

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Theses, Dissertations, and Student Research in  
Agronomy and Horticulture

Agronomy and Horticulture Department

---

4-28-2022

## Exploration of Genes Controlling Grain Yield Heterosis in Hybrid Wheat (*Triticum aestivum* L.) Utilizing 3' RNA Sequencing

Nichole Miller

University of Nebraska-Lincoln

Follow this and additional works at: <https://digitalcommons.unl.edu/agronhortdiss>



Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Agronomy and Crop Sciences Commons](#), [Bioinformatics Commons](#), [Botany Commons](#), [Genetics Commons](#), [Genomics Commons](#), [Horticulture Commons](#), [Other Plant Sciences Commons](#), and the [Plant Biology Commons](#)

---

Miller, Nichole, "Exploration of Genes Controlling Grain Yield Heterosis in Hybrid Wheat (*Triticum aestivum* L.) Utilizing 3' RNA Sequencing" (2022). *Theses, Dissertations, and Student Research in Agronomy and Horticulture*. 232.

<https://digitalcommons.unl.edu/agronhortdiss/232>

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Theses, Dissertations, and Student Research in Agronomy and Horticulture by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

EXPLORATION OF GENES CONTROLLING GRAIN YIELD HETEROSIS IN  
HYBRID WHEAT (*Triticum aestivum* L.) UTILIZING 3' RNA SEQUENCING

by

Nichole Lynn Miller

A THESIS

Presented to the Faculty of  
The Graduate College at the University of Nebraska  
In Partial Fulfillment of Requirements  
For the Degree of Master of Science

Major: Agronomy

Under the Supervision of Professor P. Stephen Baenziger

Lincoln, Nebraska

May 2022

EXPLORATION OF GENES CONTROLLING GRAIN YIELD HETEROSIS IN  
HYBRID WHEAT (*Triticum aestivum* L.) UTILIZING 3' RNA SEQUENCING

Nichole Lynn Miller, M.S.

University of Nebraska, 2022

Advisor: P. Stephen Baenziger

The implementation and future success of hybrid wheat (*Triticum aestivum* L.) is impacted by breeders' inability to create consistent high yielding, high heterosis hybrids. This research addresses this problem by conducting an exploration of transcriptomes from hybrids and parent lines to determine what genes are active in heterotic or non-heterotic hybrids and how their level of expression can explain the phenotype of grain yield heterosis. Using hybrids that showed positive mid-parent heterosis (MPH), classified as heterotic in our study, and negative or no difference MPH hybrids, classified as non-heterotic, differentially expressed genes (DEGs) potentially related to heterosis and hybrid yield response can be identified. Differential gene expression analysis found that more genes are differentially expressed in the non-heterotic hybrid to parent comparisons than in the heterotic hybrid to parent comparisons. Another important aspect of conducting a transcriptome study is adequately preserving the RNA for extraction and sequencing. Previous work has used liquid nitrogen to preserve samples taken out in the field, but this is dangerous and cumbersome. *RNAlater*<sup>®</sup> has been used as an alternative to liquid nitrogen but is not as consistent at preservation compared to liquid nitrogen. Another study to investigate this problem was conducted by sampling leaf and immature kernels from wheat, storing the samples at two temperatures for up to six months, extracting the RNA, and testing the quality parameters associated with using RNA for

sequencing. The results showed that the lower storage temperature had a negative impact on the parameters while storage time only negatively affected the purity. Both studies can be applied to research conducted on the transcriptome of wheat and allow for differences to be detected to explain heterosis.

## 1 ACKNOWLEDGEMENTS

2           This research would not have been possible without the help and support of  
3 family, friends, and the faculty, staff, and students at the University of Nebraska-Lincoln.  
4 First, I would like to thank my advisor, Dr. P. Stephen Baenziger for his support,  
5 wisdom, and the knowledge he has shared with me. Secondly, Dr. Katherine Frels for  
6 being another great mentor and pillar of support to finish my degree. Thank you to Dr.  
7 Vikas Belamkar for being the backbone of this research project from designing it, to  
8 having faith in me as the student to complete it as well as being my sounding board,  
9 teacher, mentor, and friend. Thank you to Dr. Scott Sattler for serving on my committee  
10 for this project as well as the insight you have on RNA sequencing, quality results, and  
11 for the use of your lab. A major source of support for this project also came from Dr.  
12 Keenan Amundsen who I cannot thank enough for the class he taught as well as support  
13 he offered outside of class on the computational side of this project and helping me get  
14 through all the little mistakes in my scripting. Thank you to Greg Dorn and Mitch  
15 Montgomery for making me feel like a member of your family, the good times spent  
16 chatting, taking me under your wings and teaching me the ropes of small plot field work  
17 and greenhouse management, and especially being the ones to plant and harvest my trials.  
18 Thank you to all the members of the Department of Agronomy and Horticulture, your  
19 classes, mentoring, and wisdom have made everything I have accomplished here possible  
20 and thank you for making Nebraska feel like home for the past few years.

21           Thank you to my partners at the University of Nebraska Medical Center  
22 Genomics Core group: James Eudy, Jennifer Bushing, Gregory Kubik, Alok Dhar, and  
23 Meng Niu. Without their collaboration, this project would not exist.

24           Thank you to Sakia Fields for being my right-hand man! I could not have  
25   completed this research without you! Also, thank you to Michaela Benes and Deanna  
26   Montanez Mendoza for your help in the wet lab! To the students I have had the great  
27   opportunity to work with in our small grains group, Fang Wang, Sheryl Sierra, Juan  
28   Pardo, Maria Sciencia, Emre Karahan, and Betul Kayitmazbatir, I am so thankful for our  
29   time together and friendship. You all made every day of work fun and enjoyable! I would  
30   like to especially thank Dr. Nick Garst and Hannah Stoll for taking me under their wings  
31   when I first arrived, making me feel welcome, showing me the ins and outs of the inbred  
32   and hybrid program and being a friend! I would also like to thank Dr. Sarah Blecha for  
33   taking the lead in this research before I arrived. Without you overseeing the sampling, I  
34   would have had nothing to show that first year. Thank you to all my friends both here at  
35   UNL and back in Illinois. Your friendships, calls, texts, and support have helped keep my  
36   spirits up while I am away while also knowing that I have love and support in what I am  
37   doing.

38           I would not be here today without a great support system back home! Thank you  
39   to my parents Sam and Ann Miller for the sacrifices you have made to help me realize  
40   my dreams, even if that dream was 8 hours from home and did not always make sense.  
41   Thank you for listening to my ramblings about RNA, quality, scripts, super computers,  
42   and everything else that did not make sense! Knowing that you both supported me helped  
43   me remember that I could not fail. Thank you to my sister Lindsey for being a phone call  
44   away and reminding me to take it easy occasionally and that I can do anything I set my  
45   mind to. To my grandparents Dale and Rosemary Hughes, my heavenly grandparents  
46   Raymond and Mary Miller and my numerous aunts, uncles, and cousins! Through our

47 daily phone calls, check ins, and interest in what I was doing, I always felt supported.  
48 You did everything you could to help, even listening to me explain what I am was doing  
49 just so I could get a fresh perspective on it. All of you have supported me my entire life  
50 and always taught me I could do anything I wanted to do, and I will treasure that my  
51 entire life. Finally, to my wonderful boyfriend Tanner Stratton! I could not be here today  
52 without you either as you have made the decision to go on for my master's degree easier  
53 in knowing that I have your love and support every step of the way. Thank you for being  
54 that constant in my life and driving the five hours to Lincoln every few weeks so that we  
55 did not forget what each other look like.

## 56 RESEARCH FUNDING

57           The research presented in this paper was made possible through the following  
58 funding sources: grant funding from the University of Nebraska Research Initiative,  
59 USDA-NIFA International Wheat Yield Partnership 2017-67007-25931, USDA-AFRI  
60 2020-67013-30872, and the Hatch Project NEB-22-361.

61           The University of Nebraska Genomics Core receives partial support from the  
62 National Institute for General Medical Science (NIGMS) INBRE - P20GM103427-19  
63 grant as well as The Fred & Pamela Buffett Cancer Center Support Grant - P30  
64 CA036727. This publication's contents are the sole responsibility of the authors and do  
65 not necessarily represent the official views of the NIH or NIGMS.



67	<b>Chapter 1: EXPLORATION OF GENES CONTROLLING GRAIN YIELD</b>	
68	<b>HETEROSIS IN HYBRID WHEAT (<i>Triticum aestivum</i> L.) UTILIZING 3' RNA</b>	
69	<b>SEQUENCING.....</b>	<b>1</b>
70	1. ABSTRACT.....	1
71	2. INTRODUCTION .....	2
72	3. MATERIALS AND METHODS .....	8
73	3.1 Plant Material.....	8
74	3.2 RNA Extractions .....	11
75	3.3 RNA Sequencing and Read Processing .....	12
76	3.4 Gene Expression Analysis .....	13
77	4. RESULTS .....	14
78	4.1 Plant Material.....	14
79	4.2 RNA Extraction.....	15
80	4.3 RNA Sequencing and Read Processing .....	15
81	4.4 Gene Expression Analysis .....	16
82	5. DISCUSSION .....	17
83	6. CONCLUSIONS.....	21
84	7. REFERENCES .....	22
85	<b>Chapter 2: OPTIMIZING SAMPLE PRESERVATION METHODS FOR</b>	
86	<b>SEQUENCING USING RNeasy® IN WHEAT .....</b>	<b>37</b>
87	1. ABSTRACT.....	37
88	2. INTRODUCTION .....	38
89	3. MATERIALS AND METHODS .....	41
90	3.2 RNA Extraction.....	42
91	3.3 RNA Quantity, Quality, and Integrity Testing .....	42
92	3.4 Statistical Analysis.....	43
93	4. RESULTS .....	44
94	5. DISCUSSION .....	47
95	6. CONCLUSIONS.....	48
96	7. REFERENCES .....	50
97		
98		

**List of Tables**

100	Table 1.1: Grain yield data from hybrid and parental line (triads) replicated trials grown	
101	at Lincoln, NE in 2019 or 2021. ....	25
102	Table 1.2: Comparison of RNA Quality Parameters .....	26
103	Table 1.3: <i>STAR</i> Aligner Mapping Results.....	26
104	Table 1.4: DEG Analysis-Heterotic vs Non-Heterotic Pairwise Comparison.....	27
105	Table 1.5: DEG Analysis-Leaf vs Seed Pairwise Comparison.....	27
106	Table 1.6: DEG Analysis-2019 vs 2021 Pairwise Comparison.....	27
107	Table 1.7: DEG Analysis-Leaf Triad Pairwise Comparisons .....	28
108	Table 1.8: DEG Analysis-Seed Triad Pairwise Comparisons .....	29
109	Table 2.1: Treatment Structure .....	52
110	Table 2.2: Leaf ANOVA.....	52
111	Table 2.3: Seed ANOVA .....	52
112		

## List of Figures

114	Figure 1.1: Kernel Age Comparison.....	30
115	Figure 1.2: Buffer RPE Leaf Tissue Washing Stages.....	30
116	Figure 1.3: PCA Analysis of Sources of Variation in Dataset.....	31
117	Figure 1.4: PCA Analysis of Triad x Type Interaction as Source of Variation.....	32
118	Figure 1.5: PCA Analysis of Sample Age as Source of Variation .....	33
119	Figure 1.6: Shared DEGs Across Hybrids-Leaf .....	34
120	Figure 1.7: Shared DEGs Across Hybrids-Seed .....	35
121	Figure 1.8: PCA Analysis of Triad 5 to Evaluate Selfing.....	36
122	Figure 2.1 Buffer RPE Leaf Tissue Washing. ....	53
123	Figure 2.2: LS-Means for RQN Response to Treatments-Leaf .....	54
124	Figure 2.3: LS-Means for Fragment Analyzer Concentration Response to Treatments-	
125	Leaf .....	55
126	Figure 2.4: LS-Means for 28S/18S Ratio Response to Treatments-Leaf .....	56
127	Figure 2.5: LS-Means for NanoDrop Concentration Response to Treatments-Leaf .....	57
128	Figure 2.6: LS-Means for $A_{260/280}$ Ratio Response to Treatments-Leaf.....	58
129	Figure 2.7: LS-Means for $A_{260/230}$ Ratio Response to Treatments-Leaf.....	59
130	Figure 2.8: LS-Means RQN Response to Treatments-Seed Samples.....	60
131	Figure 2.9: LS-Means for Fragment Analyzer Concentration Response to Treatments-	
132	Seed.....	61
133	Figure 2.10: LS-Means for 28S/18S Ratio Response to Treatments-Seed .....	62
134	Figure 2.11: LS-Means for NanoDrop Concentration Response to Treatments-Seed.....	63
135	Figure 2.12: LS-Means for $A_{260/280}$ Ratio Response to Treatments-Seed.....	64
136	Figure 2.13: LS-Means for $A_{260/230}$ Ratio Response to Treatments-Seed .....	65
137		

138 **CHAPTER 1: EXPLORATION OF GENES CONTROLLING GRAIN YIELD**  
139 **HETEROSIS IN HYBRID WHEAT (*Triticum aestivum* L.) UTILIZING 3' RNA**  
140 **SEQUENCING**

141 Abbreviations: HRWW: hard red winter wheat; MPH: mid-parent heterosis; CHA:  
142 chemical hybridizing agent; DAF: days after flowering; DEG: differentially expressed  
143 genes

144 **1. ABSTRACT**

145 Pure-line wheat cultivar development has not kept pace with increasing consumer  
146 demand. Hybrid wheat could meet the growing demand due to higher yields, but more  
147 research is needed to allow breeders to take advantage of the heterosis (hybrid vigor) seen  
148 in other hybrid crops. The goal of this research was to conduct a comprehensive  
149 transcriptomic study on hybrids and their corresponding parents to identify mechanisms  
150 that impact heterosis through differentially expressed genes (DEGs). Immature kernel  
151 and flag leaf tissue from ten triads (a hybrid and its parents) were sampled from two  
152 years. Five triads were identified as heterotic with positive mid-parent heterosis (MPH)  
153 and five were identified as non-heterotic with negative MPH. Biological replicates were  
154 sequenced using 3' RNA-Seq, transcript sequences were mapped to the IWGSC RefSeq  
155 2.0 reference genome, and DEG identification was done using *DESeq2*. The results found  
156 that gene expression was related to the year it was sampled as well as what tissue was  
157 sampled. However, we identified an average of 13 DEGs in the heterotic/non-heterotic  
158 hybrid comparisons of both leaf and seed tissue. We also found that more genes were  
159 differentially expressed in the non-heterotic hybrids when compared to the heterotic

160 hybrids. 600 and 402 genes were found to be differentially expressed across the ten  
161 hybrids for the seed and leaf tissues respectively. The results suggest that there are unique  
162 genes expressed in specific hybrids that could explain the differences in heterosis.  
163 Therefore, researchers should identify what role these genes have in the phenotypic  
164 response to heterosis.

## 165 2. INTRODUCTION

166 Bread wheat, *Triticum aestivum* L., is an important cereal crop in the United States  
167 for both human consumption and producers' income. In 2020, U.S. bread wheat  
168 production at 49,690,680 tons (Food and Agriculture Organization of the United Nations,  
169 2020) was the third largest grain crop behind maize (*Zea mays* subsp. *mays* L.) and  
170 soybeans (*Glycine max* L. Merrill). Hard red winter wheat (HRWW) grown in the Great  
171 Plains region accounts for 40% of the United States wheat production and is used to  
172 make bread flour, a staple in human diets (*USDA ERS - Wheat Sector at a Glance*, n.d.).  
173 Graybosch & Peterson (2012) found that, between 1987 and 2010, the genetic potential  
174 for grain yield for wheat varieties increased at a rate of ~1% every year in the Great  
175 Plains region. During that time, demand for HRWW had grown by 1.3% annually and  
176 has continued to grow by 1.4% in 2021 (USDA-ERS, 2021). While production has kept  
177 up with the current growth in demand, wheat acreage and overall production of wheat has  
178 decreased from the previous years (Food and Agriculture Organization of the United  
179 Nations, 2020). Decreasing wheat production trends are incompatible with increasing  
180 world population and caloric demands (Vespa et al., n.d.).

181 Yield genetic gain must increase for wheat production to increase to meet consumer  
182 demand. Yield could be improved more rapidly in hybrid wheat than in inbred cultivars

183 due to heterosis. Heterosis can be classified in three ways: (1) hybrid compared to the  
184 average of the parents, defined as mid-parent heterosis (MPH), (2) hybrid compared to  
185 the best parent, known as high parent heterosis, and (3) hybrid compared to the best  
186 commercial check, known as commercial heterosis (Bernardo, 2010). These estimates  
187 can be classified as either positive heterosis,  $> 0\%$ , no change, heterosis = 0, or negative  
188 heterosis,  $< 0\%$  based on the calculated performance. Previous estimates of grain yield  
189 increase due to heterosis in wheat ranged from 3.5 to 15% (Longin et al., 2012). In  
190 addition to heterosis, hybrid wheat benefits producers through yield stability, especially  
191 in marginal environments, with increased grain weight and tillering, improved biomass  
192 production and rooting depth, and increased biotic and abiotic stress tolerance (Longin et  
193 al., 2012; Tadesse et al., 2019; Tester & Langridge, 2010).

194 However, it is difficult to create wheat hybrids due to floral structure that generates  
195 inbreeding and challenges related to creating a cost-effective and easy to use hybrid  
196 fertility system (Whitford et al., 2013). Within a breeding program, a hybrid's heterosis  
197 expression could be decreased due to the genetic similarity between parental lines. Maize  
198 breeders observed that genetic similarity can decrease yield heterosis. The breeders found  
199 that by crossing two parents with optimized genetic differences, hybrids yielded more  
200 than hybrids with parents having less or more extreme genetic differences (Koekemoer et  
201 al., 2011; Melchinger, 2015). To increase hybrid wheat yield potential, the genetic  
202 mechanisms that impact heterosis should be investigated and divergent populations of  
203 inbred parents that maximize heterosis need to be created (Whitford et al., 2013).

204 The genetic differences that lead to heterosis could also lead to gene expression  
205 differences that help explain heterosis. Differentially expressed genes (DEGs) can be

206 categorized into different expression gene groups. The expression gene groups are  
207 defined as follows: over high parent, where there are more gene transcripts present in the  
208 hybrid compared to the parent with the highest expression level for that gene; high  
209 parent, where the gene expression of the hybrid is equal to the parent with the highest  
210 expression level of the specific gene; between parent, which is less than the high parent  
211 but more than the low parent; low parent, where a gene's expression level in the hybrid is  
212 equal to the parent with the lowest expression level; or under low parent, where the gene  
213 expression present in the hybrid is lower than the parent with the lowest expression level  
214 (Yang et al., 2018). Researchers found up to 82% of DEGs expressed in maize F<sub>1</sub> hybrids  
215 derived from crossing B73 and MO17, two maize lines that are known to produce hybrids  
216 with high positive heterosis, were classified as between parents. However, classification  
217 of expression differs across species and even within species (Stupar & Springer, 2006).  
218 For example, Yang et al., (2018) showed different gene action classifications in Brazilian  
219 rubber tree, (*Hevea brasiliensis* (Willd. Ex A. Juss.) Müll. Arg.) seedlings. Three rubber  
220 tree hybrids were created from two phenotypically distinct, heterozygous parents, and  
221 DEGs were identified and classified using the above gene expression groups. The  
222 researchers found that for the high performing, high heterosis hybrids, many DEGs were  
223 grouped as having expression over the high parent or under the low parent. In contrast,  
224 the low performing, low heterosis, hybrid showed minimal DEGs categorized into under  
225 the low parent or over high parent while 78% of the DEGs were classified as greater than  
226 the high parent or equal to the low parent's expression (Yang et al., 2018).

227 The gene expression groupings discussed in Yang et al., (2018) were used to  
228 introduce three common genetic models to explain the phenotypic response to heterosis

mechanisms on the allelic level: dominance, overdominance, and epistasis. The dominance model describes heterosis as the function of complementing action of marginally deleterious recessive alleles. An example of the allelic interaction would be when the gene contains the two dominant alleles, AA, the phenotypic response is equal to the heterozygote, Aa, but both produce a phenotype that is better than the recessive alleles, aa. The overdominance model explains how a favorable allelic interaction at a heterozygous locus is responsible for heterosis, for example the phenotypic performance is higher when the gene has heterozygotic loci Aa compared to AA or aa. Epistasis models describe heterosis resulting from gene-by-gene interactions at different loci. For example, if a hybrid contains the alleles A\_B\_ for two different loci, the presence of both dominant alleles produces a better phenotypic response than when there is only one dominant allele at one locus, A\_bb (Birchler et al., 2006). In rubber trees, traits such as improved hybrid yield, seedling growth, and vigor, found in the high-yielding hybrids, are modeled by overdominance, while the phenotypic trait of decreased hybrid yield found in the low-yielding hybrids is explained by the dominance genetic model. The presence of heterozygote alleles at loci made the high heterosis hybrids perform better than hybrids with loci containing both dominant alleles. In both the high yielding and low yielding rubber tree hybrids, it was concluded that DEGs can be used to explain observed heterosis (Yang et al., 2018). It is apparent that gene action classification differences are present between and within species, so DEGs should be studied for each crop species.

Because gene expression can be used to explain heterosis mechanisms in a hybrid, RNA-Seq, a method for quantifying gene expression, can be used to explore the genetic basis of heterosis. RNA-Seq uses mRNAs extracted from tissues to create cDNAs that are



252 then sequenced, producing transcript information to identify DEGs that are used to  
253 explain the phenotypic differences between samples. The mRNA is sheared at random  
254 locations using enzymes, the fragments are converted to cDNA libraries which are then  
255 sequenced, and the total number of reads produced for a certain transcript is proportional  
256 to the expression level of the corresponding gene. However, the proportion can be biased  
257 towards longer transcripts. This bias can be partially corrected by knowing the size of the  
258 transcript though transcript size can be inaccurate in non-model (less characterized)  
259 species. 3' RNA-Seq was introduced to correct for the bias in standard RNA-Seq. 3'  
260 RNA-Seq only sequences the transcript fragment from the 3' end. Transcript size does  
261 not affect fragment sequencing. This method produces the number of transcripts directly  
262 related to the expression level of the gene (Tandonnet & Torres, 2017). Another study  
263 conducted by Lohman et al., (2016) compared RNA-Seq and 3' RNA-Seq and concluded  
264 that the main difference was the expense of the two methods but both methods identify  
265 similar DEGs. RNA-Seq is more expensive than 3' RNA-Seq due to sequencing the full  
266 lengths of the transcripts and requiring a high depth of coverage. ENCODE Consortium  
267 suggests the best practice for an RNA-Seq study is to have ~30 million raw reads per  
268 sample (The ENCODE Consortium, 2011), which limits the level of multiplexing. The  
269 high cost of creating cDNA libraries and sequencing make RNA-Seq unaffordable for  
270 researchers to use this method on a large scale. However, small sample numbers  
271 negatively impact the statistical power of the research design and run the risk of missing  
272 nuanced ways that heterosis can be quantified. 3' RNA-Seq libraries cost less to generate  
273 and do not require as many raw reads to be produced for each sample. Therefore, more  
274 samples can be run per lane. Drawbacks to the 3' method include the inability to

275 distinguish alternatively spliced transcripts from a locus or identify polymorphisms or  
276 allele-specific expression (Lohman et al., 2016). However, 3' RNA-Seq was able to  
277 increase sample size, increase experimental power to improve the ability to detect  
278 differences in DEGs, and identify low abundance transcripts more accurately than RNA-  
279 Seq (Lohman et al., 2016; Tandonnet & Torres, 2017).

280 While yield heterosis exists in hybrid wheat, it has not been studied to the same extent  
281 as it has been in maize, rice (*Oryza sativa* L.), or even rubber trees. While in the early  
282 stages of developing a hybrid wheat breeding program, the University of Nebraska-  
283 Lincoln small grains breeding program has shown that yield heterosis varies over year  
284 and environment (Easterly et al., 2020). Therefore, we decided to investigate the genes  
285 that were expressed in positive heterosis (e.g., heterotic) hybrids and genes that were  
286 expressed in negative heterosis, (e.g., non-heterotic or decreased yield) hybrids. We hope  
287 to determine what controls hybrid performance and how those hybrids differ from their  
288 parents to understand the genetic mechanisms that impact observed yield heterosis.

289 To explore the gene expression of heterosis, our study was designed to explore the  
290 following objectives: 1) determine if the identified DEGs are expressed in novel  
291 groupings that may help explain the observed phenotypic differences among parents and  
292 the hybrid, 2) determine how gene expression can be classified in relation to the parents'  
293 gene expression by taking hybrid purity into account and 3) to determine if the  
294 environment impacted heterosis expression through DEG analysis between the two  
295 sampling years.

### 296 3. MATERIALS AND METHODS

#### 297 3.1 *Plant Material*

298 Every year, ~350 wheat hybrids are created in the UNL small grains hybrid  
299 research project, in collaboration with TAMU, using Nebraska- and Texas-developed  
300 inbred lines that are crossed in a field crossing block. The female lines were sterilized  
301 using Croisor 100<sup>®</sup>, (active ingredient sintofen; 1-(4- chlorophenyl)-5-(2-  
302 methoxyethoxy)-4-oxo-1,4-dihydrocinnoline-3-carboxylic acid) (Asur Plant Breeding,  
303 Estrées-Saint-Denis, France), a chemical hybridizing agent. The resulting seed was  
304 harvested and the F<sub>1</sub> progeny is grown in three sites across the state of Nebraska, Lincoln  
305 (Havelock Research Farm, UNL), North Platte (West Central Research and Extension  
306 Center, UNL), and near Alliance (collaborator's field site) to evaluate yield and  
307 agronomically important traits. The F<sub>1</sub> trial at Lincoln served as the source material for  
308 this study. An average of 17 unique triads (defined as an F<sub>1</sub> hybrid and both parent  
309 genotypes) were collected each year for two years, 2019, Year 1, and 2021, Year 2, and  
310 to find a representative five heterotic triads and five non-heterotic triads.

311 In Year 1, the hybrids were grown in an augmented design and the parents were  
312 grown in a randomized complete block design (RCBD). Triads were selected for tissue  
313 sampling based on positive and negative MPH calculated with grain yield data from the  
314 previous year's hybrid yield trial (MPH equation shown below). MPH was used in this  
315 research instead of high parent heterosis or commercial heterosis because it is the  
316 standard in genetic research studies. Twenty triads were identified to sample.

$$\%MPH = \left[ \frac{\left( \text{Yield}_{\text{Hybrid}} - \left( \frac{\text{Yield}_{\text{Female Parent}} + \text{Yield}_{\text{Male Parent}}}{2} \right) \right)}{\left( \frac{\text{Yield}_{\text{Female Parent}} + \text{Yield}_{\text{Male Parent}}}{2} \right)} \right] \times 100$$

The selected genotypes were scored for flowering date (BBCH stage 65, (German Federal Biological Research Centre for Agriculture and Forestry, 2001)), on a plot basis. Eight single stems were randomly tagged within the plot two to four days after flowering (DAF) representing the biological replicates to be sampled. Flag leaf and immature kernels were sampled from each replicate. Samples were collected on either May 31 or June 5 depending on the plot's flowering date as well as staging the stems that were previously tagged. All tissues of a triad were collected on the same day. Sampling began around 11:00 a.m. and ended by 1:00 p.m. to minimize circadian rhythm effects on gene expression. Seed tissue age ranged between five and 12 DAF with an average of nine DAF, falling within the 69-71 stages on the BBCH scale. Differences in seed age can be seen in Figure 1.1. From each stem, 10 immature seeds were collected followed by cutting off and storing half of the flag leaf. The samples were flash frozen in liquid nitrogen in the field and moved to a -80°C freezer for storage.

In Year 2, the hybrid trial with a subset of parents was grown in an alpha lattice field design with each genotype replicated three times. 16 triads were selected for sampling based on high parent heterosis calculations from previous yield trials and flowering dates were taken (high parent heterosis equation shown below). High parent heterosis was used to pick triads for sampling because we found it increased the probability of identifying triads with the desired heterosis levels compared to MPH. All

337 final triad selections were made using MPH based on the yield of the actual hybrid and  
 338 parent plots sampled.

$$339 \quad \% \text{High Parent Heterosis} = \left[ \frac{(\text{Yield}_{\text{Hybrid}} - \text{Yield}_{\text{Highest Yielding Parent}})}{\text{Yield}_{\text{Highest Yielding Parent}}} \right] \times 100$$

340 Six individual stems in each plot were tagged to mark the biological replicates  
 341 two to four DAF. Sampling was done on May 29 and June 1 based on plot flowering  
 342 dates and individual stem age. Sampling was conducted using the same method used in  
 343 2019 with the exception that *RNAlater*<sup>®</sup> (Thermo Fisher Scientific, Waltham,  
 344 Massachusetts), an RNase inactivator and RNA stabilizer for tissues and cells collected in  
 345 the field, was used to preserve the samples. The sampled tissues were stored at 4°C  
 346 overnight, and then moved to -80°C freezer for longer storage following the  
 347 manufacturer's protocol. Seed tissue age ranged between 6 and 14 DAF with an average  
 348 of 9 DAF, again falling within the 69-71 stages of the BBCH scale (Figure 1.1).

349 The 2019 and 2021 trials were harvested for grain yield and the raw values were  
 350 adjusted for spatial variation using the ASReml-R v4.1.0.160 to produce the associated  
 351 BLUPs (Butler et al., 2007). The statistical model of the yield response variable,  $Y_{ijk}$ , of  
 352 the  $i^{\text{th}}$  number of treatments,  $j^{\text{th}}$  number of times each treatment is replicated, and  $k^{\text{th}}$   
 353 number of blocks found in each replicate, was analyzed as:

$$354 \quad Y_{ijk} = \mu + \tau_i + \gamma_j + \rho_{k(j)} + \epsilon_{ijk}$$

355 Where  $\mu$  is the grand mean for the entire experiment,  $\tau_i$  is the effect of the  $i^{\text{th}}$  treatment,  
 356  $\gamma_j$  is the effect for the  $j^{\text{th}}$  replicate,  $\rho_{k(j)}$  is the block within replicate effect, which is a  
 357 nested variable in this model, and  $\epsilon_{ijk}$  is the error term. This model produced the BLUPs  
 358 for each genotype in the trial. The adjusted BLUPs were used to calculate MPH for each

359 sampled triad. In Year 1, four triads were selected for RNA extraction, two heterotic and  
360 two non-heterotic. In Year 2, six triads were selected for RNA extraction, with three as  
361 heterotic and three as non-heterotic.

### 362 ***3.2 RNA Extractions***

363       After identifying the appropriate triads, total RNA was extracted from five  
364 biological replicates from the flag leaf and immature seed tissue using the Qiagen  
365 RNeasy<sup>®</sup> Plant Mini Kit (Qiagen<sup>®</sup>, Hilden, Germany) and its protocol with a few  
366 modifications. The remaining three replicates were stored at -80°C for future testing if  
367 needed. The provided RLT buffer was used for the leaf tissue and RLC buffer was used  
368 for the seed tissue as it was superior for extracting RNA from starchy immature seed  
369 compared to the RLT buffer, as determined by optimization testing completed before  
370 RNA extractions (unpublished results). Occasionally, when the ground seed tissue was  
371 added to the RLC buffer, it would congeal and between 200-400 µL RLC buffer was  
372 added to liquify the solution. We performed the optional on-column DNase digestion step  
373 on the kit-provided spin column using the Qiagen<sup>®</sup> RNase-free DNase Kit (Qiagen<sup>®</sup>,  
374 Hilden, Germany) following the manufacturer's protocol. Leaf tissue samples required  
375 one to two additional washings with buffer RPE to remove excess chlorophyll from the  
376 column as determined by the greenish extract color (Figure 1.2). Finally, the spin column  
377 was washed with 20 µL RNase-free water and stored in a 1.5 mL collection tube, and  
378 then washed again and stored in a 1.5 mL collection tube to remove any remaining RNA  
379 for a backup sample.

### 380 **3.3 RNA Sequencing and Read Processing**

381       After extraction, the RNA samples were sent to the University of Nebraska  
382 Medical Center Genomics Core Facility in Omaha, NE. There, the samples were  
383 evaluated on an Advanced Analytical Technologies, Inc. Fragment Analyzer<sup>™</sup> (recently  
384 acquired by Agilent Technologies, Inc., Santa Clara, California) for quality, producing  
385 RQN values and sample concentrations for both years. Using the RQN quality scores and  
386 RNA concentrations received from the Genomics Core, the three best biological  
387 replicates per genotype were selected for sequencing resulting in 168 samples (ten triads  
388 x two tissues x three biological replicates x three genotypes per triad minus the common  
389 parents). All samples were diluted to 50 ng/μL and submitted to the Genomics Core for  
390 RNA sequencing. Libraries were prepared using the Lexogen<sup>®</sup> QuantSeq<sup>™</sup> 3' mRNA-Seq  
391 Library Prep Kit (Lexogen, Vienna, Austria). Library quality was checked using a Qubit<sup>®</sup>  
392 3.0 fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts) and a BioAnalyzer  
393 2100 (Agilent Technologies, Inc., Santa Clara, California). Sequencing was completed on  
394 an Illumina<sup>®</sup> NextSeq550 using 75SR High Output Flow Cell kits (Illumina<sup>®</sup>, San Diego,  
395 California) and produced 75 bp, single-end reads. In Year 1 and Year 2, we randomly  
396 assigned samples to a flowcell at 33-multiplexed and 30-multiplexed respectively.

397       The raw reads were investigated using *FastQC* v0.11 to assess the quality of the  
398 reads (Andrews et al., 2015). The results showed the first 12 bp contained high variation  
399 in per base sequence content and 10 bp that had low per base sequence quality at the tail  
400 end. *Trimmomatic* v0.38 was used to trim off the poor-quality bases and any other  
401 contaminants found in the reads using a head crop of 12 bp, a crop length of 60 bp, as  
402 well as a contaminants file that contained the most common Illumina contaminants

403 associated with 3' RNA sequencing (Bolger et al., 2014). After rerunning the trimmed  
404 reads through *FastQC*, the overall quality of the reads improved and could be mapped to  
405 the reference genome using the *STAR* v2.7 read aligner package and the IWGSC RefSeq  
406 v2.1 genome assembly and annotations (Dobin et al., 2013; T. Zhu et al., 2021).  
407 Annotation files with gff3 extension were converted to gtf files using *Cufflinks* v2.2  
408 (Trapnell et al., 2010). Gene level counts were created using the *htseq -count* command  
409 in the *HTSeq* v0.9 program and then exported into a csv file (Anders et al., 2015).

#### 410 **3.4 Gene Expression Analysis**

411 The R program *DESeq2* v1.32.0 was used to identify major sources of variation in  
412 the dataset, normalize and filter gene counts, and create contrasts to compare DEGs  
413 (Love et al., 2014). Variation was investigated using the built in “plotPCA” command in  
414 the *DeSeq2* program as well as *ggplot2* v3.3.5 (Wickham, 2009). Gene normalization and  
415 filtering was accomplished using the built-in programming of the “DESeq” command on  
416 the data set. Results were created with the “results” function using the pairwise  
417 comparisons defined in





419 Table 1.7, and Table 1.8, and an alpha of 0.05. The results' log fold change was  
420 shrunk using “lfcShrink” command with the “apeglm” option in *DeSeq2* (Anders &  
421 Huber, 2010; A. Zhu et al., 2019). The shrunken results were then exported after filtering  
422 the DEGs for an adjusted p-value less than 0.05 and an absolute log fold change of 2-fold  
423 or greater. The filtered DEGs from each parent to hybrid comparison were combined and  
424 evaluated for shared gene expression across triad hybrid/parent comparisons. All hybrids  
425 were then evaluated for differential expression trends across triads, for example Log Fold  
426 Change = Triad 1 Female Parent Expression Score – Triad 1 Hybrid Expression Score.  
427 Genes were considered up-regulated, or expressed more in the hybrid, if the log fold  
428 expression change was positive. Genes were considered downregulated, or expressed  
429 more in the parent, if the log fold change was negative.

## 430 **4. RESULTS**

### 431 ***4.1 Plant Material***

432 Trial yields were used to calculate MPH for each triad as shown in Table 1.1 with  
433 trial yields, flowering dates, and sampling dates included for year and triad comparison.  
434 Year 1 MPH calculations identified two triads with positive MPH, triads 1 and 2, and two  
435 with negative MPH, triads 6 and 7. There was one shared parent, “Panhandle”, found in  
436 the two negative MPH triads. In Year 2, we used an additional six unique triads not  
437 sampled in 2019, three with positive MPH and three with negative MPH, triads 3, 4, and  
438 5 and 8, 9, and 10 respectively. In Year 2, there was one shared parent, “SD10W153”,  
439 found in a positive MPH triad and a negative MPH triad. The average yield and percent  
440 MPH were calculated for both hybrid yield trials in Year 1 and Year 2 as well as the time  
441 of year the trials flowered. There were differences observed between the two trials. In

Year 2, flowering occurred earlier than in Year 1. There average yield for Year 2 was higher than Year 1, related to the positive MPH observed in Year 2. Lower yields and negative MPH were observed in Year 1. Triad 1, sampled in Year 1, had the best MPH of 35.67%, meaning the hybrid yielded 36% more than the average of the parents. The lowest MPH was observed in Year 1 as well in Triad 6, where the hybrid yielded 33% less than the average of the parents.

#### 4.2 RNA Extraction

After extracting the RNA from the selected triads and evaluating their quality, RQN and concentration was compared to select the best three biological replicates/genotype. Year 1 immature seed samples had an average RQN of 9.1 and concentration of 150 ng/ $\mu$ L while the leaf samples had an average RQN of 6.6 and concentration of 1,449 ng/ $\mu$ L. Year 2 immature seed samples had an average of 5.7 RQN and 337.42 ng/ $\mu$ L and leaf samples had an average of 6.1 RQN and 1,856.81 ng/ $\mu$ L (Table 1.2). By selecting the replicates with high RQN values and concentrations above 50 ng/ $\mu$ L, the sequenced samples' quality improved to 9.4 and 6.8 for Year 1 seed and leaf samples respectively and 6.4 and 6.2 for the Year 2 leaf and seed respectively (Table 1.2).

#### 4.3 RNA Sequencing and Read Processing

Sequencing produced an average of 14,660,092 sequences. *FastQC* reports found no sequencing barcode or adaptor contamination and no sequences flagged as inadequate quality in any of the 168 samples. Two issues were identified with the reads: specifically, per base sequence quality decreased (quality score less than 20) on the 5' end of the read and had poor per base sequence content on the 3' end of the reads. *Trimmomatic* kept an

465 average of 12,888,106 reads per sample, keeping 87.95% and dropping 12.05% of reads.  
466 The *STAR* aligner mapped reads back to the reference genome at an average rate of 52%  
467 uniquely mapped. However, we observed a major difference in the rate of leaf and seed  
468 samples mapping to the genome, 45% to 60% uniquely mapping, respectively (Table  
469 1.3). The program mapped read lengths within two bp of the input read lengths.

#### 470 ***4.4 Gene Expression Analysis***

471 The two main components for variation within the dataset were identified as the  
472 sample tissue (leaf vs. seed) component, representing 86% of the variation, and a  
473 secondary unidentified component in the seed tissue, which represented 6% of the overall  
474 variation (Figure 1.3). The secondary component was explored by looking at the  
475 interaction of triad and type which did not provide an explanation for the second  
476 component of variation (Figure 1.4). The age of the sample when it was collected was  
477 then investigated and showed that younger samples, seven to eight DAF, clustered  
478 together and explained the second source of variation (Figure 1.5).

479 To identify any genes that were differentially expressed between the heterotic and  
480 non-heterotic hybrids, all of the heterotic hybrids and non-heterotic hybrids pairwise  
481 comparison showed that in the leaf tissue hybrids, 9 more genes were over-expressed in  
482 the heterotic hybrids when compared to non-heterotic hybrids. Conversely, 9 more genes  
483 were expressed in the non-heterotic hybrids than the heterotic hybrids in the seed tissue (

484        Table 1.4). ~2,000 more genes were expressed in leaf tissue than in seed tissue for  
485        both heterotic and non-heterotic triads (Table 1.5). The pairwise comparison between the  
486        2019 and 2021 samples found that for both leaf and seed tissues, ~100 and 600 unique  
487        genes were expressed in 2019 compared to 2021 respectively (Table 1.6). Additional  
488        analysis on how the hybrids' gene expression compares to the parents' gene expression  
489        will be completed in the future.

490        We then conducted pairwise comparisons between each parent to the hybrid to  
491        identify DEGs. Comparisons done on the leaf samples (shown in

Table 1.7) found a total of 1,189 genes differentially expressed in the heterotic hybrids compared to their parents, while 1,707 genes were differentially expressed in the non-heterotic hybrid/parent comparisons. On average, 64 genes were identified with higher expression in the heterotic hybrids for each hybrid/parent comparisons while 98 genes were identified with higher expression in the non-heterotic hybrids for the hybrid/parent comparisons. The results of the hybrid/parent comparisons were compared and 402 shared DEGs across the 20 comparisons were identified and depicted in Figure 1.6. This figure shows the expression pattern of these DEGs and depict how the two heterotic classes share genes with opposite log fold change values. Expression analysis of the seed tissue found that 2,615 and 3,130 genes were differentially expressed in the heterotic and non-heterotic hybrid/parent comparisons respectively (Table 1.8). As was found in the leaf tissues, in the immature seed tissues more genes were over-expressed in the non-heterotic hybrids. When evaluating the DEGs across the 20 hybrid/parent combinations, the comparisons found 600 genes differentially expressed in the 20 comparisons and are shown in Figure 1.7. In the DEG analysis for hybrid/parent comparisons, we identified four triads in both seed and leaf tissue whose gene expression patterns depicted the possibility of self-pollination. Triads 5, 8, 9, and 10 are potentially self-pollinated due to the higher number of DEGs found in the male/hybrid comparisons than in the female/hybrid comparisons. This observation could indicate the hybrids are more like the female parents than the male parents because the larger number of DEGs present in the male/hybrid indicate a difference in gene expression that is not shared in the female/hybrid comparison.

## 514    **5. DISCUSSION**

515            3' RNA-Seq has proved to be a valuable tool for investigating genes being  
516 expressed in different tissues as well as different performance groups. In this study, we  
517 used 3'RNA-seq to characterize gene expression in ten triads that showed positive or  
518 negative levels of mid-parent heterosis. The triads had to be sampled over two years as  
519 we could not accurately predict if a hybrid would show positive or negative heterosis and  
520 complete the sampling in the same year, thus removing most environmental effects on the  
521 number of DEGs identified. We were able to extract quality RNA from both leaf and  
522 immature kernel tissue. Sequencing the tissues produced an average of almost 15 million  
523 sequences per sample, providing enough coverage to identify differentially expressed  
524 genes. However, with the observed variability in creating hybrids with high mid-parent  
525 heterosis in the UNL program, it was still unknown if we would find any genes that were  
526 differentially expressed (Easterly et al., 2020). While using 3' RNA-Seq reduced our  
527 mapping rate to an average of 50% of transcripts mapped to unique locations in the  
528 genome, less than 1 bp was removed from the input reads indicating the trimming  
529 removed them majority of unmappable base pairs. Also, despite reducing the mapping  
530 rate, 3' RNA-Seq still allowed us to identify DEGs and was more cost effective than  
531 traditional RNA sequencing.

532            Before identifying DEGs, we had to understand if and what kind of variation  
533 present in the dataset. While most of the variation is due to tissue type, there was a slight  
534 amount of variation found in the seed samples that could not be explained by the other  
535 obvious categories, heterotic classification (not shown), triad, or the type of line (hybrid,  
536 male, or female) (Figure 1.4). However, it was concluded that the age of the kernels when

537 they were collected explained the spread of variation in the seed samples. Figure 1.5  
538 shows that the blue dots, representing seven and eight DAF, have the greatest variation,  
539 but there was also considerable variation between the blue and red dots, 11 and 12 DAF.  
540 This observation makes sense biologically because as a kernel develops, it may have  
541 different genes expressed compared to younger kernels. The variation in sample age  
542 highlights that sampling time was very important and why we tried to sample the triads as  
543 close to the same age as possible. The variation found between the leaf and seed tissue  
544 groups suggested that the analysis needed to be done separately as they would have a  
545 confounding effect on any observed DEGs. The tissue effect was also seen in Table 1.5  
546 where there were always more genes expressed in the leaf tissue compared to the seed  
547 tissue, possibly due to the kernels being immature and still developing or the leaves are  
548 more complex organisms.

549       Although the hybrids were grown in two different years, we were able to identify  
550 DEGs associated with heterotic and non-heterotic hybrids pairwise comparison. While  
551 gene expression in the leaf and seed tissue differed significantly; more DEGs were  
552 identified in the heterotic hybrids for the leaf samples while more DEGs were found in  
553 the non-heterotic hybrids for the seed samples. We did not expect this result because the  
554 number of DEGs should be similar in both tissues when sampled from the same plant. If  
555 more DEGs were found in the heterotic hybrid leaf samples, we expected there should be  
556 more DEGs present in the heterotic hybrid seed samples. It will warrant further  
557 investigation to determine why this is occurring and its biological meaning.

558       The pairwise comparisons conducted on the parent/hybrid contrast of each triad  
559 showed that in both leaf and seed samples, more DEGs between the parents and hybrids



560 were found in the non-heterotic hybrids. We could hypothesize that the genes that show  
561 higher expression in the non-heterotic hybrids could be controlling a metabolic pathway  
562 which impacts the yield produced. However, this hypothesis needs to be explored further  
563 using pathway annotations to understand what role these genes have in expression. These  
564 gene pathways could show why the heterotic hybrids perform better than the non-  
565 heterotic group. The metabolic pathway hypothesis can be visualized in Figure 1.6 and  
566 1.7 as there are genes that are shared across the two groups while some are unique to a  
567 specific group.

568         The pairwise comparisons in both leaf and seed tissues also introduced concern  
569 that some of the hybrids were self-pollinations instead of crosses. There were large  
570 numbers of genes found to be differentially expressed in the male/hybrid comparison  
571 compared to the female/hybrid comparison, indicating that more of the hybrid's alleles  
572 are from the female. The similarity could indicate that the genes that are differentially  
573 expressed in the female/hybrid comparison are due to environmental factors and not to  
574 the hybrid being a mixture of the female's and male's alleles. This observation is also  
575 shown in Figure 1.8, showing the variation present in the three parts of Triad 5. The  
576 hybrids and female parents are clustered together while there is more gene count  
577 variation between the hybrids/females to the male genotypes for both leaf and seed tissue.  
578 While we have not been able to identify if a self-pollination occurred with this dataset,  
579 future hybrid studies should take this into consideration and conduct DNA sequencing to  
580 identify self-pollinations.

581         Because samples had to be collected in two different years, we investigated how  
582 that impacted the number of DEGs and found that more genes were expressed in 2019

583 compared to 2021 for both tissues. The two environments had significant differences  
584 between temperature and precipitation. In calendar year 2019, the yearly average overall  
585 temperature was 10.5°C, the average maximum temperature was 27.5°C, and the average  
586 minimum temperature was -5.6°C. Total precipitation that year was 917.19 mm (National  
587 Weather Service, 2020). On the other hand, in 2021 the average temperature was 12°C,  
588 average maximum temperature was 18.8°C, and the average minimum was 5.3°C with  
589 total precipitation totaling 682.24 mm (National Weather Service, 2022). We believe that  
590 more genes were expressed in 2019 as it had more extreme temperatures while 2021 was  
591 not as stressful of an environment.

## 592 **6. CONCLUSIONS**

593 The results presented in this study indicate the validity of exploring the transcriptome  
594 in hybrid wheat to identify reasons why heterotic hybrids perform better than non-  
595 heterotic hybrids. There are significant differences between the transcriptomes of the two  
596 classes of hybrids that indicate the importance of identifying the roles that these DEGs  
597 play in metabolic pathways which would allow us to identify why a heterotic hybrid is  
598 phenotypically different than a non-heterotic. Metabolic pathway identification will be  
599 conducted in future work on this project. The differences between the hybrids and their  
600 parents also imply the need to determine gene expression classifications for these triads  
601 to further tease out how the expressed genes impact the positive and negative heterosis  
602 expressed in the hybrids compared to the parents. When conducting these parent/hybrid  
603 comparisons, we will need to distinguish and remove possible self-pollinated hybrids  
604 before doing the DEG analyses. This research is the corner stone for future work done  
605 within the UNL small grains group to continue to understand heterosis in hybrid wheat.

## 606 7. REFERENCES

- 607 Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data.  
 608 *Genome Biology*, 11(10), 1–12. <https://doi.org/10.1186/gb-2010-11-10-r106>
- 609 Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with  
 610 high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169.  
 611 <https://doi.org/10.1093/BIOINFORMATICS/BTU638>
- 612 Andrews, S., Krueger, F., Secondy-Pichon, A., Biggins, F., & Wingett, S. (2015).  
 613 *FastQC. A quality control tool for high throughput sequence data*. Babraham  
 614 *Bioinformatics*. Babraham Institute.  
 615 <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- 616 Bernardo, R. (2010). *Breeding for Quantitative Traits in Plants*. Stemma Press.
- 617 Birchler, J. A., Yao, H., & Chudalayandi, S. (2006). Unraveling the genetic basis of  
 618 hybrid vigor. *Proceedings of the National Academy of Sciences of the United States*  
 619 *of America*, 103(35), 12957. <https://doi.org/10.1073/PNAS.0605627103>
- 620 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for  
 621 Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.  
 622 <https://doi.org/10.1093/bioinformatics/btu170>
- 623 Butler, D., Cullis, B. R., Gilmour, A. R., & Gogel, B. J. (2007). ASReml-R reference  
 624 manual. AsReml-R estimates variance components under a general linear mixed  
 625 model by residual maximum likelihood (REML). *The State of Queensland*,  
 626 *Department of Primary Industries and Fisheries, Brisbane, Queensland, Australia*.  
 627 <http://www.homepages.ed.ac.uk/iwhite/asreml/uop>.
- 628 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,  
 629 Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq  
 630 aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- 631 Easterly, A. C., Garst, N., Belamkar, V., Ibrahim, A. M. H., Rudd, J. C., Sarazin, J. B., &  
 632 Baenziger, P. S. (2020). Evaluation of hybrid wheat yield in Nebraska. *Crop*  
 633 *Science*, 60(3), 1210–1222. <https://doi.org/10.1002/CSC2.20019>
- 634 Food and Agriculture Organization of the United Nations. (2020). *FAOSTAT Statistical*  
 635 *Database*.
- 636 German Federal Biological Research Centre for Agriculture and Forestry. (2001). Growth  
 637 stages of mono-and dicotyledonous plants. In U. Meier (Ed.), *BBCH Monograph*.  
 638 [http://pub.jki.bund.de/index.php/BBCH/article/view/515%0Ahttp://pub.jki.bund.de/i](http://pub.jki.bund.de/index.php/BBCH/article/view/515%0Ahttp://pub.jki.bund.de/index.php/BBCH/article/download/515/464%0Ahttp://pub.jki.bund.de/index.php/BBCH/article/view/515/464)  
 639 [ndex.php/BBCH/article/download/515/464%0Ahttp://pub.jki.bund.de/index.php/BB](http://pub.jki.bund.de/index.php/BBCH/article/download/515/464%0Ahttp://pub.jki.bund.de/index.php/BBCH/article/view/515/464)  
 640 [CH/article/view/515/464](http://pub.jki.bund.de/index.php/BBCH/article/view/515/464)
- 641 Graybosch, R. A., & James Peterson, C. (2012). Specific adaptation and genetic progress  
 642 for grain yield in Great Plains hard winter wheats from 1987 to 2010. *Crop Science*,  
 643 52(2), 631–643. <https://doi.org/10.2135/cropsci2011.08.0412>
- 644 Houchmandzadeh, B., Paulsson, J., Vilar, J., Weitz, C., Young for discussions, M.,  
 645 Questembert-Balaban, N., Kussell, E., & Vallade, M. (2004). Evolutionary changes  
 646 in cis and trans gene regulation. *Nature* 2004 430:6995, 430(6995), 85–88.  
 647 <https://doi.org/10.1038/nature02698>
- 648 Koekemoer, F., van Eeden, E., & Bonjean, A. (2011). An overview of hybrid wheat  
 649 production in South Africa and review of current worldwide wheat hybrid

- developments. In A. Bonjean & W. Angus (Eds.), *The world wheat book-a history of wheat breeding* (Vol. 2, pp. 907–950). Lavoisier Publishing.
- Lohman, B. K., Weber, J. N., & Bolnick, D. I. (2016). Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Molecular Ecology Resources*, 16(6), 1315–1321. <https://doi.org/10.1111/1755-0998.12529>
- Longin, C. F. H., Mühleisen, J., Maurer, H. P., Zhang, H., Gowda, M., & Reif, J. C. (2012). Hybrid breeding in autogamous cereals. *Theoretical and Applied Genetics*, 125(6), 1087–1096. <https://doi.org/10.1007/S00122-012-1967-7/FIGURES/2>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9>
- Melchinger, A. E. (2015). Genetic Diversity and Heterosis. In *The Genetics and Exploitation of Heterosis in Crops* (pp. 99–118). John Wiley & Sons, Ltd. <https://doi.org/10.2134/1999.GENETICSANDEXPLOITATION.C10>
- National Weather Service. (2020). *2019 Annual Climate Summary-Lincoln*. <https://www.weather.gov/media/oax/Climate/AnnualLincoln2019.pdf>
- National Weather Service. (2022). *Lincoln-NWS-2021-Climate / Lincoln Weather and Climate / Nebraska*. <https://lincolnweather.unl.edu/lincoln-nws-2021-climate>
- Stupar, R. M., & Springer, N. M. (2006). Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. *Genetics*, 173(4), 2199–2210. <https://doi.org/10.1534/genetics.106.060699>
- Tadesse, W., Sanchez-Garcia, M., Assefa, S. G., Amri, A., Bishaw, Z., Ogonnaya, F. C., & Baum, M. (2019). Genetic Gains in Wheat Breeding and Its Role in Feeding the World. *Crop Breeding, Genetics and Genomics*. <https://doi.org/10.20900/cbgg20190005>
- Tandonnet, S., & Torres, T. T. (2017). Traditional versus 3' RNA-seq in a non-model species. *Genomics Data*, 11, 9–16. <https://doi.org/10.1016/j.gdata.2016.11.002>
- Tester, M., & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. In *Science* (Vol. 327, Issue 5967, pp. 818–822). American Association for the Advancement of Science. <https://doi.org/10.1126/science.1183700>
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 2010 28:5, 28(5), 511–515. <https://doi.org/10.1038/nbt.1621>
- USDA ERS - Wheat Sector at a Glance. (n.d.). Retrieved December 26, 2021, from <https://www.ers.usda.gov/topics/crops/wheat/wheat-sector-at-a-glance/#classes>
- USDA-ERS. (2021). *USDA ERS - Wheat Data*. <https://www.ers.usda.gov/data-products/wheat-data/>
- Vespa, J., Medina, L., & Armstrong, D. M. (n.d.). *Demographic Turning Points for the United States: Population Projections for 2020 to 2060 Population Estimates and Projections Current Population Reports*. Retrieved December 26, 2021, from [www.census.gov/programs-surveys/popproj](http://www.census.gov/programs-surveys/popproj)

- Whitford, R., Fleury, D., Reif, J. C., Garcia, M., Okada, T., Korzun, V., & Langridge, P. (2013). Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *Journal of Experimental Botany*, 64(18), 5411–5428. <https://doi.org/10.1093/JXB/ERT333>
- Wickham, H. (2009). ggplot2. *Ggplot2*. <https://doi.org/10.1007/978-0-387-98141-3>
- Yang, H., Wang, X., Wei, Y., Deng, Z., Liu, H., Chen, J., Dai, L., Xia, Z., He, G., & Li, D. (2018). Transcriptomic analyses reveal molecular mechanisms underlying growth heterosis and weakness of rubber tree seedlings. *BMC Plant Biology*, 18(1), 1–11. <https://doi.org/10.1186/s12870-017-1203-3>
- Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-Tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics*, 35(12), 2084–2092. <https://doi.org/10.1093/bioinformatics/bty895>
- Zhu, T., Wang, L., Rimbart, H., Rodriguez, J. C., Deal, K. R., Oliveira, R. De, Choulet, F., Keeble-Gagnère, G., Tibbits, J., Rogers, J., Eversole, K., Appels, R., Gu, Y. Q., Mascher, M., Dvorak, J., & Luo, M.-C. (2021). Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant Journal*, 107(1), 303–314. <https://doi.org/10.1111/TPJ.15289>

**Table 1.1: Grain yield data from hybrid and parental line (triads) replicated trials grown at Lincoln, NE in 2019 or 2021.**

Hybrids are in bold followed by the female then male parent. Triads 1, 2, 6, and 7 were sampled in 2019. Triads 3, 4, 5, 8, 9, and 10 were sampled in 2021.

Category	Genotype	Flowering Date (Julian)	Yield (kg/ha)	MPH (%)
Trial Mean	Year 1	145	5,462.71	-3.04
	Year 2	141	5,895.82	2.74
Triad 1	<b>NXB17-5214</b>	144	6,437.24	35.67
	TX12A001638	147	3,701.60	
	NE14663	145	5,788.08	
Triad 2	<b>NXB17-5769</b>	148	5,630.69	4.52
	NE14419	147	5,241.73	
	NI14729	147	5,532.55	
Triad 3	<b>NXB17-5818</b>	142	6,188.58	12.72
	TX14M7088	142	5,997.23	
	TX12A001420	139	4,983.05	
Triad 4	<b>NXB17-5716</b>	141	6,416.46	6.87
	TAM204	142	6,353.81	
	SD10W153	140	5,654.38	
Triad 5	<b>NXB17-5226</b>	140	5,944.51	6.45
	TAM305	139	5,655.67	
	NI15713	144	5,513.25	
Triad 6	<b>NXB17-5393</b>	145	3,639.42	-32.58
	TX14M7051	145	4,937.54	
	Panhandle	146	5,858.03	
Triad 7	<b>NXB17-5394</b>	145	4,209.29	-26.04
	Sturdy2K	145	5,525.05	
	Panhandle	146	5,858.03	
Triad 8	<b>NXB17-5375</b>	142	5,685.82	-0.28
	TX14M7290	142	5,351.59	
	TX12V7220	142	6,052.15	
Triad 9	<b>NXB17-5713</b>	142	5,376.82	-1.46
	TX12A001638	142	5,258.65	
	SD10W153	140	5,654.38	
Triad 10	<b>NXB17-5861</b>	141	5,383.60	-3.49
	TX14M7051	142	5,461.38	
	TX14A001336	141	5,694.74	

**Table 1.2: Comparison of RNA Quality Parameters**

Comparison of the average RQN quality, 28S/18S integrity ratio, and concentration of both leaf and seed samples for Year 1 triads and Year 2 triads. Within a year, the rows show how the sample quality improved from the original five reps extracted to the three reps that were selected for sequencing.

	Sample	RQN	28S/ 18S	Concentration (ng/ $\mu$ L)
Year 1	Leaf Original 5 reps	6.6	1.3	1,448.87
	Leaf Selected 3 reps	6.8	1.3	1,668.04
	Seed Original 5 reps	9.1	1.7	150.54
	Seed Selected 3 reps	9.4	1.8	167.49
Year 2	Leaf Original 5 reps	6.1	1.2	1,856.81
	Leaf Selected 3 reps	6.4	1.3	2,276.84
	Seed Original 5 reps	5.9	1.6	337.42
	Seed Selected 3 reps	6.2	1.6	347.76

**Table 1.3: STAR Aligner Mapping Results**

Comparison between the average of all samples when mapped using *STAR* and how the tissue being mapped impacted the mapping result. All samples mapped the entire input read length minus 1 bp, indicating good input quality. Percentage uniquely mapped was higher for the seed samples than the leaf samples.

Sample	# of Input Reads	Input Read Length (bp)	Mapped Read Length (bp)	Uniquely Mapped (%)	Mapped to Multiple Loci (%)	Unmapped Reads (%)
Sample Avg	12,976,532	60	58.98	52	36	12
Leaf Avg	13,110,935	60	59.02	45	45	10
Seed Avg	12,854,238	60	58.95	60	27	13

**Table 1.4: DEG Analysis-Heterotic vs Non-Heterotic Pairwise Comparison**

DEG pairwise analysis to determine the number of genes differentially expressed in the heterotic and non-heterotic group. Similar number of DEGs were found, but more genes were expressed in the heterotic hybrids for the leaf tissue while more genes were expressed in the non-heterotic seed hybrids.

Pairwise Comparison	# of Genes Higher Expression in Heterotic Hybrids	# of Genes Higher Expression in Non-Heterotic Hybrids	Total # of DEGs
Leaf Tissue: Non-Heterotic vs Heterotic	10	1	11
Seed Tissue: Non-Heterotic vs Heterotic	3	12	15

**Table 1.5: DEG Analysis-Leaf vs Seed Pairwise Comparison**

DEG pairwise analysis to determine the number of genes differentially expressed in the leaf and seed tissue groups. Similar number of DEGs were found, but more genes were expressed in the leaf tissue for both heterotic groups.

Pairwise Comparison	# of Genes Higher Expression in Leaf Tissue	# of Genes Higher Expression in Seed Tissue	Total # of DEGs
Heterotic Tissue: Seed vs Leaf	8,499	6,532	15,031
Non-Heterotic Tissue: Seed vs Leaf	8,392	6,744	15,136

**Table 1.6: DEG Analysis-2019 vs 2021 Pairwise Comparison**

DEG pairwise analysis to determine the number of genes differentially expressed in the two years samples were taken. Similar number of DEGs were found, but more genes were expressed in 2019.

Pairwise Comparison	# of Genes Higher Expression in 2019	# of Genes Higher Expression in 2021	Total # of DEGs
Leaf Samples: 2021 vs 2019	682	509	1,191
Seed Samples: 2021 vs 2019	1,020	432	1,452



**Table 1.7: DEG Analysis-Leaf Triad Pairwise Comparisons**

DEG pairwise analyses for the parental lines to the hybrid of each triad for the leaf tissue.

More DEGs were identified in the non-heterotic hybrid comparisons than the heterotic hybrid comparisons.

Pairwise Comparison	# of Genes Up-Regulated in Hybrid	# of Genes Down-Regulated in Hybrid	Total # of DEGs
Triad 1: Female vs Hybrid	53	46	99
Triad 1: Male vs Hybrid	5	13	18
Triad 2: Female vs Hybrid	3	13	16
Triad 2: Male vs Hybrid	67	34	101
Triad 3: Female vs Hybrid	5	100	105
Triad 3: Male vs Hybrid	49	71	120
Triad 4: Female vs Hybrid	133	39	172
Triad 4: Male vs Hybrid	30	31	61
Triad 5: Female vs Hybrid	2	16	18
Triad 5: Male vs Hybrid	292	187	479
Triad 6: Female vs Hybrid	0	20	20
Triad 6: Male vs Hybrid	0	40	40
Triad 7: Female vs Hybrid	0	6	6
Triad 7: Male vs Hybrid	1	37	38
Triad 8: Female vs Hybrid	1	2	3
Triad 8: Male vs Hybrid	141	44	185
Triad 9: Female vs Hybrid	8	17	25
Triad 9: Male vs Hybrid	476	149	625
Triad 10: Female vs Hybrid	5	39	44
Triad 10: Male vs Hybrid	350	362	721

**Table 1.8: DEG Analysis-Seed Triad Pairwise Comparisons**

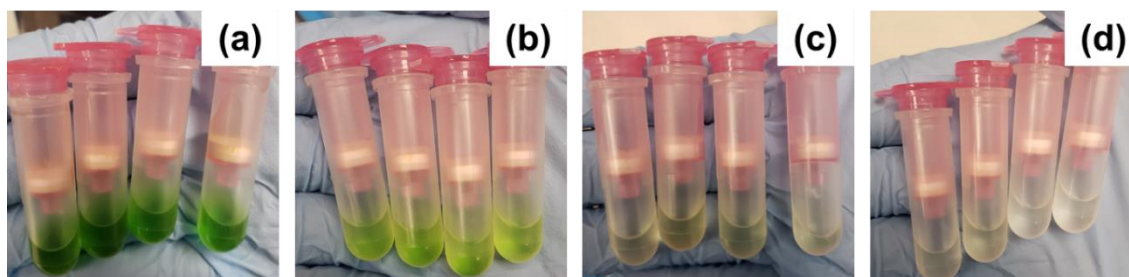
DEG pairwise analyses for the parental lines to the hybrid of each triad for seed tissue. More DEGs were identified in the non-heterotic hybrid comparisons than the heterotic hybrid comparisons.

Pairwise Comparison	# of Genes Up-Regulated in Hybrid	# of Genes Down-Regulated in Hybrid	Total # of DEGs
Triad 1: Female vs Hybrid	0	3	3
Triad 1: Male vs Hybrid	1	6	7
Triad 2: Female vs Hybrid	0	3	0
Triad 2: Male vs Hybrid	0	0	0
Triad 3: Female vs Hybrid	5	6	11
Triad 3: Male vs Hybrid	88	200	288
Triad 4: Female vs Hybrid	0	0	0
Triad 4: Male vs Hybrid	241	83	324
Triad 5: Female vs Hybrid	20	47	18
Triad 5: Male vs Hybrid	1,254	710	1,964
Triad 6: Female vs Hybrid	1	15	16
Triad 6: Male vs Hybrid	1	15	16
Triad 7: Female vs Hybrid	0	1	1
Triad 7: Male vs Hybrid	2	14	16
Triad 8: Female vs Hybrid	58	165	223
Triad 8: Male vs Hybrid	218	62	280
Triad 9: Female vs Hybrid	0	1	1
Triad 9: Male vs Hybrid	77	27	104
Triad 10: Female vs Hybrid	41	14	55
Triad 10: Male vs Hybrid	1,643	775	2,418



**Figure 1.1: Kernel Age Comparison.**

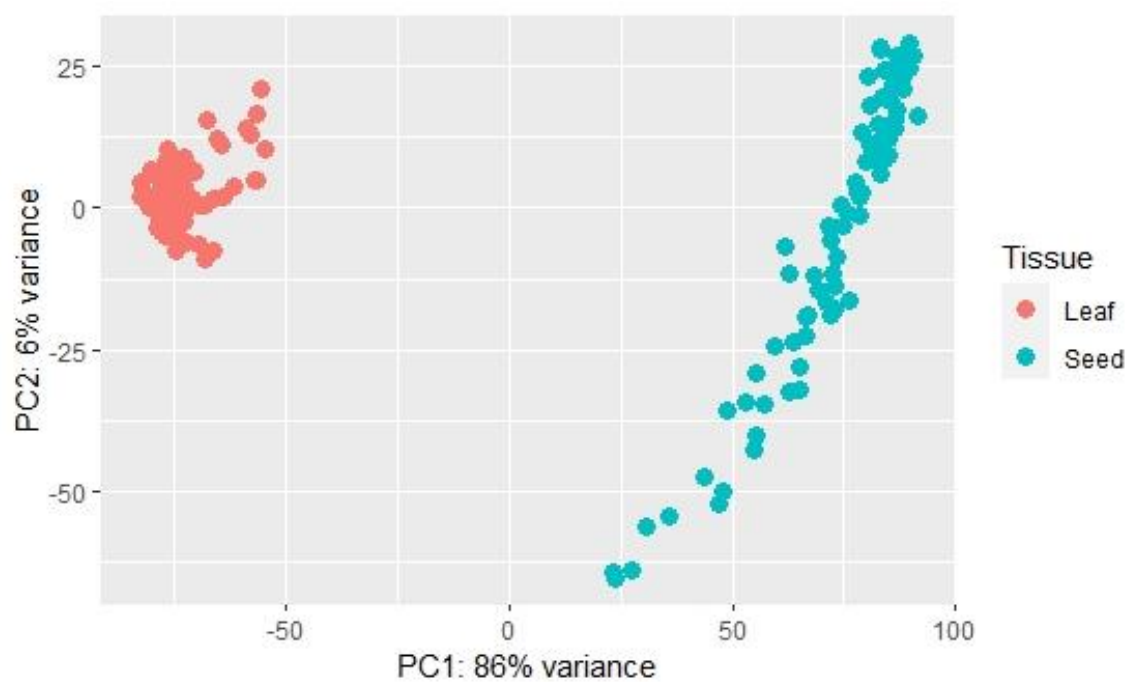
Differences in kernel size and development at the six time periods they were sampled from the plots; (a) 7 DAF; (b) 8 DAF; (c) 9 DAF; (d) 10 DAF; (e) 11 DAF; (f) 12 DAF



**Figure 1.2: Buffer RPE Leaf Tissue Washing Stages.**

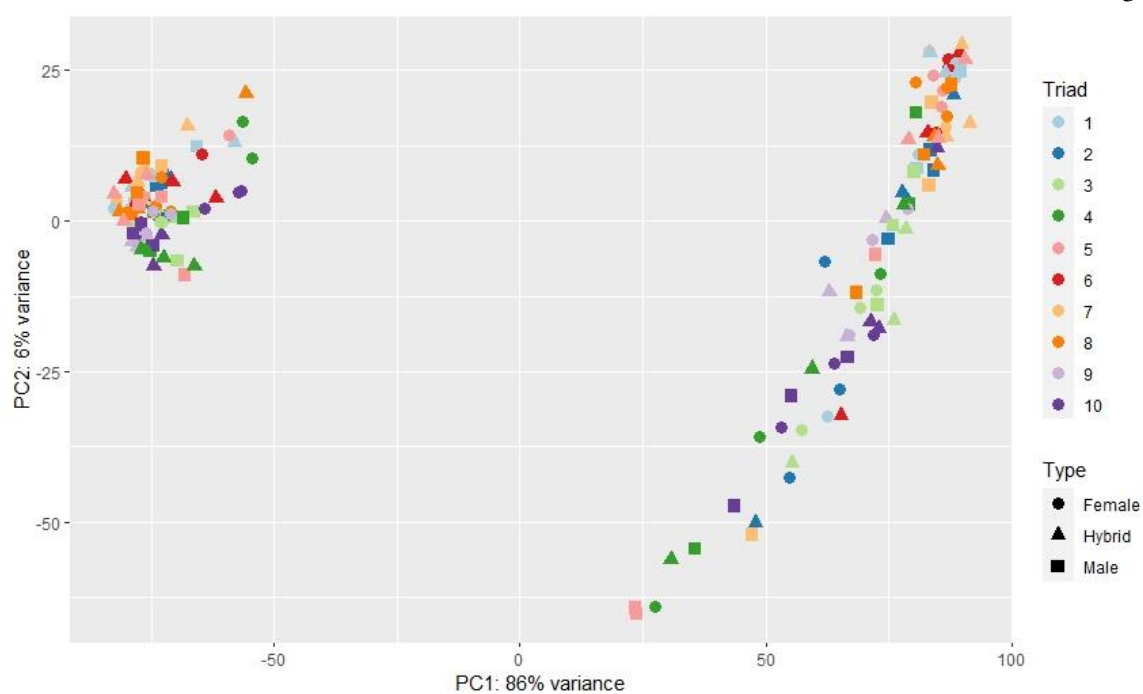
Visual representation of the washing step of leaf extractions where the chlorophyll was removed from the sample and how four washes were needed to remove the chlorophyll;

(a) Wash 1; (b) Wash 2; (c) Wash 3; (d) Wash 4



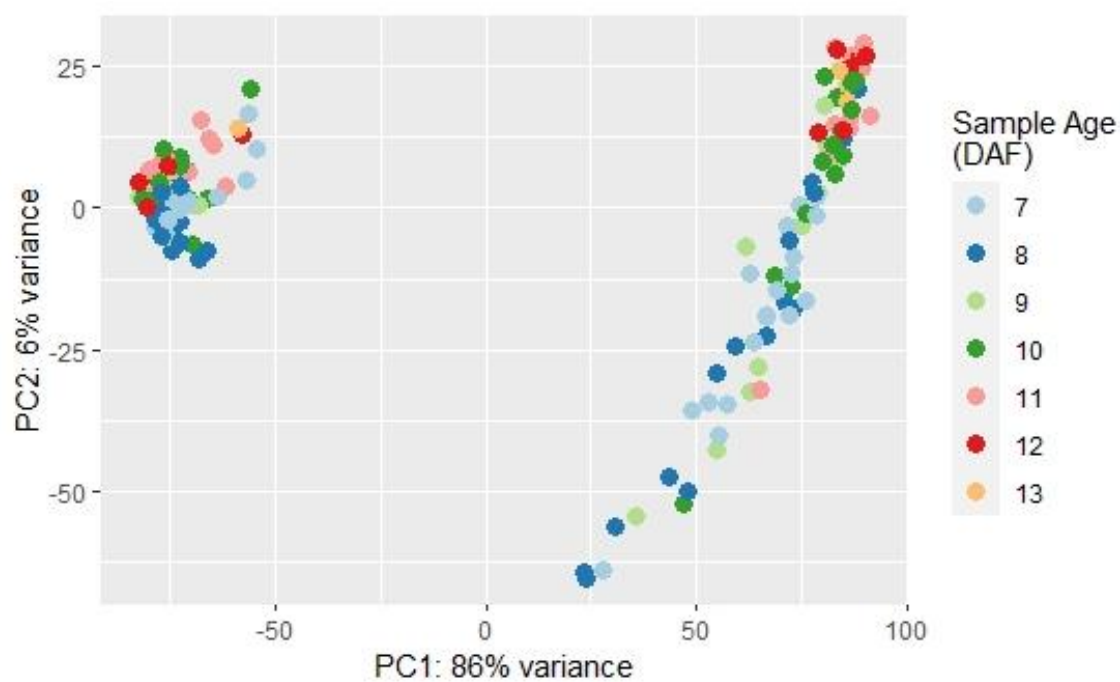
**Figure 1.3: PCA Analysis of Sources of Variation in Dataset**

Relationship between the two tissues sampled as they make up the majority (86%) of the variation in the dataset.



**Figure 1.4: PCA Analysis of Triad x Type Interaction as Source of Variation**

Relationship between the interaction of triad and type of triad on the secondary source of variation in the dataset. It does not explain the presence of variation in the seed samples on the right side of the figure.



778

779

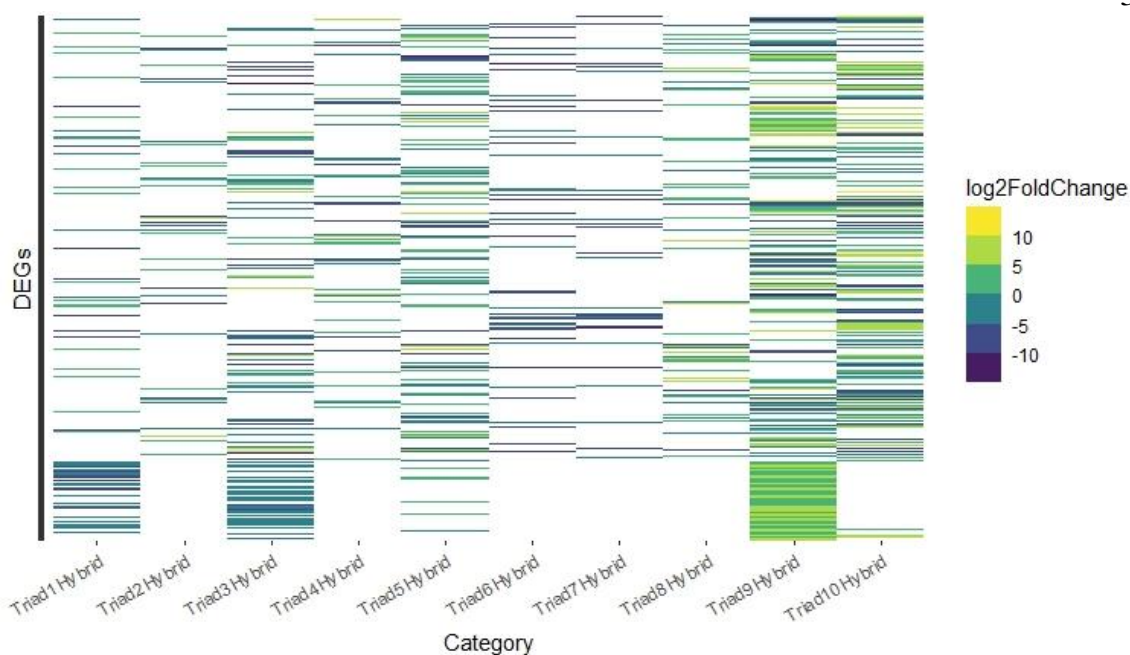
780

781

782

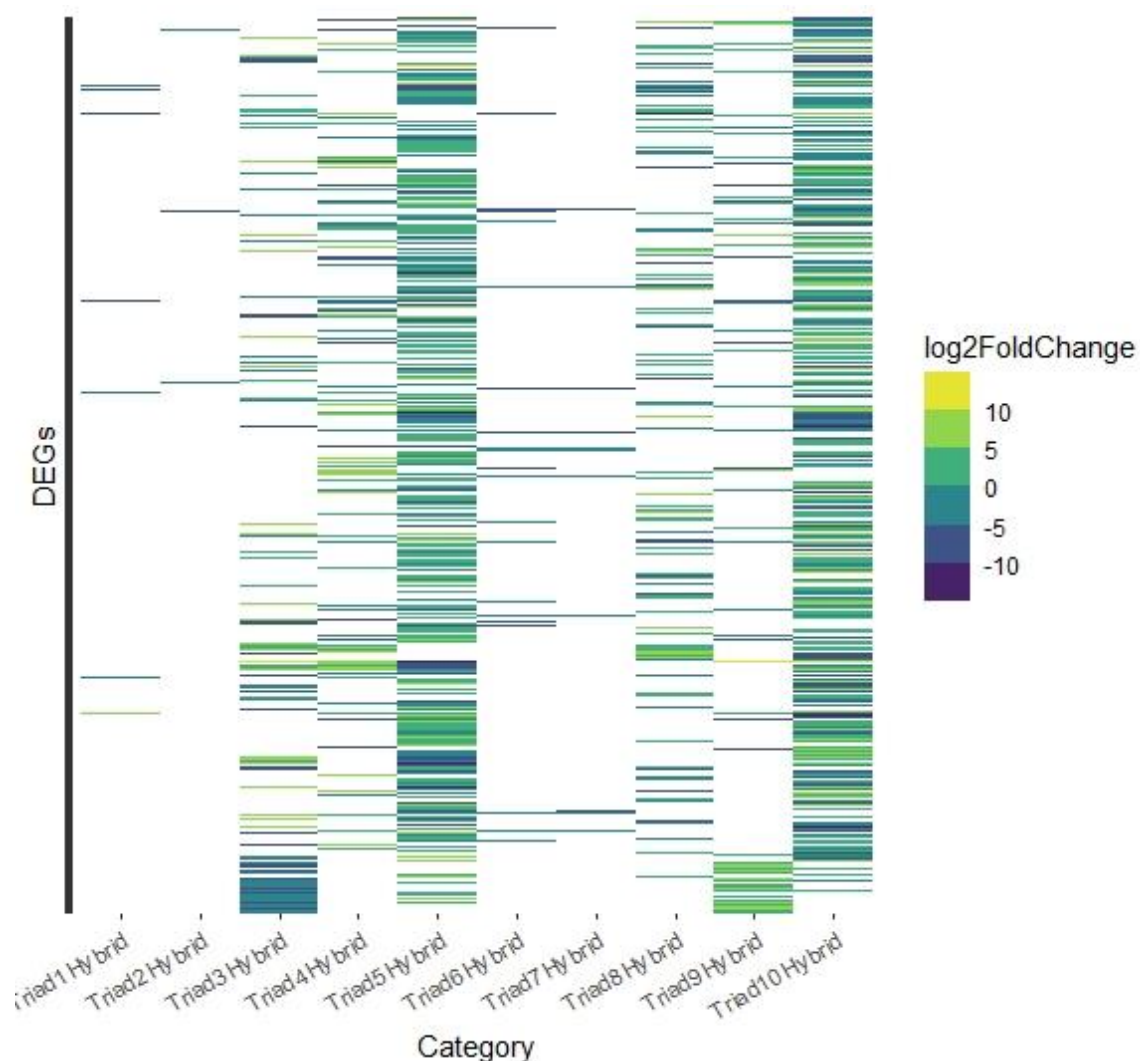
**Figure 1.5: PCA Analysis of Sample Age as Source of Variation**

Relationship of the sample's age when it was collected from the plot to the secondary source of variation in the dataset. Kernel age is the secondary component of variation in the dataset and proves the importance of sampling at the same tissue age.



**Figure 1.6: Shared DEGs Across Hybrids-Leaf**

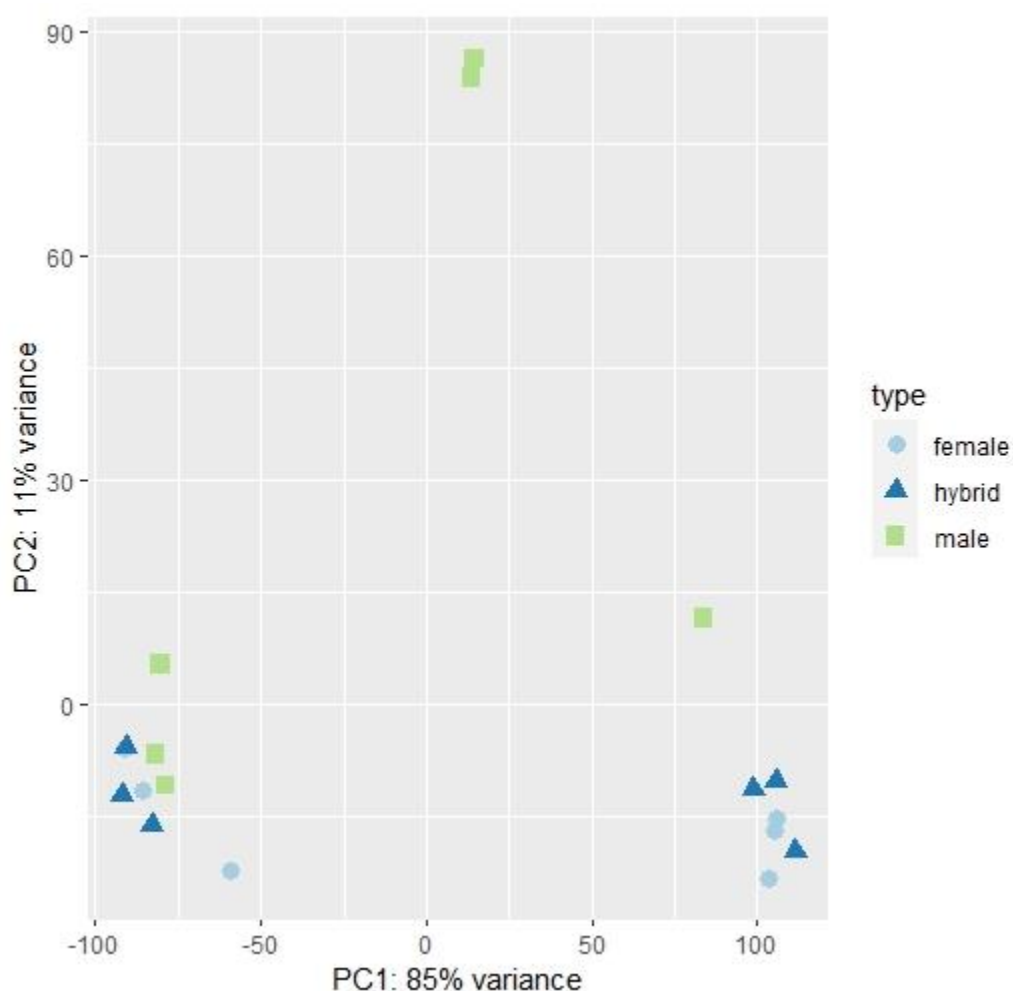
Comparison of the identified DEGs found in the leaf tissue hybrid/parent comparisons for both heterotic and non-heterotic triads, Triads 1-5 and 6-10 respectively. A positive log fold change indicates the gene is more expressed in that hybrid compared to the parents. A negative log fold change indicates the gene has higher expression in the parent than in the hybrid. Genes shared within triads in a heterotic group and among groups can also be identified.



**Figure 1.7: Shared DEGs Across Hybrids-Seed**

Comparison of the identified DEGs found in the seed tissue hybrid/parent comparisons for both heterotic and non-heterotic triads, Triads 1-5 and 6-10 respectively. A positive log fold change indicates the gene is more expressed in that hybrid compared to the parents. A negative log fold change indicates the gene has higher expression in the parent than in the hybrid. Genes shared within triads in a heterotic group and among groups can also be identified.





## 806    **CHAPTER 2: OPTIMIZING SAMPLE PRESERVATION METHODS FOR** 807    **SEQUENCING USING *RNAlater*<sup>®</sup> IN WHEAT**

### 808    **1. ABSTRACT**

809            When conducting RNA sequencing studies, researchers must ensure the RNA is  
810    intact and pure to create cDNA libraries that produce high quality sequences for  
811    transcriptome analysis. Preservation is the important step to creating high quality  
812    transcripts and the preservation standard has been using liquid nitrogen to flash freeze  
813    tissue and then keep the samples at -80°C. However, liquid nitrogen is dangerous and  
814    difficult to handle. *RNAlater*<sup>®</sup> has been created to replace liquid nitrogen as a safe to  
815    handle solution to preserve RNA, but the methods described in the protocol have  
816    produced inconsistent results in terms of sample purity and integrity, especially for long-  
817    term storage. To determine if there is a better storage temperature to keep the samples in  
818    and how much degradation occurs as the samples are stored for longer amounts of time, a  
819    study was conducted to sample leaf and immature kernel tissues from wheat plants and  
820    then place them in storage for one to six months at -20°C and -80°C. RNA was extracted  
821    from each treatment combination and was tested for the quality, quantity, and purity of  
822    the RNA. The response variables were analyzed using an analysis of variance. The results  
823    showed an improved response when samples were stored in -80°C but the RNA  
824    decreased from month one through month six. However, there was no significant  
825    decrease in sample quality and quantity as they were kept in storage. The results suggest  
826    that storing high quality samples for up to six months at -80°C was possible.

## 827 2. INTRODUCTION

828 Gene expression analysis is an important tool for characterizing the genes that control  
829 phenotypes in crop species. RNA sequencing is used to characterize gene expression  
830 through the analysis of RNA transcripts. However, it is still a relatively new high-  
831 throughput sequencing method with improvements being made to reduce the cost and  
832 read-depth requirements. These methods have allowed for faster sequencing and  
833 sequencing to be completed on more samples and transcripts simultaneously when  
834 compared to Sanger sequencing or quantitative real-time PCR (Tandonnet & Torres,  
835 2017; Torres et al., 2008). As more research projects use RNA sequencing, sample  
836 collection methods must be considered and evaluated to identify how to preserve the  
837 RNA in tissue samples during the period between collection and RNA extraction (García-  
838 Baldenegro et al., 2015).

839 High-quality sample preservation is key for ensuring RNA integrity for sequencing  
840 and transcript analyses. It is necessary to have high quality total RNA, quality score  
841 greater than four (RQN value), to make good cDNA libraries for sequencing (Sangha et  
842 al., 2010). For a sample to be considered for sequencing, it must meet a quality threshold  
843 RQN value, also known as the RIN value. RQN values are on a 1-10 scale where 10 is  
844 the intact, high quality RNA sample and a 1 is totally degraded. Each project must  
845 determine a cutoff RQN value aiming for as high as a score as possible (Schroeder et al.,  
846 2006). RNA purity is measured with  $A_{260/280}$  and  $A_{260/230}$  ratios, which measure the  
847 amount of protein contamination, and polyphenol and polysaccharide contamination  
848 respectively (de Wever et al., 2020).  $A_{260/280}$  is a ratio of the absorbance molecules in  
849 solutions at 260 nm and 280 nm on a spectrophotometer. A pure RNA sample has an

850  $A_{260/280}$  ratio ~2.0 and an  $A_{260/230}$  ratio between 2.0 and 2.2 (Thermo Fisher, 2017). In  
851 addition to sample purity, sample integrity must be evaluated. The most common method  
852 of estimating RNA integrity is to measure 28S/18S ratio which is derived from the 28S  
853 and 18S molecules of ribosomal RNA and compare that ratio to other bands on an  
854 agarose gel. A ratio close to 2.0 is desired (Schroeder et al., 2006). Total RNA  
855 concentration present in the sample will also help the researcher determine if the sample  
856 is usable. For example, the protocol for 3' RNA-Seq library preparation using Lexogen®  
857 QuantSeq™ 3' mRNA-Seq Library Prep Kit (Lexogen®, Vienna, Austria) and sequencing  
858 completed on Illumina® NextSeq550 using 75SR High Output Flow Cell kits (Illumina®,  
859 San Diego, California) requires a total RNA concentration of 50 ng/μL. Therefore, any  
860 samples with concentrations less than 50 ng/μL would not produce acceptable sequencing  
861 results.

862 Previous RNA studies in wheat have used leaf tissue as the source of RNA and used  
863 liquid nitrogen to flash freeze and preserve the sampled tissue (Amirbakhtiar et al., 2021;  
864 Chu et al., 2021; Poretti et al., 2021). Tissue preservation in liquid nitrogen after the  
865 sample is taken is considered the “gold standard” in RNA preservation (Burden, 2008;  
866 García-Baldenegro et al., 2015). However, liquid nitrogen is dangerous to handle and  
867 cumbersome to work with, especially in a field versus lab setting. To work around the  
868 complications of sampling, a tissue preservation solution, RNeasy® (Thermo Fisher  
869 Scientific, Waltham, Massachusetts), had been introduced that can inactivate RNases and  
870 stabilize the RNA for storage before RNA extractions. This preservation method has been  
871 used in different plants like annatto (*Bixa orellana* L.), cacao trees (*Theobroma cacao*  
872 L.), Arabidopsis (*Arabidopsis thaliana* L.), and wheat (*Triticum aestivum* L.), but each

873 crop required some modifications for optimal storage conditions (de Wever et al., 2020;  
874 Paul et al., 2005; Rodrigues et al., 2007).

875 Due to the logistical challenges of using liquid nitrogen in the field, the UNL small  
876 grains breeding program transitioned to using *RNAlater*<sup>®</sup> to preserve tissue collected  
877 from field trials for RNA extraction. A short test was conducted to determine the quality  
878 of *RNAlater*<sup>®</sup> preserved samples. Leaf and seed tissues were sampled from greenhouse-  
879 grown plants and treated with *RNAlater*<sup>®</sup> based on the manufacturer's protocol. Samples  
880 were stored at -20°C, and extractions were completed within a month of sampling.  
881 Quality testing showed that the RNA had been preserved adequately for down-stream  
882 sequencing with an average RQN of 6.0, 28S/18S ratio of 1.1, concentration of 522  
883 ng/μL, A<sub>260/280</sub> value of 2.13, and A<sub>260/230</sub> value of 1.78. Therefore, *RNAlater*<sup>®</sup> was used  
884 for a large field-based gene expression study in 2020 and samples were stored at -20°C.  
885 However, RNA extraction was delayed until six months after sampling due to COVID-19  
886 restrictions. When the samples were subjected to quality control analysis, the RQN values  
887 averaged 3.0-4.0 for both leaf and kernel samples, and 28S/18S ratios were also close to  
888 zero for most samples. These values were not acceptable for use in sequencing, and the  
889 study was delayed for an additional year to collect new samples.

890 This experience highlighted the need to evaluate storage temperature and storage  
891 longevity for up to six months for *RNAlater*<sup>®</sup> treated samples. Our objectives for this  
892 study were 1) to evaluate the longevity of samples preserved in *RNAlater*<sup>®</sup> up to six  
893 months and if or when major degradation occurs; and 2) to determine what storage  
894 temperature will preserve samples better. Our results are important for designing studies

895 where RNA cannot be extracted from tissue immediately due to sampling location or the  
896 need to collect additional data before selecting samples for extraction.

### 897 **3. MATERIALS AND METHODS**

#### 898 ***3.1 Plant Material and Treatment Structure***

899 Sample tissues were collected from the UNL developed wheat cultivar, “Ruth”,  
900 which is a commonly used check in the breeding program and a widely used commercial  
901 cultivar in the state (Baenziger et al., 2020). Samples were collected from six check plots  
902 in a 2021 wheat yield trial located at the Havelock Research Farm in Lincoln, NE. Four  
903 individual stems were sampled from each plot, taking both the kernels and flag leaf  
904 biological replicates from the same stem, and were immediately submerged in a 2.0 mL  
905 microcentrifuge tube filled with in *RNAlater*<sup>®</sup>. Sampling was done between 10:00 am  
906 and 12:00 pm on June 6, 2021, and tissue age was 15 days after flowering. Once the  
907 samples were collected, they were transported to the wet lab on dry ice. Following the  
908 manufacturer’s protocol, the tubes with tissue stored inside were moved to 4°C overnight.  
909 The next day, the *RNAlater*<sup>®</sup> supernatant was removed from the samples and the tissues  
910 were placed in their randomly assigned storage temperature, 24 samples in -20°C and 24  
911 samples in -80°C. Extractions were completed on four samples per tissue per storage  
912 temperature once a month for six months. The six storage lengths ranged from ideal  
913 storage length of one month to six months. This storage protocol resulted in 12 storage  
914 length by storage temperature combinations. The two storage temperatures were coded as  
915 “A\_” for -20°C and “B\_” for -80°C with the corresponding number referring to the  
916 storage length in months (Table 2.1).

### 917 **3.2 RNA Extraction**

918       Every month, RNA was extracted from the 4 biological replicates of each tissue  
919 for both storage temperatures, -20°C and -80°C on the same day. Extractions were  
920 completed using the Qiagen RNeasy® Plant Mini Kit (Qiagen®, Hilden, Germany)  
921 protocol with a few adjustments for tissue optimization. The kit provided buffer RLC was  
922 used for the seed samples as through previous testing, it was found to work better on the  
923 starchy immature kernels than the other kit buffer RLT did. RLT worked well on the leaf  
924 samples and was used for that process of extracting RNA from the ground tissue. With  
925 the additional starches found in older kernels, the initial solution of ground kernel tissue  
926 and RLC buffer would create a congealed solution in the microcentrifuge tube. To fix this  
927 problem, additional RLC buffer, 200-400 µL was added to re-liquify the solution. The  
928 on-column DNase digestion step was performed using the Qiagen® RNase-free DNase  
929 Kit (Qiagen®, Hilden, Germany). Due to the large amount of chlorophyll found in leaf  
930 tissue, those leaf samples required one to two additional washing steps using the RPE  
931 buffer on the spin column (Figure 2.1). The last adjustment to the protocol was using 20  
932 µL RNase-free water and a second 1.5 mL collection tube to complete an additional  
933 washing to remove any additional RNA found in the column and to create a backup  
934 sample. Once the extractions were completed, samples were stored at -80°C until quality  
935 testing could take place.

### 936 **3.3 RNA Quantity, Quality, and Integrity Testing**

937       After RNA samples were extracted, they were submitted to the University of  
938 Nebraska Medical Center Genomics Core in Omaha, NE for quality control analysis. The  
939 Core evaluated samples for RNA quality, quantity, and integrity using an Advanced

940 Analytical Technologies, Inc. Fragment Analyzer™ (recently acquired by Agilent  
 941 Technologies, Inc., Santa Clara, California) as well as a ThermoScientific NanoDrop™  
 942 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts). The  
 943 standard for sequencing is to use the results produced by the fragment analyzer.  
 944 However, not every research lab can access a fragment analyzer and a NanoDrop™ is  
 945 more affordable, even though its data are variable. The core returned the concentrations,  
 946 RQN scores, and 28S/18S ratios for each sample produced from the fragment analyzer.  
 947 They also returned concentrations, A<sub>260/280</sub> ratios, and A<sub>260/230</sub> ratios from the NanoDrop.

### 948 ***3.4 Statistical Analysis***

949 Standard deviations for RQN values were calculated for each tissue and any  
 950 sample whose result was two standard deviations positive or negative was considered as  
 951 an outlier and removed from the dataset. Furthermore, samples whose concentrations  
 952 were less than 50 ng/μL were also removed. These quality control measures were done to  
 953 match how RNA samples would be handled when submitting them for sequencing. The  
 954 equipment protocol requires a concentration of 50 ng/μL to ensure quality sequencing is  
 955 completed, therefore any samples below that concentration need to be discarded while  
 956 samples above that concentration were diluted to 50 ng/μL.

957 The experiment was designed in a complete randomized design using a 2x6  
 958 factorial treatment design. The statistical model of the response variable  $Y_{ijk}$ , referring to  
 959 either concentration, quality, or integrity, of the  $i^{\text{th}}$  level of storage temperature,  $j^{\text{th}}$  level  
 960 of storage time treatment, and  $k^{\text{th}}$  experimental unit, was analyzed as:

$$961 \quad Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk},$$



Where  $\mu$  is the overall mean,  $\alpha_i$  is the main effect of the  $i^{th}$  level of storage temperature treatment,  $\beta_j$  is the main effect for the  $j^{th}$  level of storage time treatment,  $\alpha\beta_{ij}$  is the interaction term between the  $i^{th}$  temperature and  $j^{th}$  time, and  $\epsilon_{ijk}$  is the error term. Tissues were analyzed separately as the leaf and seed sample's concentrations are drastically different.

A mixed linear model analysis of variance using the PROC MIXED procedure was used with the Type 3 fixed effects tests applied to the procedure in SAS 9.4 (SAS Institute, Cary, North Carolina) to evaluate the effects of the temperature and storage length treatments on concentration, quality, purity, and integrity. Least square means, LS-Means, were used to detail how the significant results in the ANOVAs were reflected biologically. An alpha level of  $\alpha < 0.1$  was used to more liberally identify differences and those differences identified can be used by researchers to improve their sample preservation. The estimates and confidence intervals created by the LS-Means test were exported to Excel files and used to create line graphs using the R program *ggplot2* v3.3.5 (Wickham, 2009).

#### 4. RESULTS

After removing outliers for leaf samples, 46 leaf tissue samples remained for the RQN, concentrations, and 28S/18S variables produced by the fragment analyzer and 47 leaf tissue samples remained for the NanoDrop<sup>TM</sup> concentrations,  $A_{260/280}$ , and  $A_{260/230}$  variables. The seed samples had more outliers removed and resulted in 30 observations used in the fragment analyzer RQN, concentrations, and 28S/18S variables and 34 in the NanoDrop<sup>TM</sup> concentrations,  $A_{260/280}$ , and  $A_{260/230}$  variables analysis. However, four of

984 these removed samples were from month three samples stored at -20°C and analyzed by  
985 the NanoDrop™. No comparisons could be made for that month.

986 The analysis of variance on the leaf samples (Table 2.2) found the storage  
987 temperature was significant for the leaf samples' RQN, 28S/18S ratio, concentration  
988 produced from the NanoDrop™, and the A<sub>260/280</sub> ratio. Storage time was significant for the  
989 fragment analyzer concentration, NanoDrop™ concentration, and A<sub>260/280</sub> ratio. The  
990 interaction of the two factors, temperature and storage length, was significant for both  
991 concentrations and purity ratios.

992 Further investigation using the LS-Means comparisons illustrate how each response  
993 variable is affected by the treatments. RQN values for leaf samples were not significant  
994 for each month except for month two compared to three through six. The -80°C storage  
995 temperature also produces a higher RQN score for every month except month five  
996 (Figure 2.2). An unexpected spike response was also seen in the concentrations produced  
997 from the fragment analyzer for month five for kernel tissues, but overall follows the same  
998 temperature pattern of leaf tissues of higher concentrations in the -80°C treatment.  
999 Concentrations are significantly lower for the values from months one and two to the  
1000 other four months (Figure 2.3). The 28S/18S ratio response shows higher ratios in the -  
1001 80°C samples but no differences in the storage length treatment (Figure 2.4). The  
1002 concentrations produced from the NanoDrop™ show a major decrease in concentration as  
1003 the samples were stored for longer lengths of time. The figure also shows that storing the  
1004 samples for one month at -20°C produced higher concentrations than storing them for six  
1005 months at -80°C (Figure 2.5). Figure 2.6 shows a decrease in A<sub>260/280</sub> values as the  
1006 samples are kept in -80°C. There was no difference in response to storage temperature

1007 except for a large decrease in the  $A_{260/280}$  ratio value at month five for the  $-80^{\circ}\text{C}$  samples.  
1008 The  $A_{260/230}$  ratio was improved in the first month when stored at  $-20^{\circ}\text{C}$  compared to -  
1009  $80^{\circ}\text{C}$  (Figure 2.7).

1010 The analysis of variance for the seed samples showed significant response for the  
1011 28S/18S ratio to storage temperature, storage time, and their interaction. The NanoDrop™  
1012 concentration and the  $A_{260/230}$  ratio was significant for the storage time and temperature  
1013 treatments. Only storage time was significant for the  $A_{260/280}$  ratio (Table 2.3).

1014 The LS-Means comparison results for the RQN values of the seed samples found no  
1015 difference among storage length treatments, but a slight increase in the score when the  
1016 samples were stored in  $-80^{\circ}\text{C}$  (Figure 2.8) compared to  $-20^{\circ}\text{C}$ . In Figure 2.9, the  
1017 concentrations from the fragment analyzer were slightly higher in samples stored at -  
1018  $80^{\circ}\text{C}$  and concentrations improved the longer they were stored. 28S/18S ratios were  
1019 higher in the samples stored in  $-80^{\circ}\text{C}$  as well (Figure 2.10). The figure also shows that the  
1020 ratio increased during the six months as the samples were stored in  $-20^{\circ}\text{C}$  but overall  
1021 remained constant in  $-80^{\circ}\text{C}$  (Figure 2.10). The response of the NanoDrop™  
1022 concentrations showed that for almost every month, they were higher in the  $-80^{\circ}\text{C}$   
1023 samples than for the  $-20^{\circ}\text{C}$  (Figure 2.11). The  $A_{260/280}$  ratio was affected by the length of  
1024 storage and the ratio decreased as the length of storage increased. While the ratio is not  
1025 significant, there was an observed increase in the ratios when stored in  $-80^{\circ}\text{C}$  (Figure  
1026 2.12). Finally, in Figure 2.13, samples had higher  $A_{260/230}$  higher ratios when stored in -  
1027  $80^{\circ}\text{C}$  and the ratios trended down in the later months, but it was not significant at the  
1028 alpha level.

## 1029 5. DISCUSSION

1030 To extract high quality RNA from plant tissue, the samples need to be preserved  
1031 at the appropriate temperature, and a researcher needs to know how long the tissue can be  
1032 kept before the samples degrade below usable limits. Before conducting the analysis on  
1033 this study, many outliers had to be removed in both the fragment analyzer and  
1034 NanoDrop™ results, indicating variation present in both datasets. The nature of these  
1035 samples means that a researcher will have a lot of variation and more reps could have  
1036 been helpful to identify significant differences.

1037 Based on the results of this study, RQN values remained constant as they were  
1038 kept in storage for both leaf and seed tissues. The values ranged between 5 and 6  
1039 meaning they could be used in sequencing. This result contradicted what was seen in our  
1040 previous experience with *RNAlater*® in 2020 and we cannot explain this difference.  
1041 However, we can conclude that the less time kept in storage, the RQN values were better.  
1042 Sample concentration ensures the researcher has enough RNA present in the sample to  
1043 produce adequate sequencing, but our research showed opposite trends for the two  
1044 methods of measuring concentrations. While the concentrations reported from the  
1045 fragment analyzer showed no significant difference in response to storage time for leaf  
1046 tissue after month three, there was an unexplained increase in concentration between  
1047 months two and three. This observation is probably due to sample-to-sample variation.  
1048 The NanoDrop™ concentrations for leaf tissue show a significant decrease between  
1049 months one and two at -20°C with a non-significant decreasing trend after that for both  
1050 temperatures. These differences could be due to many samples being removed with  
1051 concentrations less than 50 ng/μL, indicating the variability in the NanoDrop™ dataset.

1052 Sample integrity remained constant during the six months of storage for both leaf and  
1053 seed samples, indicating ribosomal RNA remained present in the samples and was not  
1054 degraded. The purity measurement  $A_{260/280}$  ratio saw the greatest response to storage  
1055 time. The ratio decreased as the number of months in storage increased, indicating that  
1056 purity suffers the longer it is stored and polyphenol contamination increases. This  
1057 response was seen in both leaf and seed tissues. The  $A_{260/230}$  ratio was not significantly  
1058 impacted by storage time, meaning polysaccharide contamination is not affected by  
1059 length of storage. For the temperature treatment, the  $-80^{\circ}\text{C}$  treatment produced higher  
1060 RQN, concentrations, and integrity and purity ratios on average. While it was not  
1061 significant across every response variable, there was an improvement observed when  
1062 samples were stored at the  $-80^{\circ}\text{C}$ , which would help researchers ensure they are getting  
1063 the highest possible values for the response variables.

## 1064 **6. CONCLUSIONS**

1065 Identifying the appropriate storage conditions for tissue samples is imperative for  
1066 RNA sequencing as RNA degrades once the tissue is taken from the host. While there  
1067 was considerable variation in the data collected for this study, important responses were  
1068 identified and can be applied to sample collection in the future. Leaf and kernel tissues  
1069 have significant differences when it comes to the value of the RQN, concentrations,  
1070 28S/18S ratio,  $A_{260/280}$  ratio, and  $A_{260/230}$  ratio variables; however, there are shared  
1071 similarities in their overall response to the treatments that a common storage treatment  
1072 could be applied to both leaf and kernel tissues and produce high quality RNA for  
1073 sequencing. Based on these results, we conclude that storing samples at  $-80^{\circ}\text{C}$  will  
1074 improve the preservation of samples compared stored in a  $-20^{\circ}\text{C}$  freezer. While we did

1075 not identify a point in time that samples are too degraded to be used, the overall quality is  
1076 nearly constant for up to six months and is above the minimum requirements for  
1077 sequencing. -80°C always produced higher response values, indicating it is the better  
1078 choice to ensure the highest quality is produced. It is our recommendation that  
1079 *RNAlater*<sup>®</sup> can be used as a viable replacement to liquid nitrogen. Researchers should  
1080 use -80°C as their storage temperature and then extract the samples as soon as reasonably  
1081 possible to ensure quality is not diminished.

1082 **7. REFERENCES**

- 1083 Amirbakhtiar, N., Ismaili, A., Ghaffari, M. R., Mansuri, R. M., Sanjari, S., & Shobbar, Z.  
 1084 S. (2021). Transcriptome analysis of bread wheat leaves in response to salt stress.  
 1085 PLoS ONE, 16(7 July), e0254189. <https://doi.org/10.1371/journal.pone.0254189>
- 1086 Baenziger, P. S., Graybosch, R. A., Rose, D. J., Xu, L., Guttieri, M. J., Regassa, T.,  
 1087 Klein, R. N., Kruger, G. R., Santra, D. K., Hergert, G. W., Wegulo, S. N., Jin, Y.,  
 1088 Kolmer, J., Hein, G. L., Bradshaw, J., Chen, M. S., Bai, G., Bowden, R. L., El-  
 1089 Basyoni, I., & Lorenz, A. (2020). Registration of ‘NE10589’ (Husker Genetics  
 1090 Brand Ruth) hard red winter wheat. *Journal of Plant Registrations*, 14(3), 388–  
 1091 397. <https://doi.org/10.1002/PLR2.20068>
- 1092 Burden, D. (2008). Guide to the homogenization of biological samples. *Random Primers*,  
 1093 7, 1–14. [http://opsdiagnostics.com/notes/ranpri/Homogenization Guide ver.1.pdf](http://opsdiagnostics.com/notes/ranpri/Homogenization%20Guide%20ver.1.pdf)
- 1094 Chu, C., Wang, S., Paetzold, L., Wang, Z., Hui, K., Rudd, J. C., Xue, Q., Ibrahim, A. M.  
 1095 H., Metz, R., Johnson, C. D., Rush, C. M., & Liu, S. (2021). RNA-seq analysis  
 1096 reveals different drought tolerance mechanisms in two broadly adapted wheat  
 1097 cultivars ‘TAM 111’ and ‘TAM 112.’ *Scientific Reports*, 11(1), 1–16.  
 1098 <https://doi.org/10.1038/s41598-021-83372-0>
- 1099 de Wever, J., Tulkens, D., Verwaeren, J., Everaert, H., Rottiers, H., Dewettinck, K.,  
 1100 Lefever, S., & Messens, K. (2020). A combined rna preservation and extraction  
 1101 protocol for gene expression studies in cacao beans. *Frontiers in Plant Science*,  
 1102 11, 992. <https://doi.org/10.3389/FPLS.2020.00992/BIBTEX>
- 1103 García-Baldenegro, C. V., Vargas-Arispuro, I., Islas-Osuna, M., Rivera-Domínguez, M.,  
 1104 Aispuro-Hernández, E., & Martínez-Tíllez, M. íngel. (2015). Total rna quality of  
 1105 lyophilized and cryopreserved dormant grapevine buds. *Electronic Journal of*  
 1106 *Biotechnology*, 18(2), 134–137. <https://doi.org/10.1016/j.ejbt.2015.01.002>
- 1107 Paul, A. L., Levine, H. G., McLamb, W., Norwood, K. L., Reed, D., Stutte, G. W., Wells,  
 1108 H. W., & Ferl, R. J. (2005). Plant molecular biology in the space station era:  
 1109 Utilization of KSC fixation tubes with RNAlater. *Acta Astronautica*, 56(6), 623–  
 1110 628. <https://doi.org/10.1016/J.ACTAASTRO.2004.10.001>
- 1111 Poretti, M., Sotiropoulos, A. G., Graf, J., Jung, E., Bourras, S., Krattinger, S. G., &  
 1112 Wicker, T. (2021). Comparative Transcriptome Analysis of Wheat Lines in the  
 1113 Field Reveals Multiple Essential Biochemical Pathways Suppressed by Obligate  
 1114 Pathogens. *Frontiers in Plant Science*, 12, 1981.  
 1115 <https://doi.org/10.3389/fpls.2021.720462>
- 1116 Rodrigues, S. M., Soares, V. L. F., de Oliveira, T. M., Gesteira, A. S., Otoni, W. C., &  
 1117 Costa, M. G. C. (2007). Isolation and purification of RNA from tissues rich in  
 1118 polyphenols, polysaccharides, and pigments of annatto (*Bixa orellana* L.).  
 1119 *Molecular Biotechnology*, 37(3), 220–224. [https://doi.org/10.1007/s12033-007-](https://doi.org/10.1007/s12033-007-0070-9)  
 1120 [0070-9](https://doi.org/10.1007/s12033-007-0070-9)
- 1121 Sangha, J. S., Gu, K., Kaur, J., & Yin, Z. (2010). An improved method for RNA isolation  
 1122 and cDNA library construction from immature seeds of *Jatropha curcas* L. *BMC*  
 1123 *Research Notes*, 3(1), 1–6. <https://doi.org/10.1186/1756-0500-3-126>
- 1124 Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M.,  
 1125 Lightfoot, S., Menzel, W., Granzow, M., & Ragg, T. (2006). The RIN: An RNA

- 1126 integrity number for assigning integrity values to RNA measurements. BMC  
1127 Molecular Biology, 7, 3. <https://doi.org/10.1186/1471-2199-7-3>
- 1128 Tandonnet, S., & Torres, T. T. (2017). Traditional versus 3' RNA-seq in a non-model  
1129 species. Genomics Data, 11, 9–16. <https://doi.org/10.1016/j.gdata.2016.11.002>
- 1130 Thermo Fisher. (2017). T042-TECHNICAL BULLETIN NanoDrop Spectrophotometers  
1131 260/280 and 260/230 Ratios. T042-Technical Bulletin, 1–2.
- 1132 Torres, T. T., Metta, M., Ottenwälder, B., & Schlötterer, C. (2008). Gene expression  
1133 profiling by massively parallel sequencing. Genome Research, 18(1), 172–177.  
1134 <https://doi.org/10.1101/gr.6984908>
- 1135 Wickham, H. (2009). ggplot2. *Ggplot2*. <https://doi.org/10.1007/978-0-387-98141-3>



**Table 2.1: Treatment Structure**

The table depicts the design structure of the treatments applied to the tissue samples. The structure was applied to both seed and leaf tissues. Samples were randomly assigned to the month they were extracted.

Storage Temperature	Storage Length					
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months
-20°C	A1	A1	A3	A4	A5	A6
-80°C	B1	B2	B3	B4	B5	B6

**Table 2.2: Leaf ANOVA**

Leaf sample ANOVA indicating the response of the variables to the treatments.

\*\* is significant at  $p < 0.1$ .

	Fragment Analyzer Results				NanoDrop Results			
		RQN	Concentration (ng/μL)	28S/18S Ratio		Concentration (ng/μL)	A <sub>260/280</sub> Ratio	A <sub>260/230</sub> Ratio
Source of Variation	Degrees of Freedom	Pr > F	Pr > F	Pr > F	Degrees of Freedom	Pr > F	Pr > F	Pr > F
Storage Temperature Treatment	1	0.0281**	0.1192	<0.0001**	1	0.0220**	<0.0001**	0.2733
Storage Time Treatment	5	0.1300	0.0019**	0.5929	5	0.0003**	<0.0001**	0.4014
Storage Temperature Treatment x Storage Time Treatment	5	0.5027	0.0005**	0.3483	5	<0.0001**	<0.0001**	0.0212*
Error	36	-	-	-	36	-	-	-

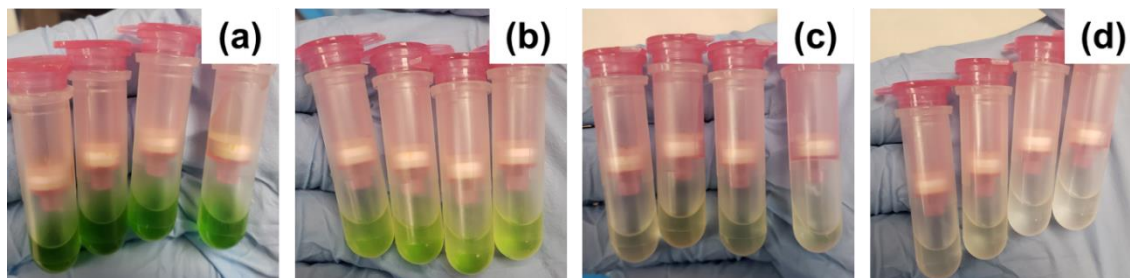
**Table 2.3: Seed ANOVA**

Seed sample ANOVA indicating the response of the variables to the treatments.

\*\* is significant at  $p < 0.1$ .

	Fragment Analyzer Results				NanoDrop Results			
		RQN	Concentration (ng/μL)	28S/18S Ratio		Concentration (ng/μL)	A <sub>260/280</sub> Ratio	A <sub>260/230</sub> Ratio
Source of Variation	Degrees of Freedom	Pr > F	Pr > F	Pr > F	Degrees of Freedom	Pr > F	Pr > F	Pr > F
Storage Temperature Treatment	1	0.3250	0.4061	0.0001**	1	0.0949**	0.1823	0.0675**
Storage Time Treatment	5	0.6769	0.6245	0.0181**	5	0.0193**	<0.0001**	0.0283**
Storage Temperature Treatment x Storage Time Treatment	5	0.1806	0.9705	0.0024**	5	0.5110	0.2892	0.5368
Error	36	-	-	-	36	-	-	-

1149



1150

1151

**Figure 2.1 Buffer RPE Leaf Tissue Washing.**

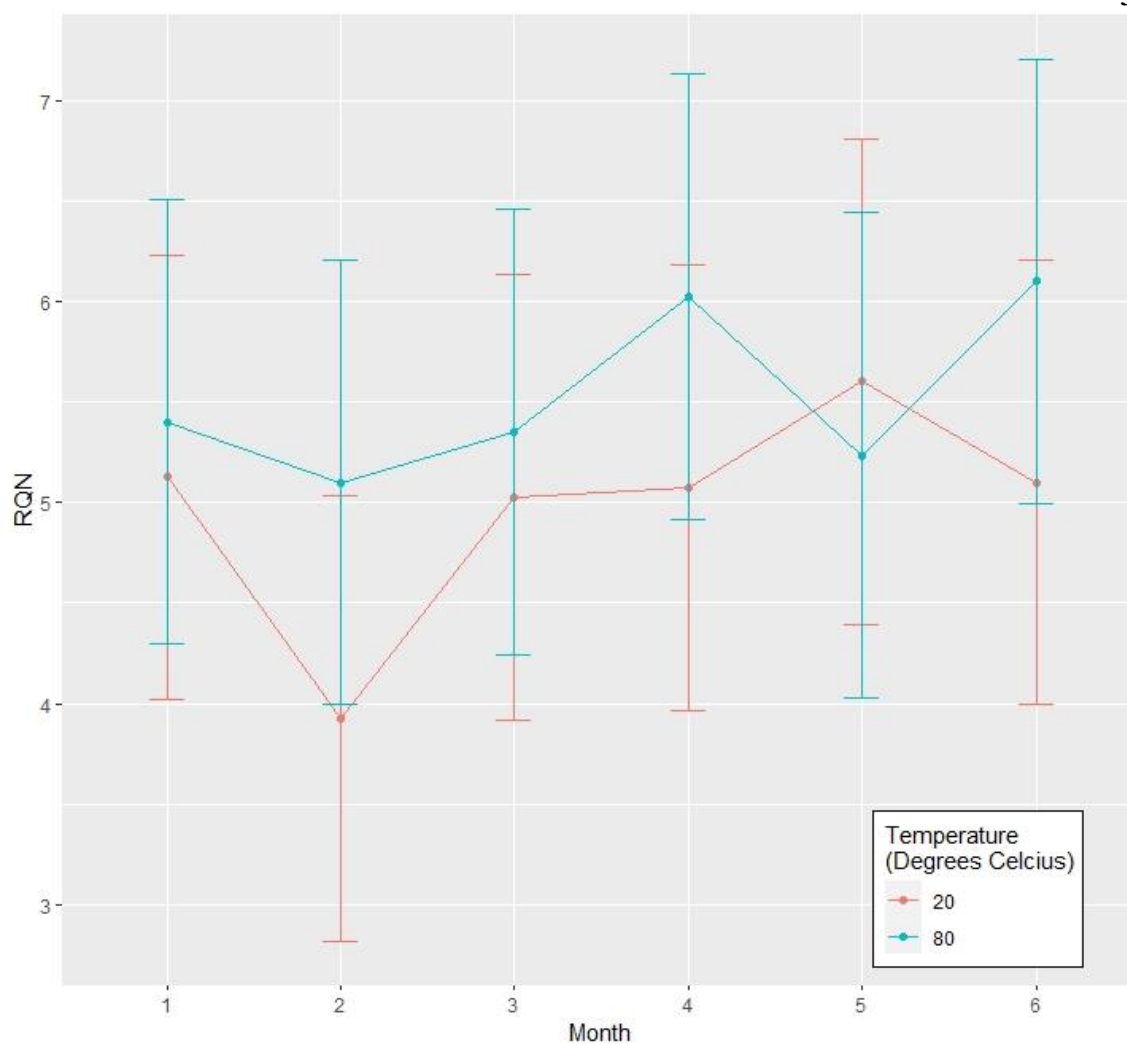
1152

Visual representation of the washing step of leaf extractions where the chlorophyll was removed from the sample and how four washes were needed to remove the chlorophyll;

1153

1154

(a) Wash 1; (b) Wash 2; (c) Wash 3; (d) Wash 4



1155

1156

**Figure 2.2: LS-Means for RQN Response to Treatments-Leaf**

1157

Comparison of the leaf tissues' RQN response to storage length (1-6 months) and storage

1158

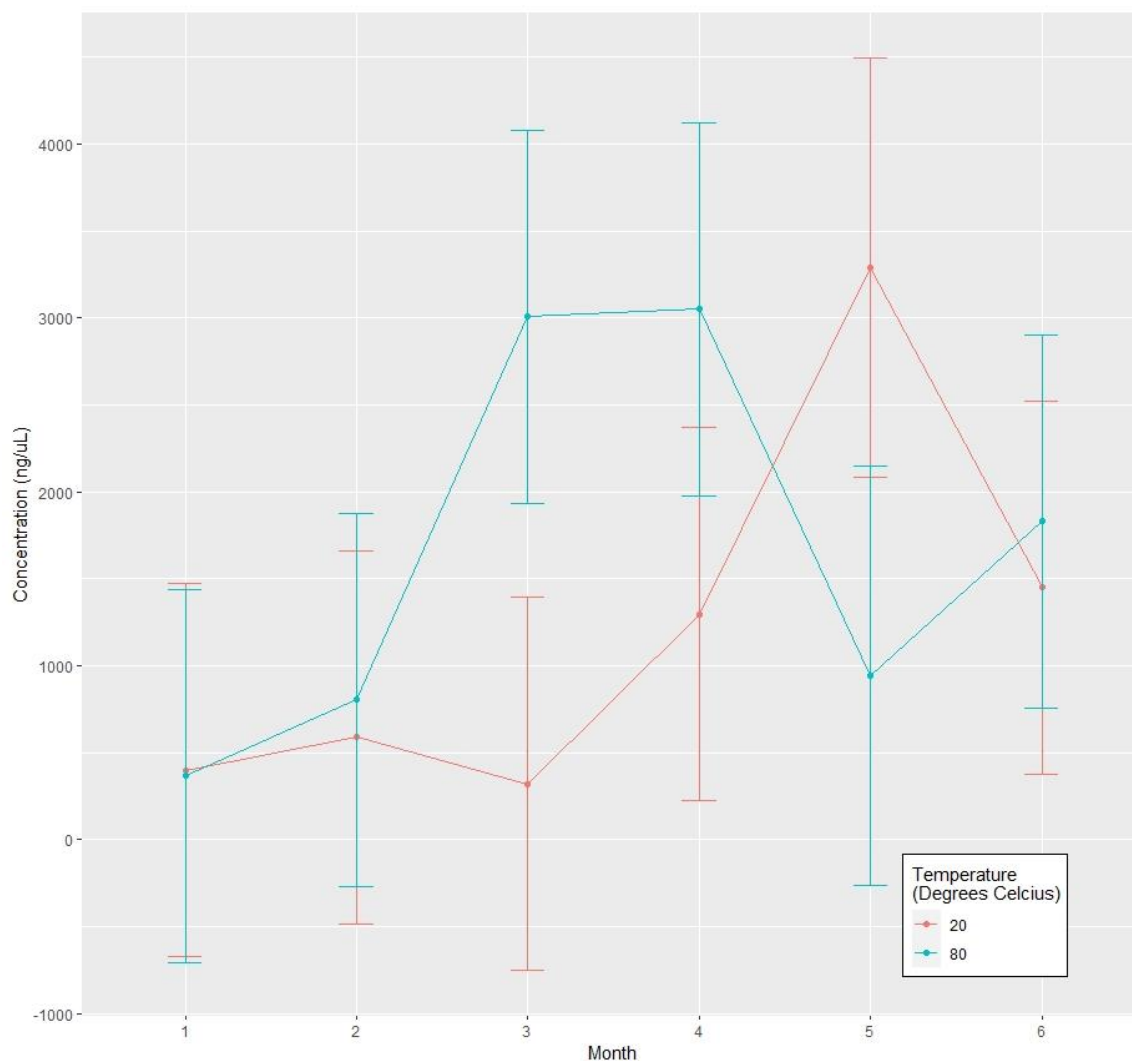
temperature (-20°C and -80°C) using 95% confidence intervals. No significant

1159

differences were identified, but samples stored in -80°C produced higher RQN values, &gt;

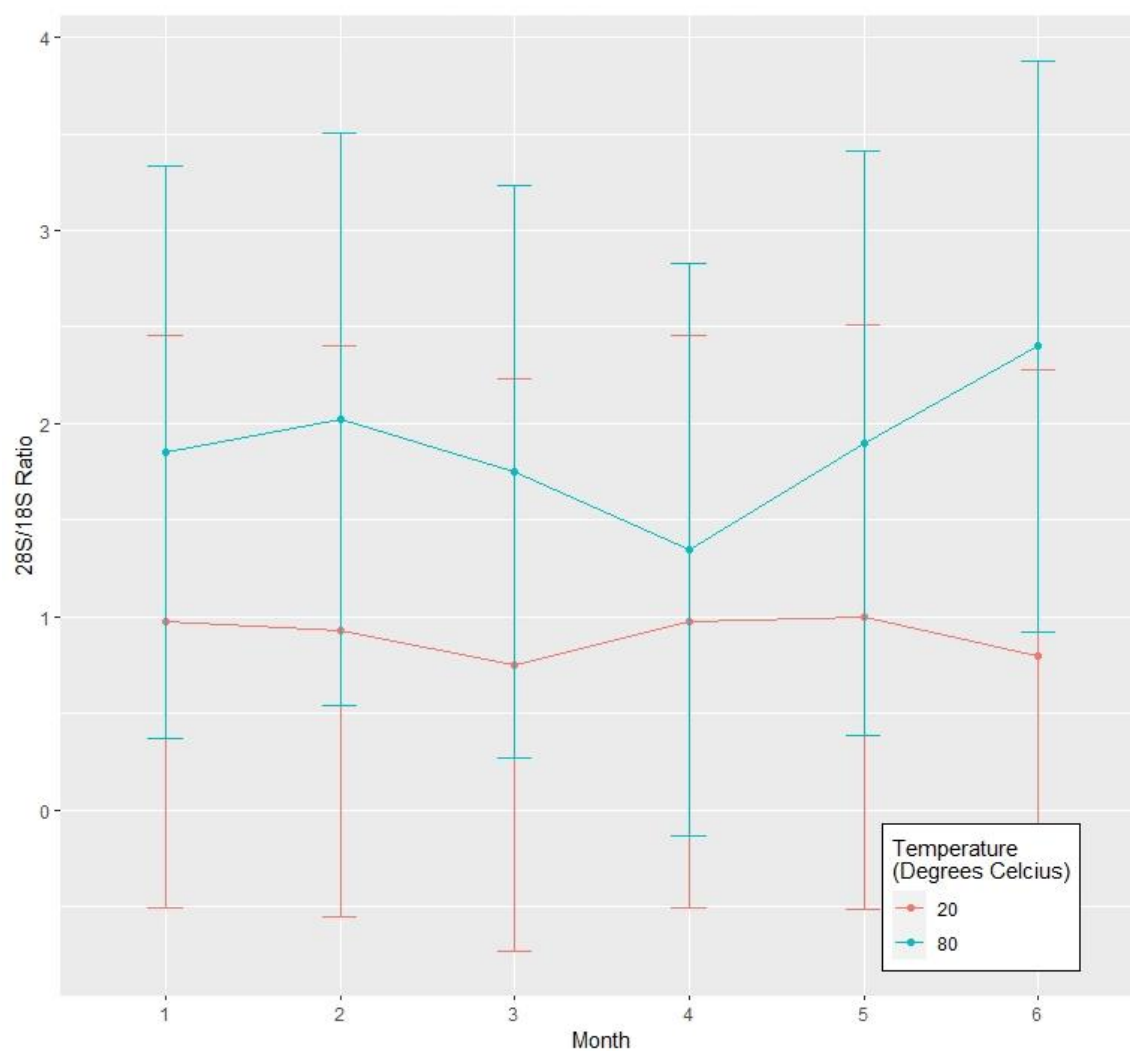
1160

5, for all six months.



**Figure 2.3: LS-Means for Fragment Analyzer Concentration Response to Treatments-Leaf**

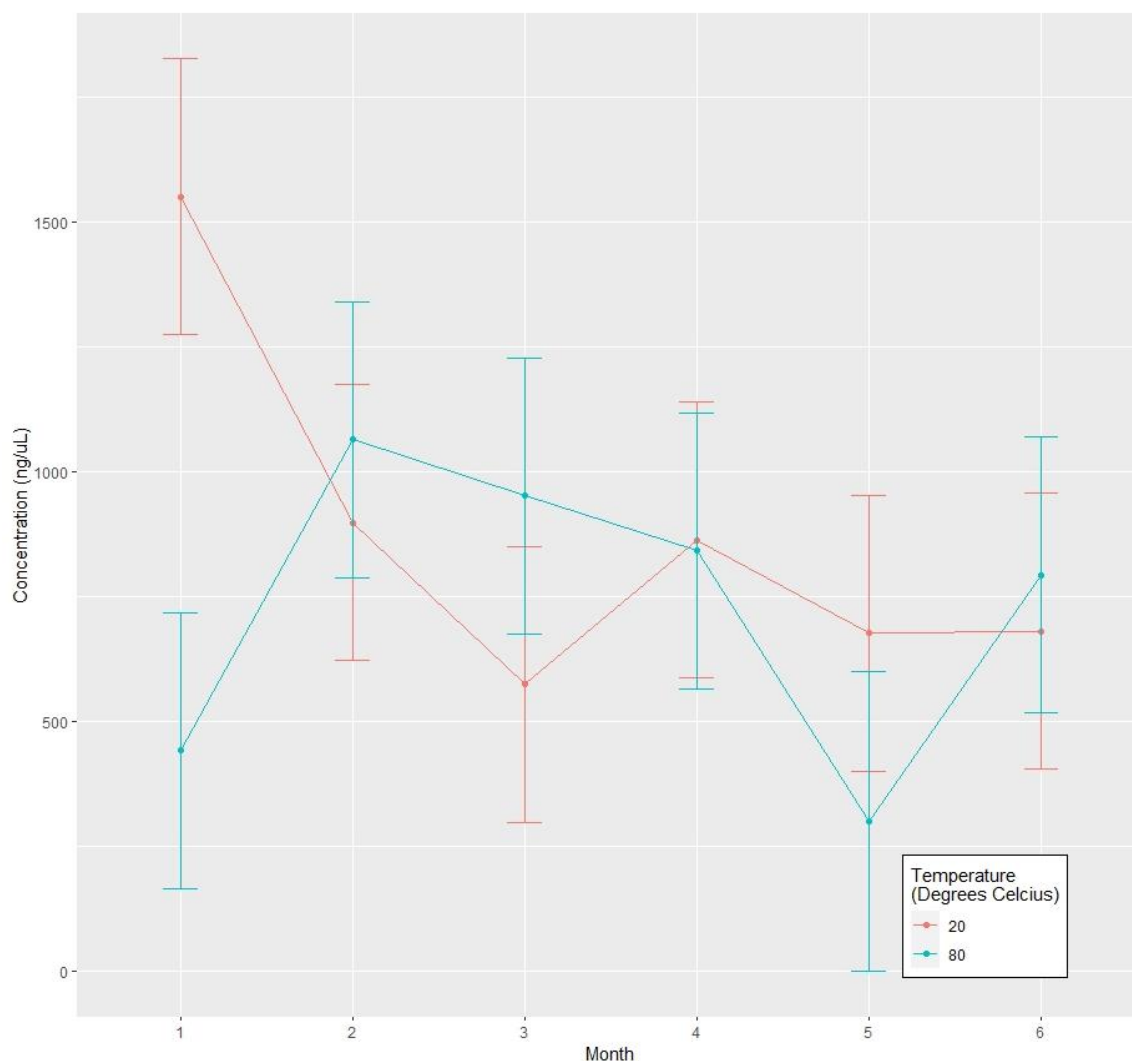
Comparison of the leaf tissues' concentration response (reported from the fragment analyzer) to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. A significantly higher concentration was produced from the samples stored in -80°C compared to those in -20°C. No other significant differences were identified, but samples stored in -80°C produced higher concentrations, > 500 ng/μL, on average for all six months.



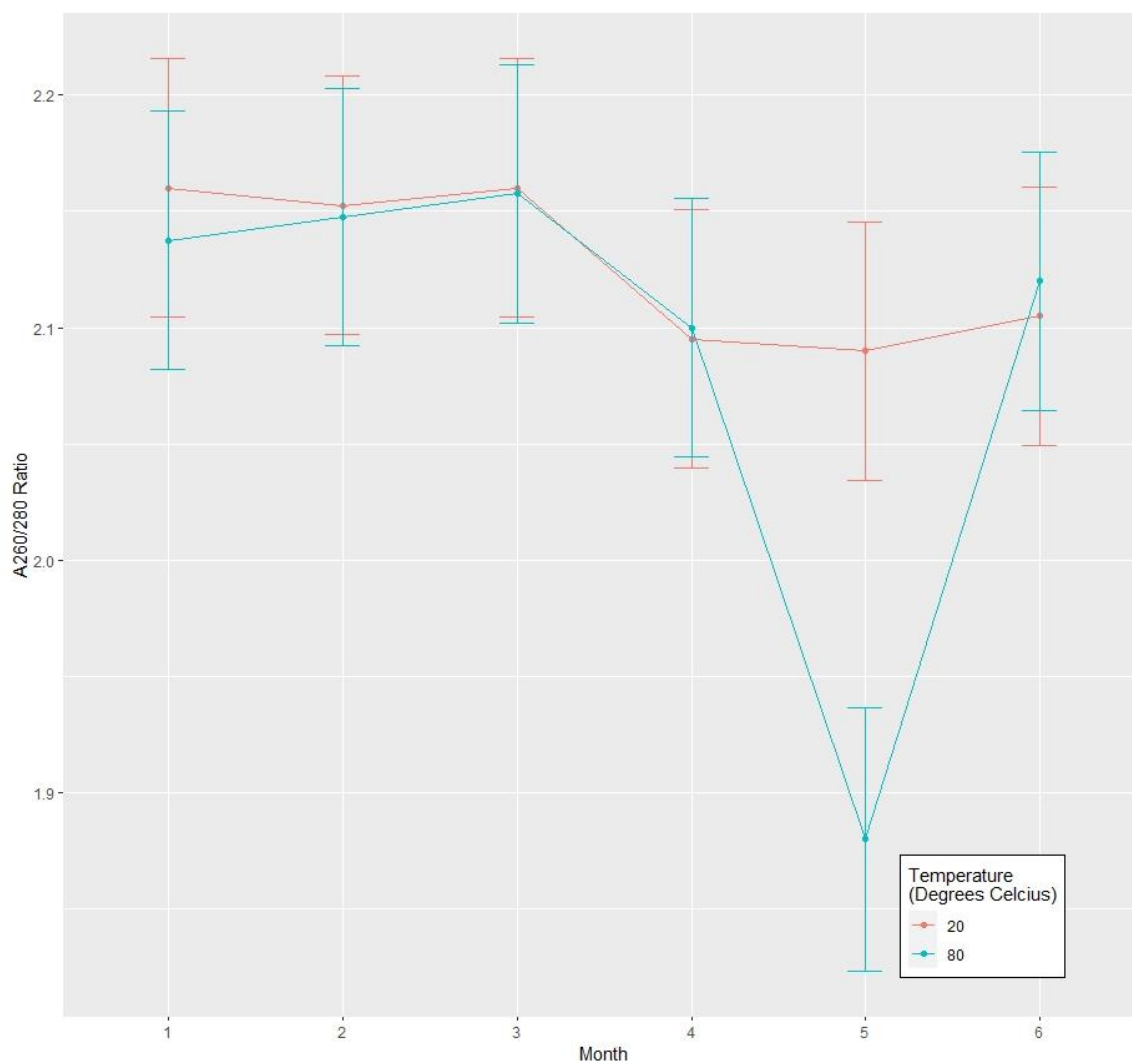
1170

### 1171 **Figure 2.4: LS-Means for 28S/18S Ratio Response to Treatments-Leaf**

1172 Comparison of the leaf tissues' 28S/18S response to storage length (1-6 months) and  
 1173 storage temperature (-20°C and -80°C) using 95% confidence intervals. No significant  
 1174 differences were identified, but samples stored in -80°C produced ratios closer to 2 for all  
 1175 six months.

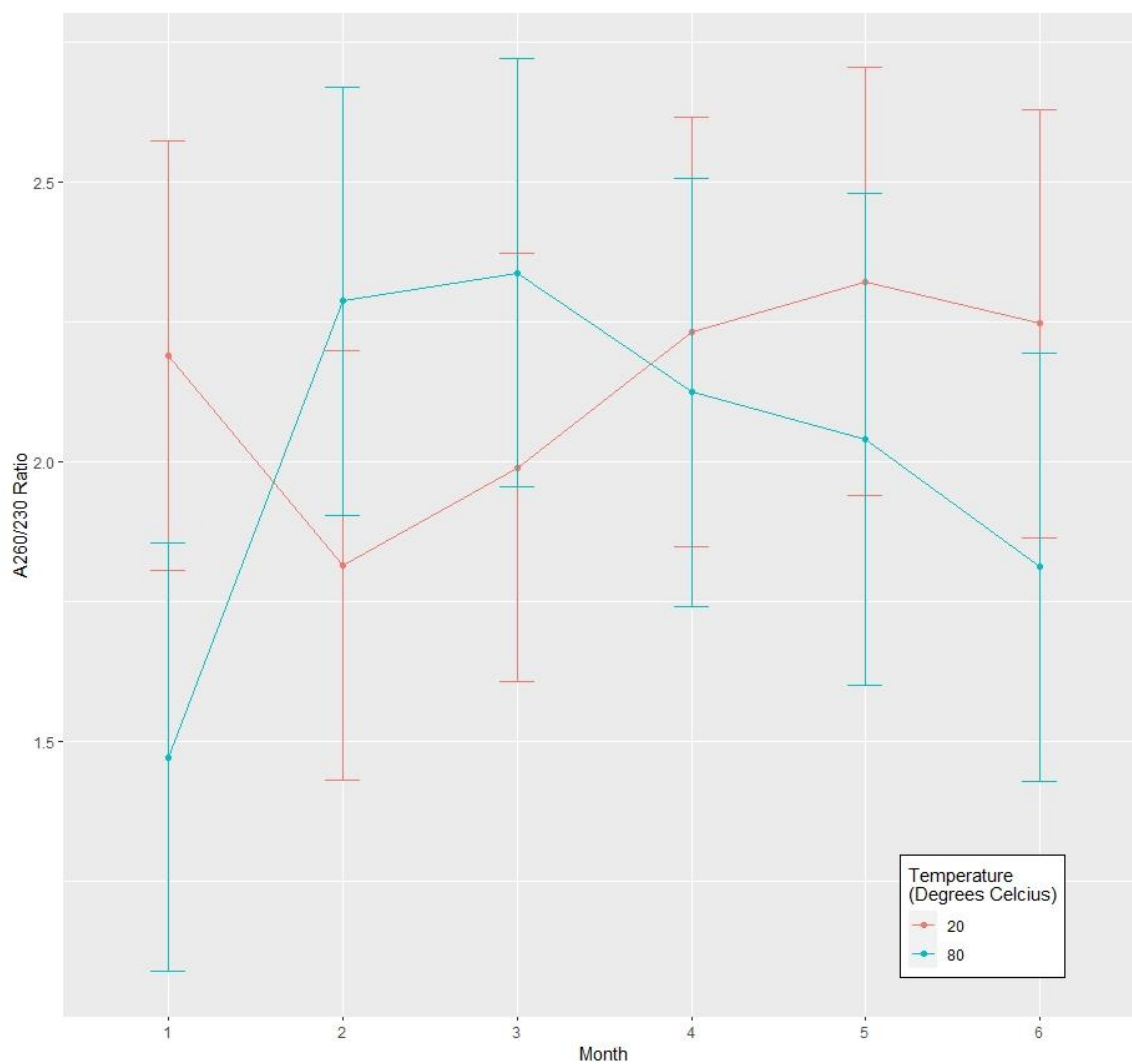


**Figure 2.5: LS-Means for NanoDrop Concentration Response to Treatments-Leaf**  
 Comparison of the leaf tissues' concentration response (reported from the NanoDrop™) to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. A significantly higher concentration was produced from the samples stored in -20°C compared to those in -80°C in month one. No other significant differences were identified, but samples stored in -80°C produced higher concentrations, > 500 ng/μL, on average for the other five months.



**Figure 2.6: LS-Means for A<sub>260/280</sub> Ratio Response to Treatments-Leaf**

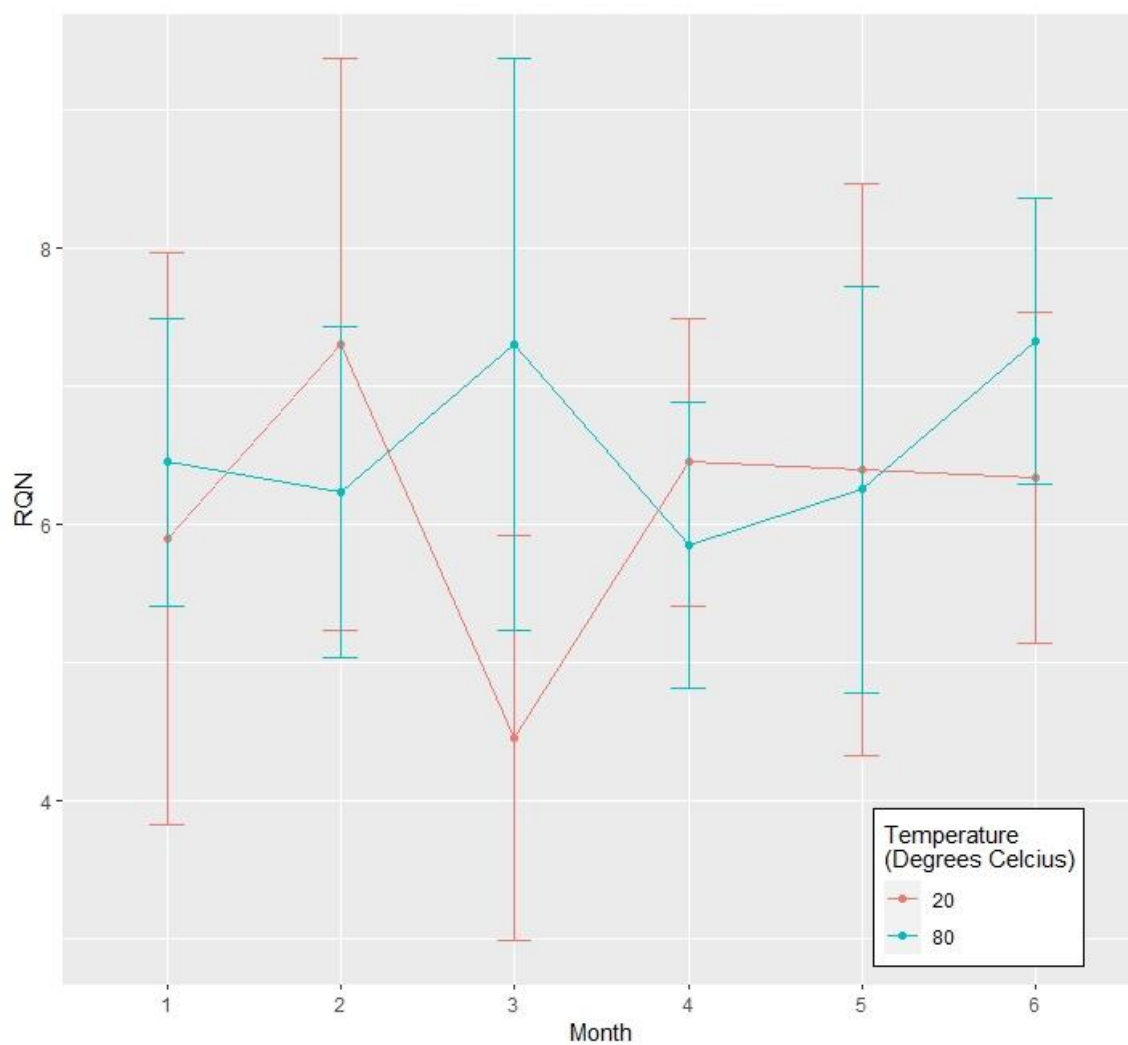
Comparison of the leaf tissues' A<sub>260/280</sub> response to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. A significant difference was seen in month five where samples stored in -20°C produced higher ratio response than those in -80°C. There was no other difference between the storage month or temperature.



**Figure 2.7: LS-Means for A<sub>260/230</sub> Ratio Response to Treatments-Leaf**

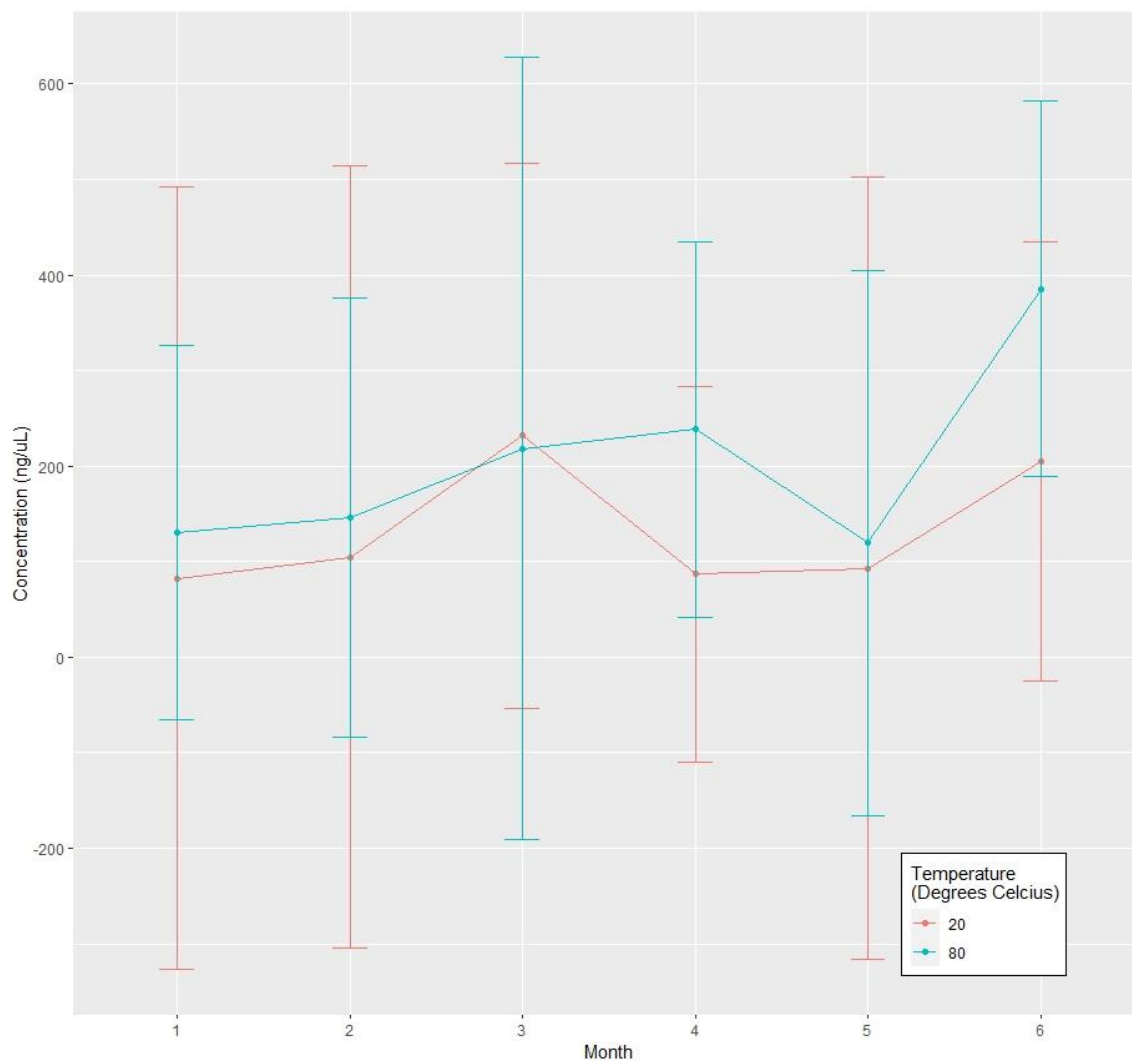
Comparison of the leaf tissues' A<sub>260/230</sub> response to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. There was no difference between the storage month or temperature and unusual variation present.





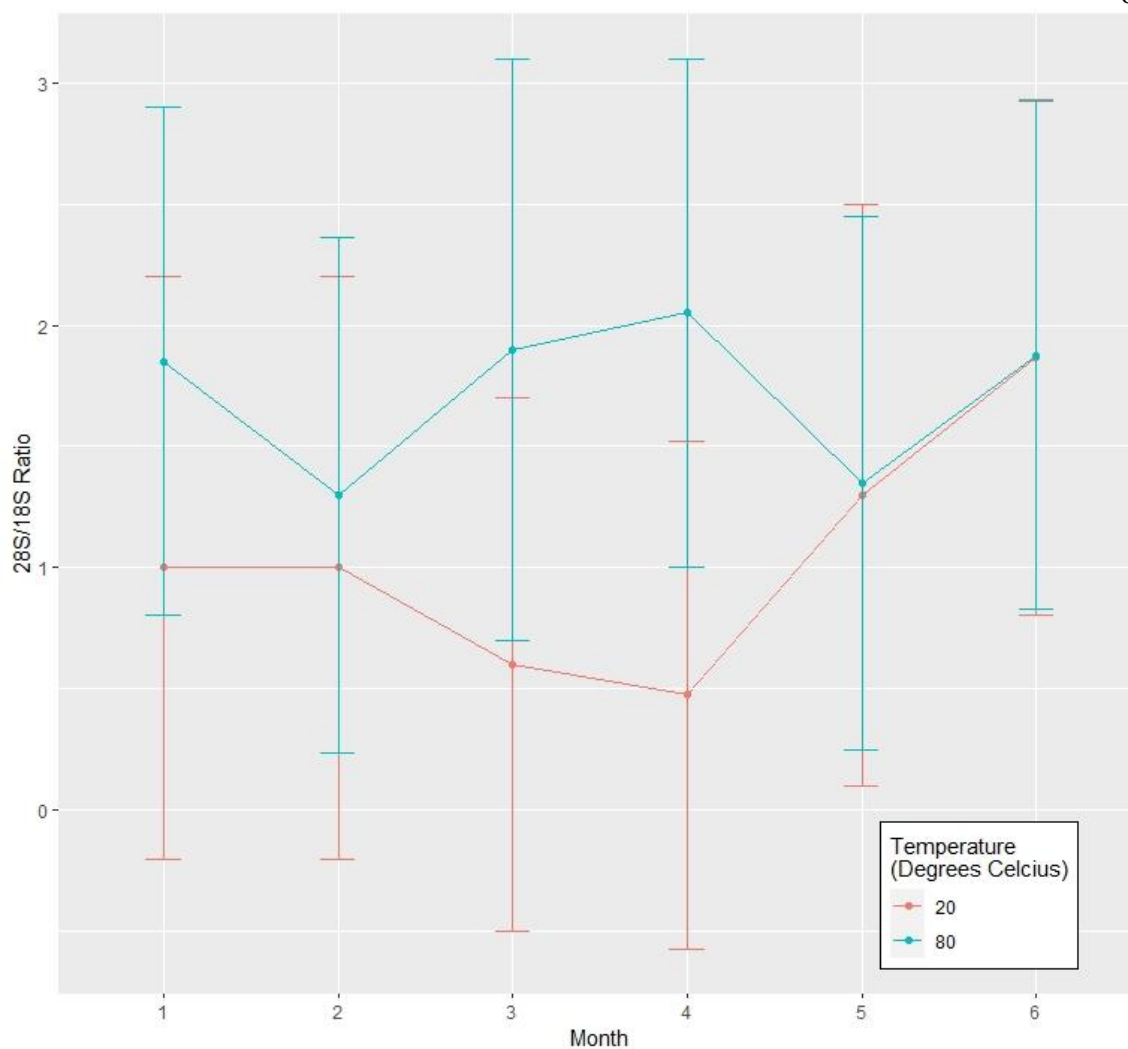
**Figure 2.8: LS-Means RQN Response to Treatments-Seed Samples**

Comparison of the seed tissues' RQN response to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. There was no difference between the storage month or temperature and unusual variation present, but most samples had an RQN > 5.



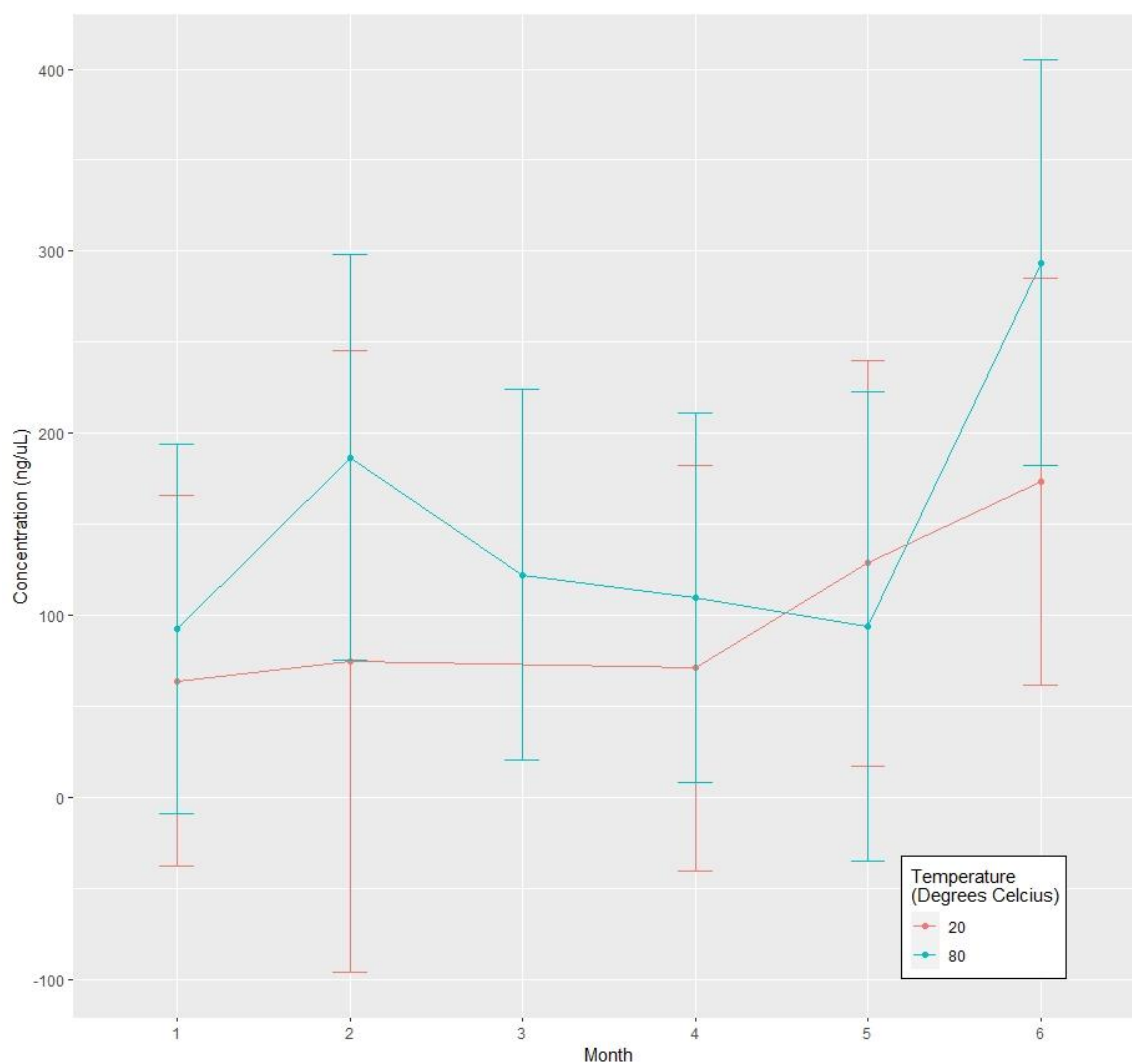
**Figure 2.9: LS-Means for Fragment Analyzer Concentration Response to Treatments-Seed**

Comparison of the seed tissues' concentration response (reported from the fragment analyzer) to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. No significant differences were identified, but samples stored in -80°C produced higher concentrations, > 100 ng/μL, for all six months.

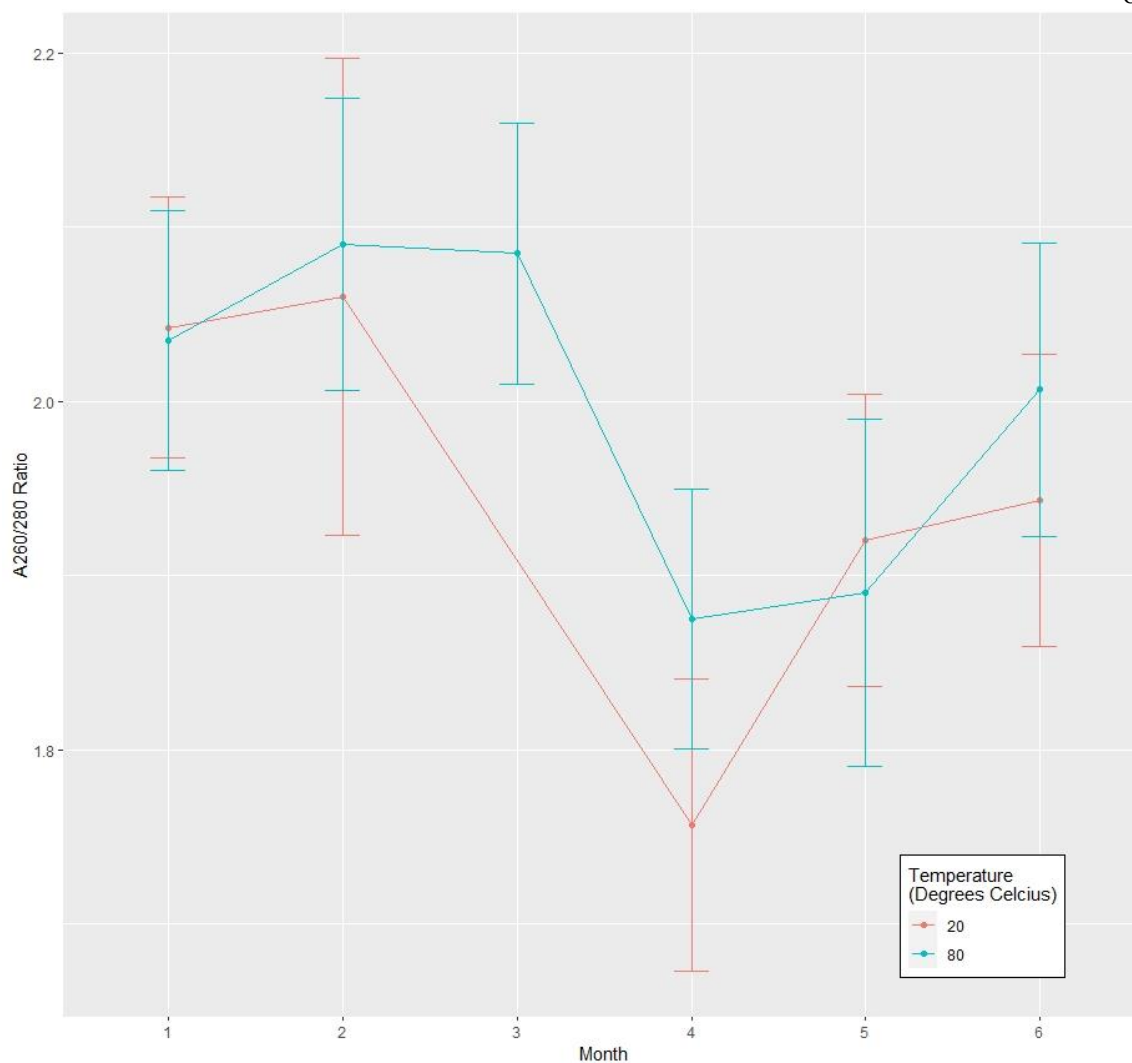


**Figure 2.10: LS-Means for 28S/18S Ratio Response to Treatments-Seed**

Comparison of the seed tissues' 28S/18S response to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. No significant differences were identified, but samples stored in -80°C produced ratios closer to 2 for all six months.

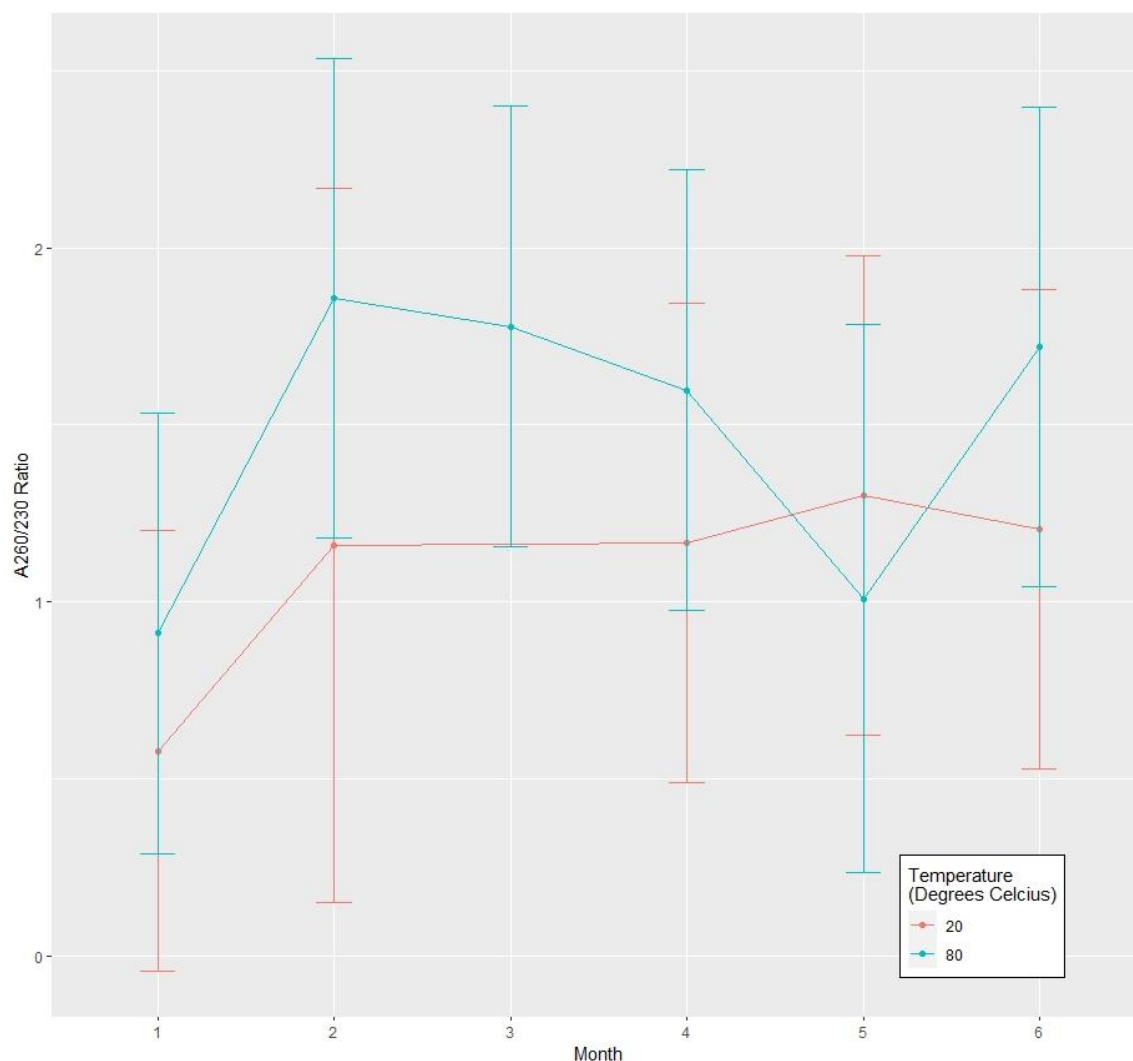


**Figure 2.11: LS-Means for NanoDrop Concentration Response to Treatments-Seed**  
 Comparison of the seed tissues' concentration response (reported from the NanoDrop™) to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. No significant differences were identified, but samples stored in -80°C produced higher concentrations, > 100 ng/μL, on average for the six months.



**Figure 2.12: LS-Means for A<sub>260/280</sub> Ratio Response to Treatments-Seed**

Comparison of the seed tissues' A<sub>260/280</sub> response to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. No significant differences were seen for storage length and temperature, but the ratio decreased from month two through six. Samples stored in -80°C produced higher ratios, > 1.8, across the six months which means the samples were better preserved at that temperature.



**Figure 2.13: LS-Means for A<sub>260/230</sub> Ratio Response to Treatments-Seed**

Comparison of the seed tissues' A<sub>260/230</sub> response to storage length (1-6 months) and storage temperature (-20°C and -80°C) using 95% confidence intervals. No significant differences were seen for storage length and temperature, but the ratio decreased from month two through six. Samples stored in -80°C produced higher ratios, > 1.0, across the six months.