2023

SEQUENCING ANTISENSE TRANSCRIPTOME OF POPULUS THROUGH ANTISENSE/SENSE TRANSCRIPT PAIR ENRICHMENT

Emma V. Burke
Michigan Technological University, evburke@mtu.edu

SEQUENCING ANTISENSE TRANSCRIPTOME OF *POPULUS* THROUGH ANTISENSE/SENSE TRANSCRIPT PAIR ENRICHMENT

By

Emma V. Burke

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Forest Molecular Genetics and Biotechnology

MICHIGAN TECHNOLOGICAL UNIVERSITY

2023

This thesis has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Forest Molecular Genetics and Biotechnology.

College of Forest Resources and Environmental Science

Thesis Advisor:	*Yinan Yuan*
Committee Member:	*Carsten Külheim*
Committee Member:	*Hairong Wei*
College Dean:	*David Flaspohler*

Table of Contents

Author Contribution Statement

This thesis contains custom tailing data that is in the process of being prepared for publishing.

RNA extractions of tissue samples, GuSCN annealing experiments, HAP column preparations, column runs, and sample concentration and desalting, sample prep and shipping for Novagene mRNA sequencing were completed by Emma Burke.

Propagation of *Populus* plants was done by Emma Burke, under the supervision of Yinan Yuan.

J2 Antibody experiments, and mRNA purifications were completed by Emma Burke and Yinan Yuan.

Library preparation, custom tailing for Nanopore sequencing, sequencing data alignment and mapping to reference genome was completed by Yinan Yuan.

Drought treatment of Populus plants was completed primarily by Emma Burke, drought treatment was supplemented by Reed Arneson on some occasions.

Quality control checks of samples including Qubit quantification, Nanodrop, and agarose gels were completed by Emma Burke, Reed Arneson, and Yinan Yuan.

Acknowledgements

Thank you to CFRES, and my advisor Dr. Yuan for the opportunity to continue my education and all the support along the way. Thank you, Reed, for being my companion in the lab these last few years and helping me through all the coding.

I would also like to thank CFRES janitorial and maintenance staff specifically for always being there when I needed them, but more importantly for always greeting me with a smiling face.

Dan and Ben, who moved with me across the country and deserve their own psychology degrees for helping me through these last 2 years, I wouldn't be here without you.

I'd like to thank GSG Softball for helping me make friends I will hold close to my heart for the rest of my life. Big shout out to TreeFolk, we weren't good, but we did have fun.

Lastly, I'd like to thank Nikita for providing me with support when I needed it, and for always having an escape plan ready if I ever wanted it.

List of Abbreviations

APA- Alternative Polyadenylation

cDNA- Complementary DNA

dsDNA- Double Stranded DNA

dsRNA- Double Stranded RNA

FC- Field Capacity

HAP- Hydroxyapatite powder

lncRNA- Long non-coding RNA

MAB- Monoclonal Antibody

mRNA- Messenger RNA

PCA- Principal components analysis

rRNA- Ribosomal RNA

RT- Reverse transcription

RTA- Reverse Transcription Adapter

ssDNA- Single Stranded DNA

ssRNA- Single Stranded RNA

Abstract

Past transcriptome research on plants focused primarily on protein coding genes, and only recently researchers began looking into the non-protein coding regions that may play significant roles in gene regulation. Antisense RNA transcripts that are found naturally in the cell share complementary sequence with sense transcripts and have been shown to regulate expression of their sense counterparts. Since antisense RNA has been largely under-studied and difficult to sequence because of their low relative abundance, new methods are needed to target antisense RNA for efficient genome-wide profiling. To address this gap in methods to efficiently and cost effectively enrich antisense RNA transcripts for sequencing, we developed methods that allow for the enrichment of antisense RNA through chemically guided annealing of sense/antisense transcript pairs of RNA and cDNA, and the separation of double stranded structures for library formation. These methods utilize guanidinium thiocyanate (GuSCN) to enrich antisense transcripts through the annealing of sense/antisense transcript pairs. Annealed RNA can then be separated through J2 monoclonal anti-dsRNA antibody binding, mRNA purification of enriched transcript pairs, and the custom tailing of sense/antisense transcript pair enriched libraries for Nanopore direct RNA sequencing. We also developed a novel method for the separation of annealed double stranded first strand cDNA using hydroxyapatite powder (HAP) chromatography to form antisense enriched Illumina libraries. The long-term goal of this project is for these methods to be used to form targeted antisense enriched libraries for the genome wide characterization of the antisense RNA response to drought stress in *Populus.*

1 Introduction

Past transcriptome research on plants has focused primarily on protein coding genes, and only recently have researchers begun looking into the non-protein coding regions that may play significant roles in gene regulation. In multicellular organisms nearly every cell contains the same genome, however not every gene is transcriptionally active in every cell, leading to high diversity in overall gene expression between different cells and tissues in an organism. It is estimated less than 5% of total RNA transcripts are translated into protein, most RNA transcripts are ribosomal RNA (rRNA) and non-coding transcripts that can play a role in gene expression and regulation but are not directly translated into proteins (Frith et al 2005). Over the last decade evidence accumulated shows long non-coding RNAs (lncRNAs) are widely expressed and play a key role in gene regulation. lncRNA are defined as RNAs longer than 200 nucleotides that are not translated into proteins. lncRNA can be sense or antisense transcripts that overlap with other gene regions. Depending on the location and specific interactions with DNA, RNA, and protein lncRNA can affect chromatin function, alter the translation and stability of mRNA, and interfere with signaling pathways. Many of these functions influence gene expression (Keller et al 2018).

Antisense RNA transcripts that are found naturally in the cell share complementary sequence with sense transcripts and have been shown to regulate expression of their sense counterparts. (Kos et al 2009) The exact mechanisms in which antisense transcripts function in the cell is still not well studied but there is evidence that antisense RNA is differentially expressed in response to stress stimuli, indicating they play a role in the response to stress in plants. (Wang et al 2005; Kos et al 2009; Xu et al 2017; Yuan & Chen 2018) The function of antisense RNA in the cell is not only to regulate the translation of mRNA but also has been shown to affect the epigenetic regulation of various genes in eukaryotic cells. There is numerous antisense RNAs with repressive functions that have been described in different organisms. Early research into antisense RNA showed in vitro and in vivo that antisense RNA caused repression of translation through complementary sense/antisense transcript pair formation. (Cui 2010; Singer et al 1963; Tomizawa et al 1981) Recent studies demonstrated that antisense RNA may silence its sense gene at the transcriptional level, affecting DNA methylation and histone modifications, and leading to epigenetic silencing. (Tufarelli et al 2003) Antisense RNA loci have been found near several tumor suppressor genes (TSG). In tumorigenesis antisense RNA can regulate the transcription of their sense counter parts. In many cancers TSG are often silenced through DNA methylation that is induced through antisense RNA binding. P15 is a TSG that is often deleted or silenced through DNA methylation in a wide variety of tumors including leukemia, melanoma, and lung cancers. A P15 antisense RNA has been identified that induced the silencing of P15 through DNA methylation (Yu et al 2008). Since in mammalian cells antisense RNA has been shown to regulate both translation and transcription, it is likely that antisense RNA also has diverse methods of regulating genes expression in plants as well.

The regulatory role of antisense RNA on gene expression on a translation and transcription level means it is crucial we fully understand the role antisense RNA plays in

the stress response, so when we are engineering trees in the future, we can more efficiently target areas of the genome that would promote stress tolerance whether that be through protein coding genes or non-coding areas of the genome that regulate other cell processes. To look specifically at the differential expression of antisense RNA, we must develop methods to enrich antisense RNA from total RNA. Since antisense RNA has been largely under-studied and difficult to sequence because of their low relative abundance, new methods are needed to target antisense RNA for efficient genome-wide profiling.

To fully study antisense RNA, it is important we can capture the full length of RNA transcripts. Nanopore direct RNA sequencing allows for the full length of RNA transcripts to be sequenced. The full-length transcripts allow us to look at alternative splicing and alternative polyadenylation isoforms of each transcript. Alternative splicing of RNA can affect localization, stability, and function of the RNA in the cell. Splicing events can differ in response to stimuli or cell cycle. In multicellular organisms many tissue or developmental changes follow specific RNA splicing patterns. Alternative splicing is tightly controlled to ensure the correct isoforms are found in the correct location and proportion. (Dujarin et al 2013) Although alternative splicing is often studied regarding mRNA transcripts it is important for us to capture the full length of non-coding antisense RNA transcripts to study how these different antisense isoforms may play a role in gene regulation. Alternative polyadenylation (APA) is a common post-transcriptional process in eukaryotes that generates altered 3' ends of transcripts. Differences in polyadenylation have effects on transcriptome diversity and regulating gene expression. APA has been found across many plant species and has been suggested to have important regulatory functions at the molecular level in plants. (Yan et al 2021) To have a better understanding of the differential expression of antisense RNA polyadenylated isoforms it is important to use methods to enrich antisense RNA, which is found at a much lower relative abundance, and use sequencing methods that capture the full length of RNA transcripts.

There are advantages to utilizing different sequencing platforms for a variety of projects. The advantages of utilizing Nanopore direct RNA sequencing are minimal sample preparation, the ability to sequence full length reads, directly sequence RNA, and the ability to detect base modifications. Using Nanopore direct RNA sequencing allows for the capture of the full length of the transcripts of interest and any base modifications in real time. Some of the drawbacks of Nanopore include its relatively lower base calling accuracy, higher cost per gene coverage compared to other sequencing platforms and the high amount of starting material required for library formation. Nanopore direct RNA sequencing was originally designed for mRNA sequencing, and uses adapters designed to capture polyadenylated RNA strands. Although Nanopore is a relatively new sequencing platform many researchers develop new techniques to utilize this technology alongside industry professionals. By adjusting the adaptors used with Nanopore direct RNA sequencing researchers have been able to develop ways to directly sequence 16s rRNA transcripts (Smith et al., 2019). These developments allowed for the direct observation of base modifications on rRNA transcripts that are erased when converted into cDNA for

Illumina sequencing. For us to use Nanopore direct RNA sequencing to target antisense lncRNA transcripts we need to pair our enrichment methods with custom tailing methods that allow Nanopore adapters that recognize polynucleotide tails to capture these antisense transcripts which are not found naturally polyadenylated in the cell. The addition of unique polynucleotide tails that can be differentiated from native polyadenylated tails when sequenced on Nanopore direct RNA can be used to capture non-polyadenylated transcripts while maintaining the ability to determine native poly(a) tail length when sequenced on Nanopore direct RNA sequencing. Inosine triphosphate (ITP), and guanosine triphosphate (GTP) are oligomeric 5'-triphosphates that can be used in place of adenosine triphosphates (ATP) to form Poly(I) and Poly(G/I) tails (Corfu et al 1990). 2'-O methyl ATP can also be used in place of ATP to form poly(mA) tails. Since Nanopore direct RNA sequencing can detect base modifications, Poly(mA), Poly (I) and Poly (G/I) tails that are added to transcripts that are naturally polyadenylated in the cell can be easily differentiated from the natural Poly(A) tail length so that the effects of alternative polyadenylation can be studied while capturing non-polyadenylated transcripts for sequencing.

When working with low concentration starting materials or limited funding, other sequencing platforms may be better suited for the project at hand. The advantages of Illumina sequencing are high read accuracy, low-cost sequencing, low starting material requirements, and well established and widely utilized protocols. Illumina sequencing uses short read sequencing that does not capture the full length of transcripts and can be difficult to study alternative isoforms. Since Illumina cannot directly sequence RNA, RNA needs to be reverse transcribed into complementary DNA (cDNA) so it can be sequenced using Illumina (Leggett and Clark 2017). Although Illumina sequencing does not give you a full picture about the target RNA species due to the loss of base modifications and short read length, it allows for you to accurately and cost effectively capture RNA transcripts even if RNA quantities are low. By utilizing both types of sequencing platforms we can enhance our ability to capture antisense RNA species using a variety of methods.

We have found guanidinium thiocyanate (GuSCN) at high concentrations promotes DNA-DNA, DNA-RNA, and RNA-RNA hybridization, thus we employ GuSCN to anneal antisense/sense transcript pairs to enrich for antisense RNA for library formation. (Gillespie et al 1987) GuSCN annealing is used to enrich antisense RNA transcripts, found in low relative abundance and that can be difficult to isolate for targeted sequencing, through the formation of transcript pairs. The enriched sense/antisense transcript pairs need to be separated from total RNA species before they can be sequenced. The enrichment step using GuSCN can be utilized for RNA-RNA binding for mRNA purification with the annealed antisense transcript, or separation of dsRNA using J2 antibodies as well as DNA-DNA binding for HAP column separation.

J2 anti-dsRNA monoclonal antibody (MAB) can be used to isolate dsRNA. They have a high specificity for natural dsRNA transcripts and have been used to target dsRNA in vivo and invitro. (Schönborn et al., 1991) J2 antibodies can be used to isolate natural

dsRNA and we aim to employ this method to isolate annealed sense/antisense transcript pairs for the formation of antisense enriched libraries. Due to their nonspecific binding with ribosomal RNA (rRNA) which is highly abundant in total RNA populations, we use J2 antibodies after rRNA removal or mRNA purification to increase the total yield of target antisense transcripts.

Hydroxyapatite powder (HAP) chromatography has been routinely used to separate double stranded DNA molecules from single stranded DNA molecules (Smith 1998). Although HAP chromatography has been around since the 1950s there are now better ways to do most of the tasks it was originally designed for and has all but disappeared from standard laboratory techniques. We aimed to use a modified version of these established methods to isolate double stranded first strand cDNA (ds-cDNA) after chemical annealing with GuSCN. The eluted duplex of antisense and sense transcript pairs can then be used to form antisense RNA enriched libraries for sequencing. This method would provide a novel technique for antisense transcript enrichment for sequencing.

Our lab is looking into the functional role of antisense RNA in the drought response of the *Populus* genus. Poplar trees are grown around the world and are primarily used to produce wood-based products but have also been used as a bioenergy crop in recent years. (Port and El-Kassaby 2015; Simmons et al 2008) Changing climate has been shown to have a major effect on forest trees and poses a serious threat to the future productivity and use of forest trees for environmental and industrial uses. (Allen 2010) The *Populus* genus is used as a model organism to study transcriptional response due to its sequenced genome and the abundance of genetic resources available for the genus. (Tuskan et al 2006; Viger et al 2016) *Populus* has a high water demand to maintain high productivity, and past research has identified many drought resistant genes that are responsible for drought regulation, many of which being transcription factors. (Chen et al 2013; Shuai et al 2016) Past transcriptome research on *Populus* has focused mainly on protein coding genes, and only recently have researchers begun looking into the non-protein coding regions that may play significant roles in gene regulation in response to drought stress. (Yuan & Chen 2018) To better study the antisense RNA response to drought stress we must first develop methods that allow for the enrichment of antisense RNA transcripts for sequencing. To do this, we developed methods that would allow for the enrichment of antisense RNA through chemically guided annealing of sense/antisense transcript pairs, and the separation of double stranded structures for library formation.

The objective of this project is to develop methods that allow for targeted sequencing of antisense RNA transcripts using chemically guided annealing of sense/antisense transcript pairs, and the separation of these annealed sense/antisense transcript pairs to form antisense RNA enriched libraries for both direct RNA sequencing and cDNA sequencing. To meet these objectives, we used GuSCN to chemically anneal sense and antisense transcript pairs using both total RNA, and first strand cDNA. We then use a variety of methods to separate these annealed transcript pairs from total RNA. These methods include using J2 antibodies to target and separate dsRNA from total RNA for Nanopore direct sequencing, custom tailing enriched RNA libraries to increase overall

antisense RNA reads with Nanopore sequencing and developing a novel approach to isolate annealed sense/antisense transcript pairs of first strand cDNA with HAP chromatography. The next step for this project is to confirm the HAP column method with Illumina sequencing. If sequencing confirms GuSCN annealing paired with HAP chromatography enriches antisense/sense transcript pairs then this would provide a novel method for targeted antisense RNA library formation that can be widely applicable to further the study of antisense RNA. The long-term goal of this project is for these enrichment methods to be used to form sense/antisense transcript pair enriched libraries to study the antisense RNA response to drought stress in *Populus*.

2 Materials and Methods

2.1 Developing methods to Enrich Antisense RNA

2.1.1 Total RNA extraction, mRNA purification and ribosome RNA depletion (rRNA-)

Total RNA from *Populus* leaf tissue from drought experiments described in section 2.3, was extracted by CTAB method (Chang et al. 1993), digested with DNase I (Thermo Fisher, CAT#EN0525), and used for mRNA purification, rRNA removal and other downstream applications. Poly(A) RNA was purified from total RNA using Oligo d(T)25 Magnetic Beads from New England Biolabs (NEB, CAT# S1419S), and rRNA removal was done with riboPOOLs (siTOOLS BIOTECH, CAT#dp-K012-000101) following the manufacturer's instructions.

2.1.2 In vitro transcription template generation

DNA template for λRNA transcripts was generated from λDNA (NEB, N3011S) through PCR amplification with a forward primer (5'-ATTTAGGTGACACTATAGAAGGTTC AGGGTTGTCGGACTTG) containing a SP6 promoter region, and a reverse primer (5'-TGGCGAACAACAAGAAACTG) terminating the targeted region. PCR amplification was obtained using Phusion™ Hot Start II DNA Polymerase (Thermo Scientific: F549S) in a reaction containing 1 x Phusion HF Buffer, 2 ng/ul λDNA, 0.2 mM dNTP mix, 0.5 uM forward primer, 0.5 uM reverse primer, 0.02 U/μL of Phusion Hot Start II DNA Polymerase., with initial denaturation of 98 °C for 30s, and 33 cycles of [98°C 5s, 58°C 10s, 72°C 15s], then 72°C 10 min. The amplified PCR product was then loaded on 1 % agarose gel to verify the length and purity. Finally, the PCR amplified template was purified using AMPure XP beads (Beckman, A63880) following the manufacturer's instructions.

2.1.3 λRNA SP6 in vitro transcription

The template generated from above was then transcribed using the HiScribe® SP6 RNA Synthesis Kit (NEB, E2070S) in a 25 ul reaction containing 1 x SP6 Reaction Buffer, 5 mM of each ATP, CTP, GTP and UTP, 1 ug of template DNA, and 2.5 ul SP6 RNA Polymerase Mix. The reaction was incubated at 37°C for 2 hours. The transcribed RNA was then separated on a 6 % 7M urea polyacrylamide gel (Invitrogen: EC6865BOX) alongside with a ssRNA ladder (NEB: N0364S). The gel was post stained with GelRed (Biotium: 41003), and the expected size of transcribed RNA was identified and excised from gel. Gel slices were mixed with RNA elution buffer (0.3 M NaOAc, pH 5.2, 0.2 % SDS, 1 mM EDTA, 10 ug/mL proteinase K) at 4°C overnight. The eluted RNA was purified with 25:24:1 phenol:chloroform:isoamyl alchohol and precipitated with 100% ethanol and 1/10 volume of NaOAc. The pelleted RNA was further washed with 70% ethnol twice and dissolved in RNase-free H_2O for 3' end custom tailing experiments.

2.1.4 GuSCN Annealing

Samples that are annealed with 4M GuSCN are prepared by mixing 6M GuSCN with cDNA/RNA sample to bring the total concentration of the solution to 4M GuSCN. The solution is then incubated for 3 days at room temperature on a Hula mixer. After incubation is complete excess GuSCN is removed using CleanNA Clean NGS beads (Ref No. CNGS-0050) and desalted using Sigma GenElute -E RNA single spin column (CAT#EC800) following manufactures protocol. Thermo Fisher RNase T1 (CAT# EN0541) digestion following manufacturer protocol was used to confirm the presence of dsRNA. Thermo Fisher dsDNase (CAT# EN0771) digestion following manufacturer protocol was used to confirm the presence of ds-cDNA.

2.1.5 J2 Binding

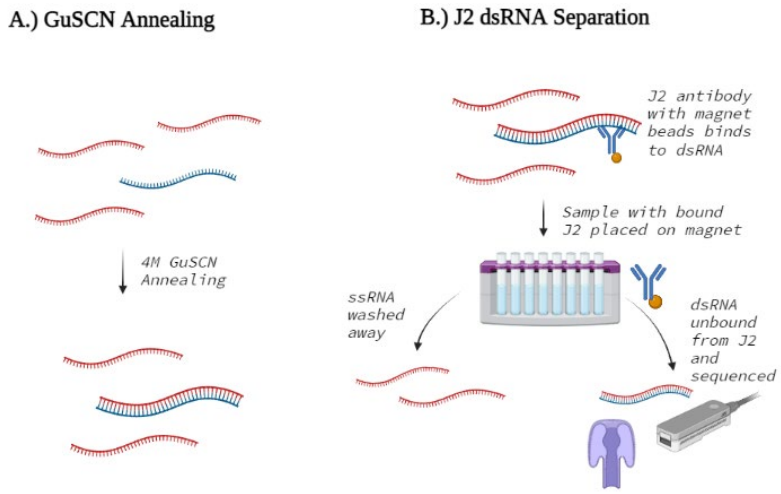

Fig 2.1.5. Workflow for J2 antibody dsRNA separation. A.) GuSCN Annealing method. B.) J2 dsRNA separation. Created with BioRender.com

GuSCN annealed antibody was mixed with 50 µl of Jena BioScience J2 anti-dsRNA MAB (CAT#RNT-SCI-10010500) and allowed to incubate overnight on Hula mixer at 4° C. 50 µl Thermo Fisher Protein G Dynabeads (CAT# 10003D) that were prepared with manufacturers recommendations are added to sample mixer, and allowed to bind to J2 antibody for 4 hours at 4 degrees C. After incubation J2 dynabead sample mixer is placed on magnet, and supernatant is removed to new tube. 250 µl 1x PBS + 0.1 Tween20 is used to wash beads 4x. For each wash beads are resuspended and incubated for 5 minutes at room temperature. After washing beads are resuspended in 100ul 1x PBS + 0.1 Tween20 and 300 µl Trizol reagent. Mix is vortexed 1 min and incubated on hula mixer for 20 minutes at room temperature. 60 µl chloroform is added to mixer and vortexed 1 min. Entire solution mixture is transferred to QuantaBio Phase Lock Heavy

Gel (CAT# 10847-802) and centrifuges 5 min 14k RPM for 5 min. Upper phase containing dsRNA is transferred to new tube (Fig 2.1.5).

The final elution containing dsRNA is DMSO treated using 80% DMSO at 65 ° C for 60 minutes to denature annealed sense/antisense pairs. DMSO is removed using CleanNA Clean NGS beads (Ref No. CNGS-0050) with an adjusted protocol. Our adjusted protocol allowed for 45 minutes for initial incubation to allow for RNA binding to the beads in the presence of high concentrations of DMSO. During the wash steps beads were resuspended in 70% ethanol and incubated for 5 minutes between washes. Final elute is custom tailed following protocol described in 2.1.5 prior to library formation.

To confirm J2 ability to bind dsRNA New England BioLabs dsRNA ladder (CAT# N036S) was bound with J2 antibody and Dynabeads.

2.1.6 Polynucleotide tailing of 3' end of RNA

To add poly(nucleotide) tails to cellular RNAs, 100 to 500 ng of target RNAs are incubated with 5 Units of E. coli Poly(A) Polymerase (NEB, CAT# M0276), 1 μM ATP, and 1 x E. coli Poly(A) Polymerase Reaction Buffer in 20 μl reaction volume, at 37°C for 3 min. For poly(G/I) tailing, same amount of RNA is used with 1 μM of GTP and ITP in 20 μl reaction containing 4U of poly(U) polymerase (NEB, CAT#M0337S), and 1 x NEBuffer™ 2, and incubated for 60 min at 37°C. The standard reaction for adding 2'-O-methyl ATP to form Poly(mA) tail is as follows: various amount of RNAs in 5 x yeast poly(A) polymerase buffer in volume up to 20 μl containing 600 Units of yeast poly(A) polymerase (Thermo Fisher, CAT#74225Z25KU), 1 μM of 2'-O-methyl ATP, was incubated for 60 min at 37°C.

2.1.7 Nanopore direct RNA library construction and sequencing

RNAs from various sources are used for nanopore direct RNA library construction. To sequence poly(A) tailed or poly(mA) tailed RNAs on nanopore, around 300 to 500 ng of mRNA enriched through oligo(dT) or custom tailed RNAs with either poly(A) or poly(mA), are used and libraries are constructed following exactly as Oxford Nanopore Technologies (ONT) recommendation with kit SQK-RNA002. To capture poly(G/I) tailed RNA for nanopore sequencing, SQK-RNA002 kit is used with an exception that a custom oligo(dC) adapter was used replacing the olio(dT) RTA adapter provided with the kit. Custom synthesized oligo(dC) adapters are annealed 1:1 at 1.4 μM in buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl) by heating to 95°C for 2 min and letting them cool down slowly at RT for 30 min. Libraries were sequenced on the MinION using ONT R9.4.1 flow cell (FLO-MIN106D) running the standard MinKNOW software with Basecalling tab set to OFF. For some libraries run on the same flow cell, the ONT flow cell wash kit (EXPWSH004) is used between runs to remove the previous library from flow cell.

2.1.8 Base-calling and mapping to Populus reference genome

Raw fast5 files generated from MinION sequencer are base called with Guppy Basecalling Software, version 6.4.6+ae70e8f, provided by Oxford Nanopore

Technologies. with the configuration file “rna_r9.4.1_70bps_hac” (Wick et al. 2019). The base called reads in fastq format are then mapped to *Populus trichocarpa* reference genome v3.0 through minimap2 version 2.17-r941 (Li. 2019) with the parameters: -a - x splice -k14 -uf. The aligned sam file is further filtered with samtools: samtools view -h -F 2324 (Danecek et al., 2021), to remove unmapped, secondary, and supplementary reads, and the filtered sam/bam is then used for downstream analysis.

2.1.9 Quantification of sense and antisense gene expression with featureCounts

FeatureCounts (Liao et al., 2014) is used to count the reads mapped to either sense or antisense strand of annotated Populus genes with the parameters set as following for counting reads mapped to either antisense orientation: featureCounts -a Populus_annotation.gtf -g gene_id -L -s 2, or sense orientation: featureCounts -a Populus_annotation.gtf -g gene_id -L -s 1.

2.2 Developing methods to Enrich Antisense cDNA

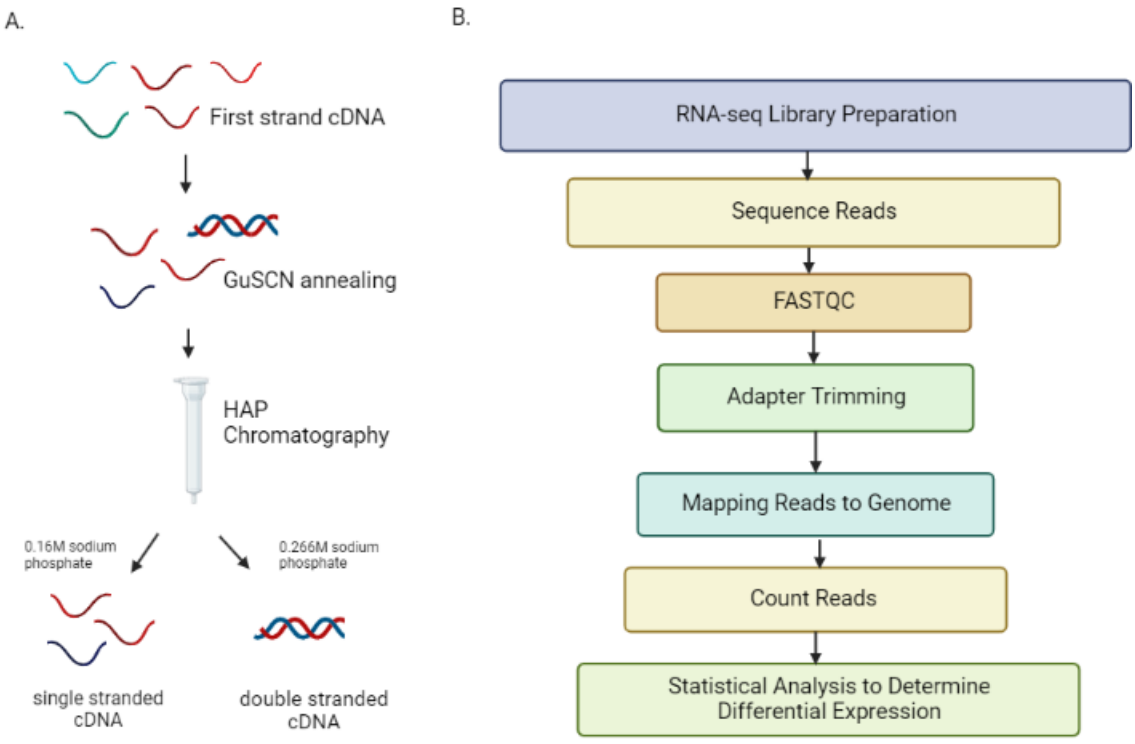

Fig 2.2.1. Workflow for HAP column separation and sequencing. A. GuSCN annealing and HAP chromatography separation B. Sequencing and analysis. Created with BioRender.com

2.2.1 RNA Fragmentation and DMSO treatment

Dnase1 treated RNA from *Populus* mature leaf tissue was fragmented using NEBNext Magnesium RNA fragmentation module (CAT#E6150S) following manufacturers protocol. For both control and drought treated samples RNA was DMSO treated with 80% DMSO at 65 ° C for 60 minutes to denature natural dsRNA transcript pairs, and then CleanNA Clean NGS beads (Ref No. CNGS-0050) using adjusted protocol described in section 2.1.3 and run through a Performa EdgeBio column (CAT#73328) to remove remaining DMSO.

2.2.2 First Strand cDNA Synthesis and GuSCN Annealing

Fragmented RNA both DMSO treated and control non-DMSO treated were reverse transcribed (RT) into first strand cDNA. First strand cDNA was annealed with 4M GuSCN for 3 days at room temperature. Excess GuSCN was removed using CleanNA Clean NGS beads (Ref No. CNGS-0050) following manufacturers protocol and eluted in 200 µl nuclease free water.

2.2.3 HAP Column Prep

1g of Bio-Rad Bio-Gel HTP hydroxyapatite powder (CAT#130-0520) is suspended in 15ml 0.01M sodium phosphate buffer. Powder is allowed to settle for 10 minutes and is then decanted and resuspended in 15ml 0.01M sodium phosphate buffer. This step is repeated 3 times. After the final decantation HAP suspended is allowed to soak overnight stored at 4 ° C. The next day 2ml of resuspended HAP is gravity packed into column. HAP is washed 3x with 1ml 0.01M sodium phosphate buffer. After the final wash column can be loaded with sample or stored at 4 ° C stopped and capped.

2.2.4 HAP Elution of Sense/Antisense Transcript Pairs

800 µl 0.01M sodium phosphate buffer was added to the 200 µl annealed cDNA samples. 1ml total solution was loaded onto stopped HAP column and allowed to bind for 15 minutes. Solution is allowed to flow through after 15 minutes. Elutes are loaded onto the stopped column and allowed to incubate 10 minutes before flow through of the column. Each elute is collected in 10ml tube after each flow through and stored for concentration and quantification. Elutes that are ran through the column are 0.05M (x2), 0.12M (x3), 0.16M (x4), 0.266M (x4), and 0.4M (x2) sodium phosphate buffer at 65 ° C (Fig 2.2.1).

2.2.5 Purification of ds-cDNA Elutes

Elutes are concentrated using butanol sec. 1 volume butanol is added to the sample and spun in centrifuge for 30s at 14,000rpm. The top layer is removed, and this step is repeated until sample reaches 100 µl. Samples are then quantified using Qubit ssDNA kit. Concentrated dsDNA samples are desalted using CleanNA Clean NGS beads (Ref No. CNGS-0050) following standard protocol. After bead purification they are run through a Performa EdgeBio column (CAT#73328) to remove remaining sodium phosphate. Samples are then concentrated using centrivac. Elutes were visualized on agarose gel to observe the size differences in elution concentrations, and dsDNase, which digested

dsDNA, digestions were done on elutes to confirm the presence of dsDNA in 0.266M buffer concentration elutes.

2.3 Drought experiment setup and design

2.3.1 Propagation

Drought experiments were done on 2 species of Poplars: *P. trichocarpa* and 717 hybrids (*P. alba* x *P. tremula*).

All *P. trichocarpa* plants were propagated using a stem cutting method previously used in other publications from our lab (Yuan & Chen 2018). The rooted cuttings were grown in the greenhouse with a temperature range of 20–35 °C. Cuttings were allowed to establish in misting chamber for 5 weeks prior to being transferred into fertilized soil. 717 hybrid plants were propagated using tissue culture method, once plants were established in medium they were moved to soil and kept on a light rack for 2 weeks to adjust to soil and humidity before being transferred to fertilized soil in the main green house. Methods for design and collection for both species were the same.

2.3.2 Drought Treatment

All plants were watered to field capacity (FC) for 1 month prior to the start of drought experiments. At the beginning of each of the experiments all plants were measured, pots at full FC were weighed, and 40% FC based off those weights was calculated. Plants were randomly separated into control and drought treated groups.

Control plants were weighed and watered every day at the same time back up to the full FC weight. Drought treated plants were weighed every day and withheld water until they reached 40% FC. For short term drought treated samples the plants were held at 40% FC for 24 hours before tissue collection. For long term drought treated samples plants were held at 40% FC for 21 days prior to tissue collection.

2.3.3 Tissue Collection

Upon the completion of drought treatment plant tissue was collected from apex (newly emerged folded leaf shoots), young leaf (unfolded expanding leaves), mature leaf (fully expanded leaves), and root tissue. Samples were flash frozen with liquid N2 once removed from the plant, and were either extracted directly after, or were stored in -80 ° C freezer until extraction. Extractions of *Populus* RNA was done following the same protocol described in section 2.1.1.

2.3.4 Sample Prep for Sequencing

Dnase1 treated RNA was prepared and shipped for mRNA Illumina sequencing using Novogene services. 2 control and 2 long term drought treated samples were selected for both *P. trichocarpa* as well as 717 hybrid mature leaf tissue samples.

2.3.5 Illumina Data Analysis

The sequencing data that we obtained was then mapped to *P. trichocarpa* reference genome v3.0 genome using STAR version 2.7.5b, SAMtools (Li et al. 2009) was used for bam file manipulation and related analysis. FeatureCounts (Liao et al., 2014) is used to count the reads mapped to either sense or antisense strand of annotated *P. trichocarpa* reference genome v3.0 genes with the parameters set as following for counting reads mapped to either antisense orientation: featureCounts –p -F GTF -t exon -g gene_id -a annotation.gtf -o counts.txt mapping_results.bam

Differential analysis was done using DESeq2 which is included in the Bioconductor (version 3.2.2) package , and PCA following Github RNAseq script (https://gist.github.com/stephenturner/f60c1934405c127f09a6.js).

3 Results

3.1 Sense/Antisense Annealing with GuSCN

To confirm the results from past publications that found that GuSCN is effective in annealing RNA-RNA and DNA-DNA in high concentrations we used synthesized lambda RNA containing both sense and antisense transcripts and tested GuSCN's effect on RNA annealing and its efficiency compared to annealing sense and antisense transcripts without chemical facilitation.

In an experiment which used lambda sense and antisense RNA that was denatured and allowed to hybridize at room temperature, as well as lambda sense and antisense RNA annealed with GuSCN, we observed that GuSCN promotes a more complete annealing between sense/antisense transcripts and worked quicker than allowing the RNA to hybridize at room temperature. In Fig 3.1.1 you can see that dsRNA upper bands get stronger between control and 3-day room temperature hybridized RNA. RNA that was annealed using 4M GuSCN shows a very strong dsRNA upper band, which is confirmed to be double stranded through a RNase T1 digestion of the sample, which digested single stranded RNA while leaving dsRNA undigested.

We observed that sense and antisense lambda RNA did anneal when denatured and allowed to hybridize at room temperature, but the annealing was not as complete as when hybridized in the presence of GuSCN and after 3 days hybridizing at room temperature did not reach the level of annealing as was observed from chemically guided annealing with GuSCN for 1 hour. When sense an antisense lambda RNA was not denatured prior to incubation at room temperature very minimal hybridization occurred. (Fig 3.1.2) These results support that GuSCN is a faster and more effective way to anneal sense and antisense transcripts than denaturing and renaturing alone. The lack of hybridization without denaturing of sense and antisense transcripts also supports the idea that these transcripts would not naturally hybridize in the cell without some form of enrichment to create these transcript pairs.

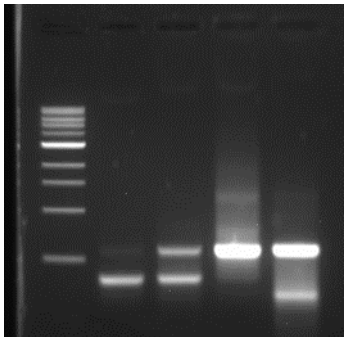

Fig 3.1.1. Agarose gel showing GuSCN annealing. Lane1: 1 kb DNA ladder, lane 2: sense/antisense lambda control RNA, lane 3: 3-day annealed sense/antisense room temperature annealed lambda RNA, lane 4: GuSCN sense/antisense annealed lambda RNA, lane 5: GuSCN annealed sense/antisense lambda RNA digested with RNase T1.

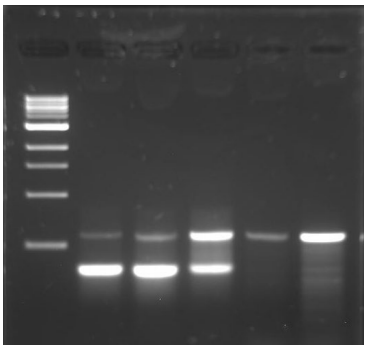

Fig 3.1.2. Agarose gel showing digestion of room temperature annealed lambda RNA without chemical annealing. Lane 1: 1 kb DNA adder, lane 2; control sense and antisense lambda RNA, lane 3: not denatured sense and antisense lambda RNA room temperature annealed overnight, lane 4: 4 day denatured and room temperature annealed sense and antisense lambda RNA, lane 5: not denatured sense and antisense lambda RNA room temperature annealed overnight digested with RNase T1, lane 6: 4 day denatured and room temperature annealed sense and antisense lambda RNA digested with RNase T1.

When *Populus* RNA was annealed for 24 hours with GuSCN and digested with RNase T1 the same effects were observed. Unlike with sense/antisense lambda annealed RNA total RNA has a very low level of antisense RNA compared to total RNA. After annealing 717 hybrid mature leaf RNA with GuSCN and digesting both control and GuSCN annealed RNA we observed that annealed RNA was not fully digested by RNase T1, indicating the annealed mature leaf sample has a larger proportion of dsRNA compared with the control. (Fig 3.1.3) These results show that GuSCN worked to anneal sense and antisense transcripts into duplexed sense/antisense transcript pairs both with synthesized lambda RNA as well as with native RNA found in plant tissue. Results with native RNA from plant tissue is less effective than when used with synthesized lambda RNA which is to be expected since antisense RNA is found in a very low abundance compared to total RNA and the presence of antisense and sense complementary sequences are needed for annealing to occur. We then utilized this method to enrich sense/antisense transcript pairs that could be separated from total RNA through other methods including mRNA purification and J2 antibody separation.

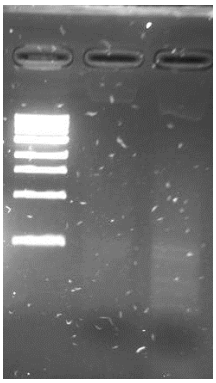

Fig 3.1.3. Agarose Gel showing RNase T1 digestion of control and GuSCN annealed RNA. Lane 1: 1 kb DNA ladder, lane 2: 717 hybrid mature leaf control RNA RNase T1 digested, lane 3: 717 hybrid mature leaf GuSCN annealed RNA RNase T1 digested.

3.2 Custom Tailing

To increase the amount of antisense RNA sequenced using Nanopore direct RNA sequencing, we made a variety of Nanopore direct RNA libraries using the same starting material from *P. trichocarpa* 28T mature leaf tissue. Nanopore direct RNA sequencing was originally designed for mRNA sequencing and uses adapters designed to capture polyadenylated RNA strands, so to utilize this sequencing method we had to develop methods for adding custom nucleotide tails to non-polyadenylated transcripts. We tested a variety of alternative poly nucleotide tail types including those that use ITP to add Poly(I) tails, ITP +GTP to add Poly (GI) tails, and methyl ATP to add 2'O Poly(mA) tails. By using alternative nucleotide tails we can differentiate between native poly adenylated mRNA and determine their original Poly(A) tail length, and sequence non-polyadenylated transcripts that have a poly nucleotide addition.

DirectRNAlib library was made using total 28T RNA without the addition of any custom tails. *28T_mRNA* library was made from mRNA purified 28T RNA without the addition of any custom tails. *mRNA_2'O* library was made from mRNA purified from 28T with the addition of Poly(mA) tails. *mRNA_ITP* library was made from mRNA purified 28T RNA with the addition of Poly(I) tails. *mRNA_G/I* library was made from mRNA purified 28T RNA with the addition of Poly (GI) tails. *rRNA-_G/I* library was made from 28T RNA with rRNA depletion and the addition of Poly (GI) tails. *rRNA-_2'-O* library was made from rRNA depleted 28T RNA with the addition of Poly(mA) custom tailing. (Table 3.2.1)

Table 3.2.1. Custom tailing library Nanopore sequencing data.

28T library	**Reads Uniquely Mapped**	**Reads Mapped to sense strand**	**Reads Mapped to Antisense strand**	**Percent (%) antisense/sense**
DirectRNAlib	1135214	1054203 (92.9%)	60575 (5.3%)	5.7%
28T_mRNA	1469115	1420920 (96.7%)	115385 (7.9%)	8.2%
mRNA_2'O	388226	281184 (72.4%)	53832 (13.9%)	19%
mRNA_G/I	63815	49947 (78.3%)	5918 (9.3%)	11.9%
mRNA_ITP	33609	29763 (88.6%)	3721 (11.1%)	13%
rRNA-_2'O	682255	233846 (34.3%)	81755 (12%)	35%
rRNA-_G/I	11948	6986 (58.5%)	1241 (10.4%)	18%

We observed that the addition of Poly(mA) tails had the highest percentage of antisense/sense reads compared to Poly(I) and Poly (GI) tail additions with the same treatment. Poly(mA) mRNA library had 19% antisense/sense reads while mRNA libraries that used Poly(I) and Poly (GI) tails had 13%, and 11.9% antisense/sense reads respectively. rRNA depleted Poly(mA) library has 35% antisense/sense reads, while rRNA depleted Poly (GI) tailed library only had 18% antisense/sense reads. This result indicated that the use of Poly(mA) tails leads to a better sequencing yield of antisense RNA when used with mRNA purification and rRNA depletion methods and is the best custom tailing method to use for antisense RNA enrichment. (Fig 3.2.1)

Although sequencing data from DirectRNAlib and 28T_mRNA libraries without custom tailing are both composed of only mRNA transcripts the increased percentage of antisense/sense reads in the mRNA purified library (8.2%) compared with total RNA (5.7%) is likely due to the overall percentage of loaded library being polyadenylated antisense transcripts compared to that of a total RNA library. The addition of all 3 kinds of custom tailing increased the percentage of antisense/sense reads due to non-polyadenylated RNA contamination being tailed and allowing for it to be sequenced using Nanopore direct RNA sequencing. (Fig 3.2.1) Libraries made from rRNA depleted total RNA had the highest percentage of antisense/sense reads when compared with samples libraries with the same tailing method. (Fig 3.2.1). This result fit our expected outcomes because most antisense RNA is not polyadenylated and by removing rRNA and adding custom tails you increase the relative abundance of antisense reads which are then able to be sequenced with the addition of poly nucleotide tails.

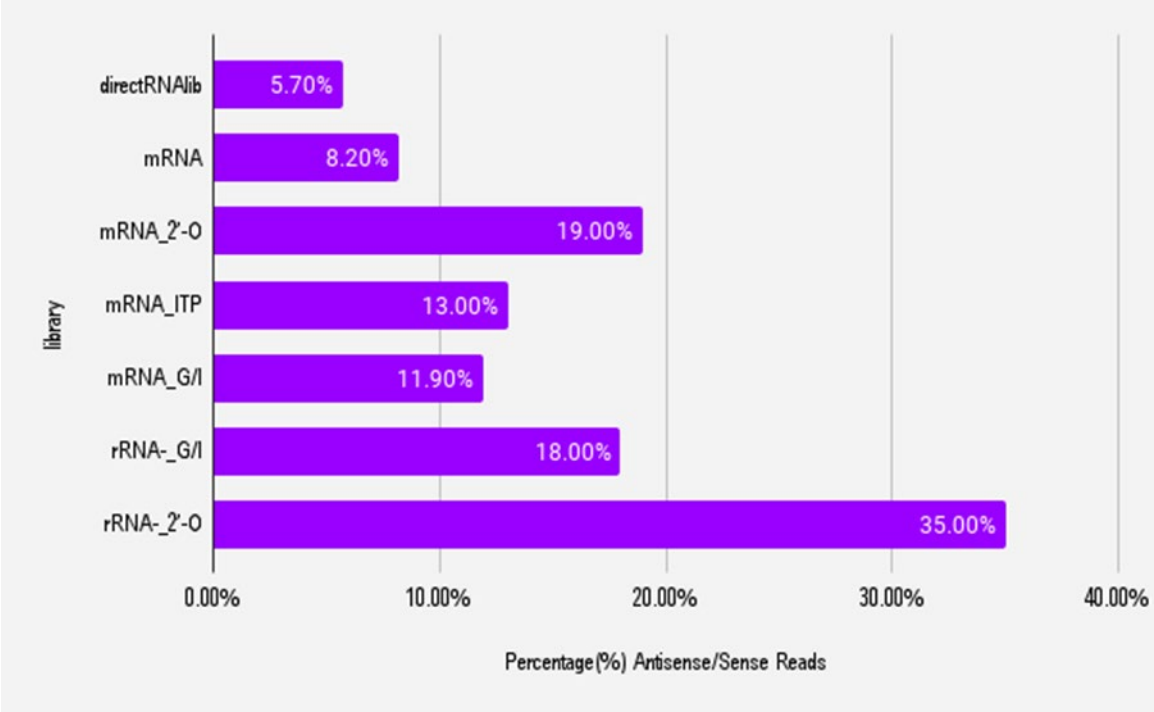

Fig 3.2.1. Comparison of ratio percentage antisense/sense reads for custom tailed Nanopore libraries.

3.3 J2 Antibody dsRNA Separation

When testing the efficiency of J2 MAB when binding dsRNA, we used dsRNA ladder. J2 antibody elute, which should contain the dsRNA, was resistant to RNase T1 digestion, which digests ssRNA, but was completely digested by Rnase3,which digests dsRNA, this shows that the elute from J2 binding contained dsRNA. These results showed that J2 antibody binds to dsRNA and can be used to isolate dsRNA (Fig 3.3.1).

When this method was tested on real RNA annealed with GuSCN, enrichment of antisense RNA transcripts was lower than expected. 13.2% of total reads mapping to the antisense strand of DNA for Poly(A) tailed BCW mRNA, and only 9.5% of total reads mapping to antisense strand of DNA for Poly (GI) tailed 717 mRNA. (Table 3.3.1) These results suggest that J2 non-specific binding with the highly abundant rRNA is likely responsible for the lower-than-expected enrichment of antisense RNA, which we would expect to be close to 50% of total uniquely mapped reads if J2 binding method was fully successful in isolating only dsRNA species. For both libraries only 63% of total reads were uniquely mapped to either strand of DNA (Fig 3.3.2). This low number of uniquely mapped reads can also be explained by J2 non-specific binding with rRNA.

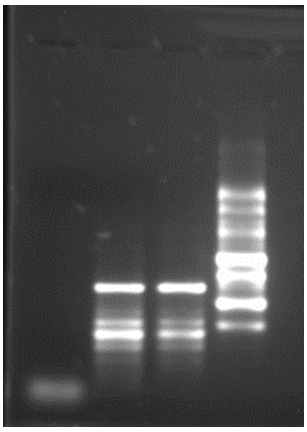

Fig 3.3.1. Agarose gel of J2 elute of dsRNA ladder digested with RNase 3 and RNase T1. Lane 1: J2 elute RNase3 digested, lane 2: J2 elute RNase T1 digested, lane 3: J2 elute control, lane 4: ssRNA ladder.

Table 3.3.1. J2 dsRNA Nanopore sequencing data.

Library	**Reads Uniquely Mapped**	**Reads Mapped to sense strand**	**Reads Mapped to Antisense strand**	**Percent (%) antisense/sense**
BCW18c_mRNA_J2_A	36172	17842 (49.3%)	4786 (13.2%)	26.8%
717_mRNA_J2_G/I	20580	11032 (53.6%)	1958 (9.5%)	18%

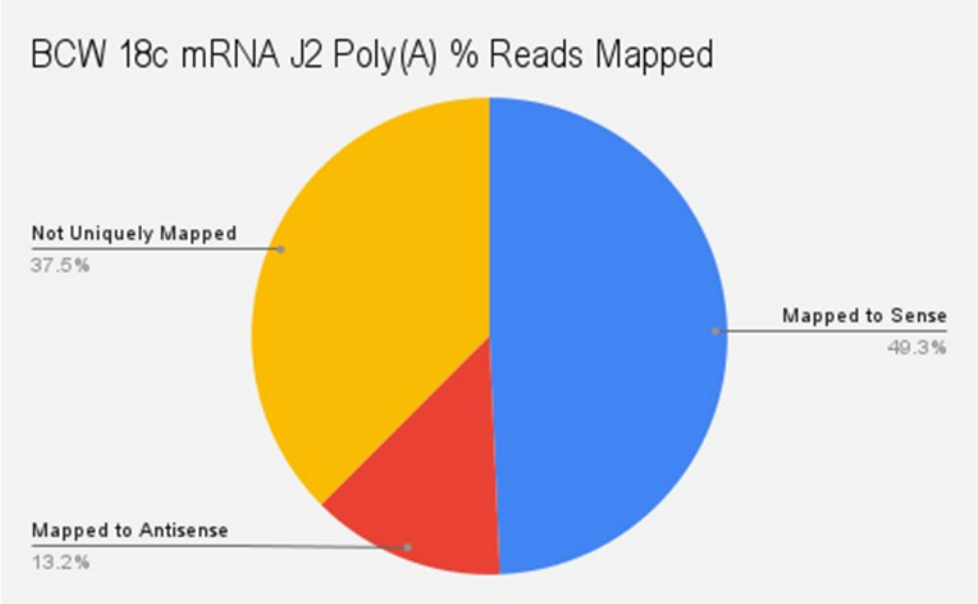

Fig 3.3.2. Percent (%) of reads uniquely mapped to sense and antisense strand from BCW 18c mRNA Poly(A) J2 library.

3.4 HAP column Separation

Early results from the HAP column separation experiments showed there is size differentiation between single stranded and double stranded elutes for cDNA experiments. Results with cDNA from total RNA showed consistent size differences when checked on agarose gel with 0.266M sodium phosphate elute having the largest size compared with 0.12M and 0.16M elutes from the same sample (Fig 3.4.1). These results suggest that larger RNA structures, assumingly including dsRNA, are eluted from HAP column at higher sodium phosphate concentrations.

We confirmed that 0.266M sodium phosphate elutes dsDNA by using dsDNase to digest 0.16M, and 0.266M elutes. The results of this experiment showed the 0.16M sodium phosphate elutes showed little to no digestion from the dsDNase while 0.266M elutes showed complete digestion from dsDNase treatment. (Fig 3.4.2) These results fit with past findings that 0.16M sodium phosphate unbinds ssDNA from HAP, whereas 0.266M sodium phosphate unbinds dsDNA from HAP.

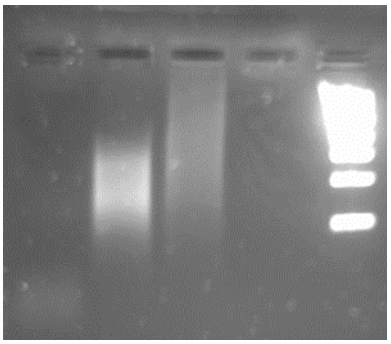

Fig 3.4.1. Agarose gel shows the size differences between concentrations of HAP sodium phosphate elutes. Lane 1: 0.12M, lane 2: 0.16M, lane 3: 0.266M, lane 4: 0.4M, lane 5: 1 kb DNA ladder

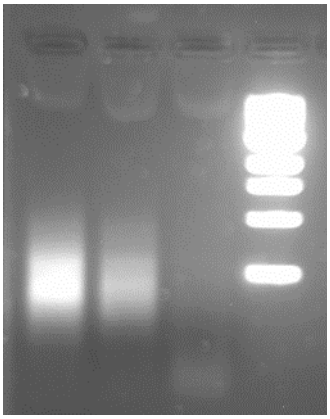

Fig 3.4.2. Agarose gel showing HAP elutes digested with ds DNase. Lane 1: 0.16-1 Poly(A) (4/29/22) digested with ds DNase, Lane 2: 0.16-1 Poly (GI) (5/5/22) digested with ds DNase, lane 3: 0.266-1 Poly (GI) (5/5/22) digested with ds DNase, lane 4: 1 kb ladder.

Total RNA that was DMSO treated prior to first strand cDNA synthesis both for control and treated RNA samples showed a similar increase in percentage of dsDNA recovered from HAP column elutes. We hypothesized natural dsRNA transcripts found in the cell would be resistant to reverse transcription and by DMSO treating the RNA prior to RT would allow for RNA transcripts that are found naturally in duplex structures to be converted into cDNA after denaturing with DMSO.

Table 3.4.1. HAP elution data for DMSO and non-DMSO treated samples.

Sample	**%ss-cDNA recovered**	**% ds-cDNA recovered**	**% difference ds-cDNA DMSO**
18c Frag	26.04%	15.91%	
18c Frag DMSO	37.53%	16.56%	0.65%
28T Frag	10.98%	3.33%	
28T Frag DMSO	33.42%	3.73%	0.40%

HAP chromatography has not been confirmed with Illumina sequencing to determine whether the method is useful for enriching antisense RNA for library formation. It has been shown to be an effective method to separate ds-cDNA from total cDNA in a very cost-effective manner. Once confirmed with sequencing this novel method could be useful for creating antisense RNA enriched cDNA libraries. The future goal for this project is to use the isolated ds-cDNA that have been concentrated and desalted, to form Illumina libraries for sequencing.

3.5 Drought Experiment

To look at the effectiveness of drought treatments on *P. trichocarpa* (Black Cottonwood) and 717 (*P. alba x P. tremula*) long term drought treated samples we submitted 2 control and 2 drought treated mature leaf samples from both *Populus* species to Novogene for mRNA Illumina sequencing. After mapping and counting reads (Table 3.5.1), we ran DESeq2 and principal component analysis (PCA) for both sets of Illumina sequencing data. DESeq2 analysis of drought treated mRNA seq libraries showed the drought treatment for *P. trichocarpa* was not successful, with no differentially expressed genes between the control and treated samples (Fig S.1). 717 libraries showed 552 differentially expressed genes with a P value below 0.1. (Fig 3.5.1)

PCA showed there is no correlation between the control and drought treated samples of black cottonwood (Fig 3.5.2). PCA analysis of 717 showed that there is a weak correlation between drought treated samples (Fig 3.5.3). The low number of differentially expressed genes is due to the weak correlation of drought treated samples. Although 717 shows some differential expressions, there are issues with the drought treatment of both species of *Populus.*

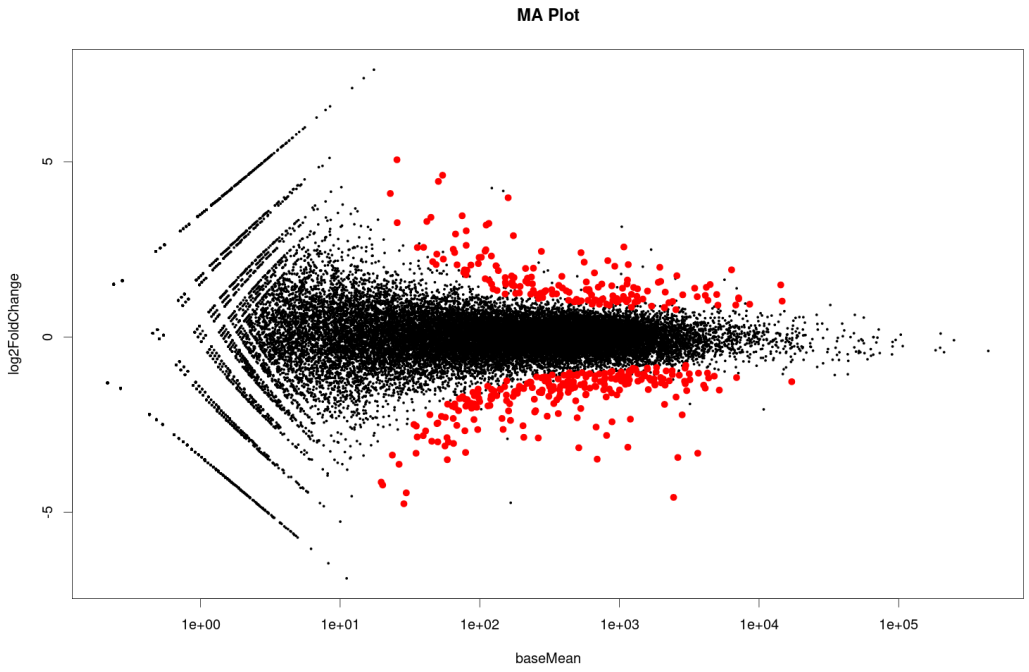

Fig 3.5.1. DESeq2 plot of differentially expressed genes for 717 hybrid drought and control mRNA.

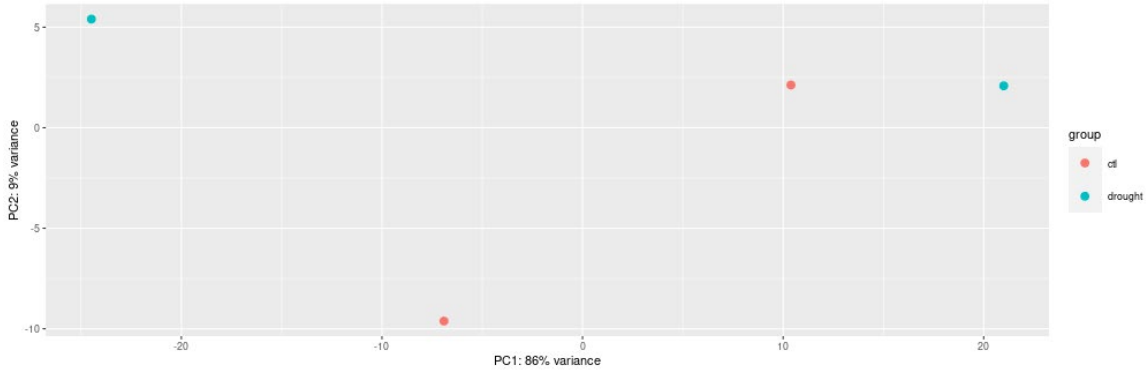

Fig 3.5.2. PCA plot showing correlation between control and drought treated *P. trichocarpa* mRNA libraries.

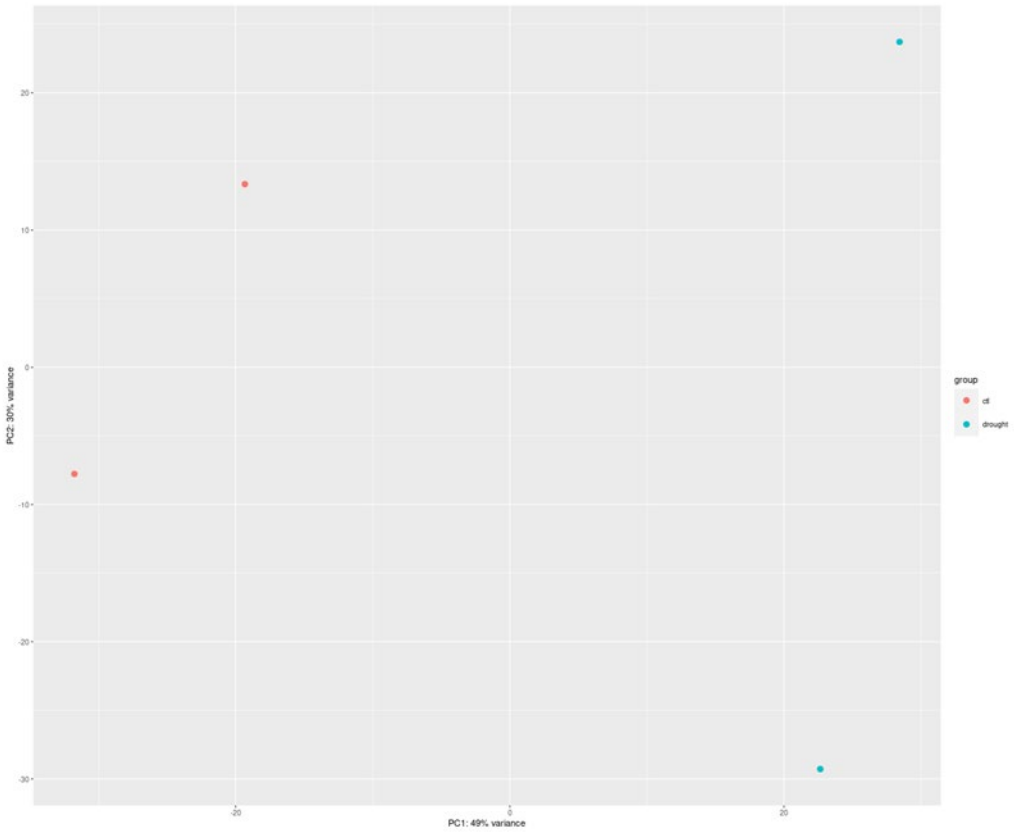

Fig 3.5.3. PCA plot showing correlation between control and drought treated 717 hybrids mRNA libraries.

Table 3.5.1. Novogene mRNA seq read data.

Sample Name	**Total Input reads**	**Uniquely Mapped reads**	**Uniquely mapped read %**	**Multiple Mapped Reads**	**Multiple Mapped Reads %**
16c 717LT	23287917	19701920	84.6	1678252	7.21
22c 717 LT	26165700	22329563	85.34	1845759	7.05
31t 717LT	20769023	17672755	85.09	1478191	7.12
13t 717LT	23165426	19828730	85.6	1537717	6.64
18c BCWLT	24570212	22232784	90.49	1225097	4.99
36c BCWLT	21542510	19672696	91.32	994970	4.62
4t BCWLT	20061511	17949132	89.47	1184973	5.91
28t BCWLT	19706151	18061551	91.65	829450	4.21

4 Discussion

4.1 J2 Binding

When total RNA from mature leaf *Populus* tissue was used to test the efficiency of J2 binding at isolating dsRNA only 63% of total reads were uniquely mapped to either strand of DNA in both Nanopore libraries sequenced. (Table 3.3.1) (Fig 3.3.2) We would expect close to 50% of total uniquely mapped reads to be antisense RNA if this J2 binding method was fully successful in isolating only dsRNA species. An explanation for the lower overall antisense RNA reads for 717_mRNA_J2-G/I library compared with BCW_mRNA_J2_A library is as is shown by our results with different custom tailing methods, Poly(G/I) tails result in a lower antisense read percentage compared to other tailing methods. This does not account for the low number of uniquely mapped reads for these libraries, but it does explain the variation between similarly treated J2 libraries. This lower-than-expected number of both total uniquely mapped reads and reads mapped to the antisense strand suggest that J2 MAB binds non-specifically to rRNA at a higher rate than was expected. (Schönborn et al., 1991; Lybecker et al., 2013; Werner et al., 2021)

To reduce the amount of non-specific binding samples could be treated with RNase T1 to digest any single stranded RNA that may be present in the sample. The use of RNase T1 would digest all ssRNA, including the overhanging segments of annealed dsRNA structures. As we observed when GuSCN annealed total RNA was digested using RNase T1 (Fig 3.1.3), when annealed total RNA is digested with RNase T1 the resulting dsRNA remaining after digestion is very short (below 250 bp). These small dsRNA fragments are likely due to natural sense and antisense complementary pairs being of varied length, or incomplete complementary sequences resulting in the double stranded portion of the transcript pairs to be a shorter length. Since we are interested in developing methods that allow us to enrich and sequence the full length of antisense lncRNAs we do not use RNase T1 to decrease the effects of rRNA non-specific binding. To reduce the amount of non-specific binding without loss of the full length of RNA transcripts samples can be rRNA depleted prior to annealing to decrease the amount of rRNA in the sample and increase the proportion of dsRNA after GuSCN annealing. One explanation we had is that J2 does not bind well to dsRNA when there is only a small amount of dsRNA in the total sample. Since antisense RNA is found at a very low relative abundance compared to total RNA, unless you use a very high quantity of total RNA there are not a lot of dsRNA transcript pairs for J2 to bind to. If this theory is correct then using a sample that has been either depleted of most rRNA before J2 binding to increase the overall proportion of dsRNA in the sample, or increasing the starting amount of total RNA before annealing to increase the overall quantity of dsRNA after annealing could both improve the overall yield of dsRNA after J2 MAB isolation.

A method we used to avoid J2 nonspecific binding with rRNA is to mRNA purify samples after annealing with GuSCN. Since oligo(dt) beads bind to Poly(A) tails of mRNA annealed antisense transcripts should not interfere with mRNA purification. This allows for mRNA and any complementary antisense transcripts to be isolated for

sequencing. mRNA purification after annealing with GuSCN can be followed by J2 binding to isolate only mRNA that has formed a duplex with annealed antisense transcripts to isolate a higher proportion of antisense reads, or the annealed mRNA can be sequenced directly. Since the J2 data that we collected using mRNA purified RNA gave us a lower antisense yield than we had been expecting we branched out to other potential ways that GuSCN enrichment could be used to enrich antisense RNA with or without the use of J2 MAB. Sequencing annealed mRNA directly should give you a lower proportion of antisense RNA since mRNA without complementary antisense transcripts will also be present. Whether sequenced directly or after J2 dsRNA isolation the sample will need to be custom tailed prior to library formation for Nanopore to capture non-polyadenylated antisense transcripts enriched with GuSCN annealing.

4.2 Custom Tailing

Custom tailing results show that by adding modified tails to total RNA you can capture non-coding RNA, including antisense RNA that cannot be captured using conventional Nanopore direct RNA sequencing. This method enabled us to sequence the full length of all RNA types and allows for a wider understanding of RNA expression that was not possible before, including different isoforms of non-coding RNA that have been largely ignored when looking at transcriptomes in the past.

The addition of custom nucleotide tails lead to an increase in the percentage of antisense/sense reads in all treatments, even those that were already mRNA purified libraries. (Fig 3.2.1) The increase of antisense/sense read percentage in these mRNA libraries is due to RNA contamination that is not polyadenylated. By adding custom tails to these mRNA libraries non-polyadenylated RNA that was purified along with mRNA can be sequenced. Most antisense RNA is non-coding and is not polyadenylated, the addition of custom tails to any library that is not 100% mRNA will lead to an increase in the percentage of antisense/sense reads sequenced due to the sequencing of non-polyadenylated RNA contamination. An increase in antisense/sense read percentage correlated with not only the addition of custom tailing, but also the relative abundance of non-coding non-rRNA RNA found in the library. The DirectRNAlib library without any tail additions had the lowest antisense/sense read percentage due to the lack of custom tailing to allow for non-polyadenylated transcripts to be sequenced as well as the low abundance of mRNA compared to other RNA species in the library. This percentage increased when mRNA libraries from the same starting material were sequenced without the addition of custom tails due to the higher relative abundance of antisense Poly(A) transcripts in the library. In the DirectRNA library as well as the mRNA library the only antisense RNA that was sequenced was antisense RNA that is naturally found polyadenylated in the cell. rRNA depleted libraries increase the relative abundance of antisense RNA by removing rRNA which makes up most of the total RNA in the cell. rRNA depleted libraries with the addition of custom tails increase the relative abundance of antisense RNA and allow for these non-coding non-polyadenylated transcripts to be sequenced using Nanopore direct RNA sequencing, which resulted in rRNA depleted custom tailed libraries having the highest percentage of antisense/sense reads compared to other libraries with the same custom tailing method used. Our results showed that 2’O

Poly(mA) custom tailing was the most effective custom tailing method to increase percentage antisense/sense RNA reads when compared with Poly(I) and Poly (GI) tailing both when used with rRNA depleted libraries as well as mRNA purified libraries. (Fig 3.3.1)

Interestingly rRNA depleted custom tailed libraries both have a lower uniquely mapped read count than those of mRNA purified libraries (Table3.2.1). These results are similar to those observed in our J2 MAB libraries (Table 3.3.1). These results are likely due to the riboPOOLs rRNA removal kit having a low efficiency for the depletion of rRNA. This low efficiency for depleting rRNA allows for these rRNA transcripts to be sequenced when paired with custom tailing methods, resulting in a lower than expected uniquely mapped read count.

By using Poly(mA) custom tailing Nanopore direct RNA sequencing can capture non-coding RNA that is not polyadenylated to get a full picture of the total RNA population. Nanopore direct RNA sequencing can detect base modifications so using Poly(mA) tails allows for you to differentiate between native Poly(A) tailed transcripts and those with Poly(mA) additions due to the presence of a methyl group. This method also allows for the differentiation between Poly(mA) additions to mRNA and the natural Poly(A) tail so you can still observe the effect of alterative polyadenylation isoforms while allowing for the sequencing of non-polyadenylated transcripts. This method is useful for enrichment of antisense RNA since antisense RNA is primarily non-polyadenylated, but it can also be used for a variety of projects that investigate any non-polyadenylated RNA and its isoforms.

4.3 HAP Column Separation

Although our HAP chromatography method still needs to be confirmed with Illumina sequencing it is a novel method for antisense transcript enrichment.

This method was originally intended to be used with Nanopore direct RNA sequencing but our early experiments using hybridized RNA did not show a lot of consistency or clear size separation between different concentrations of elutes like we observed when conducting similar experiments with first strand synthesized cDNA (Fig 3.4.1). We hypothesized that this lack of consistency or clear size differentiation between elute concentrations was likely due to RNA secondary structures confounding the data and effecting what concentration of sodium phosphate eluted these duplex structures.

The clear size differences between 0.12M, 0.16M, and 0.266M sodium phosphate elutes (Fig 3.4.1) leads us to believe that smaller cDNA species, including ss-cDNA are eluted at lower concentrations of sodium phosphate. Although manufacturers recommendations for HAP only mention 0.16M and 0.266M sodium phosphate buffer we tested a wider range of buffer concentrations to see if they might elute different types of duplex structures. Although neither 0.12M nor 0.16M sodium phosphate elutes should contain ds-cDNA that we are aiming to isolate, sequencing these elutes as well may give us a better understanding of how the different concentrations of buffer effects what size and

types of cDNA are eluted. When 0.4M elutes were checked on gel there is not any clear size differentiation between 0.4M and 0.266M elutes, so it is possible that 0.4M elutes contain ds-cDNA that was not fully eluted from previous washes. For each concentration of buffer, we did multiple washes to ensure that the quantity eluted was due to an increase in concentration rather than cDNA that was not fully eluted. In most cases 0.4M elutes did increase in concentration from 0.266M elutes from the 3rd wash (Fig S.2), which suggests that although on agarose gel these concentrations of buffer seem to elute the same size of cDNA that when sequenced there may be some notable difference between the cDNA eluted at each concentration.

We aim to use this method with Illumina sequencing rather than Nanopore cDNA sequencing because the overall recovery of ds-cDNA after HAP chromatography is low. To better utilize the ds-cDNA recovered from this method we decided to utilize Illumina sequencing which requires a much smaller amount of genetic material for sequencing. By using Illumina, we will not get information about potential antisense isoforms that may exist, but it does give a cost-effective way to get an idea of the differential expression of antisense RNA with a small amount of starting material.

4.4 Drought Experiment

The results from the Novogene mRNA sequencing show that the drought experiments on the 717 hybrid plants were partly successful and showed 552 differentially expressed genes between control and treated samples, that the *P. trichocarpa* drought experiments did not show any differential expression between drought and control groups. Although the methods for each of these experiments were the same, the differences in success of the drought treatment likely are due to the differences in growth between 717 and *P. trichocarpa* plants. 717 hybrid plants consistently did not grow secondary growth until after the 24th internode, while *P. trichocarpa* plants had very inconsistent and often very early occurrence of secondary growth. (Fig S.3) This early branching and irregular growth lead to inconsistencies between plants under the same treatment. Other differences in growth could be due to the differences in propagation between the two species. 717 hybrid plants were propagated using tissue culture method which allows for more uniform growth from the beginning than stem cutting propagation. Stem cutting propagation of *P. trichocarpa* lead to higher variability in size and growth rate than that of 717 plants that were propagated using tissue culture method.

The low number of differentially expressed genes even for the 717 hybrid plants could be because of secondary stressors. In long term drought treated plants drought stress makes plants more susceptible to pest infestation. In both *P. trichocarpa* and 717, drought treated plants had a higher rate of spider mite infection and suffered more severe symptoms from infection than control plants. These secondary stressors are likely the reason why differential expression in both species of *Populus* was low or undetectable. Drought conditions have often been linked to pest outbreaks. While under stress plants reduce production of protective metabolites and become more susceptible to insect predation (Mattson & Haack 1987). Although the presence and severity of infection increased when plants were under drought stress, spider mite infection was observed in

most plants in both control and drought treated to some degree. Insect herbivory is most likely one of the major factors impacting the low correlation between drought treated samples in both species of *Populus*. In future experiments I would suggest that plants are sprayed with pesticides on a bi-monthly basis to stay ahead of spider mite life cycles. Reducing long term drought treatment time from 21 days to 7 days to limit the effects of insect herbivory on experimental plants would also improve the quality of the drought experiment in the future.

This drought experiment served as a basis for what future drought experiments in our lab group should look like but was largely unsuccessful due to a high number of confounding factors that made drought treatment ineffective. The long-term goal of this project is to use methods developed to enrich sense/antisense transcript pairs to study the antisense response to drought stress in the *Populus* genus. For this goal to be met, execution of a more effective and controlled drought experiment will be a necessary first step. Since the objectives of this project were to develop methods to enrich sense/antisense transcript pairs for sequencing the results of these drought experiments do not affect the results of this project. RNA used to antisense RNA enrichment experiments used tissue from drought treatments, but these experiments did not compare control and drought treated samples but rather only different enrichment or separation methods on the same starting materials.

4.5 Conclusion, limitations, outlooks

Drought has major impacts on the health and productivity of forest trees, and in coming years it is predicted that drought conditions will increase both in severity and frequency. (Martignago et al 2020) In order to engineer and produce drought resistant trees adapted for climate drought stress, we must understand the molecular mechanisms controlling plant response to drought. Recent biotechnology work focuses on engineering drought resistant plants through sense gene manipulation in drought response genes (Wight et al 2013) but fail to investigate regulatory noncoding transcripts that may play a critical role in their response to drought. To better study the antisense RNA response to drought stress we must first develop methods that allow for the enrichment of antisense RNA transcripts for sequencing. To address this gap in methods to efficiently and cost effectively enrich antisense RNA transcripts for sequencing, we developed methods that would allow for the enrichment of antisense RNA through chemically guided annealing of sense/antisense transcript pairs, and the separation of double stranded structures for library formation.

The use of GuSCN to facilitate chemically guided annealing of sense/antisense transcript pairs resulted in a faster and more complete hybridization than denaturing and renaturing alone. The formation of sense/antisense transcript pairs allows for these double stranded duplex structures to be isolated through methods that separate double stranded structures from total RNA/cDNA.

Some of the limitations and challenges faced in this project include non-specific binding with rRNA when isolated dsRNA with J2 MAB. To fully utilize this method of dsRNA isolation from total RNA it is essential for methods that limit the number of reads

mapped to multiple loci. To do this, we need to refine methods to remove rRNA without the degradation of full-length transcripts. Once methods have been established to consistently limit the amount of rRNA nonspecific binding of J2 MAB then the use of GuSCN annealing with J2 MAB dsRNA isolation should be a very effective method for antisense RNA enrichment. While J2 MAB results were not as good as we had hoped, using GuSCN annealing with mRNA purification is also an alternative method to isolate antisense transcripts annealed to sense mRNA molecules. mRNA purification method does not give as high of a proportion of antisense reads as when used with J2 MAB binding as well, but still does show some enrichment of antisense reads sequenced.

Although the data that we collected point to HAP chromatography being an ideal way to enrich for antisense/sense transcript pairs for cDNA sequencing, this method still needs to be confirmed with Illumina sequencing. If this method is confirmed, then this will be a cost-effective novel method that uses a well-established DNA separation technique to target antisense transcripts for sequencing. Since this method utilizes Illumina sequencing it does not require large amounts of starting material and can be more easily utilized for a wide range of applications.

The long-term goal of this project is for these antisense enrichment methods to be utilized within our lab group to form antisense enriched libraries to study the antisense RNA response to drought stress in *Populus*. The results of this project and addressing some of the challenges and limitations that occurred should allow for cost effective formation of antisense enriched libraries which will allow for a targeted genome wide characterization of antisense transcriptome of *Populus* in response to drought stress.

5 Literature Cited

Allen, C., Macalady, A., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., . . . Cobb, N. (2010). A Global Overview of Drought and Head-Induced Tree Mortality Reveal Emerging Climate Change Risks for Forests. Forest Ecology and Management, 259(4), 660:684. doi:https://doi.org/10.1016/j.foreco.2009.09.001

Chen, J., Song, Y., Zhang, H., & Zhang, D. (2013). Genome-Wide Analysis of Gene Expression in Response to Drought Stress in *Populus simonii*. *Plant Molecular Biology Reporter, 31*. doi:10.1007/s11105-013-0563-6

Corfù NA, Tribolet R, Sigel H. (1990) Comparison of the self-association properties of the 5'- triphosphates of inosine (ITP), guanosine (GTP), and adenosine (ATP). Further evidence for ionic interactions in the highly stable dimeric [H2(ATP)] 2(4-) stack. *Eur J Biochem.* ;191(3):721-35. doi: 10.1111/j.1432-1033.1990.tb19181.x.

Cui. (2010). Antisense RNAs and epigenetic regulation. *Epigenomics.,* 2(1), 139–150. https://doi.org/10.2217/EPI.09.46

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., . . . Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience, 10*(2). doi:10.1093/gigascience/giab008

Dujardin, G., Lafaille, C., Petrillo, E., Buggiano, V., Gómez Acuña, L. I., Fiszbein, A., . . . Kornblihtt, A. R. (2013). Transcriptional elongation and alternative splicing. *Biochim Biophys Acta, 1829*(1), 134-140. doi:10.1016/j.bbagrm.2012.08.005

Frith, M. C., Pheasant, M., & Mattick, J. S. (2005). The amazing complexity of the human transcriptome. *Eur J Hum Genet, 13*(8), 894-897. doi:10.1038/sj.ejhg.5201459

Garg, S., Fungtammasan, A., Carroll, A., Chou, M., Schmitt, A., Zhou, X., . . . Li, H. (2021). Chromosome-scale, haplotype-resolved assembly of human genomes. *Nat Biotechnol, 39*(3), 309-312. doi:10.1038/s41587-020-0711-0

Keller, M. W., Rambo-Martin, B. L., Wilson, M. M., Ridenour, C. A., Shepard, S. S., Stark, T. J., . . . Barnes, J. R. (2018). Direct RNA Sequencing of the Coding Complete Influenza A Virus Genome. *Sci Rep, 8*(1), 14408. doi:10.1038/s41598-018-32615-8

Klos, R. J., Wang, G. G., Bauerle, W. L., & Rieck, J. R. (2009). Drought impact on forest growth and mortality in the southeast USA: an analysis using Forest Health and Monitoring data. *Ecol Appl, 19*(3), 699-708. doi:10.1890/08-0330.1

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics, 30*(7), 923-930. doi:10.1093/bioinformatics/btt656

Martignago, D., Rico-Medina, A., Blasco-Escámez, D., Fontanet-Manzaneque, J. B., & Caño-Delgado, A. I. (2019). Drought Resistance by Engineering Plant Tissue-Specific Responses. *Front Plant Sci, 10*, 1676. doi:10.3389/fpls.2019.01676

Mattson, W. J., & Haack, R. A. (1987). The Role of Drought in Outbreaks of Plant-eating Insects. *BioScience, 37*(2), 110-118. doi:10.2307/1310365

Porth, I., & El-Kassaby, Y. A. (2015). Using *Populus* as a lignocellulosic feedstock for bioethanol. *Biotechnol J, 10*(4), 510-524. doi:10.1002/biot.201400194

Schönborn, J., Oberstrass, J., Breyel, E., Tittgen, J., Schumacher, J., & Lukacs, N. (1991). Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts. *Nucleic Acids Res, 19*(11), 2993-3000. doi:10.1093/nar/19.11.2993

Shuai, P., Liang, D., Tang, S., Zhang, Z., Ye, C. Y., Su, Y., . . . Yin, W. (2014). Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J Exp Bot, 65*(17), 4975-4983. doi:10.1093/jxb/eru256

Shuai, P., Su, Y., Liang, D., Zhang, Z., Xia, X., & Yin, W. (2016). Identification of phasiRNAs and their drought- responsiveness in *Populus trichocarpa*. *FEBS Lett, 590*(20), 3616-3627. doi:10.1002/1873-3468.12419

Simmons, B. A., Loque, D., & Blanch, H. W. (2008). Next-generation biomass feedstocks for biofuel production. *Genome Biology, 9*(12), 242. doi:10.1186/gb-2008-9-12-242

Singer MF, Jones OW, Nirenberg MW. (1963) The effect of secondary structure of the template activity of polyribonucleotides. *Proc. Natl Acad. Sci. USA* 49, 392–399.

Smith, A. M., Jain, M., Mulroney, L., Garalde, D. R., & Akeson, M. (2019). Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One, 14*(5), e0216709. doi:10.1371/journal.pone.0216709

Smith DC. (1998) Development of glass-ionomer cement systems. *Biomaterials*. Mar;19(6):467-78. doi: 10.1016/s0142-9612(97)00126-9. PMID: 9645552.

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol, 22*(2), 96-118. doi:10.1038/s41580-020-00315-9

Thompson, J., & Gillespie, D. (1987). Molecular hybridization with RNA probes in concentrated solutions of guanidine thiocyanate. *Anal Biochem, 163*(2), 281-291. doi:10.1016/0003-2697(87)90225-9

Tomizawa J, Itoh T, Selzer G, Som T.(1981) Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc. Natl Acad. Sci. USA* 78(3), 1421–1425.

Tufarelli C, Stanley JA, Garrick D et al. (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* 34(2), 157–165.

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., . . . Rokhsar, D. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science, 313*(5793), 1596-1604. doi:10.1126/science.1128691

Viger, M., Smith, H. K., Cohen, D., Dewoody, J., Trewin, H., Steenackers, M., . . . Taylor, G. (2016). Adaptive mechanisms and genomic plasticity for drought tolerance identified in European black poplar (Populus nigra L.). *Tree Physiology, 36*(7), 909-928. doi:10.1093/treephys/tpw017

Wang, X.-J., Gaasterland, T., & Chua, N.-H. (2005). Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. *Genome Biology, 6*(4), R30. doi:10.1186/gb-2005-6-4-r30

Werner A, Clark JE, Samaranayake C, Casement J, Zinad HS, Sadeq S, Al-Hashimi S, Smith M, Kotaja N, Mattick JS. (2021) Widespread formation of double-stranded RNAs in testis. *Genome Res.* 31(7):1174–86. doi: 10.1101/gr.265603.120. Epub ahead of print. PMID: 34158368; PMCID: PMC8256860.

Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol, 20*(1), 129. doi:10.1186/s13059-019-1727-y

Wight, M., & Werner, A. (2013). The functions of natural antisense transcripts. *Essays Biochem, 54*, 91-101. doi:10.1042/bse0540091

Xu, J., Wang, Q., Freeling, M., Zhang, X., Xu, Y., Mao, Y., . . . Lu, Y. (2017). Natural antisense transcripts are significantly involved in regulation of drought stress in maize. *Nucleic Acids Res, 45*(9), 5126-5141. doi:10.1093/nar/gkx085

Yan C, Wang Y, Lyu T, Hu Z, Ye N, Liu W, Li J, Yao X, Yin H. (2021) Alternative Polyadenylation in response to temperature stress contributes to gene regulation in Populus trichocarpa. BMC Genomics. 22(1):53. doi: 10.1186/s12864-020-07353-9.

Yu W, Gius D, Onyango P (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451(7175), 202–206.

Yuan, Y., & Chen, S. (2018). Widespread antisense transcription of Populus genome under drought. *Mol Genet Genomics, 293*(4), 1017-1033. doi:10.1007/s00438-018-1456-z

6 Supplementary data

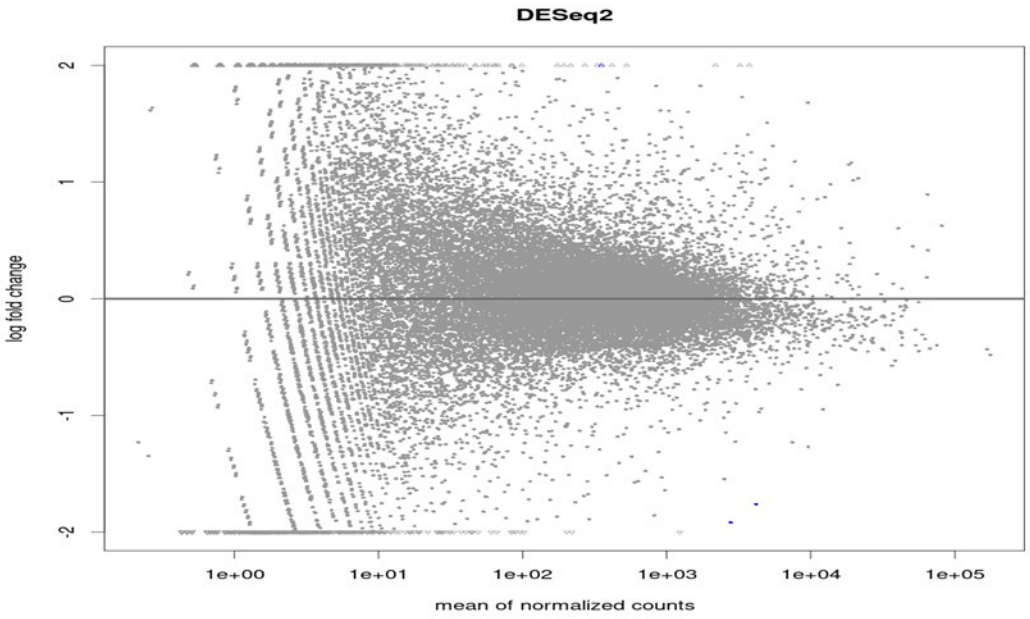

Fig S.1 DESeq2 plot of differentially expressed genes for *P. trichocarpa* drought and control mRNA

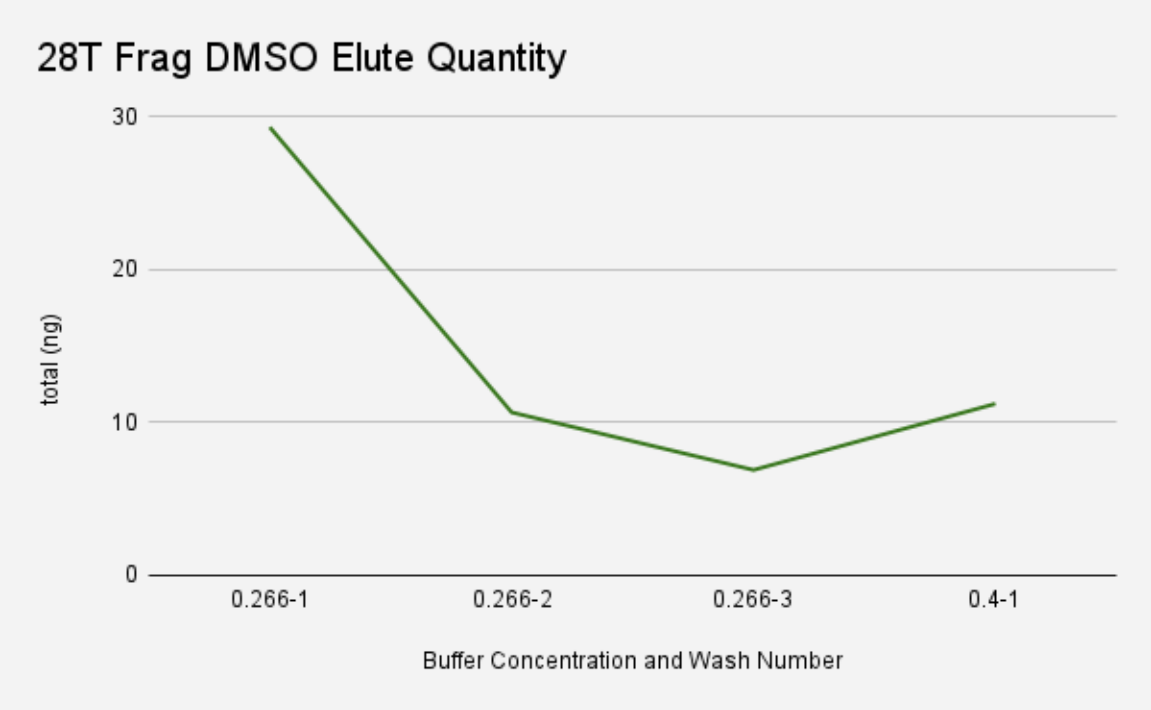

Fig S.2. 28T Fragmented DMSO treated cDNA elute quantity. Shows an increase in RNA recovered between 0.266M sodium phosphate 3rd wash and 0.4M sodium phosphate 1st wash.

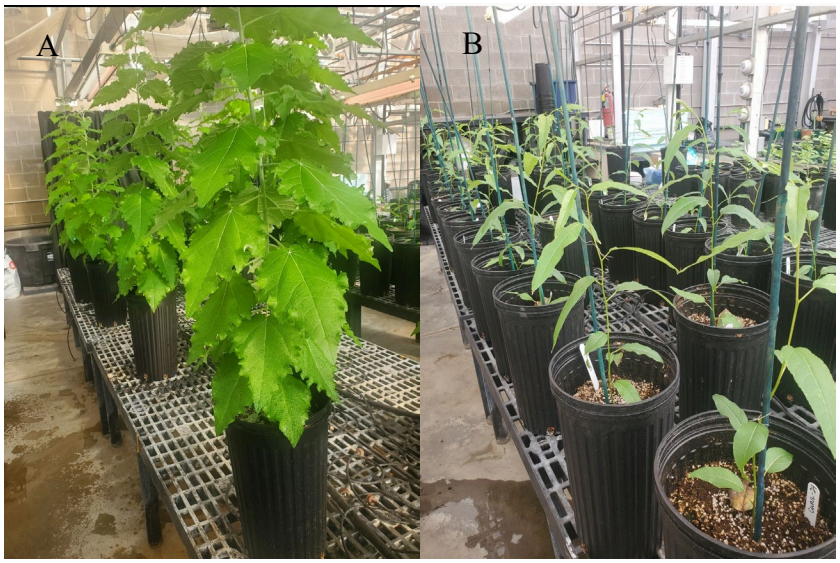

Fig S.3. Pictures of 717 hybrid (*P. alba* x *P. tremula)* and *P. trichocarpa.* A. 717 plants in Michigan Tech forestry green house. B. *P. trichocarpa* plants in Michigan Tech forestry green house. Both images were taken by Emma Burke.