified to play an important function in drought response in flowers in *Arabidopsis* (Su et al., 2013). Water stress in polyclonal plantings of *P. deltoides* would induce cellular dehydration. A strategy to avoid dehydration stress and to protect the plant's cells is to increase the levels of H_2O_2 (Sun & Yu, 2015). A drastic change in expression levels for *DEHYDRATION RESPONSE ELEMENT B1A*, and *WRKY* TFs within polyclonal plantings indicates a stress-signaling response and multiple potential stress coping mechanisms within polyclonal plantings. Additionally, the upregulation of the *PROLYL OLIGOPEPTIDASE FAMILY PROTEIN* was identified. A previous study in

Escherichia coli identified a gene in a member of the *PROLYL OLIGOPEPTIDASE FAMILY PROTEIN* to play an important role in enhancing the tolerance of *E. coli* to high salinity, high temperature and simulated drought (Tan et al., 2013). The *PROLYL OLIGOPEPTIDASE FAMILY PROTEIN* had the second largest fold change in Polyclonal vs. Monoclonal , with the *DEHYDRATION RESPONSE ELEMENT B1A* displaying the largest fold change. *PROLYL OLIGOPEPTIDASE FAMILY PROTEIN* may be an additional gene for future testing to see if it plays similar roles in stress tolerance in *P. deltoides.*

4.2 Planting Schemes of *P. deltoides* Influence Metabolic and Biosynthesis Pathways in Response to Water and Heat Stress

Although I only identified one significant functional annotation from the GO analysis, I identified downregulation of five metabolic pathways for Clone S7C8 polyclonal plantings: arginine and proline metabolism, galactose metabolism, pyruvate metabolism, phenylalanine metabolism, and tyrosine metabolism from the KEGG enrichment pathway analysis, and four biosynthesis processes: tropane, piperdine, and pyridine alkaloid biosynthesis, o-glycan biosynthesis, isoquinoline alkaloid biosynthesis, and phenylalanine, tyrosine and tryptophan biosynthesis. Amino acids such as arginine and proline are synthesized during abiotic stress to act as precursors for secondary metabolites or storage forms of organic nitrogen (Baumberg & Klingel, 1993; Hildebrandt, 2018). The metabolite proline is produced from the reallocation of nitrogen under stress conditions in plants (Majumdar et al., 2016). Proline has been shown to accumulate in plants experiencing water limitation (Verslues & Sharma, 2010). Arginine is a major storage and transport form for organic nitrogen in plants and serves as a precursor for nitric oxide, an essential metabolite for developmental processes (Majumdar et al., 2016; Winter et al., 2015). Similarly, arginine and proline metabolic pathways, including galactose metabolism

have been previously shown to be significantly impacted by heat stress (Ren et al., 2019) Like water stress, heat stress increases the production of H_2O_2, and proline accumulation (Ren et al., 2019). Heat and water stress could potentially correlate as transpiration cools leaves down, with decreased transpiration leading to heat stress. Although arginine and proline metabolism are downregulated in Clone S7C8 polyclonal plantings, the enhanced upregulation of this secondary metabolic pathway in Clone S7C8 monoclonal plantings could signify greater sensitivity to abiotic stress, specifically, water and heat stress. Genetically homogeneous populations, such as monoclonal plantations are typically more susceptible to pathogens, pests and stress due to low genetic diversity, making infection more likely to spread between the similar hosts, and inbreeding of individuals with low genetic heterozygosity could increase host susceptibility to disease and stress (King & Lively, 2012; Richards et al., 2010). Plants respond to stress through the regulation of secondary metabolism for the direct cessation of stress factors (Meraj et al., 2020). Heat stress has been known to significantly impair plants' physiological reactions, reducing biomass production and yield (Ferus et al., 2020). Similarly, photosynthesis is known to be a heat sensitive process, and can be completely inhibited by high temperatures (Zhou et al., 2010). Biomass estimates from experimental data support the idea of greater sensitivity in Clone S7C8 monoclonal plantings to abiotic stress as Clone S7C8 monoclonal plantings show reduced biomass production compared to Clone S7C8 polyclonal plantings. Although identified in the enrichment pathway analysis but not significant at the FDR level, the upregulation in biosynthesis of secondary metabolites (p-value: <0.001, FDR: 0.085) in Clone S7C8 monoclonal plantings could further support enhanced secondary metabolism in monoclonal plantings but additional testing is needed. Further, the enhanced upregulation of secondary metabolism of monoclonal plantings in the Polyclonal vs. Monoclonal model was not identified, but this could

be due to timing of sample collection, as samples were flash frozen, only capturing the transcriptome at a specific timepoint, possibly leaving out transcripts to show enhanced secondary metabolism for this model (Lowe et al., 2017). Similarly, the combination of Clone S7C8 and Clone 110412 in polyclonal plantings could affect gene expression of secondary metabolism at the pathway level.

In Clone S7C8 polyclonal plantings, the upregulation of oxidative stress genes and the downregulation of arginine, proline and pyruvate metabolism could indicate a transcriptomic response to maintain cellular homeostasis imposed by heat and water stress. One study found a dramatic upregulation of gene expression for *DREB1* or *DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN* in response to heat stress, and proline was increased, and pyruvate was decreased by heat stress (Ren et al., 2019). From the experimental analysis, I found *DEHYDRATION RESPONSE ELEMENT B1A* was the highest upregulated gene in polyclonal plantings in Polyclonal vs. Monoclonal, and arginine and proline metabolism were decreased in Clone S7C8 polyclonal plantings. Pyruvate is a citric acid cycle (TCA) metabolite. Although identified in the enrichment pathway analysis but not significant, pyruvate metabolism and citric acid cycle were downregulated for polyclonal plantings in the Polyclonal vs. Monoclonal model. The decrease in these pathways would support the response to heat stress identified by Ren et al., 2019. Although most notably, I saw contrasting responses for Clone S7C8 polyclonal plantings as proline and arginine metabolism was reduced, which could imply Clone S7C8 polyclonal plantings were better suited in coping with heat stress. This could potentially be the case for polyclonal plantings in the Polyclonal vs. Monoclonal model, although further testing on enrichment analysis, specifically primary metabolic pathways, such as pyruvate metabolism, and secondary metabolic pathways and their relationship to abiotic stress in *P. deltoides* is needed.

Downregulation of phenylalanine metabolism and phenylalanine, tyrosine and tryptophan biosynthesis were also identified within Clone S7C8 polyclonal plantings. Phenylpropanoid metabolism starts with the deamination of phenylalanine to eventually yield *p*-coumaroyl CoA, which is intermediately involved in the biosynthesis of secondary metabolites (Ma et al., 2018). Phenylalanine plays an essential role in the phenylpropanoid pathway and derived compounds such as the abundance of phenolic glycosides can influence tree growth and are a significant determinant of *Populus* productivity (Tsai et al., 2006). One study found that younger trees tend to have higher concentrations of phenolic glycosides and phenolic glycosides were the most abundant secondary metabolite in the foliage of Salicaceae (Boeckler et al., 2011). Phenolic glycosides have been shown to provide protection against pathogens, pests, and abiotic stress in clones of *P. tremuloides* (Lindroth & Hwang, 1996). Secondary metabolic pathways produce many compounds that are necessary in plants for adaptation to their environment (Movahedi et al., 2021). The upregulation of phenylalanine metabolism and biosynthesis in Clone S7C8 monoclonal plantings, but not in Clone 110412 may indicate Clone S7C8 was better adapted to stress. Similarly, Clone S7C8 displayed greater estimated aboveground biomass than Clone 110412. Clone S7C8 is a high nitrogen use efficiency clone, requiring less nitrogen for growth compared to Clone 110412. The upregulation the *Exo70 H7* active growth gene within this clone specifically, as well as secondary metabolism including phenylalanine metabolism and biosynthesis at the pathway level analysis supports why I witnessed increased estimated aboveground biomass in this clone. Other identified secondary metabolic pathways, for example, tyrosine metabolism has been shown to increase as a result of salt stress in *P. simonii* (Meng et al., 2016). Downregulation of other secondary metabolites in Clone S7C8 polyclonal plantings were also identified, including nitrogenous isoquinoline alkaloid biosynthesis, o-glycan

biosynthesis and tropane, piperdine and pyridine alkaloid biosynthesis. Alkaloids and phenolics do not play an active role in growth and development of plants but are concerned with defense and interaction with other organisms, including other plants (Inayat et al., 2020). Downregulation in Clone S7C8 polyclonal plantings could be due to synergetic effects between Clone 110412, although this pathway was not downregulated in the Polyclonal vs. Monoclonal model and would require additional analysis. Similarly, the upregulation in Clone S7C8 monoclonal plantings could indicate antagonistic effects between individuals of Clone S7C8. However, no significant pathways were identified for the Polyclonal vs. Monoclonal model or for the Clone 110412 Polyclonal vs. Clone 110412 Monoclonal model and would require additional transcriptomic and pathway analysis to better understand why these metabolic pathways are elucidated at the pathway level.

The overall objective states that differentially expressed genes are an underlying molecular mechanism that may explain differences in productivity and resource use between monoclonal and polyclonal plantings, through morphological and physiological plasticity was supported by the results. The hypothesis that differential gene expression (DGE) is an underlying molecular mechanism contributing to overyielding (polycultures are more productive than the average of their monocultures), increased resource utilization, and increased stress tolerance in polycultures through regulation of morphology and physiology was supported by the results. For this project, I observed greater estimated aboveground biomass for polyclonal plantings and an increase in gene regulation for an active growth gene in polyclonal plantings. Similarly, I identified differentially expressed genes that were associated with morphological and physiological plasticity for *Populus spp.*, with polyclonal plantings displaying an increase in gene regulation (decrease in monoclonal plantings) for oxidative stress genes, and active growth

genes *(EXO70 H7),* and a decrease in gene regulation (increase in monoclonal plantings) for enrichment secondary metabolic pathways associated with plant defense and stress responses. Although stress genes were upregulated in polyclonal plantings, these genes did not interact with or did not have a large enough influence to display a stress response at the pathway level in polyclonal plantings. Although polyclonal plantings displayed upregulation for abiotic stress, previous literature suggests that monoclonal plantings were more sensitive to abiotic stress, affecting productivity (King & Lively, 2012; Richards et al., 2010). While most identified SNP markers associated with biomass traits were identified in *P. trichocarpa* (Chhetri et al., 2019; Fahrenkrog et al., 2017; McKown, Klápště, et al., 2014), *P. trichocarpa* has been extensively studied and represents a phenotypically diverse genus (Tuskan et al., 2004). Similar family groups of genes identified in the genome wide association study for *P. trichocarpa* and the same genes with matching locus names were identified within our *P. deltoides* planting schemes. Although some genes specific functionality in *P.deltoides,* such as the *Exo70 H7, WRKY 40,* and *WRKY60* genes, along with enrichment primary and secondary metabolic pathways require further testing to fully understand the role they play in different planting schemes.

While it can be more difficult to identify phenotypic traits in field trials of *Populus spp.,* transcriptomic approaches have greatly contributed to the understanding of plasticity effects in *Populus* (Fahrenkrog et al., 2017; Lv et al., 2020; McKown, Klápště, et al., 2014; Meraj et al., 2020; Ren et al., 2019). Transcriptomic analysis is a tool to help researchers better understand genotypic and phenotypic variation in different tissues, conditions or in time within an organism (Lowe et al., 2017). Although this study is incredibly informative for preliminary understanding of biological processes within *P. deltoides* planting schemes, a drawback is the transcriptome only captures a snapshot in time of the total transcripts present within a cell (Lowe et al., 2017),

providing information only in a distinct timeframe, with the ultimate outcome in protein expression varying. Certain mRNA transcripts that are still being transcribed may not be captured at that timepoint, while others have been fully degraded. For this study, we sampled in the morning, although sampling at different time points throughout the day may elucidate different mRNA transcripts, providing more information on gene expression effects for the day sampling occurred. Although we minimized temporally related gene expression changes with a small timeframe for sample collection, samples collected from the first and last plots had the largest time difference, and this could potentially affect gene expression output. This could be one explanation as to why I am observing similar gene expression patterns but different enriched pathways within models. However, comparative analysis of transcriptomics and proteomics across multiple tissue types in *P. deltoides* will provide a panoramic view of gene expression and regulation, uncovering new results that may have been missed in my preliminary findings (Hu et al., 2023).

4.3 Implications

Since field varieties of *Populus spp.* produce diverse phenotypes (Brunner et al., 2004), transcriptomic analysis informs selection of genes, or candidate genes for specific traits, such as *EXOCYST SUBUNIT EXO70 FAMILY PROTEIN H7, WRKY40,* and *WRKY60* for increased productivity and tolerance to oxidative stress. For this study, polyclonal plantings were more productive than the average of each monoculture (overyielding) and more productive than either monoculture (transgressive overyielding). Testing of potential candidate genes identified from this study is essential, as this may help researchers, breeders, and silviculturists better select traits in *P. deltoides* that are not only more productive, but more resilient to climate change.

CHAPTER V

CONCLUSIONS

5.1 Final Conclusions

The research presented here suggested that *WRKY DNA-Binding proteins* and TFs may play an important role in stress responses, specifically oxidative stress including water and heat stress in *P. deltoides* clonal varieties. Additionally, the downregulation of secondary metabolic pathways has been previously associated with heat stress. *PROLYL OLIGOPEPTIDASE FAMILY PROTEIN*, *EXOCYST SUBUNIT EXO70 FAMILY PROTEIN H7*, OXIDATIVE STRESS 3 LIKE 1, *DEHYDRATION RESPONSE ELEMENT B1A, WRKY40*, and *WRKY60* are all potential candidate genes related to productivity, water and heat stress that should be further tested in *P. deltoides* to identify their specific functions. The leaf transcriptome raw dataset produced by this study will be made publicly available, further promoting research, helping geneticists, trees breeders, and silviculturists make informed decisions for future field trials of *P. deltoides* clonal varieties.

5.2 Future Directions

This study sheds light on new preliminary transcriptomic data and genomic information for *P. deltoides* clonal varieties. Although this study only provides one transcriptome in one tissue type for this study, omics technologies allow for individual experiments to be normalized across full databases, allowing researchers to compare gene expression across diverse experiments. Because of this, a potential direction that may arise from this project is larger

transcriptomic analysis of multiple clonal varieties of *Populus spp.*, including hybrid poplars. In addition, new spatial transcriptomic approaches, such as single-cell RNA sequencing for gene expression profiling, can explore genotype-phenotype relationships at the cellular level, unveiling new information on pathway level changes (Dong & Chen, 2013). This study is important as it highlights early gene expression effects and gene regulatory networks, providing a supplement approach for researchers to select traits in for future *P. deltoides* plantings that are more productive, more tolerant to stress and more resilient to climate change.

REFERENCES

Aban, I. B., Cutter, G. R., & Mavinga, N. (2008). Inferences and Power Analysis Concerning Two Negative Binomial Distributions with an Application to MRI Lesion Counts Data. *Computational Statistics & Data Analysis*, *53*(3), Article 3. https://doi.org/10.1016/j.csda.2008.07.034

Agilent Technologies, Inc. (2023). *RNA Electrophoresis, RNA Integrity, Bioanalyzer RIN | Agilent*. https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-rna-kits-reagents/bioanalyzer-rna-analysis-228256

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., & Robinson, M. D. (2013). Count-Based Differential Expression Analysis of RNA Sequencing Data Using R and Bioconductor. *Nature Protocols*, *8*(9), 1765–1786. https://doi.org/10.1038/nprot.2013.099

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the Unification of Biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Babu, S., & Gassmann, M. (2016). *Assessing Integrity of Plant RNA with the Agilent 2100 Bioanalyzer System*. Agilent Technologies, Inc. https://www.agilent.com/cs/library/applications/5990-8850EN.pdf

Bauer, H., & Thöni, W. (1988). Photosynthetic Light Acclimation in Fully Developed Leaves of the Juvenile and Adult Life Phases of *Hedera helix*. *Physiologia Plantarum*, *73*(1), 31–37.

Baumberg, S., & Klingel, U. (1993). Biosynthesis of Arginine, Proline, and Related Compounds in *Bacillus subtilis* and Other Gram-Positive Bacteria (pp. 299–306). John Wiley & Sons, Ltd. https://doi.org/10.1128/9781555818388.ch21

Boeckler, G. A., Gershenzon, J., & Unsicker, S. B. (2011). Phenolic Glycosides of the Salicaceae and Their Role as Anti-Herbivore Defenses. *Phytochemistry*, *72*(13), 1497–1509. https://doi.org/10.1016/j.phytochem.2011.01.038

Bita, C. E., & Gerats, T. (2013). Plant Tolerance to High Temperature in a Changing Environment: Scientific Fundamentals and Production of Heat Stress-tolerant Crops. *Frontiers in Plant Science*, *4*, 273.

Braatne, J. H., Rood, S. B., & Heilman, P. E. (1996). Life History, Ecology, and Conservation of Riparian Cottonwoods in North America. *Biology of Populus and Its Implications for Management and Conservation, Part I*, 57–85.

Bradshaw, H. D., Ceulemans, R., Davis, J., & Stettler, R. (2000). Emerging Model Systems in Plant Biology: Poplar (*Populus*) as a Model Forest Tree. *Journal of Plant Growth Regulation*, *19*(3), 306–313.

Brunner, A. M., Busov, V. B., & Strauss, S. H. (2004). Poplar Genome Sequence: Functional Genomics in an Ecologically Dominant Plant Species. *Trends in Plant Science*, *9*(1), 49–56. https://doi.org/10.1016/j.tplants.2003.11.006

Bunn, S. M. (2004). Leaf-level Productivity Traits in *Populus* Grown in Short Rotation Coppice for Biomass Energy. *Forestry*, *77*(4), 307–323. https://doi.org/10.1093/forestry/77.4.307

Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., the AmiGO Hub, & the Web Presence Working Group. (2009). AmiGO: Online Access to Ontology and Annotation Data. *Bioinformatics*, *25*(2), 288–289. https://doi.org/10.1093/bioinformatics/btn615

Carlson, M., Falcon, S., Pages, H., & Li, N. (2019). GO. db: A Set of Annotation Maps Describing the Entire Gene Ontology. *R Package Version*, *3*(2), 10–18129.

CD Genomics. (2023). *KEGG Pathway Enrichment Analysis*. https://bmb.cd-genomics.com/kegg-pathway-enrichment-analysis.html

Ceulemans, R. (1990). *Genetic Variation in Functional and Structural Productivity Determinants in Poplar*. Master's thesis, University of Antwerp, Dept. of Biology.

Chen, H., Lai, Z., Shi, J., Xiao, Y., Chen, Z., & Xu, X. (2010). Roles of *Arabidopsis* WRKY18, WRKY40 and WRKY60 Transcription Factors in Plant Responses to Abscisic Acid and Abiotic Stress. *BMC Plant Biology*, *10*(1), 281. https://doi.org/10.1186/1471-2229-10-281

Chen, J., Yin, W., & Xia, X. (2014). Transcriptome Profiles of *Populus euphratica* upon Heat Shock stress. *Current Genomics*, *15*(5), 326–340. https://doi.org/10.2174/1389202915051411061018 35

Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From Reads to Genes to Pathways: Differential Expression Analysis of RNA-Seq Experiments Using Rsubread and the edgeR Quasi-Likelihood Pipeline. *F1000Research*, *5*, 1438. https://doi.org/10.12688/f1000research.8987.2

Chen, Y., Lun, A. T., McCarthy, D. J., Ritchie, M. E., Phipson, B., Hu, Y., Zhou, X., Robinson, M. D., & Smyth, G. K. (2023). *edgeR: Empirical Analysis of Digital Gene Expression Data in R* (3.40.2). Bioconductor version: Release (3.16). https://doi.org/10.18129/B9.bioc.edgeR

Chen, Y., McCarthy, D., Ritchie, M., Robinson, M., & Smyth, G. (2022). *EdgeR: Differential Analysis of Sequence Read Count Data User's Guide*.

Chhetri, H. B., Macaya-Sanz, D., Kainer, D., Biswal, A. K., Evans, L. M., Chen, J.-G., Collins, C., Hunt, K., Mohanty, S. S., Rosenstiel, T., Ryno, D., Winkeler, K., Yang, X., Jacobson, D., Mohnen, D., Muchero, W., Strauss, S. H., Tschaplinski, T. J., Tuskan, G. A., & DiFazio, S. P. (2019). Multitrait Genome-wide Association Analysis of *Populus trichocarpa* Identifies Key Polymorphisms Controlling Morphological and Physiological Traits. *New Phytologist*, *223*(1), 293–309. https://doi.org/10.1111/nph.15777

Compton, J. E., Harrison, J. A., Dennis, R. L., Greaver, T. L., Hill, B. H., Jordan, S. J., Walker, H., & Campbell, H. V. (2011). Ecosystem Services Altered by Human Changes in the Nitrogen Cycle: A New Perspective for US Decision Making. *Ecology Letters*, *14*(8), 804–815. https://doi.org/10.1111/j.1461-0248.2011.01631.x

Dahal, B., Poudel, K. P., Renninger, H. J., Granger, J. J., Leininger, T. D., Gardiner, E. S., Souter, R. A., & Rousseau, R. J. (2022). Aboveground Biomass Equations for Black Willow (*Salix nigra* Marsh.) and Eastern Cottonwood (*Populus deltoides* Bartr. Ex Marsh.). *Trees, Forests and People*, *7*, 100195. https://doi.org/10.1016/j.tfp.2022.100195

Danquah, A., de Zelicourt, A., Colcombet, J., & Hirt, H. (2014). The Role of ABA and MAPK Signaling Pathways in Plant Abiotic Stress Responses. *Biotechnology Advances*, *32*(1), 40–52. https://doi.org/10.1016/j.biotechadv.2013.09.006

Day-Richter, J., Harris, M. A., Haendel, M., The Gene Ontology OBO-Edit Working Group, & Lewis, S. (2007). OBO-Edit—An Ontology Editor for Biologists. *Bioinformatics*, *23*(16), 2198–2200. https://doi.org/10.1093/bioinformatics/btm112

de Klerk, E., & 't Hoen, P. A. C. (2015). Alternative mRNA Transcription, Processing, and Translation: Insights from RNA Sequencing. *Trends in Genetics*, *31*(3), 128–139. https://doi.org/10.1016/j.tig.2015.01.001

Dong, Z., & Chen, Y. (2013). Transcriptomics: Advances and Approaches. *Science China Life Sciences*, *56*(10), 960–967. https://doi.org/10.1007/s11427-013-4557-2

Duan, B., Dong, T., Zhang, X., Zhang, Y., & Chen, J. (2014). Ecophysiological Responses of Two Dominant Subalpine Tree Species *Betula albo-sinensis* and *Abies faxoniana* to Intra- and Interspecific Competition Under Elevated Temperature. *Forest Ecology and Management*, *323*, 20–27. https://doi.org/10.1016/j.foreco.2014.03.036

Erskine, P. D., Lamb, D., & Bristow, M. (2006). Tree Species Diversity and Ecosystem Function: Can Tropical Multi-Species Plantations Generate Greater Productivity? *Forest Ecology and Management*, *233*(2), 205–210. https://doi.org/10.1016/j.foreco.2006.05.013

Fahrenkrog, A. M., Neves, L. G., Resende, M. F. R., Vazquez, A. I., Campos, G. de los, Dervinis, C., Sykes, R., Davis, M., Davenport, R., Barbazuk, W. B., & Kirst, M. (2017). Genome-wide Association Study Reveals Putative Regulators of Bioenergy Traits in *Populus deltoides*. *New Phytologist*, *213*(2), Article 2. https://doi.org/10.1111/nph.14154

Farraji, H., Zaman, N., Tajuddin, R., & Faraji, H. (2016). Advantages and Disadvantages of Phytoremediation a Concise Review. *International Journal of Environmental & Technological Science*, *2*, 69–75.

Ferus, P., Hniličkа, F., Hniličková, H., Kurjak, D., Kmeť, J., Otepka, P., Gubiš, J., Havrlentová, M., Malbeck, J., & Konôpková, J. (2020). Productivity and Heat-Stress Tolerance in Canadian Poplar (*Populus* × *canadensis* Moench) Clones with Different Ecological Optimum. *Biomass and Bioenergy*, *138*, 105605. https://doi.org/10.1016/j.biombioe.2020.105605

Forrester, D. I., Bauhus, J., & Cowie, A. L. (2005). Nutrient Cycling in a Mixed-Species Plantation of *Eucalyptus globulus* and *Acacia mearnsii*. *Canadian Journal of Forest Research*, *35*(12), 2942–2950. https://doi.org/10.1139/x05-214

Forrester, D. I., Bauhus, J., Cowie, A. L., & Vanclay, J. K. (2006). Mixed-Species Plantations of Eucalyptus with Nitrogen-Fixing Trees: A Review. *Forest Ecology and Management*, *233*(2), 211–230. https://doi.org/10.1016/j.foreco.2006.05.012

Fu, X., Fu, N., Guo, S., Yan, Z., Xu, Y., Hu, H., Menzel, C., Chen, W., Li, Y., Zeng, R., & Khaitovich, P. (2009). Estimating Accuracy of RNA-Seq and Microarrays with Proteomics. *BMC Genomics*, *10*(1), 161. https://doi.org/10.1186/1471-2164-10-161

Gail, M. (1974). Power Computations for Designing Comparative Poisson Trials. *Biometrics*, *30*(2), Article 2. https://doi.org/10.2307/2529645

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M. C., Fröberg, M., Stendahl, J., Philipson, C. D., Mikusiński, G., Andersson, E., Westerlund, B., Andrén, H., Moberg, F., Moen, J., & Bengtsson, J. (2013). Higher Levels of Multiple Ecosystem Services are Found in Forests with More Tree Species. *Nature Communications*, *4*(1), 1340. https://doi.org/10.1038/ncomms2328

Gene Ontology overview. (n.d.). Gene Ontology Resource. Retrieved 28 April 2023, from http://geneontology.org/docs/ontology-documentation/

Gilroy, S., & Jones, D. L. (2000). Through Form to Function: Root Hair Development and Nutrient Uptake. *Trends in Plant Science*, *5*(2), 56–60. https://doi.org/10.1016/S1360-1385(99)01551-4

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A Comparative Platform for Green Plant Genomics. *Nucleic Acids Research*, *40*(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Han, X., An, Y., Zhou, Y., Liu, C., Yin, W., & Xia, X. (2020). Comparative Transcriptome Analyses Define Genes and Gene Modules Differing Between Two *Populus* Genotypes with Contrasting Stem Growth Rates. *Biotechnology for Biofuels*, *13*(1), 139. https://doi.org/10.1186/s13068-020-01758-0

Hefting, M., Beltman, B., Karssenberg, D., Rebel, K., van Riessen, M., & Spijker, M. (2006). Water Quality Dynamics and Hydrology in Nitrate Loaded Riparian Zones in the Netherlands. *Environmental Pollution*, *139*(1), 143–156. https://doi.org/10.1016/j.envpol.2005.04.023

Hefting, M., Clement, J.-C., Bienkowski, P., Dowrick, D., Guenat, C., Butturini, A., Topa, S., Pinay, G., & Verhoeven, J. T. A. (2005). The Role of Vegetation and Litter in the Nitrogen Dynamics of Riparian Buffer Zones in Europe. *Ecological Engineering*, *24*(5), 465–482. https://doi.org/10.1016/j.ecoleng.2005.01.003

Hildebrandt, T. M. (2018). Synthesis Versus Degradation: Directions of Amino Acid Metabolism During *Arabidopsis* Abiotic Stress Response. *Plant Molecular Biology*, *98*(1–2), 121–135. https://doi.org/10.1007/s11103-018-0767-0

Hong, G., Zhang, W., Li, H., Shen, X., & Guo, Z. (2014). Separate Enrichment Analysis of Pathways for Up- And Downregulated genes. *Journal of the Royal Society, Interface / the Royal Society*, *11*, 20130950. https://doi.org/10.1098/rsif.2013.0950

Hu, Y., Zhang, Y., Šmarda, P., Bureš, P., & Guo, Q. (2023). Transcriptome and Proteome Associated Analysis of Flavonoid Metabolism in Haploid *Ginkgo biloba*. *International Journal of Biological Macromolecules*, *224*, 306–318. https://doi.org/10.1016/j.ijbiomac.2022.10.125

Illumina | Sequencing and Array-Based Solutions for Genetic Research. (n.d.). Retrieved 15 April 2023, from https://www.illumina.com/

Inayat, N., Zahir Muhammad, R., & Majeed, A. (2020). 90. Phytochemical Screening and Allelopathic Evaluation of Aqueous and Methanolic Leaf Extracts of *Populus nigra* L. *Pure and Applied Biology (PAB)*, *9*(1), 956–962.

Inzé, D., & Montagu, M. V. (1995). Oxidative Stress in Plants. *Current Opinion in Biotechnology*, *6*(2), 153–158. https://doi.org/10.1016/0958-1669(95)80024-7

Izawa, T. (2015). Deciphering and Prediction of Plant Dynamics Under Field Conditions. *Current Opinion in Plant Biology*, *24*, 87–92. https://doi.org/10.1016/j.pbi.2015.02.003

Jansson, S., & Douglas, C. (2007). *Populus*: A Model System for Plant Biology. *Annual Review of Plant Biology*, *58*, 435–458. https://doi.org/10.1146/annurev.arplant.58.032806.103956

Jeffreys, J. (2005). *Performance of Selected Eastern Cottonwood Clones from the Southeastern United States through Two Years of Age on Two Sites,* Master's thesis, Mississippi State University, Dept. of Forestry.

Jia, H., Wang, L., Li, J., Sun, P., Lu, M., & Hu, J. (2020). Comparative Metabolomics Analysis Reveals Different Metabolic Responses to Drought in Tolerant and Susceptible Poplar Species. *Physiologia Plantarum*, *168*(3), 531–546. https://doi.org/10.1111/ppl.13036

Jiang, J., Ma, S., Ye, N., Jiang, M., Cao, J., & Zhang, J. (2017). WRKY Transcription Factors in Plant Responses to Stresses. *Journal of Integrative Plant Biology*, *59*(2), 86–101. https://doi.org/10.1111/jipb.12513

Jiang, Y., Duan, Y., Yin, J., Ye, S., Zhu, J., Zhang, F., Lu, W., Fan, D., & Luo, K. (2014). Genome-wide Identification and Characterization of the *Populus* WRKY Transcription Factor Family and Analysis of their Expression in Response to Biotic and Abiotic Stresses. *Journal of Experimental Botany*, *65*(22), 6629–6644. https://doi.org/10.1093/jxb/eru381

Jiang, Z., Zhou, X., Li, R., Michal, J. J., Zhang, S., Dodson, M. V., Zhang, Z., & Harland, R. M. (2015). Whole Transcriptome Analysis with Sequencing: Methods, Challenges and Potential Solutions. *Cellular and Molecular Life Sciences*, *72*, 3425–3439.

Jose, S., Williams, R., & Zamora, D. (2006). Belowground Ecological Interactions in Mixed-Species Forest Plantations. *Forest Ecology and Management*, *233*(2), 231–239. https://doi.org/10.1016/j.foreco.2006.05.014

Kanehisa, M. (2019). Toward Understanding the Origin and Evolution of Cellular Organisms. *Protein Science: A Publication of the Protein Society*, *28*(11), 1947–1951. https://doi.org/10.1002/pro.3715

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023a). KEGG for Taxonomy-Based Analysis of Pathways and Genomes. *Nucleic Acids Research*, *51*(D1), Article D1. https://doi.org/10.1093/nar/gkac963

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023b). KEGG for Taxonomy-Based Analysis of Pathways and Genomes. *Nucleic Acids Research*, *51*(D1), D587–D592. https://doi.org/10.1093/nar/gkac963

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Keim, P., Paige, K. N., Whitham, T. G., & Lark, K. G. (1989). Genetic Analysis of an Interspecific Hybrid Swarm of *Populus*: Occurrence of Unidirectional Introgression. *Genetics*, *123*(3), 557–565.

Kelty, M. J. (2006). The Role of Species Mixtures in Plantation Forestry. *Forest Ecology and Management*, *233*(2), 195–204. https://doi.org/10.1016/j.foreco.2006.05.011

Khan, M. I. R., Fatma, M., Per, T. S., Anjum, N. A., & Khan, N. A. (2015). Salicylic Acid-Induced Abiotic Stress Tolerance and Underlying Mechanisms in Plants. *Frontiers in Plant Science*, *6*, Article 462. https://www.frontiersin.org/articles/10.3389/fpls.2015.00462

King, K. C., & Lively, C. M. (2012). Does Genetic Diversity Limit Disease Spread in Natural Host Populations? *Heredity*, *109*(4), 199–203. https://doi.org/10.1038/hdy.2012.33

Kuchma, O., Rebola-Lichtenberg, J., Janz, D., Krutovsky, K. V., Ammer, C., Polle, A., & Gailing, O. (2022). Response of Poplar Leaf Transcriptome to Changed Management and Environmental Conditions in Pure and Mixed with Black Locust Stands. *Forests*, *13*(2), Article 2. https://doi.org/10.3390/f13020147

Langholtz, M. H., Stokes, B. J., & Eaton, L. M. (2016). *2016 Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy* (DOE/EE-1440). EERE Publication and Product Library, Washington, D.C. (United States). https://doi.org/10.2172/1271651

Li, H., Li, M., Luo, J., Cao, X., Qu, L., Gai, Y., Jiang, X., Liu, T., Bai, H., Janz, D., Polle, A., Peng, C., & Luo, Z.-B. (2012). N-fertilization has Different Effects on the Growth, Carbon and Nitrogen Physiology, and Wood Properties of Slow- and Fast-Growing *Populus* species. *Journal of Experimental Botany*, *63*(17), 6173–6185. https://doi.org/10.1093/jxb/ers271

Lindroth, R. L., & Hwang, S.-Y. (1996). Clonal Variation in Foliar Chemistry of Quaking Aspen (Populus tremuloides Michx.). *Biochemical Systematics and Ecology*, *24*(5), 357–364. https://doi.org/10.1016/0305-1978(96)00043-9

Liu, B., Wu, J., Yang, S., Schiefelbein, J., & Gan, Y. (2020). Nitrate Regulation of Lateral Root and Root Hair Development in Plants. *Journal of Experimental Botany*, *71*(15), 4405–4414. https://doi.org/10.1093/jxb/erz536

Liu, L., Jiang, L.-G., Luo, J.-H., Xia, A.-A., Chen, L.-Q., & He, Y. (2021). Genome-wide Association Study Reveals the Genetic Architecture of Root Hair Length in Maize. *BMC Genomics*, *22*(1), 664. https://doi.org/10.1186/s12864-021-07961-z

Loreau, M., & Hector, A. (2001). Partitioning Selection and Complementarity in Biodiversity Experiments. *Nature*, *412*(6842), Article 6842. https://doi.org/10.1038/35083573

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics Technologies. *PLOS Computational Biology*, *13*(5), Article 5. https://doi.org/10.1371/journal.pcbi.1005457

Luo, J., Zhou, J., Li, H., Shi, W., Polle, A., Lu, M., Sun, X., & Luo, Z.-B. (2015). Global Poplar Root and Leaf Transcriptomes Reveal Links Between Growth and Stress Responses Under Nitrogen Starvation and Excess. *Tree Physiology*, *35*(12), Article 12. https://doi.org/10.1093/treephys/tpv091

Luo, J., & Zhou, J.-J. (2019). Growth Performance, Photosynthesis, and Root Characteristics are Associated with Nitrogen Use Efficiency in Six Poplar Species. *Environmental and Experimental Botany*, *164*, 40–51. https://doi.org/10.1016/j.envexpbot.2019.04.013

Lv, B., Wu, Q., Wang, A., Li, Q., Dong, Q., Yang, J., Zhao, H., Wang, X., Chen, H., & Li, C. (2020). A WRKY Transcription Factor, FtWRKY46, from Tartary Buckwheat Improves Salt Tolerance in Transgenic *Arabidopsis thaliana*. *Plant Physiology and Biochemistry*, *147*, 43–53.

Ma, D., Reichelt, M., Yoshida, K., Gershenzon, J., & Constabel, C. P. (2018). Two R2R3-MYB Proteins are Broad Repressors of Flavonoid and Phenylpropanoid Metabolism in Poplar. *The Plant Journal*, *96*(5), 949–965. https://doi.org/10.1111/tpj.14081

Majumdar, R., Barchi, B., Turlapati, S. A., Gagne, M., Minocha, R., Long, S., & Minocha, S. C. (2016). Glutamate, Ornithine, Arginine, Proline, and Polyamine Metabolic Interactions: The Pathway is Regulated at the Post-Transcriptional Level. *Frontiers in Plant Science*, *7*, Article 78. https://www.frontiersin.org/articles/10.3389/fpls.2016.00078

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012a). Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation. *Nucleic Acids Research*, *40*(10), 4288–4297. https://doi.org/10.1093/nar/gks042

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012b). Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation. *Nucleic Acids Research*, *40*(10), 4288–4297. https://doi.org/10.1093/nar/gks042

McDermaid, A., Monier, B., Zhao, J., Liu, B., & Ma, Q. (2019). Interpretation of Differential Gene Expression Results of RNA-Seq Data: Review and Integration. *Briefings in Bioinformatics*, *20*(6), 2044–2054. https://doi.org/10.1093/bib/bby067

McKown, A. D., Guy, R. D., Klápště, J., Geraldes, A., Friedmann, M., Cronk, Q. C., El-Kassaby, Y. A., Mansfield, S. D., & Douglas, C. J. (2014). Geographical and Environmental Gradients Shape Phenotypic Trait Variation and Genetic Structure in *Populus trichocarpa*. *New Phytologist*, *201*(4), 1263–1276.

McKown, A. D., Klápště, J., Guy, R. D., Geraldes, A., Porth, I., Hannemann, J., Friedmann, M., Muchero, W., Tuskan, G. A., Ehlting, J., Cronk, Q. C. B., El-Kassaby, Y. A., Mansfield, S. D., & Douglas, C. J. (2014). Genome-wide Association Implicates Numerous Genes Underlying Ecological Trait Variation in Natural Populations of *Populus trichocarpa*. *New Phytologist*, *203*(2), 535–553. https://doi.org/10.1111/nph.12815

Meng, S., Su, L., Li, Y., Wang, Y., Zhang, C., & Zhao, Z. (2016). Nitrate and Ammonium Contribute to the Distinct Nitrogen Metabolism of *Populus simonii* During Moderate Salt Stress. *PLOS ONE*, *11*(3), e0150354. https://doi.org/10.1371/journal.pone.0150354

Meraj, T. A., Fu, J., Raza, M. A., Zhu, C., Shen, Q., Xu, D., & Wang, Q. (2020). Transcriptional Factors Regulate Plant Stress Responses Through Mediating Secondary Metabolism. *Genes*, *11*(4), Article 4. https://doi.org/10.3390/genes11040346

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: More Genomes, a New PANTHER GO-slim and Improvements in Enrichment Analysis tools. *Nucleic Acids Research*, *47*(D1), Article D1. https://doi.org/10.1093/nar/gky1038

Miller, A,, Bowan, W., & Suding, K. (2007). *PLANT UPTAKE OF INORGANIC AND ORGANIC NITROGEN: NEIGHBOR IDENTITY MATTERS*. *Ecology*, 88(7), 1832-1840. https://esajournals.onlinelibrary.wiley.com/doi/epdf/10.1890/06-0946.1?src=getftr

Moreno-Hagelsieb, G., & Latimer, K. (2008). Choosing BLAST Options for Better Detection of Orthologs as Reciprocal Best Hits. *Bioinformatics*, *24*(3), 319–324. https://doi.org/10.1093/bioinformatics/btm585

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), Article 7. https://doi.org/10.1038/nmeth.1226

Movahedi, A., Almasi Zadeh Yaghuti, A., Wei, H., Rutland, P., Sun, W., Mousavi, M., Li, D., & Zhuge, Q. (2021). Plant Secondary Metabolites with an Overview of *Populus*. *International Journal of Molecular Sciences*, *22*(13), 6890.

MRNA Sequencing (mRNA-Seq). (2023). *Novogene*. https://www.novogene.com/us-en/services/research-services/transcriptome-sequencing/mrna-sequencing/

NGS Workflow Steps | Illumina sequencing workflow. (n.d.). Retrieved 28 April 2023, from https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html

Niemczyk, M., Hu, Y., & Thomas, B. R. (2019). Selection of Poplar Genotypes for Adapting to Climate Change. *Forests*, *10*(11), Article 11. https://doi.org/10.3390/f10111041

Pagès, H., Carlson, M., Falcon, S., & Li, N. (2023). *AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor* (1.60.2). Bioconductor version: Release (3.16). https://bioconductor.org/packages/AnnotationDbi/

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon Provides Fast and Bias-aware Quantification of Transcript Expression. *Nature Methods*, *14*(4), Article 4. https://doi.org/10.1038/nmeth.4197

Pauli, A., Rinn, J. L., & Schier, A. F. (2011). Non-coding RNAs as Regulators of Embryogenesis. *Nature Reviews Genetics*, *12*(2), Article 2. https://doi.org/10.1038/nrg2904

Pečenková, T., Potocká, A., Potocký, M., Ortmannová, J., Drs, M., Janková Drdová, E., Pejchar, P., Synek, L., Soukupová, H., Žárský, V., & Cvrčková, F. (2020). Redundant and Diversified Roles Among Selected *Arabidopsis thaliana* EXO70 Paralogs During Biotic Stress Responses. *Frontiers in Plant Science*, *11*, Article 960. https://www.frontiersin.org/articles/10.3389/fpls.2020.00960

Piotto, D. (2008). A Meta-Analysis Comparing Tree Growth in Monocultures and Mixed Plantations. *Forest Ecology and Management*, *255*(3–4), 781–786.

Popko, J., Hänsch, R., Mendel, R.-R., Polle, A., & Teichmann, T. (2010). The Role of Abscisic Acid and Auxin in the Response of Poplar to Abiotic Stress. *Plant Biology*, *12*(2), 242–258. https://doi.org/10.1111/j.1438-8677.2009.00305.x

Pretzsch, H. (2005). Diversity and Productivity in Forests: Evidence from Long-term Experimental Plots. *Ecological Studies*, *176*, 41–64.

Pretzsch, H., & Schütze, G. (2009). Transgressive Overyielding in Mixed Compared with Pure Stands of Norway Spruce and European Beech in Central Europe: Evidence on Stand Level and Explanation on Individual Tree Level. *European Journal of Forest Research*, *128*, 183–204.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing. R foundation for Statistical Computing, Vienna, Austria.* https://www.r-project.org/.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C. E., Socci, N. D., & Betel, D. (2013). Comprehensive Evaluation of Differential Gene Expression Analysis Methods for RNA-Seq Data. *Genome Biology*, *14*(9), Article 9. https://doi.org/10.1186/gb-2013-14-9-r95

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., Merico, D., & Bader, G. D. (2019). Pathway Enrichment Analysis and Visualization of Omics Data Using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature Protocols*, *14*(2), Article 2. https://doi.org/10.1038/s41596-018-0103-9

Ren, S., Ma, K., Lu, Z., Chen, G., Cui, J., Tong, P., Wang, L., Teng, N., & Jin, B. (2019). Transcriptomic and Metabolomic Analysis of the Heat-Stress Response of *Populus tomentosa* Carr. *Forests*, *10*(5), Article 5. https://doi.org/10.3390/f10050383

Renninger, H. J., Stewart, L. F., Freeman, J. L., & Rousseau, R. J. (2022). Physiological Functioning and Productivity in Eastern Cottonwood and Hybrid Poplars on Contrasting Sites in the Southeastern US. *BioEnergy Research*, 1–14.

Richards, A. E., Forrester, D. I., Bauhus, J., & Scherer-Lorenzen, M. (2010). *The Influence of Mixed Tree Plantations on the Nutrition of Individual Species: A Review*. *Tree Physiology*, *30*, 17.

RNeasy Mini Handbook—(EN)—QIAGEN. (2019, October). https://www.qiagen.com/us/resources/resourcedetail?id=14e7cf6e-521a-4cf7-8cbc-bf9f6fa33e24&lang=en

RNeasy Plant Mini Kit. (n.d.). Retrieved 15 April 2023, from https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/rna-purification/total-rna/rneasy-plant-mini-kit

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson, M. D., & Smyth, G. K. (2007). Small-sample Estimation of Negative Binomial Dispersion, with Applications to SAGE Data. *Biostatistics*, *9*(2), 321–332. https://doi.org/10.1093/biostatistics/kxm030

Robinson, M. D., & Smyth, G. K. (2008). Small-sample Estimation of Negative Binomial Dispersion, with Applications to SAGE Data. *Biostatistics*, *9*(2), 321–332. https://doi.org/10.1093/biostatistics/kxm030

Rowe, E. C., Noordwijk, M. V., Suprayogo, D., & Cadisch, G. (2005). Nitrogen Use Efficiency of Monoculture and Hedgerow Intercropping in the Humid Tropics. *Plant and Soil*, *268*, 61–74.

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2021). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Schmid, I., & Kazda, M. (2002). Root Distribution of Norway Spruce in Monospecific and Mixed Stands on Different Soils. *Forest Ecology and Management*, *159*(1), 37–47. https://doi.org/10.1016/S0378-1127(01)00708-3

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M., & Barton, G. J. (2016). How Many Biological Replicates are Needed in an RNA-Seq Experiment and Which Differential Expression Tool Should You Use? *RNA*, *22*(6), Article 6. https://doi.org/10.1261/rna.053959.115

Sequencing Platforms | Illumina NGS platforms. (n.d.). Retrieved 28 April 2023, from https://www.illumina.com/systems/sequencing-platforms.html

Shim, D., Kim, S., Choi, Y.-I., Song, W.-Y., Park, J., Youk, E. S., Jeong, S.-C., Martinoia, E., Noh, E.-W., & Lee, Y. (2013). Transgenic Poplar Trees Expressing Yeast Cadmium Factor 1 Exhibit the Characteristics Necessary for the Phytoremediation of Mine Tailing Soil. *Chemosphere*, *90*(4), 1478–1486. https://doi.org/10.1016/j.chemosphere.2012.09.044

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential Analyses for RNA-Seq: Transcript-level Estimates Improve Gene-level Inferences. *F1000Research*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.1

Srivastava, A., Sarkar, H., Gupta, N., & Patro, R. (2016). RapMap: A Rapid, Sensitive and Accurate Tool for Mapping RNA-Seq Reads to Transcriptomes. *Bioinformatics*, *32*(12), i192–i200. https://doi.org/10.1093/bioinformatics/btw277

Stanton, B. J., & Gustafson, R. R. (2019). Advanced Hardwood Biofuels Northwest: Commercialization Challenges for the Renewable Aviation Fuel Industry. *Applied Sciences*, *9*(21), Article 21. https://doi.org/10.3390/app9214644

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA Sequencing: The Teenage Years. *Nature Reviews Genetics*, *20*(11), Article 11. https://doi.org/10.1038/s41576-019-0150-2

Strauss, S. H., Ma, C., Ault, K., & Klocko, A. L. (2016). Lessons from Two Decades of Field Trials with Genetically Modified Trees in the USA: Biology and Regulatory Compliance. In C. Vettori, F. Gallardo, H. Häggman, V. Kazana, F. Migliacci, G. Pilate, & M. Fladung (Eds.), *Biosafety of Forest Transgenic Trees* (Vol. 82, pp. 101–124). Springer Netherlands. https://doi.org/10.1007/978-94-017-7531-1_5

Strauss, S. Y., & Irwin, R. E. (2004). Ecological and Evolutionary Consequences of Multispecies Plant-Animal Interactions. *Annual Review of Ecology, Evolution, and Systematics*, *35*, 435–466.

Su, Z., Ma, X., Guo, H., Sukiran, N. L., Guo, B., Assmann, S. M., & Ma, H. (2013). Flower Development Under Drought Stress: Morphological and Transcriptomic Analyses Reveal Acute Responses and Long-term Acclimation in *Arabidopsis*. *The Plant Cell*, *25*(10), 3785–3807.

Sun, Y., & Yu, D. (2015). Activated Expression of AtWRKY53 Negatively Regulates Drought Tolerance by Mediating Stomatal Movement. *Plant Cell Reports*, *34*(8), 1295–1306. https://doi.org/10.1007/s00299-015-1787-8

Synek, L., Schlager, N., Eliáš, M., Quentin, M., Hauser, M.-T., & Žárský, V. (2006). AtEXO70A1, a Member of a Family of Putative Exocyst Subunits Specifically Expanded in Land Plants, is Important for Polar Growth and Plant Development. *The Plant Journal*, *48*(1), 54–72. https://doi.org/10.1111/j.1365-313X.2006.02854.x

Tan, C.-M., Chen, R.-J., Zhang, J.-H., Gao, X.-L., Li, L.-H., Wang, P.-R., Deng, X.-J., & Xu, Z.-J. (2013). OsPOP5, a Prolyl Oligopeptidase Family Gene from Rice Confers Abiotic Stress Tolerance in *Escherichia coli*. *International Journal of Molecular Sciences*, *14*(10), 20204–20219.

Taylor, G. (2002). *Populus: Arabidopsis* for Forestry. Do We Need a Model Tree? *Annals of Botany*, *90*(6), 681–689. https://doi.org/10.1093/aob/mcf255

Thermo Fisher Scientific. (2023). *NanoDrop One Microvolume UV-Vis Spectrophotometers*.

Tsai, C.-J., Harding, S. A., Tschaplinski, T. J., Lindroth, R. L., & Yuan, Y. (2006). Genome-wide Analysis of the Structural Genes Regulating Defense Phenylpropanoid Metabolism in *Populus*. *New Phytologist*, *172*(1), 47–62. https://doi.org/10.1111/j.1469-8137.2006.01798.x

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., … Rokhsar, D. (2006). The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, *313*(5793), Article 5793. https://doi.org/10.1126/science.1128691

Tuskan, G. A., DiFazio, S. P., & Teichmann, T. (2004). Poplar Genomics is Getting Popular: The Impact of the Poplar Genome Project on Tree Research. *Plant Biology*, *6*(1), 2–4. https://doi.org/10.1055/s-2003-44715

Valenzuela, S., & Strauss, S. H. (2005). Lost in the Woods. *Nature Biotechnology*, *23*(5), Article 5. https://doi.org/10.1038/nbt0505-532

Verlinden, M., Van Kerkhove, A., & Nijs, I. (2013). Effects of Experimental Climate Warming and Associated Soil Drought on the Competition between Three Highly Invasive West European Alien Plant Species and Native Counterparts. *Plant Ecology*, *214*, 243–254.

Verslues, P. E., & Sharma, S. (2010). Proline Metabolism and Its Implications for Plant-Environment Interaction. *The Arabidopsis Book / American Society of Plant Biologists*, *8*, e0140. https://doi.org/10.1199/tab.0140

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Ward, N., & Moreno-Hagelsieb, G. (2014). Quickly Finding Orthologs as Reciprocal Best Hits with BLAT, LAST, and UBLAST: How Much Do We Miss? *PloS One*, *9*(7), Article 7. https://doi.org/10.1371/journal.pone.0101850

Wei, H., Yordanov, Y., Kumari, S., Georgieva, T., & Busov, V. (2013). Genetic Networks Involved in Poplar Root Response to Low Nitrogen. *Plant Signaling & Behavior*, *8*(11), e27211. https://doi.org/10.4161/psb.27211

Wei, H., Yordanov, Y. S., Georgieva, T., Li, X., & Busov, V. (2013). Nitrogen Deprivation Promotes *Populus* Root Growth through Global Transcriptome Reprogramming and Activation of Hierarchical Genetic Networks. *New Phytologist*, *200*(2), 483–497. https://doi.org/10.1111/nph.12375

Winter, G., Todd, C. D., Trovato, M., Forlani, G., & Funck, D. (2015). Physiological Implications of Arginine Metabolism in Plants. *Frontiers in Plant Science*, *6*. https://www.frontiersin.org/articles/10.3389/fpls.2015.00534

Wullschleger, S. D., Weston, D. J., DiFazio, S. P., & Tuskan, G. A. (2013). Revisiting the Sequencing of the First Tree Genome: *Populus trichocarpa*. *Tree Physiology*, *33*(4), 357–364. https://doi.org/10.1093/treephys/tps081

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., Huang, J., Li, M., Wu, X., Wen, L., Lao, K., Li, R., Qiao, J., & Tang, F. (2013). Single-Cell RNA-seq Profiling of Human Preimplantation Embryos and Embryonic Stem Cells. *Nature Structural & Molecular Biology*, *20*(9), Article 9. https://doi.org/10.1038/nsmb.2660

Yang, J., Chen, X., Zhu, C., Peng, X., He, X., Fu, J., Ouyang, L., Bian, J., Hu, L., & Sun, X. (2015). RNA-Seq Reveals Differentially Expressed Genes of Rice (*Oryza sativa*) Spikelet in Response to Temperature Interacting with Nitrogen at Meiosis Stage. *BMC Genomics*, *16*, 1–18.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene Ontology Analysis for RNA-Seq: Accounting for Selection Bias. *Genome Biology*, *11*(2), R14. https://doi.org/10.1186/gb-2010-11-2-r14

Yu, Y., Zhang, Y., Chen, X., & Chen, Y. (2019). Plant Noncoding RNAs: Hidden Players in Development and Stress Responses. *Annual Review of Cell and Developmental Biology*, *35*(1), Article 1. https://doi.org/10.1146/annurev-cellbio-100818-125218

Zalesny, R. S., Donner, D. M., Coyle, D. R., & Headlee, W. L. (2012). An Approach for Siting Poplar Energy Production Systems to Increase Productivity and Associated Ecosystem Services. *Forest Ecology and Management*, *284*, 45–58. https://doi.org/10.1016/j.foreco.2012.07.022

Zalesny, R. S., Headlee, W. L., Gopalakrishnan, G., Bauer, E. O., Hall, R. B., Hazel, D. W., Isebrands, J. G., Licht, L. A., Negri, M. C., Nichols, E. G., Rockwood, D. L., & Wiese, A. H. (2019). Ecosystem Services of Poplar at Long-Term Phytoremediation Sites in the Midwest and Southeast, United States. *WIREs Energy and Environment*, *8*(6), Article 6. https://doi.org/10.1002/wene.349

Zhang, L., Du, J., Ge, X., Cao, D., & Hu, J. (2021). Leaf Size Development Differences and Comparative Transcriptome Analyses of Two Poplar Genotypes. *Genes*, *12*(11), Article 11. https://doi.org/10.3390/genes12111775

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS ONE*, *9*(1), e78644. https://doi.org/10.1371/journal.pone.0078644

Zhou, H. H., Chen, Y. N., Li, W. H., & Chen, Y. P. (2010). Photosynthesis of *Populus euphratica* in Relation to Groundwater Depths and High Temperature in Arid Environment, Northwest China. *Photosynthetica*, *48*, 257–268.

APPENDIX A

RNA QUALITY CONTROL

Table A.1 Absorbance Ratios for RNA Samples

Planting and Plot Number	**Clone**	**Treatment**	**Date**	**Biosample**	**Nucleic Acid (ng/uL)**	**A260/A280**	**A260/A230**	**A260**	**A280**
Monoclonal 7	S7C8	Non Endophyte	8/3/2022	4NES_1	1013.54	2.096	2.007	25.338	12.089
Monoclonal 7	S7C8	Non Endophyte	8/3/2022	4NES_2	973.854	2.093	1.96	24.346	11.631
Monoclonal 7	S7C8	Non Endophyte	8/22/2022	4NES_3	1104.498	2.106	2.025	27.612	13.11
Monoclonal 17	S7C8	Non Endophyte	8/22/2022	4NES_4	399.459	2.1	1.865	9.986	4.755
Polyclonal 8	110412	Non Endophyte	8/22/2022	4NEM1_1	1582.357	2.151	2.273	39.559	18.395
Polyclonal 8	110412	Non Endophyte	8/22/2022	4NEM1_2	1534.692	2.108	2.272	38.367	18.198
Polyclonal 8	110412	Non Endophyte	8/22/2022	4NEM1_3	1693.083	2.123	2.305	42.327	19.935

Table A.1 (Continued)

Planting and Plot Number	Clone	Treatment	Date	Biosample	Nucleic Acid (ng/uL)	A260/A280	A260/A230	A260	A280
Polyclonal 8	110412	Non Endophyte	8/22/2022	4NEM1_4	1234.677	2.092	2.31	30.867	14.753
Polyclonal 8	S7C8	Non Endophyte	8/22/2022	4NEMS_1	1039.235	2.092	2.218	25.981	12.417
Polyclonal 8	S7C8	Non Endophyte	8/22/2022	4NEMS_2	262.127	2.032	1.727	6.553	3.226
Polyclonal 8	S7C8	Non Endophyte	8/22/2022	4NEMS_3	286.422	2.042	1.422	7.161	3.506
Polyclonal 8	S7C8	Non Endophyte	8/22/2022	4NEMS_4	2413.697	2.151	2.339	60.342	28.056
Monoclonal 10	110412	Non Endophyte	8/22/2022	4NE1_1	1712.317	2.13	2.326	42.808	20.096
Monoclonal 10	110412	Non Endophyte	8/22/2022	4NE1_2	996.789	2.112	2.127	24.92	11.799
Monoclonal 10	110412	Non Endophyte	8/22/2022	4NE1_3	1644.412	2.163	2.362	41.11	19.004
Monoclonal 10	110412	Non Endophyte	8/22/2022	4NE1_4	1477.658	2.167	2.162	36.94	17.048

Table A.1 (Continued)

Planting and Plot Number	Clone	Treatment	Date	Biosample	Nucleic Acid (ng/uL)	A260/A280	A260/A230	A260	A280
Polyclonal 15	110412	Endophyte	8/22/2022	4EM1_1	2074.356	2.172	2.372	51.859	23.88
Polyclonal 15	110412	Endophyte	8/22/2022	4EM1_2	1543.945	2.164	2.387	38.599	17.841
Polyclonal 15	110412	Endophyte	8/25/2022	4EM1_3	2107.706	2.177	2.34	52.693	24.203
Polyclonal 15	110412	Endophyte	8/25/2022	4EM1_4	1644.301	2.105	2.311	41.108	19.527
Polyclonal 15	S7C8	Endophyte	8/25/2022	4EMS_1	367.82	2.052	1.759	9.196	4.48
Polyclonal 15	S7C8	Endophyte	8/25/2022	4EMS_2	800.755	2.132	2.011	20.019	9.392
Polyclonal 15	S7C8	Endophyte	8/25/2022	4EMS_3	1628.71	2.167	2.325	40.718	18.79
Polyclonal 15	S7C8	Endophyte	8/25/2022	4EMS_4	410.995	2.098	1.888	10.275	4.897
Monoclonal 18	110412	Endophyte	8/25/2022	4E1_1	1152.1	2.149	2.312	28.802	13.4

Table A.1 (Continued)

Planting and Plot Number	Clone	Treatment	Date	Biosample	Nucleic Acid (ng/uL)	A260/A280	A260/A230	A260	A280
Monoclonal 18	110412	Endophyte	8/25/2022	4E1_2	1548.744	2.165	2.333	38.719	17.885
Monoclonal 18	110412	Endophyte	8/26/2022	4E1_3	1597.524	2.106	2.334	39.938	18.962
Monoclonal 18	110412	Endophyte	8/26/2022	4E1_4	822.463	2.123	2.213	20.562	9.686
Monoclonal 22	S7C8	Endophyte	8/26/2022	4ES_1	1244.811	2.156	2.316	31.12	14.434
Monoclonal 22	S7C8	Endophyte	8/26/2022	4ES_2	1214.612	2.138	2.275	30.365	14.206
Monoclonal 22	S7C8	Endophyte	8/26/2022	4ES_3	606.955	2.105	1.975	15.174	7.209
Monoclonal 22	S7C8	Endophyte	8/26/2022	4ES_4	114.727	2.074	1.498	2.868	1.383

Date indicates the date RNA was extracted and checked for quantification. Biosample is labeled based on replicate plot number, whole plot treatment factor, clone, and biological replicate. Nucleic acid concentration is measured in nanograms per milliliter.

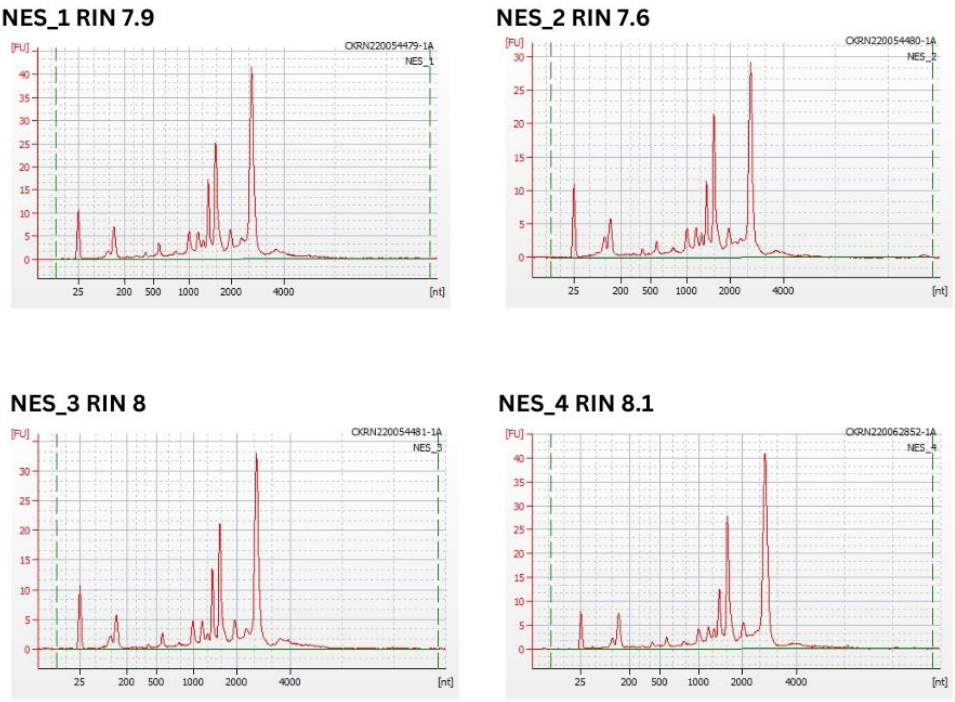
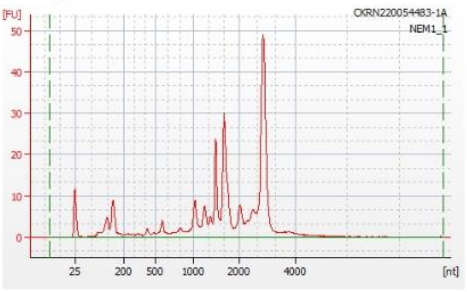


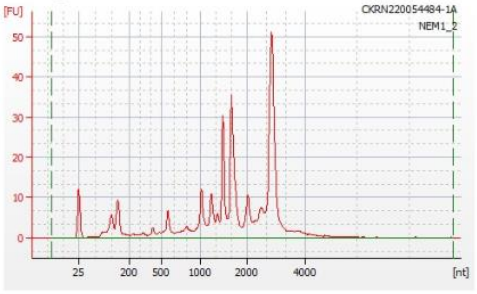
Figure A.1 RNA Integrity Numbers (RIN) for Leaf Tissue Samples

Figure A.1 (Continued)

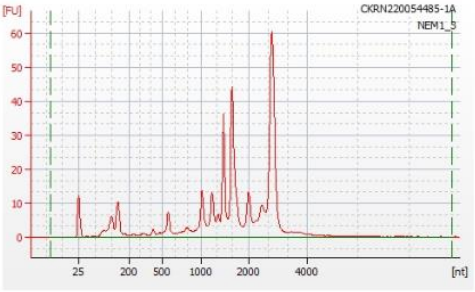
NEM1_1 RIN 7.9



NEM1_2 RIN 7.6



NEM1_3 RIN 7.5



NEM1_4 RIN 7.7

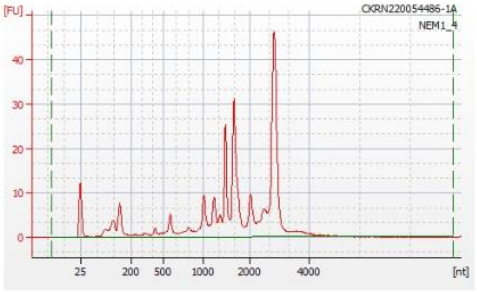
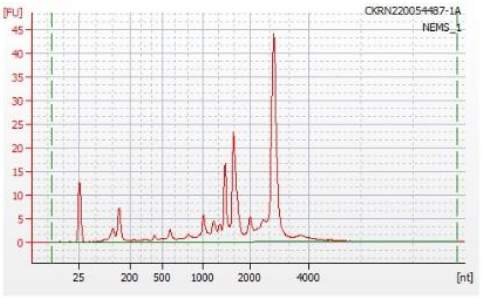
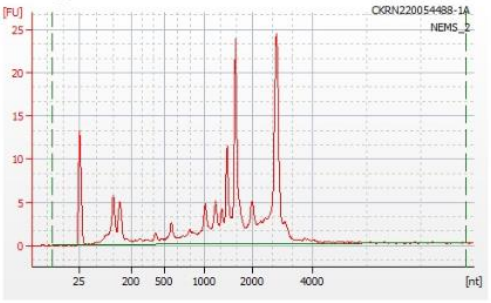


Figure A.1 (Continued)

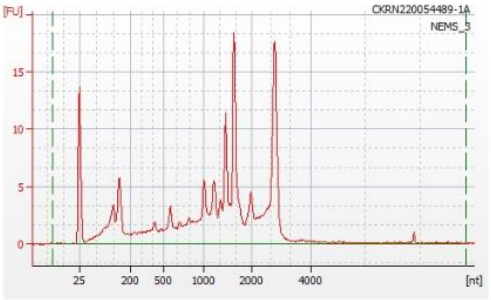
NEMS_1 RIN 8.2



NEMS_2 RIN 7.1



NEMS_3 RIN 6.6



NEMS_4 RIN 7.6

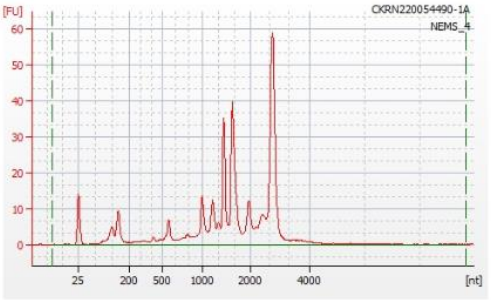
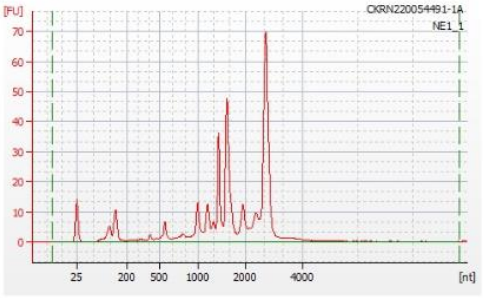
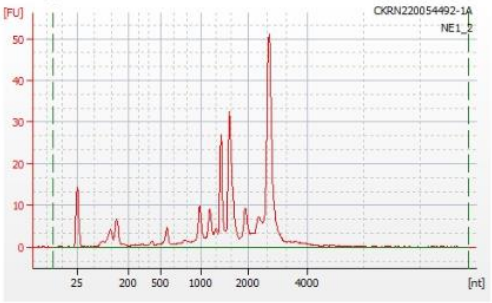


Figure A.1 (Continued)

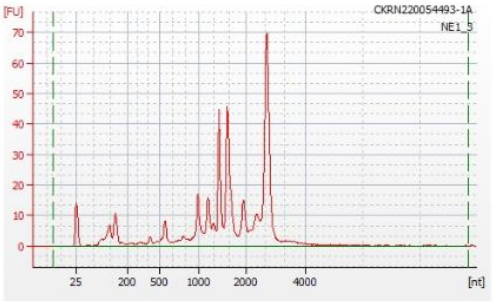
NE1_1 RIN 7.8



NE1_2 RIN 7.9



NE1_3 RIN 7.6



NE1_4 RIN 7.5

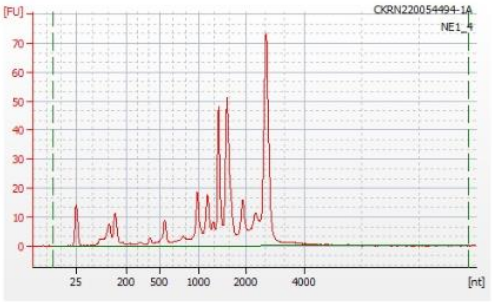
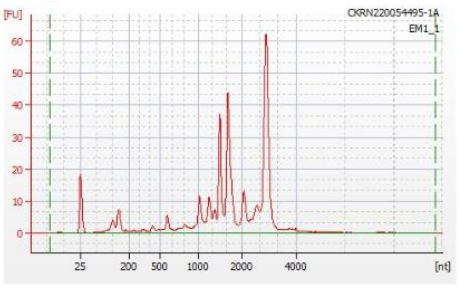
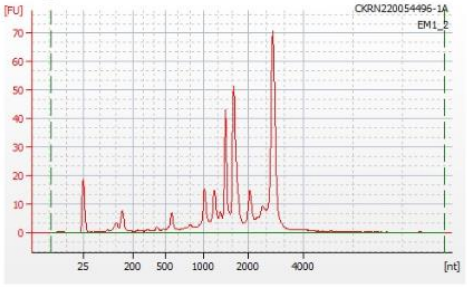


Figure A.1 (Continued)

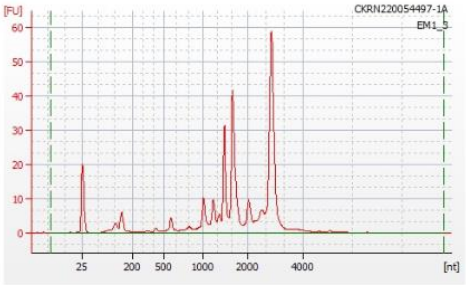
EM1_1 RIN 7.7



EM1_2 RIN 7.6



EM1_3 RIN 7.9



EM1_4 RIN 8

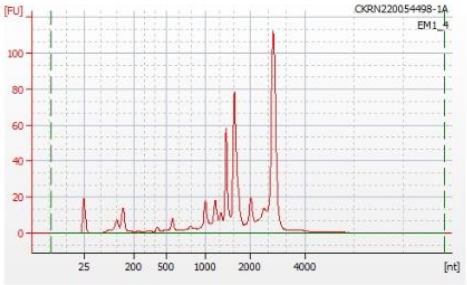
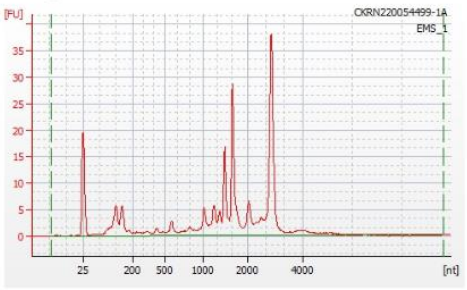
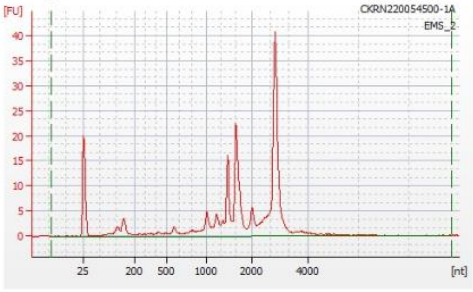


Figure A.1 (Continued)

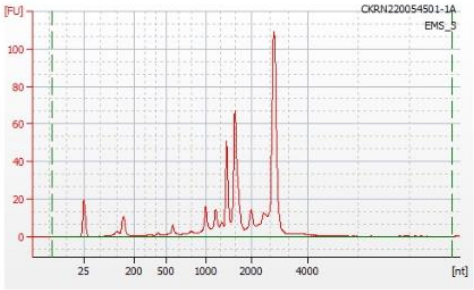
EMS_1 RIN 7.5



EMS_2 RIN 8.3



EMS_3 RIN 8.2



EMS_4 RIN 8.2

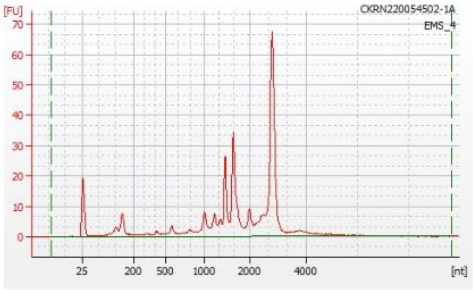


Figure A.1 (Continued)

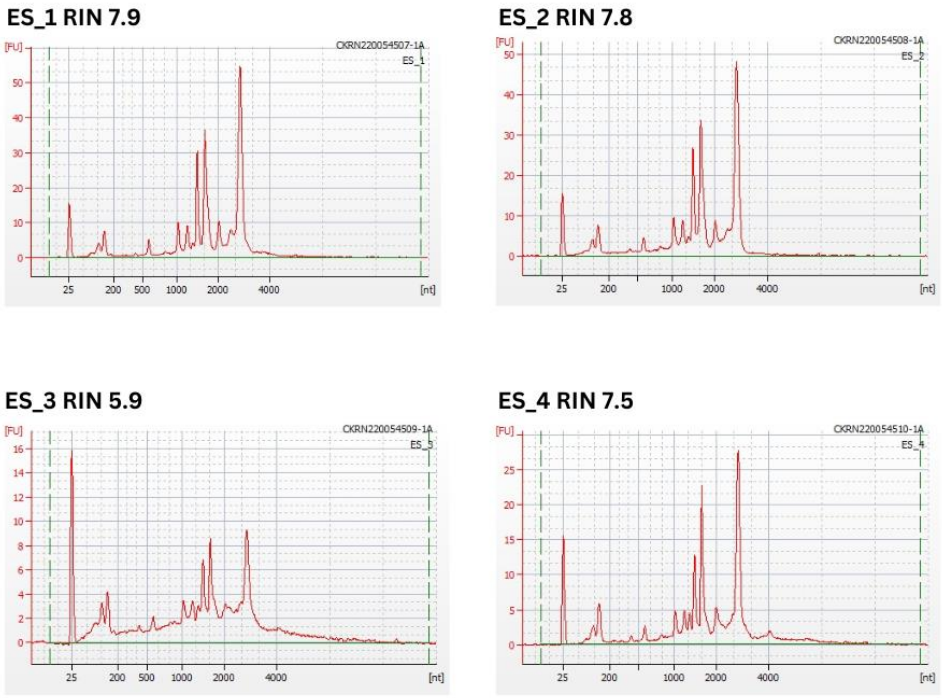
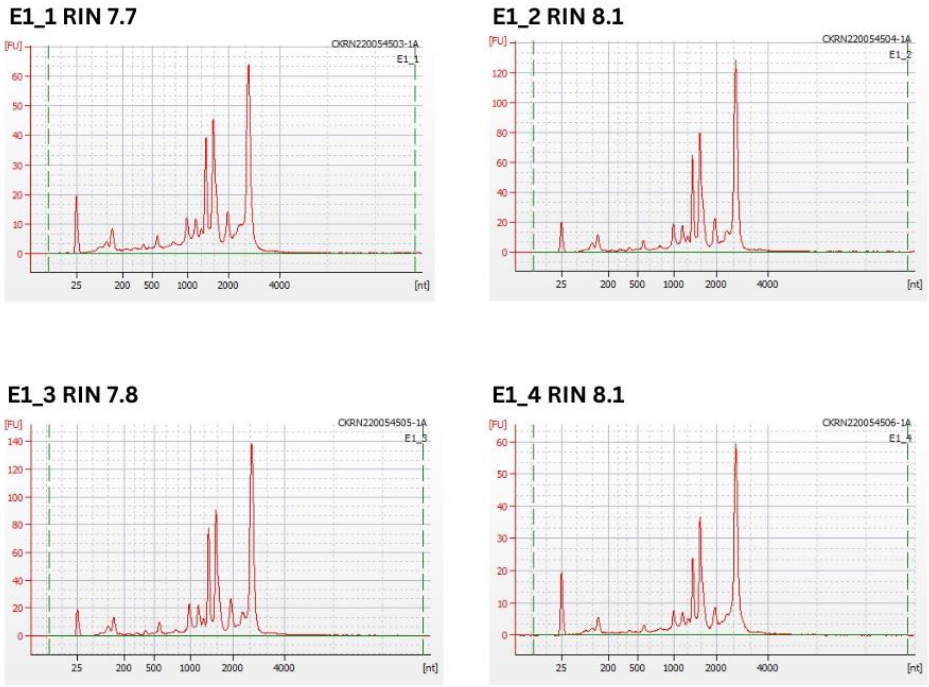


Figure A.1 (Continued)



APPENDIX B

P. DELTOIDES METADATA, ABOVEGROUND BIOMASS AND DIFFERENTIAL GENE EXPRESSION ANALYSIS RESULTS

Table B.1 Metadata for Transcript Quantification, Differential Gene Expression, Gene Ontology and KEGG Enrichment Pathway Analysis

Biosample	Salmon Output	Clone	Color	Planting	Plot	Block	Site	Treatment
NES_1	NES_1_RUN	S7C8	Red	Monoclonal	7	4	POA	NE
NES_2	NES_2_RUN	S7C8	Red	Monoclonal	7	4	POA	NE
NES_3	NES_3_RUN	S7C8	Red	Monoclonal	7	4	POA	NE
NES_4	NES_4_RUN	S7C8	Red	Monoclonal	7	4	POA	NE
NEM1_1	NEM1_1_RUN	110412	White	Polyclonal	8	4	POA	NE
NEM1_2	NEM1_2_RUN	110412	White	Polyclonal	8	4	POA	NE
NEM1_3	NEM1_3_RUN	110412	White	Polyclonal	8	4	POA	NE
NEM1_4	NEM1_4_RUN	110412	White	Polyclonal	8	4	POA	NE
NEMS_1	NEMS_1_RUN	S7C8	Red	Polyclonal	8	4	POA	NE
NEMS_2	NEMS_2_RUN	S7C8	Red	Polyclonal	8	4	POA	NE
NEMS_3	NEMS_3_RUN	S7C8	Red	Polyclonal	8	4	POA	NE
NEMS_4	NEMS_4_RUN	S7C8	Red	Polyclonal	8	4	POA	NE

Table B.1 (Continued)

NE1_1	NE1_1_RUN	110412	White	Monoclonal	10	4	POA	NE
NE1_2	NE1_2_RUN	110412	White	Monoclonal	10	4	POA	NE
NE1_3	NE1_3_RUN	110412	White	Monoclonal	10	4	POA	NE
NE1_4	NE1_4_RUN	110412	White	Monoclonal	10	4	POA	NE
EM1_1	EM1_1_RUN	110412	White	Polyclonal	15	4	POA	E
EM1_2	EM1_2_RUN	110412	White	Polyclonal	15	4	POA	E
EM1_3	EM1_3_RUN	110412	White	Polyclonal	15	4	POA	E
EM1_4	EM1_4_RUN	110412	White	Polyclonal	15	4	POA	E
EMS_1	EMS_1_RUN	S7C8	Red	Polyclonal	15	4	POA	E
EMS_2	EMS_2_RUN	S7C8	Red	Polyclonal	15	4	POA	E
EMS_3	EMS_3_RUN	S7C8	Red	Polyclonal	15	4	POA	E
EMS_4	EMS_4_RUN	S7C8	Red	Polyclonal	15	4	POA	E
E1_1	E1_1_RUN	110412	White	Monoclonal	18	4	POA	E
E1_2	E1_2_RUN	110412	White	Monoclonal	18	4	POA	E

Table B.1 (Continued)

E1_3	E1_3_RUN	110412	White	Monoclonal	18	4	POA	E
E1_4	E1_4_RUN	110412	White	Monoclonal	18	4	POA	E
ES_1	ES_1_RUN	S7C8	Red	Monoclonal	22	4	POA	E
ES_2	ES_2_RUN	S7C8	Red	Monoclonal	22	4	POA	E
ES_3	ES_3_RUN	S7C8	Red	Monoclonal	22	4	POA	E
ES_4	ES_4_RUN	S7C8	Red	Monoclonal	22	4	POA	E

Biosample is labeled based on replicate plot number, whole plot treatment factor, clone, and biological replicate. Salmon Output indicates transcript quant.sf files produced from *Salmon* transcript quantification.

Table B.2 Aboveground Biomass Estimations for the Second Growing Season

Planting Schemes	Aboveground Biomass (AGB) (Dry Mg/ha)	Mean for Each Planting Scheme	Standard Error
110412 Monoclonal	5.46537	5.56593	0.100558
110412 Monoclonal	5.66649	5.56593	0.100558
S7C8 Monoclonal	6.99914	7.175163	0.176021
S7C8 Monoclonal	7.351183	7.175163	0.176021
Polyclonal	12.0664	11.46569	0.60066
Polyclonal	10.865	11.46569	0.60066

Estimated aboveground biomass (AGB) between planting schemes of *P. deltoides* for the second growing season.

Table B.3 Differential Gene Expression Analysis for Model 1: Polyclonal vs. Monoclonal

Contrast	locusName	logFC	logCPM	Likelihood Ratio	PValue	False Discovery Rate	Description
Polyclonal.vs. Monoclonal	Potri.015G136400	7.2392	-0.1959	15.6592	7.58E-05	0.02508	dehydration response element B1A
Polyclonal.vs. Monoclonal	Potri.002G014000	5.7673	6.7781	20.0384	7.59E-06	0.00683	Prolyl oligopeptidase family protein
Polyclonal.vs. Monoclonal	Potri.016G126900	4.5261	5.0416	17.6812	2.61E-05	0.01457	receptor like protein 34
Polyclonal.vs. Monoclonal	Potri.003G138600	4.422	4.6409	33.1609	8.48E-09	8.85E-05	WRKY family transcription factor
Polyclonal.vs. Monoclonal	Potri.016G127000	4.2895	3.0195	17.893	2.34E-05	0.01361	receptor like protein 34
Polyclonal.vs. Monoclonal	Potri.003G072200	4.146	-2.7654	17.4842	2.90E-05	0.01477	Protein of unknown function (DUF567)
Polyclonal.vs. Monoclonal	Potri.006G070000	4.0738	0.9408	15.3211	9.07E-05	0.02825	

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.002G001400	4.0292	4.589	15.8315	6.92E-05	0.02378	calmodulin like 37
Polyclonal.vs. Monoclonal	Potri.012G078900	3.8867	2.4817	16.5852	4.65E-05	0.01859	plant U-box 18
Polyclonal.vs. Monoclonal	Potri.009G057900	3.82	-0.8132	14.5231	0.0001384	0.04075	alpha/beta-Hydrolases superfamily protein
Polyclonal.vs. Monoclonal	Potri.018G011950	3.7092	1.5884	16.9439	3.85E-05	0.01719	
Polyclonal.vs. Monoclonal	Potri.016G127101	3.6354	2.7457	16.7906	4.17E-05	0.01781	receptor like protein 34
Polyclonal.vs. Monoclonal	Potri.017G045200	3.5901	3.3342	23.9957	9.66E-07	0.002396	
Polyclonal.vs. Monoclonal	Potri.007G113000	3.4678	1.1922	17.5442	2.81E-05	0.01477	uncoupling protein 5
Polyclonal.vs. Monoclonal	Potri.009G073000	3.4466	4.0639	29.9071	4.53E-08	0.0002428	Protein phosphatase 2C family protein
Polyclonal.vs. Monoclonal	Potri.004G072900	3.3574	3.6934	18.3367	1.85E-05	0.0118	Protein of unknown function (DUF3464)

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.001G218800	3.2975	3.5991	23.8707	1.03E-06	0.002396	NAC domain containing protein 61
Polyclonal.vs. Monoclonal	Potri.017G047800	3.2901	3.1808	20.8567	4.95E-06	0.005524	uncoupling protein 5
Polyclonal.vs. Monoclonal	Potri.018G038100	3.2732	2.2456	21.4818	3.57E-06	0.004556	
Polyclonal.vs. Monoclonal	Potri.017G044300	3.2142	-1.4155	16.727	4.32E-05	0.01804	
Polyclonal.vs. Monoclonal	Potri.017G045400	3.0771	3.6319	18.7557	1.49E-05	0.01047	
Polyclonal.vs. Monoclonal	Potri.001G234600	3.0696	5.5493	22.9707	1.65E-06	0.002546	exocyst subunit exo70 family protein H7
Polyclonal.vs. Monoclonal	Potri.016G134000	2.9954	0.7253	14.3145	0.0001547	0.04354	ARM repeat superfamily protein
Polyclonal.vs. Monoclonal	Potri.019G078300	2.9402	3.3827	24.2697	8.38E-07	0.002396	CRINKLY4 related 4
Polyclonal.vs. Monoclonal	Potri.005G214000	2.9369	0.5538	15.957	6.48E-05	0.02316	Bax inhibitor-1 family protein

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.003G160400	2.9267	2.472	23.1428	1.50E-06	0.002518	
Polyclonal.vs. Monoclonal	Potri.014G122000	2.8943	4.3509	16.9571	3.82E-05	0.01719	Adenine nucleotide alpha hydrolases-like superfamily protein
Polyclonal.vs. Monoclonal	Potri.003G081200	2.7714	3.2042	14.192	0.0001651	0.04511	ethylene responsive element binding factor 1
Polyclonal.vs. Monoclonal	Potri.009G034800	2.7236	2.2748	18.381	1.81E-05	0.0118	zinc finger (C3HC4-type RING finger) family protein
Polyclonal.vs. Monoclonal	Potri.006G219800	2.683	3.9982	47.7859	4.75E-12	1.27E-07	
Polyclonal.vs. Monoclonal	Potri.016G128300	2.666	4.3066	18.7677	1.48E-05	0.01047	WRKY DNA-binding protein 33

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.010G116000	2.6487	3.7948	31.8048	1.71E-08	0.0001141	(1 of 28) PF14009 - Domain of unknown function (DUF4228) (DUF4228)
Polyclonal.vs. Monoclonal	Potri.014G087700	2.5897	3.0152	19.6824	9.14E-06	0.007653	brassinosteroid-responsive RING-H2
Polyclonal.vs. Monoclonal	Potri.009G141400	2.5695	2.1907	20.0949	7.37E-06	0.00683	expansin-like A3
Polyclonal.vs. Monoclonal	Potri.014G096200	2.5211	6.1857	27.0095	2.03E-07	0.0009037	WRKY family transcription factor
Polyclonal.vs. Monoclonal	Potri.010G129200	2.4531	3.2522	18.4336	1.76E-05	0.01178	related to ABI3/VP1 2
Polyclonal.vs. Monoclonal	Potri.018G048100	2.4489	3.2386	25.0993	5.45E-07	0.002083	
Polyclonal.vs. Monoclonal	Potri.004G206625	2.3995	-1.281	17.53	2.83E-05	0.01477	
Polyclonal.vs. Monoclonal	Potri.018G019800	2.366	5.768	21.686	3.21E-06	0.0043	WRKY DNA-binding protein 60

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.008G069400	2.355	8.4283	19.2061	1.17E-05	0.009243	salt-inducible zinc finger 1
Polyclonal.vs. Monoclonal	Potri.008G066000	2.3498	4.4501	15.9101	6.64E-05	0.02341	Hs1pro-1 protein
Polyclonal.vs. Monoclonal	Potri.001G243300	2.3321	4.2502	15.4817	8.33E-05	0.02721	zinc finger (C3HC4-type RING finger) family protein
Polyclonal.vs. Monoclonal	Potri.011G087000	2.2776	1.3856	16.2617	5.52E-05	0.02036	
Polyclonal.vs. Monoclonal	Potri.001G092900	2.223	1.467	16.8287	4.09E-05	0.01781	WRKY family transcription factor
Polyclonal.vs. Monoclonal	Potri.008G186000	2.2213	5.2322	17.4668	2.92E-05	0.01477	beta-1,4-N-acetylglucosaminyltransferase family protein
Polyclonal.vs. Monoclonal	Potri.015G074200	2.1443	7.0037	16.7833	4.19E-05	0.01781	plant U-box 18
Polyclonal.vs. Monoclonal	Potri.007G099400	2.1123	7.265	20.0233	7.65E-06	0.00683	

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.012G007500	2.0989	4.7793	32.8591	9.91E-09	8.85E-05	TCV-interacting protein
Polyclonal.vs. Monoclonal	Potri.008G117100	2.0886	4.0146	16.7003	4.38E-05	0.01804	related to ABI3/VP1 2
Polyclonal.vs. Monoclonal	Potri.010G191300	2.0392	5.8002	17.8381	2.41E-05	0.01371	Hs1pro-1 protein
Polyclonal.vs. Monoclonal	Potri.012G145800	1.9952	3.1505	23.7515	1.10E-06	0.002396	salt-inducible zinc finger 1
Polyclonal.vs. Monoclonal	Potri.004G208500	1.9691	6.1314	24.4926	7.46E-07	0.002396	
Polyclonal.vs. Monoclonal	Potri.011G055200	1.9671	1.9813	15.3635	8.87E-05	0.02794	
Polyclonal.vs. Monoclonal	Potri.019G116800	1.9244	3.7045	20.913	4.81E-06	0.005524	Wound-responsive family protein
Polyclonal.vs. Monoclonal	Potri.003G182200	1.9186	5.3929	17.0124	3.71E-05	0.01719	WRKY DNA-binding protein 40

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.011G162400	1.906	1.691	15.4248	8.59E-05	0.02757	ethylene-responsive nuclear protein / ethylene-regulated nuclear protein (ERT2)
Polyclonal.vs. Monoclonal	Potri.019G131300	1.8674	6.3667	18.5007	1.70E-05	0.01166	
Polyclonal.vs. Monoclonal	Potri.005G204600	1.848	4.5127	14.0533	0.0001777	0.04666	Eukaryotic aspartyl protease family protein
Polyclonal.vs. Monoclonal	Potri.015G103900	1.7571	2.8031	16.2501	5.55E-05	0.02036	matrix metalloproteinase
Polyclonal.vs. Monoclonal	Potri.017G128700	1.7527	2.3917	16.2882	5.44E-05	0.02036	
Polyclonal.vs. Monoclonal	Potri.008G192000	1.7511	3.9598	15.6742	7.52E-05	0.02508	Domain of unknown function (DUF23)

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.001G099400	1.7049	3.9871	14.3536	0.0001515	0.04333	1-aminocyclopropane-1-carboxylic acid (acc) synthase 6
Polyclonal.vs. Monoclonal	Potri.002G147100	1.681	3.0568	20.3455	6.47E-06	0.006413	Protein of unknown function (DUF1442)
Polyclonal.vs. Monoclonal	Potri.005G020200	1.6765	2.3759	23.2695	1.41E-06	0.002514	
Polyclonal.vs. Monoclonal	Potri.005G259600	1.6711	3.4105	17.9406	2.28E-05	0.01356	Protein kinase superfamily protein
Polyclonal.vs. Monoclonal	Potri.008G119600	1.6481	3.6157	15.0263	0.000106	0.03227	Glycolipid transfer protein (GLTP) family protein
Polyclonal.vs. Monoclonal	Potri.009G019200	1.645	5.7913	15.8644	6.81E-05	0.02367	NAC domain containing protein 61
Polyclonal.vs. Monoclonal	Potri.015G086800	1.6206	8.0984	19.7893	8.65E-06	0.00747	Leucine-rich repeat protein kinase family protein

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.009G100200	1.6061	4.619	20.9877	4.62E-06	0.005524	plant U-box 25
Polyclonal.vs. Monoclonal	Potri.001G079900	1.5595	2.5343	15.9555	6.49E-05	0.02316	
Polyclonal.vs. Monoclonal	Potri.009G033300	1.5221	7.0428	19.0087	1.30E-05	0.009958	
Polyclonal.vs. Monoclonal	Potri.009G141600	1.5134	4.6546	14.9408	0.0001109	0.03338	NAC domain containing protein 36
Polyclonal.vs. Monoclonal	Potri.017G137801	1.4575	2.4142	14.8878	0.0001141	0.03396	
Polyclonal.vs. Monoclonal	Potri.008G127600	1.4475	5.2747	16.2931	5.43E-05	0.02036	ribose-5-phosphate isomerase 2
Polyclonal.vs. Monoclonal	Potri.010G189600	1.3145	3.351	14.3462	0.0001521	0.04333	Alpha/beta hydrolase related protein
Polyclonal.vs. Monoclonal	Potri.005G240700	1.3026	5.8919	16.43	5.05E-05	0.01988	
Polyclonal.vs. Monoclonal	Potri.006G202200	1.2671	5.2708	14.2977	0.0001561	0.04354	syntaxin of plants 121
Polyclonal.vs. Monoclonal	Potri.012G043200	1.257	6.5853	22.5853	2.01E-06	0.002834	MAP kinase kinase 7

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.001G297200	1.2448	5.1137	17.467	2.92E-05	0.01477	Protein phosphatase 2C family protein
Polyclonal.vs. Monoclonal	Potri.005G257900	1.2322	6.2633	16.3013	5.40E-05	0.02036	basic region/leucine zipper motif 60
Polyclonal.vs. Monoclonal	Potri.018G099800	1.2095	7.3369	15.2259	9.54E-05	0.02936	tetraspanin8
Polyclonal.vs. Monoclonal	Potri.017G041600	1.1966	3.9159	23.6375	1.16E-06	0.002396	Phototropic-responsive NPH3 family protein
Polyclonal.vs. Monoclonal	Potri.002G005500	1.1939	5.9204	15.6953	7.44E-05	0.02508	phosphate transporter 1;8
Polyclonal.vs. Monoclonal	Potri.006G273100	1.1555	5.8456	22.8948	1.71E-06	0.002546	(1 of 9) KOG4210 - Nuclear localization sequence binding protein

Table B.3 (Continued)

Polyclonal.vs. Monoclonal	Potri.008G198000	1.1392	6.8142	14.2773	0.0001578	0.04356	(1 of 2) PF15365 - Proline-rich nuclear receptor coactivator (PNRC)
Polyclonal.vs. Monoclonal	Potri.010G182200	1.1183	6.543	16.9811	3.78E-05	0.01719	senescence associated gene 20
Polyclonal.vs. Monoclonal	Potri.013G132300	1.1114	5.1066	18.2324	1.96E-05	0.0119	RING membrane-anchor 1
Polyclonal.vs. Monoclonal	Potri.004G236100	1.0281	6.8632	15.4115	8.65E-05	0.02757	
Polyclonal.vs. Monoclonal	Potri.019G063500	-1.6353	5.3675	20.4831	6.02E-06	0.006197	root hair specific 2
Polyclonal.vs. Monoclonal	Potri.009G139800	-2.635	3.6364	17.3131	3.17E-05	0.01572	O-methyltransferase family protein
Polyclonal.vs. Monoclonal	Potri.001G389400	-3.0498	2.5485	16.947	3.84E-05	0.01719	pathogenesis-related family protein

Locus name: Location of Differentially Expressed Gene, LogFC(Fold Change), and LogCPM(Counts per Million). Descriptions are blank if no annotation was available.

Table B.4 Differential Gene Expressional Analysis for Model 3: Clone S7C8 Polyclonal vs. Clone S7C8 Monoclonal

Contrast	locusName	logFC	logCPM	Likelihood Ratio	PValue	False Discovery Rate	Description
Potri.009G115800	0.7015	5.1185	27.9318	1.26E-07	0.001593		
Red.Poly.vs.Mono	Potri.002G014000	4.9181	6.7781	22.301	2.33E-06	0.004802	Prolyl oligopeptidase family protein
Red.Poly.vs.Mono	Potri.016G126900	4.1193	5.0416	26.9808	2.06E-07	0.001593	receptor like protein 34
Red.Poly.vs.Mono	Potri.016G127000	3.7136	3.0195	25.7053	3.98E-07	0.00213	receptor like protein 34
Red.Poly.vs.Mono	Potri.003G072200	3.6772	-2.7654	23.4772	1.26E-06	0.003385	Protein of unknown function (DUF567)
Red.Poly.vs.Mono	Potri.018G011950	3.2168	1.5884	23.106	1.53E-06	0.003733	

Table B.4 (Continued)

Red.Poly.vs. Mono	Potri.016G127101	3.095	2.7457	23.7186	1.12E-06	0.003318	receptor like protein 34
Red.Poly.vs. Mono	Potri.004G072900	2.5335	3.6934	19.613	9.48E-06	0.01327	Protein of unknown function (DUF3464)
Red.Poly.vs. Mono	Potri.014G122000	2.3052	4.3509	20.5549	5.80E-06	0.0097	Adenine nucleotide alpha hydrolases-like superfamily protein
Red.Poly.vs. Mono	Potri.019G034000	2.2954	5.1385	17.4455	2.96E-05	0.0228	NDH-dependent cyclic electron flow 5
Red.Poly.vs. Mono	Potri.003G138600	2.2159	4.6409	16.5989	4.62E-05	0.02748	WRKY family transcription factor
Red.Poly.vs. Mono	Potri.012G026900	2.0857	2.2072	15.136	0.0001	0.04122	
Red.Poly.vs. Mono	Potri.017G045200	2.0681	3.3342	15.1446	9.96E-05	0.04122	

Table B.4 (Continued)

Red.Poly.vs. Mono	Potri.010G135100	1.9171	4.3266	15.4762	8.36E-05	0.0373	TCP-1/cpn60 chaperonin family protein
Red.Poly.vs. Mono	Potri.001G459200	1.8264	2.0145	18.9339	1.35E-05	0.0145	FAD-binding Berberine family protein
Red.Poly.vs. Mono	Potri.009G073000	1.8081	4.0639	16.3571	5.25E-05	0.0281	Protein phosphatase 2C family protein
Red.Poly.vs. Mono	Potri.001G234600	1.7297	5.5493	14.4884	0.000141	0.04905	exocyst subunit exo70 family protein H7
Red.Poly.vs. Mono	Potri.006G145600	1.7012	2.3788	16.566	4.70E-05	0.02748	(1 of 1) PF10998 - Protein of unknown function (DUF2838) (DUF2838)
Red.Poly.vs. Mono	Potri.008G031601	1.6002	-0.7101	16.8678	4.01E-05	0.02612	

Table B.4 (Continued)

Red.Poly.vs. Mono	Potri.010G129200	1.5118	3.2522	14.4463	0.0001442	0.04952	related to ABI3/VP1 2
Red.Poly.vs. Mono	Potri.011G087000	1.5027	1.3856	14.3961	0.0001481	0.04958	
Red.Poly.vs. Mono	Potri.011G157100	1.4951	2.5337	14.5264	0.0001382	0.04871	WRKY DNA-binding protein 35
Red.Poly.vs. Mono	Potri.010G116000	1.4517	3.7948	19.1605	1.20E-05	0.01392	(1 of 28) PF14009 - Domain of unknown function (DUF4228) (DUF4228)
Red.Poly.vs. Mono	Potri.001G066400	1.3727	9.0393	15.7212	7.34E-05	0.03391	CCR-like
Red.Poly.vs. Mono	Potri.012G123600	1.3543	4.2654	17.2797	3.23E-05	0.0228	Tetratricopeptide repeat (TPR)-like superfamily protein
Red.Poly.vs. Mono	Potri.019G131300	1.3525	6.3667	19.5288	9.91E-06	0.01327	

Table B.4 (Continued)

Red.Poly.vs. Mono	Potri.012G104300	1.3431	2.6095	20.6557	5.50E-06	0.0097	Uncharacterised protein family (UPF0497)
Red.Poly.vs. Mono	Potri.006G219800	1.3342	3.9982	23.7714	1.09E-06	0.003318	
Red.Poly.vs. Mono	Potri.018G048100	1.287	3.2386	14.4191	0.0001463	0.04958	
Red.Poly.vs. Mono	Potri.002G186200	1.2788	1.8054	19.0892	1.25E-05	0.01392	(1 of 1) PTHR36027: SF1 - ASYNAPTIC 3
Red.Poly.vs. Mono	Potri.015G008000	1.2058	6.3588	17.3715	3.07E-05	0.0228	Rhodanese/Cell cycle control phosphatase superfamily protein
Red.Poly.vs. Mono	Potri.012G007500	1.1243	4.7793	19.3938	1.06E-05	0.01356	TCV-interacting protein

Table B.4 (Continued)

Red.Poly.vs. Mono	Potri.002G158200	1.1202	4.9521	14.5438	0.0001369	0.04871	AT hook motif DNA-binding family protein
Red.Poly.vs. Mono	Potri.001G079900	1.119	2.5343	16.56	4.71E-05	0.02748	
Red.Poly.vs. Mono	Potri.002G010100	1.1063	6.7817	16.3887	5.16E-05	0.0281	BTB/POZ domain-containing protein
Red.Poly.vs. Mono	Potri.001G249800	1.0364	6.0297	18.1155	2.08E-05	0.01989	phosphate transporter 4;1
Red.Poly.vs. Mono	Potri.002G147100	1.0337	3.0568	15.6225	7.73E-05	0.0351	Protein of unknown function (DUF1442)
Red.Poly.vs. Mono	Potri.002G093300	-1.0016	3.5956	14.8062	0.0001191	0.04564	Phosphoglycerate mutase family protein
Red.Poly.vs. Mono	Potri.018G005300	-1.8199	2.7478	14.7749	0.0001211	0.0457	

Table B.4 (Continued)

Red.Poly.vs. Mono	Potri.003G066400	-2.0722	7.6333	17.7335	2.54E-05	0.02195	
Red.Poly.vs. Mono	Potri.009G139800	-2.1589	3.6364	22.7131	1.88E-06	0.004198	O-methyltransferase family protein
Red.Poly.vs. Mono	Potri.011G031700	-2.4352	0.4082	24.7787	6.43E-07	0.00287	NAD(P)-binding Rossmann-fold superfamily protein
Red.Poly.vs. Mono	Potri.018G051300	-3.056	4.2649	18.3921	1.80E-05	0.01784	
Red.Poly.vs. Mono	Potri.010G000600	-3.2184	2.2757	15.7433	7.26E-05	0.03391	
Red.Poly.vs. Mono	Potri.001G239700	-3.4988	3.207	15.9002	6.68E-05	0.03312	
Red.Poly.vs. Mono	Potri.003G066800	-3.8073	4.9522	24.0867	9.21E-07	0.003318	
Red.Poly.vs. Mono	Potri.001G259904	-4.5437	-2.0371	19.2498	1.15E-05	0.01392	

Locus name: Location of Differentially Expressed Gene, LogFC(Fold Change), and LogCPM(Counts per Million). Descriptions are blank if no annotation was available

APPENDIX C

GENE ONTOLOGY AND KYOTO ENCYCLOPEDIA OF GENES AND GENOMES

SUPPLEMENTARY DATA

Table C.1 Gene Ontology Analysis Output for All Three Models

Contrast	GO ID	Number of Genes	Direction	PValue	False Discovery Rate	TERM	ONTOLOGY
Polyclonal.vs. Monoclonal	GO:0009690	6	Down	0.00038	0.37	cytokinin metabolic process	BP
Polyclonal.vs. Monoclonal	GO:0019139	6	Down	0.00038	0.37	cytokinin dehydrogenase activity	MF
Polyclonal.vs. Monoclonal	GO:0004108	1	Down	0.0014	0.61	citrate (Si)-synthase activity	MF
Polyclonal.vs. Monoclonal	GO:0006101	1	Down	0.0014	0.61	citrate metabolic process	BP
Polyclonal.vs. Monoclonal	GO:0033180	7	Down	0.0016	0.61	proton-transporting V-type ATPase, V1 domain	CC
Polyclonal.vs. Monoclonal	GO:0010338	1	Up	0.0033	0.67	leaf formation	BP
Polyclonal.vs. Monoclonal	GO:0006595	3	Down	0.0035	0.67	polyamine metabolic process	BP

Table C.1 (continued)

Polyclonal.vs.Monoclonal	GO:0034477	1	Up	0.004	0.67	U6 snRNA 3’-end processing	BP
Polyclonal.vs.Monoclonal	GO:0003923	1	Down	0.0043	0.67	GPI-anchor transamidase activity	MF
Polyclonal.vs.Monoclonal	GO:0003952	2	Down	0.0053	0.67	NAD+ synthase (glutamine-hydrolyzing) activity	MF
Polyclonal.vs.Monoclonal	GO:0004367	4	Down	0.0053	0.67	glycerol-3-phosphate dehydrogenase [NAD+] activity	MF
Polyclonal.vs.Monoclonal	GO:0006893	2	Down	0.0057	0.67	Golgi to plasma membrane transport	BP
Polyclonal.vs.Monoclonal	GO:0005875	9	Up	0.0068	0.67	microtubule associated complex	CC
Polyclonal.vs.Monoclonal	GO:0004149	3	Down	0.0072	0.67	dihydrolipoyllysine-residue succinyltransferase activity	MF

Table C.1 (continued)

Polyclonal.vs. Monoclonal	GO:0045252	3	Down	0.0072	0.67	oxoglutarate dehydrogen ase complex	CC
Polyclonal.vs. Monoclonal	GO:0009331	5	Down	0.0082	0.67	glycerol-3-phosphate dehydrogen ase complex	CC
Polyclonal.vs. Monoclonal	GO:0048037	3	Down	0.0084	0.67		
Polyclonal.vs. Monoclonal	GO:0000062	8	Down	0.0087	0.67	fatty-acyl-CoA binding	MF
Polyclonal.vs. Monoclonal	GO:0003830	5	Up	0.009	0.67	beta-1,4-mannosylglycoprotein 4-beta-N-acetylglucosaminyltransferase activity	MF
Polyclonal.vs. Monoclonal	GO:0006904	14	Down	0.0094	0.67	vesicle docking involved in exocytosis	BP
White.Poly.vs .Mono	GO:0031012	14	Up	0.0017	0.98	extracellular matrix	CC

Table C.1 (continued)

White.Poly. vs.Mono	GO:0046522	2	Down	0.002	0.98	S-methyl-5-thioribose kinase activity	MF
White.Poly. vs.Mono	GO:0008429	2	Up	0.0025	0.98	phosphatidylethanolamine binding	MF
White.Poly. vs.Mono	GO:0009909	2	Up	0.0025	0.98	regulation of flower development	BP
White.Poly. vs.Mono	GO:0048573	2	Up	0.0025	0.98	photoperiodism, flowering	BP
White.Poly. vs.Mono	GO:0008897	5	Up	0.0034	1	holo-[acyl-carrier-protein] synthase activity	MF
White.Poly. vs.Mono	GO:0017176	5	Down	0.0037	1	phosphatidylinositol N-acetylglucosaminyltransferase activity	MF
White.Poly. vs.Mono	GO:0031347	11	Up	0.0042	1	regulation of defense response	BP

Table C.1 (continued)

White.Poly.vs.Mono	GO:0006506	11	Down	0.0055	1	GPI anchor biosynthetic process	BP
White.Poly.vs.Mono	GO:0009415	4	Up	0.0061	1	response to water	BP
White.Poly.vs.Mono	GO:0008237	28	Up	0.0079	1	metallopeptidase activity	MF
White.Poly.vs.Mono	GO:0004000	1	Down	0.0093	1	adenosine deaminase activity	MF
White.Poly.vs.Mono	GO:0009001	4	Down	0.0096	1	serine O-acetyltransferase activity	MF
White.Poly.vs.Mono	GO:0007067	6	Up	0.0098	1		
Red.Poly.vs.Mono	GO:0033180	7	Down	1.80E-05	0.035	proton-transporting V-type ATPase, V1 domain	CC
Red.Poly.vs.Mono	GO:0004364	2	Up	0.0003	0.2	glutathione transferase activity	MF
Red.Poly.vs.Mono	GO:0004106	4	Down	0.00049	0.2	chorismate mutase activity	MF

Table C.1 (continued)

Red.Poly.vs.Mono	GO:0009690	6	Down	0.00052	0.2	cytokinin metabolic process	BP
Red.Poly.vs.Mono	GO:0019139	6	Down	0.00052	0.2	cytokinin dehydrogenase activity	MF
Red.Poly.vs.Mono	GO:0003983	2	Down	0.0011	0.27	UTP:glucose-1-phosphate uridylyltransferase activity	MF
Red.Poly.vs.Mono	GO:0006011	2	Down	0.0011	0.27	UDP-glucose metabolic process	BP
Red.Poly.vs.Mono	GO:0003952	2	Down	0.0012	0.27	NAD+ synthase (glutamine-hydrolyzing) activity	MF
Red.Poly.vs.Mono	GO:0004861	6	Up	0.0015	0.27	cyclin-dependent protein serine/threonine kinase inhibitor activity	MF

Table C.1 (continued)

Red.Poly.vs.Mono	GO:0009331	5	Down	0.0017	0.27	glycerol-3-phosphate dehydrogenase complex	CC
Red.Poly.vs.Mono	GO:0009029	1	Up	0.0021	0.27	tetraacyldisaccharide 4'-kinase activity	MF
Red.Poly.vs.Mono	GO:0046835	8	Down	0.0022	0.27	carbohydrate phosphorylation	BP
Red.Poly.vs.Mono	GO:0006595	3	Down	0.0022	0.27	polyamine metabolic process	BP
Red.Poly.vs.Mono	GO:0004108	1	Down	0.0022	0.27	citrate (Si)-synthase activity	MF
Red.Poly.vs.Mono	GO:0006101	1	Down	0.0022	0.27	citrate metabolic process	BP
Red.Poly.vs.Mono	GO:0046417	3	Down	0.0023	0.27	chorismate metabolic process	BP
Red.Poly.vs.Mono	GO:0004367	4	Down	0.0028	0.29	glycerol-3-phosphate dehydrogenase [NAD+] activity	MF

Table C.1 (continued)

Red.Poly.vs.Mono	GO:0005875	9	Up	0.0028	0.29	microtubule associated complex	CC
Red.Poly.vs.Mono	GO:0006749	6	Down	0.0029	0.29	glutathione metabolic process	BP
Red.Poly.vs.Mono	GO:0000266	3	Up	0.003	0.29	mitochondrial fission	BP
Red.Poly.vs.Mono	GO:0000786	73	Up	0.0032	0.29	nucleosome	CC
Red.Poly.vs.Mono	GO:0000062	8	Down	0.0033	0.29	fatty-acyl-CoA binding	MF
Red.Poly.vs.Mono	GO:0006904	14	Down	0.0038	0.32	vesicle docking involved in exocytosis	BP
Red.Poly.vs.Mono	GO:0016624	8	Down	0.0048	0.35	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	MF

Table C.1 (continued)

Red.Poly.vs.Mono	GO:0003949	2	Up	0.0051	0.35	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase activity	MF
Red.Poly.vs.Mono	GO:0008534	2	Up	0.0052	0.35	oxidized purine nucleobase lesion DNA N-glycosylase activity	MF
Red.Poly.vs.Mono	GO:0004322	1	Up	0.0053	0.35	ferroxidase activity	MF
Red.Poly.vs.Mono	GO:0008483	26	Down	0.0056	0.35	transaminase activity	MF
Red.Poly.vs.Mono	GO:0004565	5	Down	0.0058	0.35	beta-galactosidase activity	MF
Red.Poly.vs.Mono	GO:0004470	9	Down	0.0059	0.35	malic enzyme activity	MF

Table C.1 (continued)

Red.Poly.vs.Mono	GO:0004471	9	Down	0.0059	0.35	malate dehydrogenase (decarboxylating) (NAD+) activity	MF
Red.Poly.vs.Mono	GO:0006659	2	Up	0.0062	0.35	phosphatidylserine biosynthetic process	BP
Red.Poly.vs.Mono	GO:0001735	2	Down	0.0062	0.35	prenylcysteine oxidase activity	MF
Red.Poly.vs.Mono	GO:0019030	1	Up	0.0063	0.35	icosahedral viral capsid	CC
Red.Poly.vs.Mono	GO:0046982	107	Up	0.0064	0.35	protein heterodimerization activity	MF
Red.Poly.vs.Mono	GO:0004748	3	Down	0.0066	0.35	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor	MF

Table C.1 (continued)

Red.Poly.vs.Mono	GO:0006893	2	Down	0.0067	0.35	Golgi to plasma membrane transport	BP
Red.Poly.vs.Mono	GO:0006480	1	Up	0.0071	0.35	N-terminal protein amino acid methylation	BP
Red.Poly.vs.Mono	GO:0051499	1	Down	0.0071	0.35	D-aminoacyl-tRNA deacylase activity	MF
Red.Poly.vs.Mono	GO:0003923	1	Down	0.0075	0.36	GPI-anchor transamidase activity	MF
Red.Poly.vs.Mono	GO:0016992	6	Up	0.0077	0.36	lipoate synthase activity	MF
Red.Poly.vs.Mono	GO:0003830	5	Up	0.0081	0.37	beta-1,4-mannosylglycoprotein 4-beta-N-acetylglucosaminyltransferase activity	MF
Red.Poly.vs.Mono	GO:0004335	3	Down	0.0083	0.37	galactokinase activity	MF

Table C.1 (continued)

Red.Poly.vs.Mono	GO:0003856	2	Down	0.0085	0.37	3-dehydroquinate synthase activity	MF
Red.Poly.vs.Mono	GO:0050662	69	Down	0.0088	0.37		
Red.Poly.vs.Mono	GO:0000287	121	Down	0.0089	0.37	magnesium ion binding	MF
Red.Poly.vs.Mono	GO:0016872	6	Down	0.0089	0.37	intramolecular lyase activity	MF
Red.Poly.vs.Mono	GO:0030915	6	Up	0.0091	0.37	Smc5-Smc6 complex	CC
Red.Poly.vs.Mono	GO:0006072	6	Down	0.0096	0.37	glycerol-3-phosphate metabolic process	BP
Red.Poly.vs.Mono	GO:0000123	7	Up	0.0098	0.37	histone acetyltransferase complex	CC

Number of Genes signifies the number of genes in that gene set associated with the Gene Ontology ID. Blank descriptions had no annotations available for that specific GO ID.

Table C.2 Kyoto Encyclopedia of Genes and Genomes Pathway Analysis Output for All Models

Contrast	PathwayID	NGenes	Direction	PValue	FDR	PValue.Mixed	FDR.Mixed	Description
White.Poly. vs.Mono	pop00920	40	Down	0.004	0.55	0.34	1	Sulfur metabolism
White.Poly. vs.Mono	pop00543	4	Down	0.0087	0.61	1	1	Exopolysaccharide biosynthesis
Red.Poly.vs. Mono	pop00330	63	Down	0.00033	0.026	0.093	0.26	Arginine and proline metabolism
Red.Poly.vs. Mono	pop00960	36	Down	0.0005	0.026	0.011	0.19	Tropane, piperidine and pyridine alkaloid biosynthesis

Table C.2 (continued)

Red.Poly.vs. Mono	pop00514	16	Down	0.00056	0.026	0.046	0.25	Other types of O-glycan biosynthesis
Red.Poly.vs. Mono	pop00950	31	Down	0.001	0.033	0.009	0.19	Isoquinoline alkaloid biosynthesis
Red.Poly.vs. Mono	pop00400	54	Down	0.0012	0.033	0.0011	0.077	Phenylalanine, tyrosine and tryptophan biosynthesis
Red.Poly.vs. Mono	pop00052	59	Down	0.0015	0.035	0.046	0.25	Galactose metabolism
Red.Poly.vs. Mono	pop00620	121	Down	0.0019	0.038	0.072	0.26	Pyruvate metabolism
Red.Poly.vs. Mono	pop00360	45	Down	0.0027	0.044	0.0054	0.19	Phenylalanine metabolism

Table C.2 (continued)

Red.Poly.vs.Mono	pop00350	56	Down	0.0028	0.044	0.023	0.19	Tyrosine metabolism
Red.Poly.vs.Mono	pop00380	52	Down	0.0065	0.085	0.076	0.26	Tryptophan metabolism
Red.Poly.vs.Mono	pop01110	1334	Down	0.0071	0.085	0.046	0.25	Biosynthesis of secondary metabolites
Red.Poly.vs.Mono	pop00905	15	Down	0.0074	0.085	0.025	0.19	Brassinosteroid biosynthesis
Red.Poly.vs.Mono	pop01200	320	Down	0.0093	0.099	0.023	0.19	Carbon metabolism
Polyclonal.vs.Monoclonal	pop00620	121	Down	0.0047	0.25	0.16	0.67	Pyruvate metabolism

Table C.2 (continued)

Polyclonal.vs.Monoclonal	pop00514	16	Down	0.0054	0.25	0.089	0.67	Other types of O-glycan biosynthesis
Polyclonal.vs.Monoclonal	pop00020	68	Down	0.0075	0.25	0.008	0.67	Citrate cycle (TCA cycle)

Table C.3 Kyoto Encyclopedia of Genes and Genomes Mixed Analysis Output

Contrast	PathwayID	NGenes	Direction	PValue	FDR	PValue.Mixed	FDR.Mixed	Description
Red.Poly.vs.Mono	pop00941	46	Down	0.027	0.18	0.0069	0.19	Flavonoid biosynthesis
Red.Poly.vs.Mono	pop04146	114	Down	0.74	0.89	0.0035	0.16	Peroxisome
Red.Poly.vs.Mono	pop00760	26	Down	0.75	0.89	0.00061	0.077	Nicotinate and nicotinamide metabolism