ELUCIDATION OF MIRNAS IN *AVENA SATIVA*

by

Jaime Lynn Sheridan

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Bioinformatics and Computational Biology

Charlotte

2014

Approved by:

_____
Dr. Jessica Schlueter

_____
Dr. Shannon Schlueter

_____
Dr. Jennifer Weller

_____
Dr. Cory Brouwer

_____
Dr. Todd Steck

ABSTRACT

JAIME LYNN SHERIDAN. Elucidation of miRNAs in *Avena sativa*. (Under the direction of DR. JESSICA SCHLUETER)

Micro RNAs (miRNAs) are a class of small non-coding RNAs with post-transcriptional gene regulation responsibilities. In plants, miRNAs regulate development of leaves, roots, shoots, and flowers through targeting families of transcription factors and through targeting genes in the signal transduction and hormone pathways involved in plant growth and development. In addition to plant development, miRNAs are also involved in biotic and abiotic stress responses (Mallory & Vaucheret, 2006; Zhang, Pan, & Anderson, 2006), Plant disease and environmental stress reduce crop yield, which is a global problem as reduced crop yield struggles to meet world food production demand. With the majority of our world population's diet being composed of grains, there is a critical need to understand the epigenetic influence of miRNAs in cereal crops, especially in *Avena sativa.* There is currently no annotated miRNAs in oat.

The role of miRNAs in post-transcriptional regulation in plants is only beginning to be understood. MiRNA annotation in large, complex, un-sequenced genomes is challenging due to the lack of existing software and pipelines that incorporate sequence and structure-based approaches to effectively predict miRNAs. Integrating computational and molecular biology techniques, this study provides the first annotated miRNA library in hexaploid oat using a pipeline for annotation of miRNAs in the absence of a reference genome for use in other large, complex genomes. As an allopolyploid, the annotation of miRNAs in a diploid, tetraploid, and synthetic hexaploid exists as the first investigation of genome specific miRNAs. This study also proposes a

new model to explain the conditioning of the hulless trait in oat by miR172, enabling hulless oats to be bred more efficiently to fulfill the needs of an emerging market. In its entirety, this work improves the sequencing resources of oat and provides the framework for further inquiries into the effect of miRNAs on plant genomes.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

1.1 MiRNAs 2

1.1.1 Nuclear Processing of the Pre-miRNA 3

1.1.2 MiRNA Targeting in the Cytoplasm 4

1.2 SiRNAs 6

1.3 MiRNA Annotation 8

1.3.1 Criteria for MiRNA Annotation 8

1.3.2 Annotation Tools and Limitations 10

1.4 MiRNA Target Validation 13

1.5 Conserved MiRNAs 16

1.6 MiR172 in the ABCDE Model of Floral Development 17

1.7 Oat 21

1.7.1 Health Claims 23

1.8 Hulless Oat 24

1.8.1 Deficiencies of Hulless Oat 25

1.8.2 Markets for Hulless Oat 28

1.9 Research Directions 31

CHAPTER 2: ANNOTATION OF MIRNAS IN HEXAPLOID OAT 40

2.1 Introduction 40

2.1.1 Classification and Evolution of MiRNAs 40

2.1.2 Small RNA Sequencing 42

2.1.3 MiRNA Nuclear Processing 42

2.1.4 MiRNA Duplexes and the RISC 44

2.1.5 MiRNA Annotation 45

2.1.6 MiRNA Annotation Tools 46

2.1.7 miRDeep and miRDeep-P for MiRNA Annotation 49

2.2 Materials and Methods 52

2.2.1 Testing miRDeep-P on Known MiRNA Precursors 52

2.2.2 Tissue and RNA Isolation 53

2.2.3 Transcriptome Sequencing and Assembly 54

2.2.4 Small RNA Sequencing and Analysis 55

2.2.5 Identification of Known MiRNAs 56

2.2.6 MiRNA Discovery in the HiSeq Transcriptome 56

2.3 Results and Discussion 57

2.3.1 Testing miRDeep-P on Plant MiRNA Precursors 57

2.3.2 MiRNA Predictions in Leggett, Jordan and Betania 58

2.3.2.1 Leggett: Predicted Previously Annotated MiRNAs 58

2.3.2.2 Leggett: Predicted Novel Candidate MiRNAs 60

2.3.2.3 Jordan: Predicted Known MiRNAs 61

2.3.2.4 Jordan: Predicted Novel Candidate MiRNAs 62

2.3.2.5 Betania: Predicted Previously Annotated MiRNAs 63

2.3.2.6 Betania: Predicted Novel Candidate MiRNAs 65

2.3.2.7 Stem-to-loop and Loop-to-base Processing of Bet_cand1x12674 67

2.3.2.8 Betania: Novel Candidate MiRNA Families 67

2.4 Conclusions 68

CHAPTER 3: MIRNA DISCOVERY IN AVENA SPECIES OF THREE PLOIDY LEVELS 99

3.1 Introduction 99

3.2 Materials and Methods 103

3.2.1 Total RNA Sequencing Tissue and RNA Extraction 103

3.2.2 Total RNA Sequencing and Assemblies 103

3.2.3 Small RNA Sequencing Tissue Collection and RNA Extraction 104

3.2.4 Small RNA Sequencing and Sequence Pre-processing 105

3.2.5 Identification of Known MiRNAs 106

3.2.6 Identification of Candidate MiRNA Precursors 106

3.2.7 Multiple Sequence Alignment of Small RNAs and Transcript Assemblies 107

3.3 Results and Discussion 107

3.3.1 Identification of Known MiRNAs in *Avena strigosa* 108

3.3.2 Prediction of Previously Annotated MiRNAs in *Avena strigosa* 110

3.3.3 Predicted Novel Caniddate MiRNAs in *Avena strigosa* 111

3.3.4 Identification of Known MiRNAs in *Avena magna* 112

3.3.5 Predicted Conserved and Non-conserved MiRNAs in *Avena magna* 114

3.3.6 Predicted Novel Candidate MiRNAs in *Avena magna* 115

3.3.7 Identification of Known MiRNAs in the Synthetic Hexaploid 115

3.3.8 Predicted Known MiRNAs in the Synthetic Hexaploid 116

3.3.9 Predicted Novel Candidate MiRNAs in the Synthetic Hexaploid 117

3.3.10 A Novel Candidate MiRNA Family in the Synthetic Hexaploid 117

3.3.11 Predicted A-genome Candidate MiRNAs 119

3.4 Conclusions 120

CHAPTER 4: REGULATION OF THE HULLESS TRAIT IN AVENA SATIVA 211

4.1 Introduction 211

4.1.1 Oat 211

4.1.2 MiR172 213

4.1.3 AP2/ERF Transcription Factor Family 214

4.1.4 Interaction of MiR172 and AP2/ERF Transcription Factors in the Grasses 215

4.1.5 Hullessness in Barley 216

4.1.6 Wheat and the Q 217

4.1.7 Spikelet Abnormality in Rice 217

4.1.8 Floral Differentiation and Indeterminacy in Corn 218

4.1.9 IDS1, NUD, and MiR172 in *Avena sativa* 220

4.2 Materials and Methods 220

4.2.1 Tissue and DNA/RNA Isolation 220

4.2.2 MiR172 Genome Walking 221

4.2.3 Genomic IDS1 Amplification 221

4.2.4 Genomic NUD Amplification 222

4.2.5 Cloning and DNA Sequencing 223

4.2.6 IDS1, NUD, and Genomic MiR172 Sequence Analysis 223

4.2.7 MiR172 Sequence Amplification 224

4.2.8 Small RNA Sequencing and Processing 225

4.2.9 MiR172 Expression 226

4.3 Results and Discussion 227

4.3.1 Genomic Variation in IDS1 227

4.3.2 Coding Variation in IDS1 228

4.3.3 Genomic Variation in NUD 229

4.3.4 MiR172 Binding in NUD 229

4.3.5 Genomic MiR172 230

4.3.6 Quality Control of Log2-transformed Raw Count Data 231

4.3.7 Normalized Distributions 231

4.3.8 ANOVA of MiR172 233

4.3.9 ANOVA of Missing Data Filtered Subset of Small RNA Intensities 235

4.4 Conclusions 236

REFERENCES 279

APPENDIX: MIRNA PREDICTIONS FROM OTHER SEQUENCING SOURCES 296

A.1 Introduction 296

A.2 Materials and Methods 296

A.2.1 Tissue and Total RNA Extraction 296

A.2.2 454 Total RNA Sequencing 296

A.2.3 454 Transcriptome Assembly 297

A.2.4 HiSeq Transcriptome Sequencing 297

A.2.5 HiSeq Transcriptome Assembly 297

A.2.6 Small RNA Sequencing and Analysis 297

A.2.7 Identification of Candidate MiRNAs and Corresponding Precursors 298

A.2.8 Additional Candidate MiRNAs in *Avena strigosa* (AA) 298

A.2.9 Additional Candidate MiRNAs in *Avena magna* (CCDD) 299

A.2.10 Additional Candidate MiRNAs in *Avena sativa*, Variety Leggett (AACCDD) 300

A.3 Summary 301

CHAPTER 1: INTRODUCTION

Gene expression has many points of regulation including transcription factor binding, DNA methylation, and histone modification. At the post-transcriptional level, one of the more recently described mechanisms is the regulation of genes through small non-coding RNAs. Non-coding RNAs are functional RNAs, which are not translated into protein. This diverse classification of RNA is composed of distinctly functional families including ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs), U1-6 (spliceosome RNAs), telomerase RNAs, ribozymes, and small RNAs (sRNAs) (Griffiths-Jones et al. 2005). Of these families, only tRNAs, rRNAs, snoRNAs, and sRNAs are shared between eukaryotes and prokaryotes. sRNAs are the family of non-coding RNAs with the ability to regulate gene expression. In eukaryotes, micro RNAs (miRNAs) and short interfering RNAs (siRNAs) are the primary family members of sRNAs with post-transcriptional gene regulation responsibilities. MiRNAs and siRNAs use the RNA interference pathway to silence genes primarily through cleavage of the target mRNA in plants. MiRNAs also influence gene expression through a translational repression mechanism that has minor effect on the target mRNA level, but rather significantly reduces the target protein level. The discovery of regulatory small RNAs demonstrates a new level of complexity in the field of post-transcriptional regulation.

1.1 MiRNAs

In 1993, micro RNAs or miRNAs, were first discovered in the nematode *C. elegans* by positional cloning (Jones-Rhoades, Bartel, & Bartel, 2006; R. C. Lee, Feinbaurm, & Ambros, 1993). Lin-4 negatively regulates lin-14 mRNA between the L1 and L2 stages of nematode development. Although lin-4 was first discovered in the 1993, the second miRNA let-7 was not identified until 2000 (Reinhart et al., 2000). These regulatory small RNAs have since been discovered in other eukaryotes; however, no miRNAs have been found in prokaryotes. Bacteria have regulatory small RNAs that have an analogous, but not identical function to miRNAs (Tjaden et al., 2006; Wassarman, 2002). In eukaryotes, this class of ~22-nucleotide long RNAs is predominantly involved in gene repression and acts as translation inhibitors that regulate genes associated with development, cell differentiation and metabolism, as well as involvement in human diseases (Calin & Croce, 2006; Huntzinger & Izaurralde, 2011; Y. Wang et al., 2012). A clear picture of gene regulation remains difficult due to the diverse functions of the genes miRNAs target.

Although both animals and plants have a miRNA pathway with homologous key proteins with the similar function to negatively regulate gene expression, the pathways are substantially different. Animal miRNAs function primarily by translational inhibition whereas plant miRNAs use RISC (RNA-induced silencing complex) to cleave mRNA target transcripts. The three key proteins shared by plant and animal systems are DICER-like proteins, ARGONAUTE-Piwi like proteins, and RNA-dependent polymerases (Shabalina & Koonin, 2008).

1.1.1 Nuclear Processing of the Pre-miRNA

MiRNAs have a very distinct biogenesis pathway (Naqvi, Sarwat, Hasan, & Roychodhury, 2012). In plant cell nuclei, genes that encode miRNAs are transcribed into RNA by RNA polymerase II. Originating primarily from intergenic regions (Rajagopalan, Vaucheret, Trejo, & Bartel, 2006), the transcript has secondary structure and folds into a "stem-hairpin-loop." This is the pri-miRNA. DCL1 (DICER-LIKE ENDORIBONUCLEASE 1), associated with the double-stranded RNA-binding protein HYL1 (HYPOSTATIC LEAVES 1) (Han, Goud, Song, & Fedoroff, 2003), recognizes the secondary structure through its RNA-binding domain and cleaves the pri-miRNA into a pre-miRNA. It is unclear if HYL1 is associated in both successions of cleavage: the pri-miRNA into the pre-miRNA and the pre-miRNA into the miRNA duplex. In the stem-to-loop processing, the pre-miRNA is a smaller hairpin with a shorter stem. A second, less common miRNA processing mechanism is loop-to-base processing, where the terminal loop of the pri-miRNA is cleaved first, resulting in the double-stranded stem of the pri-miRNA. (Jones-Rhoades et al., 2006; Naqvi et al., 2012; Song, Axtell, & Fedoroff, 2010). The second cleavage of the pre-miRNA by DCL1 creates a stem of 21 nucleotides, the miRNA/miR* duplex. HEN1 (HUA ENHANCER 1) methylates the 2'-OH of the 3' terminal nucleotides of the miRNA/miRNA* duplex (Z. Yang, Ebright, Yu, & Chen, 2006). The double stranded RNA duplex has the mature miRNA on one strand and the miRNA* on the opposite strand. In addition to being terminally methylated, the duplex also has a 5' monophosphate and a 2-nucleotide overhang on the 3' end. Once the processing of pri-miRNA to miRNA/miR* duplex has occurred, HASTY facilitates the

transport of the RNA duplex from the nucleus to the cytoplasm (Park, Wu, Gonzalez-Sulser, Vaucheret, & Poethig, 2005).

1.1.2 MiRNA Targeting in the Cytoplasm

Once in the cytoplasm, a RISC (RNA-induced silencing complex) unwinds the double stranded RNA through helicase activity. Although not yet discernable in plants, in humans it is the N domain of the AGO2 that initiates the unwinding of the miRNA duplex (Kwak & Tomari, 2012). AGO1, an ARGONAUTE protein (and a specific protein of the RISC in plants) associates with only the mature miRNA strand of the double stranded RNA molecule. The miR* is degraded quickly through an unknown mechanism (Bartel, 2004), possibly through SDNs (SMALL RNA DEGRADING NUCLEASE) (Ramachandran & Chen, 2008). In small RNA sequencing experiments, miR* are few in comparison to their mature miRNA complements. Once associated with AGO1, the exposed miRNA strand is scanned across mRNA transcripts.

Near complete complementary mRNA transcripts bind to the exposed mature miRNA. Mature miRNAs primarily bind to the coding region of the transcript; however, binding to the 3' and 5' untranslated regions (UTRs) does occur (Jones-Rhoades & Bartel, 2004; Rhoades et al., 2002; Sunkar, Girke, Jain, & Zhu, 2005). The high complementarity triggers the PIWI domain of AGO1 to cleave the mRNA transcript between position 10 and 11 of the miRNA (Jinek & Doudna, 2009). In the cytoplasm, the two cleaved fragments are degraded; the 5' fragment is degraded by the 3'-to-5' exonuclease activity of the exosome whereas the 3' fragment is degraded by the 5'-to-3' exonuclease activity of XRN4 (Souret, Kastenmayer, & Green, 2004). However, exactly

how these enzymes are recruited to the AGO complex remains to be determined. This is the primary mechanism by which miRNAs affect translation.

Plants also have a secondary mode of gene regulation, translational repression. This is the primary mode of animal miRNA regulation. In contrast to miRNA target transcripts being targeted for degradation causing both a reduce in transcript and protein, translational repression can result in a steady transcript level but reduced protein level such as the relationship between miR172 and AP2 (APETALA2) and AP2-like genes (Aukerman & Sakai, 2003; X. Chen, 2004). This interaction does result in detection of low levels of target cleavage; therefore, there is an overlap between cleavage and repression fates of miR172 targets (Aukerman & Sakai, 2003). The mechanism of miRNA-mediated translational repression remains unclear (Huntzinger & Izaurralde, 2011). MiRNAs post-transcriptionally regulate their corresponding targets through transcript cleavage or translational inhibition. Figure 1 (A) shows the proteins involved in and cellular location of miRNA processing.

In order to fully understand the annotation criteria for miRNAs (discussed below), it is important to understand the fundamental similarities and differences between miRNAs and another regulatory class of small RNAs, siRNAs. In contrast to miRNAs, which can translationally repress and cleave complementary targets, siRNAs (short interfering RNAs) regulate targets through cleavage mechanisms and methylation despite the inclusive use of RISCs. Plant siRNAs are classified by the proteins involved in biogenesis, their trigger, and the homologue of ARGONAUTE protein implicated in silencing.

1.2. SiRNAs

SiRNAs are either exogenous or endogenous. Exogenous siRNAs led to the discovery of double stranded RNA being the silencing trigger in *C. elegans* (Fire et al., 1998). Sources of exogenous RNA can be viral or transgenic but both methods result in the production of siRNA duplexes from double-stranded RNA. Double-stranded RNA is not the only source of siRNAs. Single-stranded RNA transcripts from transgenes or viruses can be converted to double-stranded RNA by RNA-dependent RNA polymerases (Qu, Ye, Hou, Clemente, & Morris, 2005). A DICER cleaves exogenous double-stranded RNA duplexes into 21- and 24-nucleotide siRNA duplexes where HEN1, a methyltransferase, terminally 2'-O-methylates complementary strands of the duplex (Z. Yang et al., 2006). An ARGONAUTE protein as part of RISC, directs regulation through using one stand of the siRNA duplex as the guide strand to base pair with a target; the passenger strand is destroyed. The target is cleaved by the RISC and degraded, resulting in the unavailability of mRNA transcript for translation.

Endogenous siRNAs in plants vary in biogenesis, source of the trigger, and the types of targets. Endogenous siRNAs include casiRNAs (cis-acting siRNAs), tasiRNAs (trans-acting siRNAs), and natsiRNAs (natural antisense siRNAs). After cleavage of the double-stranded RNA into a duplex and methylation, endogenous RNAs are exported and incorporated into the RISC akin to exogenous RNAs.

The siRNA and miRNA pathways converge through the action of tasiRNAs. After miRNA cleavage by AGO1, one fragment of the cleaved transcript, usually having a second miRNA or siRNA binding site, is processed into double-stranded RNA by RDR6 and the RNA binding protein SGS3 (SUPPRESSOR OF GENE SILENCING 3).

DCL4 cleaves the double-stranded RNA duplex every 21 nucleotides beginning at the miRNA cleavage site to produce siRNA duplexes. ARGONAUTE 1 and 7 are the AGOs that are involved in the RISC that initiate cleavage of the target transcript (E. Allen, Xie, Gustafson, & Carrington, 2005; Ghildiyal & Zamore, 2009; Poulsen, Vaucheret, & Brodersen, 2013).

Despite having different triggers, different DICERs and ARGONAUTE proteins, and targets, endogenous siRNAs are difficult to distinguish from each other and miRNAs in a small RNA sequencing dataset. MiRNAs and siRNAs are about 21-25 nucleotides in length as a result of DICER protein activity. The miRNA and siRNA duplexes have a 5' monophosphate, a 2-nucleotide overhang and a 2'-O-methylated on the 3' end. The methylation occurs for both siRNAs and miRNAs by HEN1. ARGONAUTE proteins as part of the RISC are responsible for target cleavage/inhibition/methylation although different variants of AGO are involved in miRNA and siRNA pathways. The primary difference used in distinguishing siRNAs from miRNAs for annotation purposes is the biogenesis of each non-coding RNA. RNA Pol II transcribes miRNA genes whereas RNA Pol IV transcribes endogenous siRNA genes. The pri-miRNA transcript forms a stem-hairpin loop secondary structure, characteristic of miRNAs. From one pri-miRNA transcript, one miRNA/miR* duplex is formed but transcripts from siRNA precursors (no hairpin loop structure) form multiple, overlapping siRNA duplexes. Table 1 summarizes (Ghildiyal & Zamore, 2009; Poulsen et al., 2013) the difference between endogenous siRNAs and miRNAs. These variations in biogenesis are used to distinguish siRNAs from miRNAs.

1.3 MiRNA Annotation

MiRNAs and siRNAs are classifications of small RNAs. Within the class of miRNAs, mature miRNAs are conserved evolutionarily. They are classified into conserved, non-conserved, and novel families. Each family may have more than one member; resulting from multiple miRNA genes processed into identical or near identical mature miRNA sequences (Jones-Rhoades et al., 2006). Conserved miRNA families are conserved across plant lineages; every plant has one or more representation of each of these families. There are 20 conserved miRNA families present in *Arabidopsis*, rice and *Populus* as described in the 2006 annual review of Plant Biology (Jones-Rhoades et al., 2006). Since, these 20 families have become the standard of classifying conserved miRNAs (Table 2). Non-conserved miRNAs are considered lineage specific. An example of this would be miR528, which is conserved in monocot plant species. Novel miRNA families are specific to one species or one genus. Figure 2 shows an interpretation of the conservation of miRNA families across bryophytes, gymnosperms, and angiosperms (Cuperus, Fahlgren, & Carrington, 2011).

1.3.1 Criteria for MiRNA Annotation

Because both siRNAs and miRNAs take advantage of AGO cleavage of transcripts, it is difficult to annotate miRNAs. Ambros proposed a set of expression and biogenesis criteria for miRNA annotation that distinguish miRNAs from siRNAs and other RNAs (Ambros, 2003). Expression criteria include Northern blot detection of a distinct ~22-nucleotide RNA transcript and identification of a distinct ~22-nucleotide sequence in a cDNA library, matching the genomic sequence. The first biogenesis criteria is a hairpin with lowest free energy predicted by an RNA folding algorithm with the ~22-nucleotide

mature miRNA sequence on one arm and the miR* complementary to the mature miRNA on the opposite arm of the 60-80 nucleotide hairpin in animals, more variable precursor length in plants. Within the hairpin there must not be large internal bulges, especially in the stem region where the mature miRNA and the miR* are located. The second biogenesis criterion is phylogenetic conservation of the miRNA and the hairpin and the third criteria is the accumulation of precursor transcript in DICER knockdowns. Annotation of a miRNA cannot be done on any one individual criterion; a combination of expression and biogenesis is required. Sufficient combinations are Northern blot, phylogenetic conservation, and precursor accumulation in knockdowns although in the absence of a DICER knockdown, meeting the first two criteria is sufficient evidence to be added to a public miRNA database such as miRBase. Other combinations are Northern blot and predicted hairpin structure of the precursor, and sequence signature in cDNA library and phylogenetic conservation of hairpin. The naming system for miRNAs is also established as the "miR" prefix and a unique number and a letter or number suffix to distinguish similar miRNA sequences from the same organism.

Revised miRNA annotation criteria for plant miRNA annotation was proposed in 2008 (Meyers et al., 2008). The primary criterion proposed is "precise excision of an ~21-nucleotide miRNA/miRNA* duplex from the stem of a single-stranded, stem-loop precursor." Within the pri-miRNA hairpin, one arm must have the miRNA that base pairs with the miRNA* with a 2nt 3' overhang. Between the miRNA/miRNA*, there must be four or fewer mismatches and fewer than two bulges of one or two bases. Here are cases that would satisfy the primary criterion: identification of both miRNA and miRNA* in a small RNA cDNA sequencing library with precursor prediction, and identification of the

miRNA in multiple small RNA cDNA sequencing libraries in the absence of the miRNA* sequence in combination with predicting the miRNA from one arm of a hairpin. The authors also proposed a set of ancillary criteria, which individually are insufficient for miRNA annotation; however meeting these ancillary criteria can greatly increase the confidence of a miRNA once the biogenesis criterion is met. The ancillary criteria include sequence alignment to establish conservation of the pri-hairpin and miRNA sequence, target cleavage or translational repression by the miRNA, accumulation of precursor in dcl1 mutants, and establishment of independence from RNA-dependent RNA polymerases and RNA polymerase IV/V. The authors include notes on special care when miRNAs overlap with repetitive sequences such as tandem repeats and transposons because siRNAs that are products of RNA polymerase IV/V originate primarily from repetitive regions such as these. True miRNAs that originate from this region must satisfy the biogenesis criteria. These biogenesis criteria exclude siRNAs from being incorrectly annotated as miRNAs.

1.3.2 Annotation Tools and Limitations

Many prediction pipelines, software, and servers have been developed to annotate and identify miRNAs from high-throughput sequencing experiments. These computational resources rely on the annotation criteria listed above in order to maintain a high quality of annotated sequences in databases. miRBase is the most comprehensive database of miRNAs with version 20 having 24,521 hairpin sequences and 30,424 mature miRNA sequences across 206 species, including 72 plant species (Griffiths-Jones, 2004; Griffiths-Jones, Grocock, van Dongen, Bateman, & Enright, 2006; Griffiths-Jones, Saini, van Dongen, & Enright, 2008; Kozomara & Griffiths-Jones, 2011).

A standard for plant miRNA annotation appears clear; but as the cost of high-throughput sequencing decreases, more miRNA discovery projects are being initiated and maintaining a quality of sequences in public databases has become a caveat of the most comprehensive miRNA database, miRBase (Griffiths-Jones, 2004; Griffiths-Jones et al., 2008). Beginning as a public repository to assign an official naming scheme for miRNAs in plants and animals, miRBase has become a sequence database including mature miRNA and miRNA precursor sequences. The previous requirement for deposition of "annotated" miRNAs has been the acceptance of a manuscript with a description of the miRNA. This has become extremely challenging when reviewers do not identify inaccurate annotations with miRNAs missing precursors and corresponding miR* sequences (Y. Meng, Shao, Wang, & Chen, 2012). In addition, miRNAs are annotated using various methods, pipelines, and tools with non-uniform stringencies on criteria. The current version of miRBase has a high confidence set, which has been validated with multiple publically available deep sequencing datasets (Kozomara & Griffiths-Jones, 2014). The high confidence classification relies on a more restrictive miRNA/miR* duplex mismatching, the number of reads mapping to the precursor, and the precursor free energy. Many species-specific, tissue-specific, stage-specific, and stress-specific miRNAs remain in the low-confidence set due to the lack of sequence support in other datasets. It is not suggested to create new criteria or a standard due to the non-canonical behavior of some miRNAs; but re-examination of miRNA datasets is necessary to ensure high quality miRNA annotations (Kozomara & Griffiths-Jones, 2014; Y. Meng et al., 2012).

Computational tools for miRNA annotation are most accurate and comprehensive when the species has a well-annotated sequenced genome. Starting with a deep sequencing of small RNAs, sequences go through a pre-processing step where sequencing adapters and barcodes are trimmed, low quality sequences are filtered out, other non-coding RNAs are removed based on sequence similarity to annotated non-coding RNAs in a database, and sequences undergo a filter for length. To meet the biogenesis criteria, these short sequences are mapped to a reference genome using a mapping software or alignment tool. Candidate miRNA precursors are run through an RNA folding software to ensure the secondary structure and location of the miRNA and miR* sequence. This can identify conserved, non-conserved, and novel miRNAs.

In the absence of a reference genome, aligning the small RNA sequences to miRBase will satisfy the ancillary phylogenetic criteria as well as identify miR* of known miRNAs; however, it is not sufficient for annotation unless the corresponding hairpin is conserved. For example, if the species of interest is a monocot and in the grass family, sequence similarity between known miRNAs conserved across grasses only identifies non-conserved miRNAs. Conserved miRNAs can be identified through sequence similarity of miRNAs conserved across more distant species such as *Arabidopsis thaliana* and *Glycine max*. However, identifying novel miRNAs becomes a challenge for two reasons. A small RNA sequence from a library may not map to an EST or genomic DNA available. These unmappable small RNAs are disregarded as sequencing errors or contamination from degraded mRNA or other non-coding RNAs. The second reason is that when a small RNA sequence does map to an EST or genomic DNA sequence, every hit must be evaluated to ensure the proper folding of the hairpin

loop. Aligning a small RNA sequence to a well-annotated genome excludes precursor candidates that originate from repetitive regions, coding regions, or introns. Special cases of miRNAs that overlap coding regions or introns (miRtrons) resulting from alternative splicing can also be predicted with a well-annotated reference genome; however, most transcriptomes are not annotated well enough to reduce the number of miRNA precursors to screen.

In summary, having a genomic reference sequence of the species from which small RNAs are sequenced is a great advantage. All conserved families, and non-conserved families (such as monocot-specific miRNAs) can be identified and annotated as well as novel miRNAs. In very well annotated genomes, non-canonical miRNAs such as miRtrons, miRNAs overlapping with coding regions, and miRNA producing tasiRNAs can be identified (Q. H. Zhu et al., 2008).

1.4 MiRNA Target Validation

MiRNA annotation is primarily based on the biogenesis of miRNAs and the establishment of miRNA function through mRNA target cleavage or translational repression is not required for annotation. This is comparable to the function of protein-coding genes not being required for annotation (Meyers et al., 2008). Predicting and validating miRNA targets is essential to understand how miRNAs epigenetically control plant genotypes and phenotypes. MiRNA targets can be predicted computationally using a scoring system with some considerations (Dai, Zhuang, & Zhao, 2011). Due to the high complementarity of the mature miRNA to its targets, the number of mismatches (including non-canonical G:U pairs) and bulges must be limited. Targets are evolutionarily conserved, especially transcription factor families targeted by conserved

miRNAs families. For target prediction, scoring systems for miRNA target prediction account for phylogenetically conserved domains. Another consideration in predicting miRNA targets is the accessibility of the miRNA target site. mRNA has secondary structure and therefore, may inhibit the binding of the miRNA despite high complementarity. Complexity is added when temperature influences secondary structure of the mRNA target. Folding under a distinct temperature may allow accessibility to the miRNA target site, which may result in these predicted targets being missed with a different temperature parameters for prediction. MiRNA/target pairs with more mismatches may also be excluded depending on the prediction parameters, but performing miRNA target predictions with less stringent parameters allows for an increase in false positive predictions. Nonetheless, any predicted target pair is considered putative until validated.

Validation methods exist for experimentally demonstrating a specific miRNA/target interaction as well as a high-throughput global sequencing of miRNA targets. For specific miRNA/target interactions, expression mutants can be used. Aside from and indirect measure of a phenotypic difference, using qPCR or probing an RNA blot of total RNA from mutants overexpressing the miRNA of interest would show a reduction in the target transcript level. Knockdown mutants of the miRNA, on the other hand, would show an increase in target transcript level compared to the wildtype. Knockdown mutants of the target may also help characterize the interaction.

Sequencing technologies are also available to validate miRNA/target interactions. To validate a specific miRNA/target interaction, a modified 5' RACE, RLM-RACE (RNA Ligase-Mediated 5' Amplification of cDNA Ends), can be used (Scotto-Lavino,

Guangwei, & Frohman, 2007). The technology to amplify cleaved miRNA targets is based on the 5' monophosphate left on the 3' fragment upon AGO cleavage of the target transcript. From total RNA, an adapter is ligated to the monophosphate followed by reverse transcription and nested PCR to amplify the specific target. Upon sequencing, the fragment should map to the known full-length transcript, validating the interaction. This can be scaled up to validate all miRNA/target interactions instantaneously using degradome sequencing or PARE (Parallel Analysis of RNA Ends). High-throughput degradome sequencing of the 3' fragment ligates an RNA adapter with a MmeI restriction site to the 5' monophosphate. cDNA is digested with MmeI, leaving a 20-21 nucleotide tag attached to the 5' RNA adapter. After sequencing, the tag can be mapped to transcriptome sequences and other expression data to validate the target of each miRNA on a whole genome scale (Addo-Quaye, Eshoo, Bartel, & Axtell, 2008; German et al., 2008).

These sequencing technologies as well as RNA blot analysis evaluate changes in transcript level; however, this would only hold if the miRNA/target interaction of interest follows the mode of cleavage. Because some miRNAs function primarily by a translational repression mechanism where only minimal changes in transcript levels are observed, expression differences at the protein level are able to validate the interaction. The approach to evaluate protein level in overexpression or knockdown mutants is more costly. An antibody specific to the target protein is required to measure the expression using a Western blot. In these cases with an antibody available, lower expression of the protein would be observed in mutants overexpressing the miRNA and mutants with a knockdown of the target. This has been the case for conserved miR172 and its targeting

of the APETALA2 transcription factor in *Arabidopsis* (X. Chen, 2004) and the conserved miR156/157 family and its translational inhibition of the SPL3, a member of the SQUAMOSA PROMOTER BINDING PROTEIN transcription factor family (Gandikota et al., 2007).

1.5 Conserved MiRNAs

Many conserved miRNAs target families of transcription factors, injecting themselves into many growth and development networks. Table 3 shows the transcription factor families targeted by a selected few of the conserved miRNAs (Jones-Rhoades et al., 2006; B. Zhang, Pan, Cobb, & Anderson, 2006). Conserved miRNAs also target non-transcription factor families. Examples are miR162, miR168, miR390, and miR399. miR162 and miR168 are involved in a feedback mechanism to regulate miRNA biogenesis and function: miR162 targets DCL1 (Xie, Kasschau, & Carrington, 2003) and miR168 targets AGO1 (Rhoades et al., 2002). miR390 and miR173 are involved in the production of tasiRNAs from TAS loci (E. Allen et al., 2005). The mechanism by which miRNAs regulate siRNA biogenesis is shown in Figure 1 (B). miR390 targets the TAS3 loci whereas miR173 targets the TAS1 and TAS2 loci; however, miR173 has been annotated in miRBase exclusively in *Arabidopsis*. Phosphate homeostasis is regulated by miR399 through target mimicry (Franco-Zorrilla et al., 2007). In phosphate starvation, miR399 cleaves the PHO2 transcript, which when translated; the protein negatively regulates inorganic phosphate transport. In addition, miR399 targets a non-protein coding gene, IPS1 (INDUCED by PHOSPHATE STARVATION1) with a bulge at position 10 and 11. Therefore, IPS1 is not cleaved by miR399 but rather behaves as a target mimic to sequester miR399, allowing phosphate transport to resume. Conserved

miRNA families regulate many growth and development processes through regulating families of transcription factors in response to abiotic and biotic stress (Mallory & Vaucheret, 2006; B. Zhang, Pan, & Anderson, 2006). Including the specialized regulation of non-transcription factor targets and even non-protein coding transcripts is part of a larger regulatory network of miRNAs to ensure homeostasis and proper development throughout the plant life cycle. Table 3 summarizes nine conserved miRNA families with transcription factor targets and their function in plant development (Jones-Rhoades et al., 2006; B. Zhang, Pan, Cobb, et al., 2006). MiR172 is one of these conserved miRNA families that participates in plant development, especially in floral development.

1.6 MiR172 in the ABCDE Model of Floral Development

Understanding floral development began with cloning genes responsible for floral development and has since expanded to a more regulative investigation (Krizek & Fletcher, 2005). ChIP (chromatin immunoprecipitation) has allowed for experimental validation of the gene targets of transcription factors responsible for floral development and in addition, post-transcriptional regulation of these transcription factors has become of interest. Floral development is explained by the ABC model (Bowman, Smyth, & Meyerowitz, 1991), which explodes into complex regulatory networks of transcription factors and other regulatory elements such as miRNAs.

The basic model begins with the organization of floral organs into whorls. In wildtype flowers, sepals (or lemma and palea in grasses) make up the outermost whorl, petals (or lodicules in grasses) the second whorl, stamen the third, and carpels the innermost whorl. The outermost whorl is controlled by class A genes only, the second by

class A and B, the third by class B and C genes, and class C genes controlling the innermost whorl (Table 4). Homeotic mutants exhibiting the replacement of one floral organ type to another were observed which allowed the assignment of class gene functions. Table 4 explains the Bowman ABC model of flower development and the assignment of Arabidopsis genes in controlling tissue development in each whorl.

The authors recognized the over-simplification of the model. This basic model did not explain the number and position of the flower, mosaic organs, missing organs, and indeterminacy in particular mutants. An expansion on the original ABC model includes class E activity (Pelaz, Ditta, Baumann, Wisman, & Yanofsky, 2000), which is required for organ type specification and class D activity which controls ovule identity (Colombo et al., 1995). Figure 3 shows a comparison of eudicot model from Arabidopsis and the monocot model in the grasses (Krizek & Fletcher, 2005). The inclusion of D and E class genes addressed a few of the basic model complications but the spatiotemporal expression pattern of class A and C genes could only be predicted.

In particular, the antagonism of class A and class C genes was unclear. Class A gene AP2 (APETALA 2) represses class C gene AG (AGAMOUS) (Drews, Bowman, & Meyerowitz, 1991) as identified by ChIP-seq (Yant et al., 2010) but how can the global AP2 gene expression result in AG repression in exclusively the outer whorls. In the ag mutants, AP2 transcripts would accumulate if AG was the only regulator. The discovery that miR172 post-transcriptionally regulates AP2 suggested a clear answer to the question (Aukerman & Sakai, 2003; X. Chen, 2004). miR172 was discovered in the model plant species *Arabidopsis thaliana* in 2002 but was not characterized until 2004 when it was associated with a phenotype.

Overexpression of miR172 in *Arabidopsis* resulted in floral organ defects as well as incorrect timing in floral development. Sharing almost perfect complementarity to its predicted target, miR172 translationally represses APETALA2 and APETALA-like transcription factors. A reverse-genetics approach was taken by Chen in 2004 to identify AP2 as a target of miR172 through translational repression. He created transgenic mutant ap2 lines, overexpressing miR172 transgenic plants, and overexpressing AP2 transgenic mutants. Over-expressing miR172 transgenic plants showed reduced AP2 protein abundance but not in the ap2 mutant line. miR172 was not able to bind to the mutant ap2 to translationally repress the transcript, causing delayed flowering times (X. Chen, 2004). Expression analysis and immunoblotting of miR172 and two AP2-like genes in the eat-D mutant (overexpressing miR172) and the ap2-like gene mutants verified that miR172 regulates AP2-like genes primarily through translational repression but also through a cleavage mechanism (Aukerman & Sakai, 2003).

In the Arabidopsis model of plant flowering, high AP2 expression and low miR172 expression was observed in the outer two whorls and the reverse true for the inner whorls, suggesting miR172 represses translation of AP2 in the inner whorls (Wollmann, Mica, Todesco, Long, & Weigel, 2010). However, the low expression of miR172 in the outer whorls emerged as the new question. Just as ChIP-seq identified binding of AP2 to the second intron of AG (Yant et al., 2010), miR172b was also found as a target of AP2 in the presence of two co-repressors (Grigorova et al., 2011). This suggests a positive feedback loop where AP2 represses its target and regulator miR172 in the outer whorls.

Another feedback mechanism involving the regulation of miR172 can be expanded to the developmental transition of juvenile to adult phase. As described earlier, miR172 targets AP2-like genes and its spatial expression regulates floral organ identity. miR156 targets SQUAMOSA PROMOTER BINDING PROTEIN LIKE (SPL) transcription factors. The SBP box DNA binding domain characterizes this family of transcription factors. Each of the four clades of SPLs has different spatiotemporal expression patterns. The feedback mechanism of juvenile to adult phase change involves miR172 acting downstream of miR156 (G. Wu et al., 2009). miR156 represses SPL3/4/5 and SPL 9/10 clades (inducers of miR172 expression) in the vegetative phase. High levels of miR156 and AP2 characterize the juvenile/vegetative phase; low levels of SPL and miR172. In the reproductive phase characterized by high SPL and miR172 and low levels of miR156 and AP2, higher miR172 indirectly induces the transcription of SPL3/4/5 to overcome a threshold to initiate adult phase change (Jung, Seo, Kang, & Park, 2011). Through promoting the expression of miR172, miR172 down-regulates AP2 protein expression. The SPL transcription factors by which miR156 indirectly regulates miR172 have not been identified in monocots (Q. H. Zhu & Helliwell, 2011).

Forward and reverse genetics approaches have been used in elucidating the miR172/AP2-like interaction in the grasses. Studies in barley have identified phenotypic variation in barley cultivars missing an AP2 domain in an ERF/AP2 transcription factor (Taketa et al., 2008) and in a mutant miR172 barley line (Brown & Bregitzer, 2011). Similarly in wheat (Abdollahi, Kamiya, Kawaura, & Ogihara, 2012; Simons et al., 2006), rice (D. Y. Lee & An, 2012; Ren et al., 2013), and corn (Chuck, Meeley, & Hake, 2008; Chuck, Meeley, & Hake, 1998), lines exhibiting mutations in the miR172 or an AP2

domain transcript exhibit floral development abnormalities. Although no mutation studies have been carried out in another grass, oat, the developmental abnormality indeterminacy is also an observed phenotype in many hulless oat cultivars.

1.7 Oat

Oat (*Avena* ssp.) is a monocot and a member of the grass family Poaceae, which is composed of over 700 genera and 10,000 species (Gaut, 2002). This family is composed of grasses with a wide range of sequenced genomes and other sequence resources available.

Within this family is the monophyletic BEP clade, which includes economically important rice, wheat, barley, and oat (Gaut, 2002). The BEP clade is made of the Bambusoideae, Ehrhartoideae, and Pooideae subfamilies. The phylogenetic tree topology suggested by Gaut in Figure 4 is based on the Grass Phylogeny Working Group's 2001 studies but the tree shares the same topology as the maximum parsimony tree for rbcL and ndhF sequences with divergence times estimated by a non-parametric rate smoothing method (Sanderson, 1997). Estimated divergence times are supported by Ks estimation of duplicated genes in the rice lineage (Paterson, Bowers, & Chapman, 2004). It is estimated that the Poaceae family originated 77 million years ago but it was not until 50 million years ago that the BEP clade diverged from the Panicoideae family containing millet, sorghum, and maize. To identify genera more closely related to oat, the divergence times of genera within the BEP clade were calculated. Rice (Ehrhartoideae subfamily) diverged from Pooideae (oats, barley, and wheat) 46 million years ago. Therefore, oats are more closely related to wheat and barley than to rice. The Triticeae

tribe (barley and wheat) diverged from the Aveneae tribe 25 million years ago with barley and wheat diverging from each other 13 million years ago.

Aside from the evolutionary relationships, these grasses are very diverse in genome size, ploidy level, and chromosome number. Table 5 shows estimations of genome size of maize, rice, barley, wheat, and oat obtained from the Gramene database (Ware et al., 2002); however, the size of the oat genome is debatable. Personal communication with Eric Jackson and Ryan Brown (General Mills) suggests the oat genome to be closer in size to 13 Gb. Sequencing of the oat genome will reveal a more accurate estimation of the hexaploid genome size.

For the grasses discussed above, many have sequenced genomes: Erhartoideae subfamily (*Oryza sativa* (Yu et al., 2002) and Panicoideae subfamily (*Zea mays* (cultivar B73 (Schnable et al., 2009)), *Sorghum bicolor* (Paterson et al., 2009), and *Setaria italica* (Bennetzen et al., 2012)). The Pooideae subfamily includes *Brachypodium distachyon*, *Triticum aestivum*, *Hordeum vulgare*, and *Avena sativa*. Of these, genome sequencing of *Brachypodium* (International Brachypodium, 2010) and barley (International Barley Genome Sequencing et al., 2012) has been completed. MiRNA annotations are also skewed towards these fully annotated genomes (Table 6). There are many more miRNAs annotated in miRBase for the grasses with sequenced genomes with the exception of *Setaria italica*. In the release of the *Setaria* genome, 48 known miRNA families were identified in satisfying phylogenetic ancillary criterion, but they are not annotated in miRBase (Bennetzen et al., 2012). There are currently no annotations of miRNAs in *Avena sativa* despite its economic importance.

Oat has had great economic importance in America's history; however, with the replacement of the farm workhorse with machinery, oat acreage declined. Workhorses consumed millions of tons of oats each year. Corn and barley as metabolizable energy and soybean as the primary protein source have replaced the demand of oats in diets of ruminants and nonruminants. Great potential does exist as research validates the nutritive advantages of oat in human and animal diets.

1.7.1 Health Claims

The health claims of oats in the human diet are, in part, due to its properties as a "whole grain". The American Association of Cereal Chemists, Whole Grains Council, FDA, and the Whole Grains Task Force have defined "whole grain"; however, few are compatible with current Western food processing technologies. Processing whole grains such as wheat, rye, corn, rice, maize, sorghum, millet, and oat results in a great loss of nutritional qualities. The best-adapted definition of "whole grain" is that the cereal may have limited processing but must result in the preservation of the nutritional value of the whole seed (endosperm, germ, and bran) in approximately the same relative proportions as the natural occurring grain. This definition is appropriate for Westernized processing practices but the definition has not yet been accepted worldwide. The reason for maintaining the nutritional qualities of the whole seed is the spectrum of health benefits of whole grain cereals. Some of these qualities include the contribution of dietary fiber to the diet, improved carbohydrate tolerance hours after a low-glycemic index meal of whole grain ("second meal effect"), source of anti-carcinogenic compounds, and source of antioxidants to name a few (Fardet, 2010).

To further the health benefits of oat as a whole grain cereal, oat contains high levels of avenanthramides, saponins, and beta-glucan (Fardet, 2010). Fardet reviews the effects of these bioactive compounds in the human diet. Avenanthramides are polyphenols with antioxidant, anti-inflammatory, and anti-atherogenic properties through protecting LDL from oxidation. Oats synthesize two classes of saponins: avenacosides (sugars are bound to a steroid with a glycosidic bond) and avenacins (glycosidic bond linking sugars to a triterpenoid), which are both localized in the endosperm. Both of these classes of saponins contribute to stimulation of the immune system, lowering of cholesterol, and anti-carcinogenesis. β (1-3)(1-4)-glucan is found in high levels in both barley and oat, depending of cultivar and presence of a hull. The hull provides higher levels of insoluble β-glucan. As a whole, β-glucans have cholesterol-lowering effects and slow the rate of nutrient absorption in the gut. This suggests foods high in β-glucan such as oat and barley have a "stay full longer effect". In addition, in vitro studies have found that β-glucans prevent mutagens from damaging DNA, establishing an anti-carcinogenic effect.

1.8 Hulless Oat

Despite the health benefits of including oats in human diets, the presence of the oat hull lowers nutrient density because it is insoluble fiber indigestible by human and animals. Before it can be used in food, hulled oats must go through extensive processing to remove the hull from the nutritive part of the seed, the groat. A new market for hulless oats has emerged. Without a hull, oats use less storage space decreasing shipping costs and do not require the extensive processing steps, such as dehulling. As this market emerges for hulless oat in animal food for horses and pets, and in our food with a niche

market for celiac and gluten-allergic individuals, the hulless trait in oat must be better understood to improve breeding hulless oat to overcome its deficiencies (V. D. Burrows, 2011).

Hulled and hulless oats differ morphologically, which influences how each is processed. Hulled oats, also described as covered oats, have a determinate spikelet, indicated by two or three florets per spikelet. The hull is made of a thick lemma and palea that cover the groat. For food use, multiple passes of the seed through dehulling machinery are performed, to remove all the hulls and sort groats from hulls, broken groats, and trichomes. Trichomes are fine hairs on the groats that are removed through the dehulling process or subsequent cleaning. Dehulling results in cracked and broken seeds, which activates lipases, peroxidases, and lipoxygenase enzyme systems. Cleaning with high heat and moisture inactivates these systems, which avoids rancidity of damaged seeds and carry-over of rancidity to intact groats. In contrast to the hulled oat determinate spikelet, hulless oats are classified as indeterminate. Indeterminate spikelets are multiflorous, containing as few as two but as many as ten florets with the primary floret producing the largest seed, subsequent florets producing smaller and smaller seeds as position in the spikelet increases. Hulless oats, also known as naked oats, have a thin and papery lemma and palea, which do not adhere tightly to the groat and are discarded upon harvesting. With non-adhering hulls, expensive dehulling machinery is unnecessary (V. D. Burrows, 2011). Figure 5 compares the morphology of hulless and hulled oat.

1.8.1 Deficiencies of Huless Oat

Naked oats were present in the United States in the late 1800s, with descent from China; however, the vast deficiencies of naked oat slowed its commercial use (Stanton,

1955; Valentine, 1995). Some of these deficiencies include mixed hulled and hulless seeds due to incomplete expression of hullessness, trichomes covering the groat that become irritants to handlers, protruding embryos easily damaged during handling, kernel size variation, and low yield potential. Breeding useful traits into hulless was attempted after an international exchange of hulless oat genetic material, but improvements in other feed sources such as barley, corn, and soy created a difficult market for oat to dominate.

Trichomes are problematic to the handler and the oat groat. These tiny hairs are removed from the groat during threshing, becoming airborne. They can be inhaled, come in contact with the skin, and transferred from skin to the eyes. This causes intense respiratory irritation and an "itchiness" of the skin and eyes. Trichomes also have a negative effect on the groat. The hairs can develop an electrostatic charge and attract fungal spores, leading to disease. The number and coarseness of trichomes on both hulless and hulled oat groats varies; however, trichomes are protected in hulled oats so the irritation threat to oat handlers is minimal. Bald groats have few and fine trichomes, which is a preferred trait. Burrows identified a bald conditioning gene, Gt-1, which has been breed into hundreds of VAO lines in Ottawa, including AC Gehl and VAO-8 (V. D. Burrows, 2011; V. D. Burrows, McDiarmid, & Marder, 2001).

Another deficiency of hulless oat is the protruding embryo of the groat. The hull protects embryos of hulled cultivars so the potential of damage during threshing and handling is reduced. Threshing hulless oat groats requires lower combine cylinder speed and cushioning of the seed as it passes between combine concaves. Aside from damage during harvest, transport of groats from storage to cleaning using pneumatic systems can cause damage when hulless oat embryos are broken off as they collide with bends in the

piping. Once the embryo is damaged, rancidity sets in if the groat is not immediately treated. If the seed is to be planted, a damaged embryo results in no germination of the seed. Solutions to the protruding embryos of older hulless cultivars has been to breed for shorter groats with recessed embryos (V. D. Burrows, 2011).

Most hulless oat cultivars are semi-hulless: the hullessness trait is incompletely expressed. In these cultivars, seeds from higher-ranked florets of the multiflorous spikelets are hulled whereas the older primary and secondary florets are hulless (V. D. Burrows, 2011). This is not always true, as more hulled seeds have been observed when herbicide is applied at the wrong stage (Marshall, McDaniel, & Creggar, 1992) and when lower temperatures are experienced during flower differentiation (Lawes & Boland, 1974). In a comparison of three temperatures (25 °C, 20 °C, and 15 °C) with 16 hour photoperiod, complete expression of hullessness in eight hulless genotypes was observed at 25 °C. The same genotypes showed both hulled and hulless phenotypes at the two reduced temperatures. Furthermore, a QTL analysis using simple interval mapping of a hulless by hulled oat mapping population identified nine QTLs mapping to the N1 locus and by using N1 as a covariate, the study identified genomic regions potentially responsible for the critical quality traits such as yield and test weight (De Koeyer et al., 2004). In addition, the N1 locus was associated with QTLs through QTL-by-environment interaction tests and tests for epistasis.

The N1 locus has been identified as the “switch” gene in the most recent model to explain the hulless trait in oat. Jenkins and Hanson explain the hulless trait as incompletely dominant, controlled by four genes. This complex four-gene model explains hullessness in oat by interaction of a “switch” gene, N1, and doses of three

"modifier" genes (Jenkins & Hanson, 1976). A hulless and a hulled oat cultivar were crossed and progeny self-crossed. The model explained the segregation ratio of the hulless trait in this F2 population. Hulled plants had panicles that produced only fully husked seeds. "Hulless" plants had panicles that produced all naked seeds, or were mosaic and produced both husked and naked seeds. This model has not been updated to explain mosaic plants. The hulless trait in oat is complex due to the epistasis of N1 with three modifier genes (see Table 7).

1.8.2 Markets for Hulless Oat

Despite the deficiencies of hulless oats, a unique food market does exist. Animal feed ingredients are used as metabolizable energy and protein. The crop primarily used for metabolizable energy is corn, but grasses such as sorghum, wheat, and barley may be substituted (V. D. Burrows, 2011). Protein sources include soybean, canola, and alfalfa. Hulless oat is a good substitute in cattle, horse, pig, and poultry feed due to its nutritional profile. As a metabolizable energy substitute, hulless oat is comparable to corn, a much better substitute than barley and wheat. The amino acid profile of hulless oat is better balanced than corn, wheat, and barley (Welch, 1995), which reduces the amount of amino acids needed as supplements in these animal diets. Positive effects of a hulless oat diet have been found in hens, swine, beef cattle, rabbits, and horses. Hens fed hulless oats produced larger eggs with a better nutritional value due to increased levels of linoleic acid in oat oil (Cave, Hamilton, & Burrows, 1989). The Avian Feed Efficiency from Naked Oats Project found that these eggs also required more time to spoil, which was attributed to the high level of antioxidants in hulless oat. Difference in the taste of eggs and quality of shell remained insignificant as found by the Agriculture and Agri-Food

Canada. However, hulless oat diets have not shown as promising results when related to weight gain in chickens and turkeys. This may be due to high levels of beta glucan that interact with the gut micro biome to reduce nutrient absorption in these poultry (Broz & Frigg, 1986).

Hulless oats with supplemented lysine fed to swine showed better weight gains than swine fed a control diet composed of corn and soybean (Friend et al., 1989). Aside from weight gains, a sensory panel concluded that swine fed an all-hulless oat diet produced meat with better flavor intensity.

Similar effects on weight gain were found in finishing steers in studies at North Dakota State University (Anderson, Bauer, Canton, Berg, & Dhuyvetter, 1997) and in rabbits (Drevjany & Drevjany, 1988). In contrast to weight gain requirement for poultry, swine, and steers, racehorses are judged on their performance. Horse diets include hulled oats. Although the hulls are indigestible, the horse requires high levels of insoluble fiber to keep an efficient digestive system. With hay providing cheaper insoluble fiber, the need for the hull is eliminated. The company Equavena claims their hulless oats contain twenty to thirty percent more calories per pound than hulled oats and in turn, trainers/breeders are instructed to feed horses one-fourth less ("Equavena Hulless Oats,"). The company has also claimed improved performance of racehorses fed Equavena hulless oats in 2006 and 2007 although no further study to establish statistical significance has been published (V. D. Burrows, 2011).

While most of animal feed relies on corn ingredients, the human diet relies more on rice and wheat as staples. Oat has greater health benefits than both grains. In foods such as plain rice, soups, casseroles, and desserts, oat can be used as a rice substitute.

Aside from containing almost twice as much protein as rice (Fric, Gabrovska, & Nevoral, 2011), rice contains high levels of a miRNA that targets the mRNA of a human liver enzyme (L. Zhang et al., 2012). This enzyme is low-density lipoprotein receptor adapter protein 1 (LDLRAP1), which is responsible for removal of LDL from plasma. Through targeting this enzyme, rice miR168 binds to exon 4 and reduces LDLRAP1 protein level in vitro. Lower levels of this protein reduce liver cells' endocytosis of LDL (Wilund et al., 2002; Zuliani et al., 1999). Lower levels of miR168 in oat may be an added health benefit of oat and make hulless oat a good substitute in rice dishes.

Hulless oat may also be used as "pure" oats in a gluten free diet. A gluten-free diet is attractive for health beliefs and for livelihood necessity. Avoidance of gluten through a gluten-free diet is the treatment for the near 1% of the world's population afflicted with Celiac disease ("National Institutes of Health Consensus Development Conference Statement on Celiac Disease, June 28-30, 2004," 2005). Celiac disease is an autoimmune disorder where a 33-mer polypeptide derivative of gluten causes a reaction that damages the intestinal villi (Shan et al., 2002) leading to improper absorption of food nutrients. In children, celiac disease can manifest as diarrhea and growth retardation due to vitamin, iron, and calcium deficiencies. Severe manifestations include digestive bleeding, edema, infertility, neurological disorders, and osteoporosis (Garcia-Manzanares & Lucendo, 2011). Treatment is a gluten-free diet. Identifying alternative cereal sources to replace gluten-containing wheat, barley, and rye is a challenge. Current alternatives are limited to rice and corn, but inclusion of oat in this specialized diet has been controversial because oats contain gluten. Clustered proline and glutamine residues of gluten proteins become T-cell epitopes. Oat prolamins make up only 13% of all protein

in oat (wheat prolamins make up 35% of total protein) and oat prolamins have half the number of proline residues in comparison to wheat prolamin. However, localization of these amino acid residues can resemble T-cell epitopes and potentially trigger an immune response in some Celiac patients (Vader et al., 2003). Studies by the NIH in 2004 found that 50 grams of non-contaminated oats per day can be tolerated by persons afflicted with celiac disease; however, this excludes the 5% of celiac patients who also suffer from an oat intolerance ("National Institutes of Health Consensus Development Conference Statement on Celiac Disease, June 28-30, 2004," 2005). Within these levels, the problem with using oat as a cereal alternative in gluten-free diets is the contamination of oat samples with wheat or barley. This contamination arises from barley and wheat contamination in the field or from storage facilities that have stored wheat or barley in previous growing seasons (personal communication Dr. Ryan Brown, General Mills). For persons that suffer from Celiac disease, consuming an oat product that was processed in the same facility that also processes wheat and barley can trigger the Celiac immune response. For oats to be used in a gluten free diet, “pure” handling facilities and strict inspection specifications must be met to ensure minimal presence of high gluten cereals in hulless oat processing. Due to hulless oats not requiring the intensive processing to remove the hull, separate “pure” handling facilities to process all hulless oats would not be illogical. This would allow hulless oat to thrive in a niche market for gluten-free individuals.

1.9 Research Directions

The potential of hulless oat entering into this market is hindered by two primary deficiencies: indeterminacy and incomplete expression. Through investigating the role of

miR172 in conditioning the hulless trait in oat, breeding determinate and completely hulless cultivars will allow that potential to be reached. With the long-term goal of shedding light on miRNAs' effect on plant genomes including both identification of miRNAs and their respective target, understanding miR172 is only one example of a miRNA's effect on oat genotype and phenotype. To understand the complex regulatory network of miRNAs in oat, miRNA discovery provides the foundation for complete annotation of miRNAs to include their respective targets. Integrating high-throughput next-generation sequencing and computational strategies is involved in each of the three objectives of this research. The objectives of this research are to (1) identify and annotate miRNAs present in developing oat flowers, (2) discover miRNAs in developing oat flowers in oat cultivars of different ploidy levels and identify miRNAs that originate from a particular subgenome of oat, and (3) investigate the role of miR172 in conditioning the hulless trait in oat.

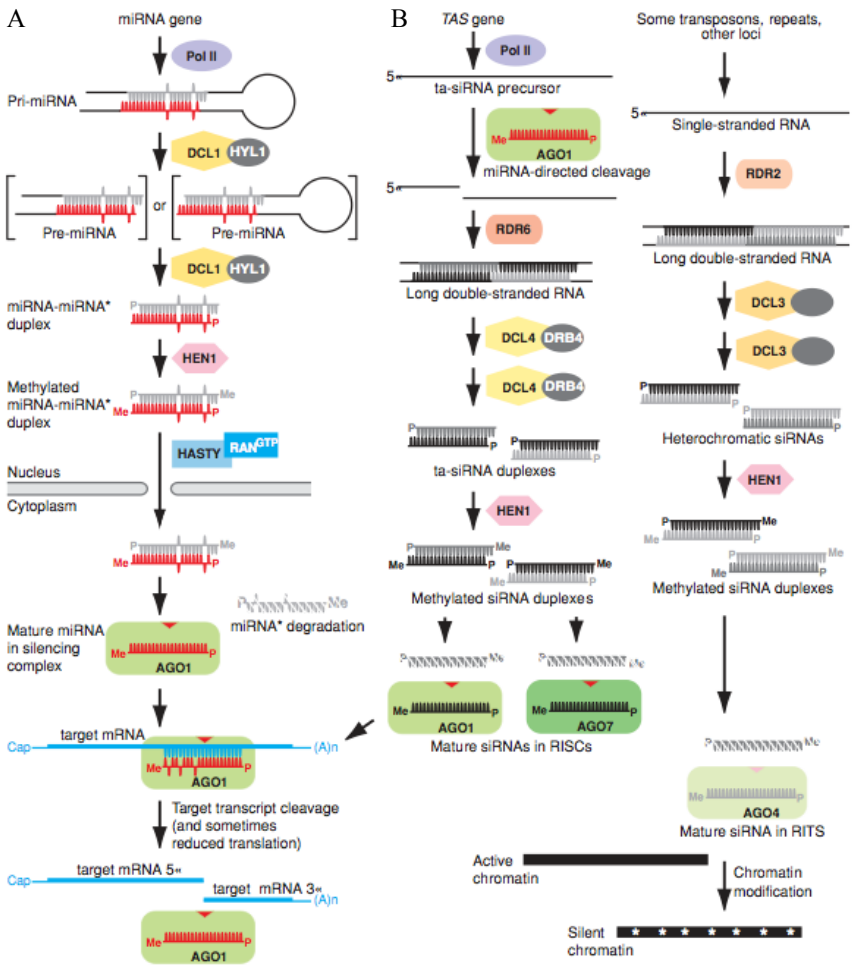


Figure 1: Processing of plant miRNAs and siRNAs from (Jones-Rhoades et al., 2006). (A) MiRNAs are processed in the nucleus and perform target regulation in the cytoplasm. (B) MiRNAs can also direct siRNA biogenesis through cleavage of a TAS gene. Other loci can produce siRNAs and can direct chromatin modification.

Table 1: Differences between miRNA and siRNAs. The main differences between miRNAs and endogenous siRNAs are the DICER homologs involved in the RNA processing, the ARGONAUTE protein as part of the RISC, the source of the trigger, and the function.

Classification	Length (nt)	Dicer	Argonaute	Trigger	Function
miRNA	20-24	DCL1	AGO1/10	RNAPol II transcription of pri-miRNAs	Growth and development, abiotic/biotic stress response
endogenous siRNA					
casiRNA	24	DCL3	AGO4	Transposons, repeats	Chromatin modification
tasiRNA	21	DCL4	AGO2/7	miRNA cleaved target from TAS loci	ARF- regulated genes, unknown
natsiRNA	21	DCL2/DCL1	AGO1/4	Stress induced anti-sense transcripts	Stress-response genes

Table 2: 20 Conserved miRNA families.

Conserved miRNA families
miR156
miR157
miR159/miR319
miR160
miR162
miR164
miR165/miR166
miR167
miR168
miR169
miR170/miR171
miR172
miR390
miR393
miR394
miR395
miR396
miR397
miR398
miR399

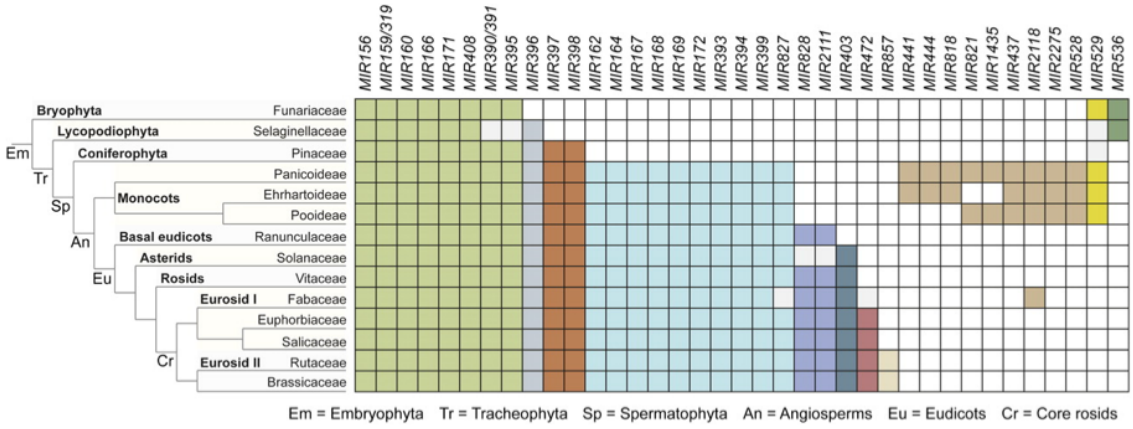


Figure 2: Phylogenetic conservation of plant miRNAs from (Cuperus et al., 2011). Most known miRNAs are conserved across all plant lineages. Non-conserved miRNAs are lineage specific. Novel miRNAs are not shown here due to their species specificity.

Table 3: Selected conserved miRNA families and their respective transcription factor family Ttrgets. Conserved miRNAs target families of transcription factors to regulate plant growth and development.

Conserved miRNA Family	Target Transcription Factor Family	Function
miR156	SBP (SQUAMOSA PROMOTER BINDING PROTEIN)	Vegetative growth
miR159/miR319	MYB, TCP	Anther and floral development
miR160	ARF (AUXIN RESPONSE FACTOR)	Hormone signal transduction
miR164	NAM/CUC	Floral and shoot development
miR165/miR166	HD-ZIP	Vascular development, meristem formation
miR167	ARF (AUXIN RESPONSE FACTOR)	Hormone signal transduction
miR170/miR171	SCL (SCARECROW-LIKE)	Flower development
miR172	AP2 (APETALA2)	Flower development
miR396	GRF (GROWTH REGULATING FACTOR)	Growth and development

Table 4: Gene and gene classes assigned to the original ABC model of floral development.

Whorl	Tissue	Gene Class Activity	Arabidopsis Gene Name
1	Sepals (grasses: Lemma and Palea)	A	AP2
2	Petals (grasses: Lodicules)	A&B	AP3, PI, AP2
3	Stamen	B&C	AP3, PI, AG
4	Carpels	C	AG

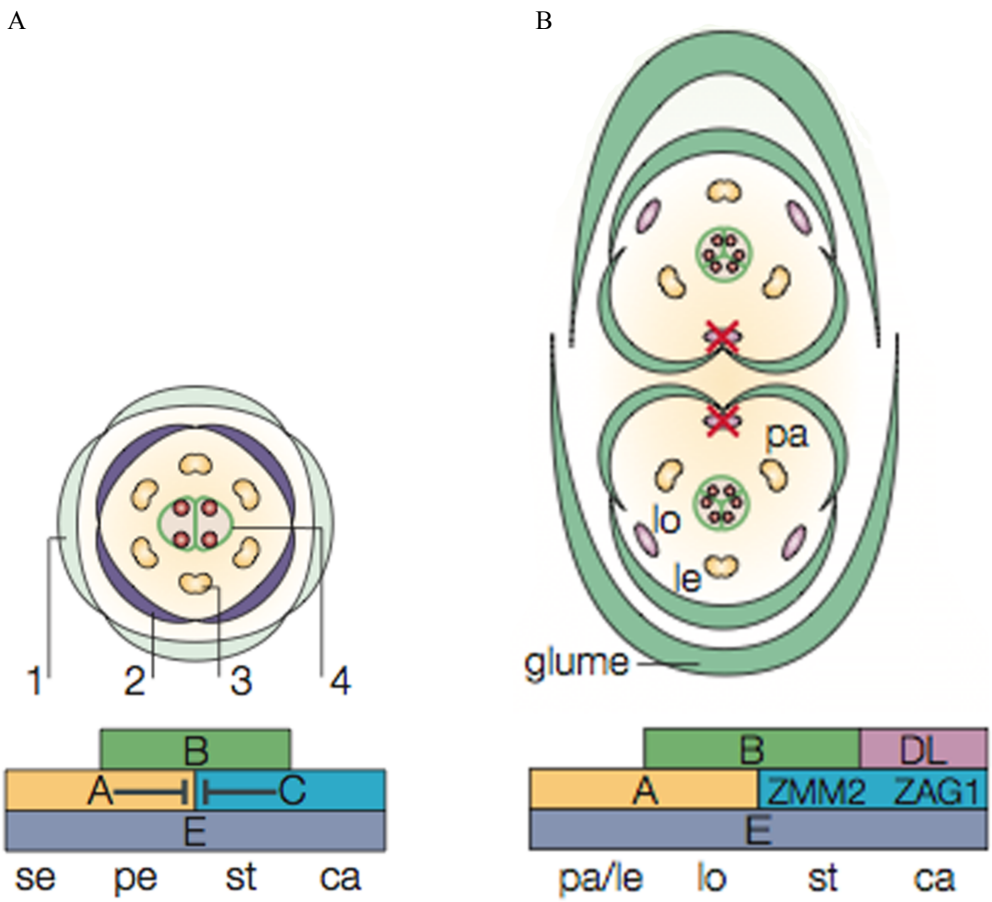


Figure 3: Models of floral development from (Krizek & Fletcher, 2005). (A) Shows the eudicot model of floral development and (B) shows floral development in the monocot grasses. Sepals (palea/lemma in grasses) are controlled by Class A and E gene functions. Petals (lodicules in grasses) are controlled by Class A, B, and E genes. Class B, C, and E genes control stamen tissue identity and development. Class E and C genes control carpel development. In grasses, carpel specification is also controlled by another factor, YABBY gene DL.

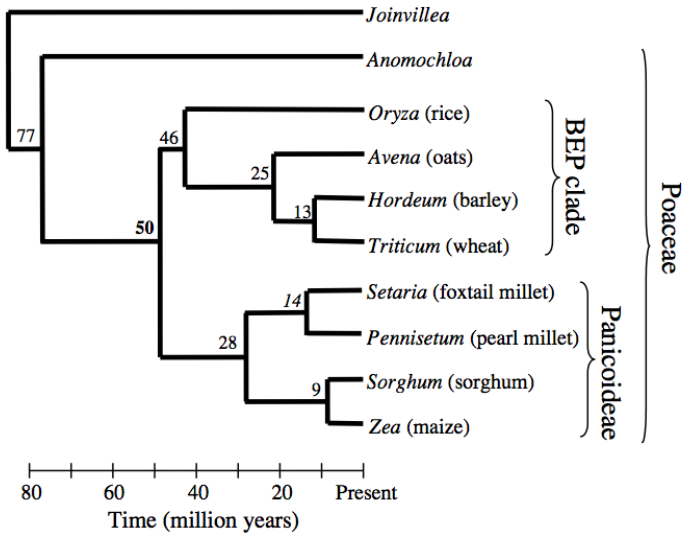


Figure 4: Evolutionary relationships of the grasses (Gaut, 2002). Oats are most closely related to barley and wheat, having diverged 25 million years ago.

Table 5: Relative comparisons of ploidy and genome sizes of common grasses. Oat and wheat have the largest genomes; however, these estimations will be refined upon completion of oat and wheat genome sequencing projects.

Grass	Ploidy	Genome Size Estimation
Maize	2n=2x=20	2.5 Gb
Rice	2n=2x=24	.43 Gb
Barley	2n=2x=14	5 Gb
Wheat	2n=6x=42	16.5 Gb
Oat	2n=6x=42	11.3 Gb

Table 6: Annotation of miRNAs in miRBase v.20 across grass species. Arabidopsis is not a grass; however, it is included for comparison purposes as a model plant.

Species	miRNA Precursors	Mature miRNAs
*Arabidopsis thaliana**	298	337
Aegilops tauschii	2	2
*Brachypodium distachyon**	258	464
Elaeis guneensis	6	6
Festuca arundinacea	15	15
Hordeum vulgare (barley)*	67	69
Oryza sativa (rice)*	592	713
Sorghum bicolor (sorghum)*	205	241
Saccharum officinarum	16	16
Saccharum ssp.	19	20
Triticum aestivum (wheat)	42	42
Triticum turgidum	1	1
Zea mays (corn)+	172	321
Setaria italica (foxtail millet)*	48 known families	
***Avena sativa* (oat)**	**0**	**0**

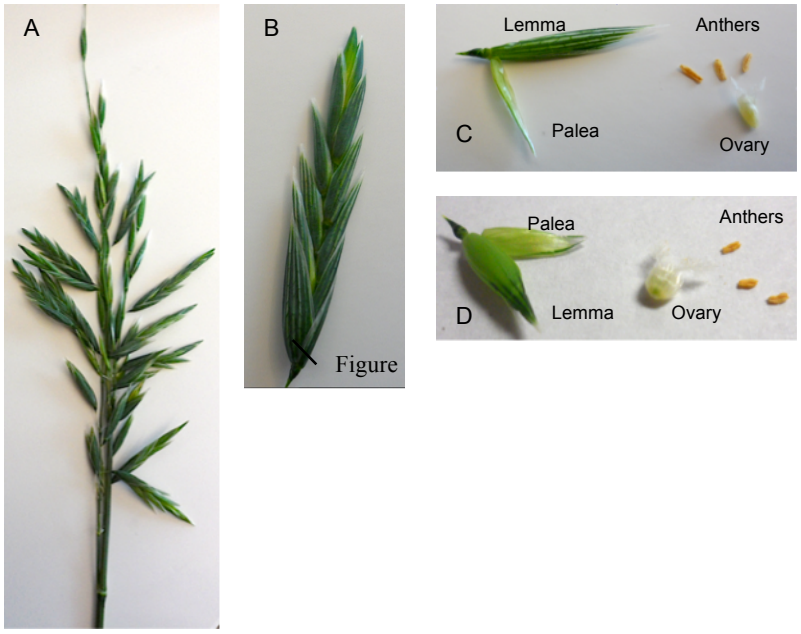


Figure 5: Morphology of oat. The panicle (A) is the branch to which spikelets are attached. A spikelet (B) is composed of individual flowers, each of which matures into a seed. Flowers are protected by sterile bracts, glumes. The flower can be hulless (C) with a papery lemma and palea or hulled (D) with a hardened lemma and palea. The panicle and spikelet (A, B) are from a hulless oat cultivar.

Table 7: Summary of 1976 model of hulless oat. The hulless trait is complex, regulated by a dominant "switch" gene and a dosage effect of three modifier genes.

Phenotype	Switch Gene	Modifier Gene dosage	Genotype
naked	homozygous dominant	any	NN
	heterozygous	4	Nn N_1N_1 N_2N_2 n_3n_3
		3	Nn N_1N_1 N_2n_2 n_3n_3
		3	Nn N_1n_1 N_2N_2 n_3n_3
		3	Nn N_1n_1 N_2n_2 N_3n_3
		3	Nn N_1N_1 n_2n_2 N_3n_3
		3	Nn n_1n_1 N_2N_2 N_3n_3
		2	Nn N_1n_1 N_2n_2 n_3n_3
		2	Nn N_1N_1 n_2n_2 n_3n_3
		2	Nn n_1n_1 N_2N_2 n_3n_3
hulled	homozygous recessive	any	nn
	heterozygous	any of N1 and N2	Nn N_3N_3
		2	Nn n_1n_1 N_2n_2 N_3n_3
		2	Nn N_1n_1 n_2n_2 N_3n_3
		1	Nn n_1n_1 n_2n_2 N_3n_3

CHAPTER 2: ANNOTATION OF MIRNAS IN HEXAPLOID OAT

2.1 Introduction

2.1.1 Classification and Evolution of MiRNAs

The role of miRNAs in post-transcriptional regulation in plants is only beginning to be understood. As newly identified small non-coding RNAs with regulatory roles in diverse pathways, the structure and origin of miRNA genes is in question. Across and within plant species, mature miRNA sequences with high sequence similarity are grouped into the same family. MiRNA families are further classified on the extent of phylogenetic conservation across plant lineages. MiRNAs that are conserved across plant lineages are classified as "conserved," conserved across particular lineages "non-conserved" or "semi-conserved," and species-specific "novel." There are approximately 20 conserved plant miRNA families (Table 1). Found to be conserved in *Arabidopsis*, rice, and *Populus*, these 20 families have become a standard of the "conserved" classification (Jones-Rhoades et al., 2006). As more non-conserved miRNAs are discovered in various plant lineages, the number of conserved miRNA classifications number will increase.

Within one species, one miRNA family may have many members. Some of the members have identical sequences but originate from distinct precursor genes, which suggest different evolution rates and selection pressures between paralogous miRNA

genes of the same family (Axtell & Bowman, 2008; B. Zhang, Pan, Cannon, Cobb, & Anderson, 2006). Slight variations in the sequences of miRNA members of the same family impact the specificity of the miRNA to its respective target leading to sub-functionalization of the miRNAs of the same family (Palatnik et al., 2007).As is true with miRNA families within species, miRNA families shared among species may vary in mature miRNA sequence. To maintain a regulatory network such as floral development, the majority of miRNAs in conserved families, despite minimal mature miRNA sequence variations, have homologous targets in all species the miRNA is present. The conserved target sites ensure regulation of the mRNA target in plants (Axtell & Bowman, 2008; Jones-Rhoades et al., 2006; Shabalina & Koonin, 2008). A number of the conserved target sites are those of transcription factors (Table 1). It can be inferred that the conservation of miRNA families and the target site conservation across plant lineages is crucial for fundamental processes traversing plant lineages.

The evolutionary model of how these families of sequences have evolved is not simple. As most plants are polyploids and have complex evolutionary histories, if a miRNA was present before polyploidization, duplicated genes would be produced. Aside from whole genome duplication events, other evolutionary forces such as segmental duplications, chromosomal rearrangements gene loss, and acquisition of new promoters could account for the random distribution of paralogous miRNA genes in plant genomes (Axtell & Bowman, 2008; Jones-Rhoades et al., 2006; A. Li & Mao, 2007). It has also been suggested some non-conserved miRNA families originated through inverted duplications of the target genes (E. Allen et al., 2004; Rajagopalan et al., 2006). The miRNA families that exist in plant genomes today have had a convoluted evolution are a

result of genomic rearrangements, co-evolution of miRNAs and their targets, and the retention of miRNA genes that give a selective advantage. Identifying the miRNAs in species whose miRNAome has never been studied is necessary to understand the complexities of plant regulatory mechanisms by miRNAs and the divergence of miRNA families.

2.1.2 Small RNA Sequencing

From the discovery of miRNAs by positional cloning (R. C. Lee et al., 1993) to microarray technology (Krichevsky, King, Donahue, Khrapko, & Kosik, 2003; Nelson et al., 2004), high throughput sequencing of small RNA libraries (Hafner et al., 2008) has become the prevalent method to create an annotated set of miRNA sequences for a species. One of the biggest pitfalls of such high throughput technologies is the inherent sequencing of RNA species from a size-selected sample (Ambros, 2003). Other non-coding RNAs such as rRNAs, snoRNAs, tRNAs, as well as degraded mRNAs are also sequenced. Sequence-based searches can filter out many of these sequences; however, the siRNAs (small interfering RNAs) are indistinguishable from miRNAs in a small RNA sequencing library. It is crucial to discriminate the two to identify miRNAs and their corresponding functions.

2.1.3 MiRNA Nuclear Processing

At the core of post-transcriptional gene regulators by miRNAs and siRNAs is the RNA silencing machinery: DICERS and ARGONAUTES. Although miRNAs and siRNAs appear indistinguishable in a small RNA sequencing dataset, the respective noncoding RNA precursor is processed differently in the nucleus. Both miRNA and endogenous siRNA genes traditionally originate from the intergenic regions with

exceptions originating from introns and coding regions (Llave, 2002; Rajagopalan et al., 2006; X. Zhang et al., 2012). RNA Polymerase II transcribes plant miRNA genes whereas RNA Polymerase IV transcribes endogenous siRNA genes. RDP (RNA-dependent RNA polymerases) generate the siRNA transcript's complement strand, forming a long double stranded RNA transcript, which is cleaved by a DCL (DICER endoribonuclease) into 21-24 nucleotide siRNA duplexes. Multiple duplexes can arise from the long ds-RNA. The DCL homologue and the RDP are specific to the subclass of siRNA (Ghildiyal & Zamore, 2009; Mallory & Vaucheret, 2006). RNA-dependent polymerase is not involved in miRNA processing.

A different yet complimentary mechanism for processing exists for miRNAs. Once transcribed, the single-stranded pri-miRNA transcript folds back on itself, forming the defining secondary structure- the stem-hairpin loop (Naqvi et al., 2012). DCL1 and associated proteins (SERRATE, a zinc-finger protein; HYL1 (HYPONASTIC LEAVES 1), an RNA-binding protein; and the nuclear cap-binding complex) recognize the hairpin loop structure and cleave the hairpin in two successions, forming a 21-nucleotide duplex with a 5' monophosphate and a 2-nucleotide overhang (Czech & Hannon, 2011; Han et al., 2003). Only one duplex is produced from the pre-miRNA. The miRNA star sequence (miR*) is the complement of the mature miRNA sequence and in the duplex. After methylation, siRNA and miRNA duplexes are transported into the cytoplasm for association with an ARGONAUTE protein of the RISC (RNA Induced Silencing Complex) (Brodersen & Voinnet, 2006; Ghildiyal & Zamore, 2009).

2.1.4 MiRNA Duplexes and the RISC

Once in the cytoplasm, the small RNA duplexes associate with an AGO (ARGONAUTE). Not all of the functions of the 10 Arabidopsis AGO homologues have been identified; however, AGO1 is known to be primarily involved in the miRNA pathway (Vaucheret, 2008). The miRNA duplex is unwound upon entry into the RISC and the miR* passenger strand is released. The mature miRNA is used as a guide strand that binds with high complementarity, three or fewer mismatches, to a mRNA transcript in the coding region, 3' untranslated region, or the 5' untranslated region (Jones-Rhoades & Bartel, 2004; Rhoades et al., 2002; Sunkar et al., 2005). The miR* can be used as a guide strand in a different RISC (Devers, Branscheid, May, & Krajinski, 2011; X. Zhang et al., 2011). Once cleaved by AGO1, the mRNA fragments are released for degradation by the exosome or another exonuclease (Souret et al., 2004). To maintain homeostatic miRNA levels, single-stranded miRNAs are degraded by SDNs (SMALL RNA DEGRADING NUCLEASE), which have 3' to 5' exonucleolytic activity (Ramachandran & Chen, 2008). The miR* may be degraded in a similar manner.

miRNAs have a second mode of target repression, also utilizing AGO1. Translational repression of target mRNA results in reduced expression at the protein level but only a small effect on the level of the target mRNA. The small effect is attributed to a crossover in the pathway of translational inhibition and target cleavage (Aukerman & Sakai, 2003). Once the passenger strand, or miR*, is released after the unwinding of the miRNA duplex in the RISC, translational inhibition takes place at the endoplasmic reticulum (S. Li et al., 2013). Although the involved subcellular compartment is known,

the mechanism of translational repression is still unknown (Huntzinger & Izaurralde, 2011).

It was originally thought miRNAs act in trans to regulate developmental pathways whereas siRNAs were involved in antiviral response. Both classes of small RNAs are more diverse. MiRNAs regulate their targets in trans by translational repression or cleavage mechanisms. Subclasses of plant siRNAs can act in cis or in trans to regulate their targets via mRNA cleavage or DNA-methylation (Carthew & Sontheimer, 2009; Mallory & Vaucheret, 2006; X. Zhang et al., 2012). The diverse classes of siRNAs can regulate gene expression under biotic and abiotic stress, just as miRNAs (Borsani, Zhu, Versiues, Sunkar, & Zhu, 2005; Katiyar-Agarwai, 2006). In addition to biotic and abiotic stress response, miRNAs have essential roles in regulating plant growth and development, hormone response, redox, metabolism, and signaling (Dietz, Vogel, & Viehhauser, 2010). High-throughput sequencing of small RNA libraries enables identification of both miRNAs and siRNAs to illustrate the complex regulatory mechanisms in plants. However, the diversity of siRNA subclasses generates a more difficult annotation. This study focuses on the discovery and annotation of miRNAs.

2.1.5 MiRNA Annotation

The annotation of miRNAs in plants relies on one primary criterion: the excision of a ~21-nucleotide long miRNA/miR* duplex from the stem of the hairpin precursor. In turn, the miRNA/miR* duplex has a 2-nucleotide 3' overhang with four or fewer mismatches between the complementary strands and one or fewer 1-2 nucleotide asymmetric bulges (Meyers et al., 2008). The excision criterion excludes endogenous siRNAs because they do not arise from a single-stranded RNA transcript that folds into a

hairpin structure. In the high-throughput sequencing of small RNA libraries, this criteria can be satisfied by the presence of both a miRNA and miR* sequence mapping to a putative miRNA precursor hairpin, predicted from the genome. However, miR* sequences are degraded quickly in the cytoplasm and appear in low read counts, if at all, in small RNA sequencing libraries. In its absence, candidate miRNA sequence presence in multiple small RNA libraries from the same species and/or same stage is necessary. These are the fundamental minimal criteria for miRNA annotation from high-throughput sequencing of small RNA libraries. The ancillary criteria of conservation of predicted miRNAs increases the confidence of miRNA annotation from small RNA libraries. Despite the presence of a small RNA sequence that is phylogenetically conserved across lineages, the corresponding hairpin must also be conserved. In the case the conserved miRNA and hairpin fail the initial biogenesis criterion, phylogenetic conservation of both the miRNA and hairpin can rescue the annotation. Other ancillary criteria relating to targets and DCL1 dependence are not applicable with the data presented in this study.

2.1.6 MiRNA Annotation Tools

For annotation of a miRNA from a high-throughput small RNA sequencing library, the biogenesis annotation criteria must be met. After pre-processing of the raw fastq files (quality, adapter-trimming, length, duplicate collapse), small RNA sequences are aligned to a reference genome, or in its absence, other genomic sequences available such as ESTs and transcriptome. Candidate precursors are excised and the integrity of the hairpin secondary structure as well as location of mapped reads is evaluated according to the annotation criteria (Meyers et al., 2008). These steps are prevalent in most tools used for evaluation of miRNAs in deep sequencing small RNA datasets.

Although there exist a common set of steps to identify miRNAs that pass biogenesis criteria for annotation, the algorithms underlying the core steps of alignment and RNA folding differ amongst available tools.

For alignment of small RNA sequences to available sequencing resources, common algorithms include SOAP2 (R. Li et al., 2009), Smith-Waterman (Gotoh, 1982; Smith & Waterman, 1981), BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990), Burrows-Wheeler Alignment (M. Burrows & Wheeler, 1994; H. Li & Durbin, 2009), and Bowtie (Langmead, Trapnell, Pop, & Salzberg, 2009). Mfold (Zuker, 1989) and RNAfold (McCaskill, 1990; Zuker & Stiegler, 1981) are the most acceptable secondary structure prediction methods. Due to the diverse underlying algorithms, detection rates of miRNAs in a deep sequencing dataset differ for each tool but also can vary with different versions of the same software. In addition, even miRNA detection rates of software tools using the same alignment algorithms may differ due to parameter sets to drive the algorithm (Williamson et al., 2013).

The field lacks a standard and many tools emerge for miRNA discovery to accommodate specific requirements and goals. For example, a tool for novel miRNA discovery differs from a tool developed for differential expression of particular miRNAs. The target user also may direct the accessibility and design of the tool. For users unfamiliar with command-line implementation, a web-based server or a GUI is a user-friendly design with more options for visualization of results. Table 2 lists three of the tools discussed here, the underlying algorithms, and utility in detecting and predicting miRNAs from small RNA sequencing datasets.

MiRNA annotation in species without a fully sequenced genome is challenging due to the lack of existing software and pipelines that incorporate sequence and structure-based approaches to effectively predict miRNAs and reduce the number of false positive miRNAs for validation. In addition, the utility of many tools is limited when a reference genome is absent. DSAP (Huang et al., 2010) and miRExpress (W. C. Wang et al., 2009) are tools designed for use in the absence of a reference genome but these tools cannot predict novel miRNAs. In a performance comparison of eight common miRNA prediction tools on three small RNA datasets from *Gallus gallus*, *Caenorhabditis elegans*, and *Homo sapiens* (Y. Li et al., 2012), DSAP and miRExpress had the highest sensitivity because these tools construct profiles by aligning the small RNA sequences to known miRNAs in miRBase (Griffiths-Jones et al., 2006; Griffiths-Jones et al., 2008; Kozomara & Griffiths-Jones, 2014). However, in assessing accuracy as the performance metric, DSAP performed poorly due to the large number of false positive predictions. Evaluating the performance of these tools in predicting novel miRNAs becomes more difficult. miRExpress and DSAP cannot predict novel miRNAs because they lack the utility to incorporate secondary structure folding into precursor identification. Novel miRNA prediction requires a reference genome or the use of available sequencing resources such as cDNA, ESTs, and draft genome assemblies to identify miRNA hairpins. A well-annotated genome can exclude coding regions, introns, non-coding RNA genes, and repetitive elements that are not canonical sources of miRNA precursors. Without a reference genome, annotation and sequencing coverage pose two challenges. Lacking filters for non-canonical miRNA precursor regions may increase the number of

false positive miRNA predictions as well as computational time. Low coverage transcriptomes result in unmappable miRNAs that are excluded from further predictions.

Six tools with novel miRNA prediction capabilities with the use of a reference genome were also compared (Y. Li et al., 2012). Novel miRNAs were categorized as true miRNAs if they had a precursor identified by multiple tools. In predicting novel miRNAs, MIReNA (Mathelier & Carbone, 2010) was recommended as first choice for novel miRNA prediction in *C. elegans* and *H. sapiens*; mireap (http://sourceforge.net/projects/mireap/) was recommended for the *G. gallus*. Of all 8 tools, miRDeep (Friedlander et al., 2008) was recommended as second choice for all three datasets, reinforcing the diverse utility of miRDeep.

2.1.7 miRDeep and miRDeep-P for MiRNA Annotation

Instead of proposing a new standard for miRNA annotation in the absence of a reference genome, we utilize an extension of the existing tool miRDeep for miRNA annotation in a species without a reference genome. To avoid a large number of false positive miRNA candidate hairpins, which need to be evaluated with further experiments, the miRDeep algorithm is based on a probabilistic scoring method of hairpins. In a real miRNA precursor, the algorithm assumes a biogenesis model: one or more sequences from the small RNA library will map to the mature miRNA, miR*, or the loop of the hairpin (Friedlander et al., 2008). The scoring algorithm is a Bayesian probabilistic scoring of the hairpin structure and signature sequence. The score is the log-odds probability of a genuine miRNA hairpin compared to the probability that the hairpin is background. In general, score increases when more small RNA reads map to the miR*

sequence predicted from the candidate precursor and when the hairpin is more energetically stable.

Despite the algorithm being initially designed for species with reference genomes and tested on *C. elegans*, human and dog, parameter estimation for plants has been established using the miRDeep probabilistic model (Thakur et al., 2011). This optimization is needed due to the difference in hairpin lengths, minimum free energy (MFE) range, and the almost complete sequence complementarity between the mature miRNA and miR* in the hairpin. Similar plant parameters are applied in an extended utility of the miRDeep algorithm: miRDeep-P, which uses a plant parameterized miRDeep algorithm, but also allows use of cDNA for precursor search (X. Yang & Li, 2011). The primary difference between miRDeep and miRDeep-P is a modified scoring of the MFE due to the longer length of miRNA precursors (Reinhart, Weinstein, Rhoades, Bartel, & Bartel, 2002; Thakur et al., 2011). In miRDeep, the longer length results in higher MFE scores, which inflates the output score and can cause the core miRDeep algorithm to collapse (X. Yang & Li, 2011). Setting a maximum score using a Gumbel distribution of the MFE as pre-miRNA length increases alleviates the potential inflated score. Minor variations different from the original miRDeep implementation include the use of Bowtie-1.0.0 for small RNA sequence alignment and allowing for a user specified precursor length. A filter step at the end filters for plant-specific criteria as reported in Meyers, et al. 2008. These criteria include a 2nt 3' overhang between the miRNA/miR* duplex, fewer than four mismatches between the miR* and miRNA sequence, and only one asymmetrical bulge of one or two bases is allowed between the miRNA/miR* sequence (Meyers et al., 2008).

miRDeep-P was trained on the Arabidopsis known 199 miRNA precursor genes (X. Yang, Zhang, & Li, 2011). The original implementation of miRDeep-P was applied to eight small RNA sequencing libraries from three plants: Arabidopsis and rice, both of which have very well annotated genomes and papaya, a plant with no miRNA gene annotation at the time. miRDeep-P identified 90 known and 18 novel miRNA genes from Arabidopsis, 158 known and 51 novel miRNA genes from rice, and 56 known and 48 novel miRNA genes from papaya. The miRDeep-P pipeline has also been used in miRNA annotation utilizing the whole genome of peach (Colaiacovo, Bernardo, et al., 2012), potato (R. Zhang, Marshall, Bryan, & Hornyik, 2013), cotton (Gong, Kakrana, Arikit, Meyers, & Wendel, 2013), sugarcane (Gentile et al., 2013), and cassava (Ballen-Taborda et al., 2013). Additionally, the utility of the miRDeep-P for miRNA annotation has been established in non-model species lacking sequenced genomes. The pipeline was able to annotate 23 known miRNA families and 161 putative novel miRNA families from an assembled transcriptome in orchid (Aceto et al., 2014) and two novel miRNAs from ESTs of a species in the same genera as the basal angiosperm *Liriodendron chinense* (K. Wang et al., 2012). In wheat, a grass with similar genome complexity to *Avena sativa*, 145 candidate miRNAs were identified in 14 small RNA libraries from wheat by applying the miRDeep-P pipeline to ESTs and a draft genome assembly of wheat (Su et al., 2014).

Although sequencing resources are not as complete in *Avena sativa*, a sibling member to wheat in the Pooideae subfamily of monocot grasses, we implement a modified miRDeep-P pipeline in the first identification and annotation of miRNAs in oat. In the annotation of miRNAs in *Avena sativa* from small RNA deep sequencing and

transcriptome assemblies from developing flowers in three hexaploid oat cultivars, we can identify conserved miRNAs, non-conserved miRNAs that may be monocot or Poaceae family specific, and predict novel miRNAs with a pipeline designed for the annotation of miRNAs in the absence of a reference genome (Figure 1).

2.2.Materials and Methods

2.2.1 Testing miRDeep-P on Known MiRNA Precursors

Because miRDeep-P was trained on Arabidopsis miRNAs deposited in miRBase, the steps of the miRDeep-P pipeline (X. Yang & Li, 2011) were tested on three datasets of plant miRNA precursors from miRBase v. 20 (Griffiths-Jones et al., 2006; Griffiths-Jones et al., 2008; Kozomara & Griffiths-Jones, 2011, 2014). This included 6,189 plant miRNA precursors, 2,778 precursors from the 20 conserved miRNA families, and 568 high confidence miRNA precursors downloaded from ftp://mirbase.org/pub/mirbase/CURRENT/. Each dataset was tested with three different precursor excision sizes. The first step is to build the precursor index using bowtie-build. Bowtie (Langmead et al., 2009) is used to align the mature, known miRNAs from miRBase v20 to the miRNA precursors also from miRBase, using one mismatch. Next, miRDeep-P converts the bowtie output to blast format using convert_bowtie_to_blast.pl. A user-specified length of precursor is excised from the alignment with excise_candidate.pl. Lengths 250, 276, and 336 were chosen for comparison and each precursor is folded, without producing graphical output, using RNAfold (McCaskill, 1990; Zuker & Stiegler, 1981) from ViennaRNA Package v1.8.5 (Gruber, Lorenz, Bernhart, Neubock, & Hofacker, 2008; Hofacker et al., 1994). The excised candidates that align with the mature sequences are indexed and aligned a second time, followed by

converting bowtie to blast format. The blast output file is sorted based on field 4 and is necessary for input into the miRDP.pl core algorithm. The core algorithm scores the folding of the precursors using plant specific parameters. rm_redundant_meet_plant.pl from miRDeep-P is the filter for precursors that meet the plant specific scoring thresholds and removes redundant precursors.

2.2.2 Tissue and RNA Isolation

Avena sativa cultivars Leggett, Betania, and Jordan were planted in a 52x18x14 cm pot with Pete's Potting Mix (Daddy Pete's Plant Pleaser, Stony Point) mixed with ¼ cup Tree & Shrub fertilizer/insect protector (Bayer Advanced, Research Triangle Park). Four seeds of each cultivar were grown in a 19°C and a 26°C Celsius growth chamber, both chambers with 18 hr. photoperiod. Plants were monitored for pollination and the desired stage (see Figure 2). The lemma, palea, anthers, and ovary of stage A (1-3 Days Post Anthesis) and stage B (4-6 DPA) of flower development were collected on ice and stored in the -80°C freezer. Two or three flowers from each plant served as internal biological replicates. A modified protocol (Emir Islamovic, UC Berkeley) using the MiraZol Small RNA Isolation Kit (BiooScientific, Austin) was used for small RNA extraction.

For total RNA, flower tissue was crushed in liquid nitrogen with a mortar and pestle and total RNA was isolated using the Qiagen Plant RNeasy kit (Qiagen, Valencia). A DNA-free step (Ambion, Austin) removed DNA contamination. For both enriched small RNA and total RNA, the RNA size, quality, and concentration assessed and measured on the Agilent Bioanalyzer 2100 using the RNA Nano 6000 chip and reagents (Agilent, Santa Clara). The Plant RNA assay calculates a RNA Integrity Number (RIN)

using the 25S and 18S rRNA peaks as reference. Using this assay, enriched small RNA has a very low integrity number (2.6) because of the small sizes of RNA are assumed to be degradation products from total RNA. Higher RIN numbers above 5 have been observed in enriched small RNA samples to be a result of contamination by large RNAs or phenol. RIN numbers above 7 have been a result of phenol contamination. For total RNA, the recommended RIN number for input into the Illumina TruSeq RNA Sample Prep Kit v2 is 7.

2.2.3 Transcriptome Sequencing and Assembly

The Illumina TruSeq RNA Sample Prep kit v2 (Illumina, San Diego) was used for library preparation. After purifying the poly-A mRNA using poly-T magnetic beads, RNA is fragmented enzymatically and undergoes reverse transcription. cDNA ends are cleaned up and indexed barcodes are ligated. PCR amplification enriches cDNA that contains adapters on both ends. The library was validated using the Agilent Bioanalyzer 2100 DNA1000 chip and reagents (Agilent, Santa Clara). A single peak size 260 base pairs was observed. Libraries were sequenced on the HiSeq2500 through Eurofins MWG Operon (Huntsville). MiRNA discovery in the transcriptomes of Leggett, Jordan, and Betania uses selected features from the miRDeep-P pipeline (X. Yang & Li, 2011) making it applicable for miRNA discovery in the absence of a reference genome and in a complicated genome such as hexaploid oat, essentially containing three subgenomes. *De novo* RNA-seq assemblies of pooled stages A and B from oat cultivars Leggett, Jordan, and Betania were done using Velvet/Oases (Schulz, Zerbino, Vingron, & Birney, 2012; Zerbino & Birney, 2008) by the Bioinformatics Service Division at the NCRC. The Velvet optimizer was used to find the optimal k-mer length and the Oases coverage cutoff

was 10. The sequencing statistics of the RNA-seq assemblies are shown in Table 3.

2.2.4 Small RNA Sequencing and Analysis

Sequencing library preparations were done per manufacturer's instructions (Illumina TruSeq sRNA Sample Prep Kit, Illumina, San Diego). The steps of the kit are ligation of 3' and 5' adapters, reverse transcription, barcode ligation, and PCR amplification. 18 cycles for PCR amplification were used instead of the recommended 11 cycles. The Agilent Bioanalyzer 2100 High Sensitivity chip and reagents (Agilent, Santa Clara) assessed the success of the library preparation as well as the concentration and molarity of each individual library. 24 libraries were pooled equimolarly based on the molarity estimated by the Bioanalyzer software of cDNA size 140-165. Small RNAs of size 147 base pairs corresponds to mature miRNAs. The pooled library was run on a 6% pre-cast TBE PAGE gel (Invitrogen, Carlsbad), stained with ethidium bromide and visualized with a UV transilluminator. A custom ladder size 145 and 160 base pairs guided the excision of library in that size range. The library was concentrated by ethanol precipitation and the Agilent Bioanalyzer DNA1000 chip and reagents (Agilent, Santa Clara) estimated total concentration and size of the final library. This library pool of 24 samples was sent for sequencing on the Illumina HiSeq2000 at the D.H. Murdock Research Institute (DHMRI) at the North Carolina Research Campus (NCRC) in Kannapolis, North Carolina or libraries were sent to Eurofins MWG Operon (Huntsville) for sequencing on the Illumina HiSeq2500.

Pre-processing of 100 base pair reads was done following the pipeline shown in Figure 1. Table 4 shows the number of reads filtered out after pre-processing steps. FASTQ files were filtered for quality using the fastx_quality_filter tool of the FASTX-

Toolkit (Hannon Lab, CSHL). The filter parameter was –q 15 –p90. These parameters correspond to 90% of the reads needing a quality of 98% or better. An in-house python script trimmed the 5' and 3' adapters to eliminate the adapter portion of the reads; a sequence must have a 3' adapter and after trimming the 3' and 5' adapters, sequences length 18-36 were retained. The collapse_reads_md.pl from miRDeep2 (Friedlander, Mackowiak, Li, Chen, & Rajewsky, 2012) collapsed duplicate reads keeping read counts and formatting for input into miRDeep-P. Small RNA library sequences from the same cultivar were pooled for annotation purposes.

2.2.5 Identification of Known MiRNAs

For the annotation of known miRNAs, mature miRNA sequences were downloaded from miRBase release 20. The alignment program used was Vmatch (Kurtz, 2011). The mkvtree program created a suffix tree from the known miRNAs and the vmatch program solved the matching tasks over the index. Parameter options –complete –d and –h were used to specify complete and direct matches between the query and indexed sequences, constraining the number of mismatches or hamming distance of the matches returned to two or fewer mismatches.

2.2.6 MiRNA Discovery in the HiSeq Transcriptome

Transcriptome assemblies were indexed by bowtie-1.0.0 (Langmead et al., 2009) and the matching small RNA sequences were aligned to the transcriptome assemblies using bowtie-1.0.0. One mismatch was allowed despite the recommendation by miRDeep-P for a perfect alignment due to the subgenomic complexity of hexaploid oat. miRDeep-P filters alignments by the size of the largest miRNA family in Arabidopsis. miR169 contains 14 members so miRDeep-P would discard a candidate small RNA if it

aligns with more than 15 contigs. In the case of hexaploid oat with three subgenomes, 14 mir169 members * 3 genomes = 42 potential miR169 members. This filtering was skipped. A second filtering step of miRDeep-P was ignored. Filtering by annotated coding regions is not relevant without an annotated genome. In addition, this filtering step would eliminate candidate miRNAs that are alternatively spliced may overlap coding regions. Filtering by non-coding RNAs such as those annotated in Rfam will reduce computational time and false positives of precursor predictions (Burge et al., 2012; Griffiths-Jones, Bateman, Marshall, Khanna, & Eddy, 2003; Griffiths-Jones et al., 2005). The Rfam webserver was used on both the precursor sequences and the small RNA sequences that passed the plant filtering criteria as predicted by the modified miRDeep-P pipeline. The Rfam webserver (Burge et al., 2013; Griffiths-Jones et al., 2005) uses cmsearch version 1.1 to search annotations using covariance models. The order of running Bowtie indexing and alignments, miRDeep scripts, and RNAfold is the same as what was mentioned in testing miRDeep-P on known precursor sequences in miRBase.

2.3 Results and Discussion

2.3.1 Testing miRDeep-P on Plant MiRNA Precursors

The length of the excised precursor has an effect on the sensitivity of the prediction algorithm. The miRDeep-P pipeline recommends an excision length of 250 bases; however, in a similar plant parameterized miRDeep, an excision length of 276 covers approximately 96% of all plant precursors (Thakur et al., 2011) and length 336 covers 98%. We define sensitivity as the number of true positives divided by the total number of true positives and false negatives. Because miRBase contains some predicted precursors that are not validated, we also looked at the sensitivity of a subset of

precursors from the 20 conserved plant families and experimentally validated high confidence precursors (Table 5). These datasets are available upon request. For the two latter datasets, there was not as much of an improvement on sensitivity by increasing the excision length from 276 to 336. No improvement was observed in the high confidence precursor set. The trade-off in increasing the excision length is the computational time for folding of these precursors in the later step of the miRDeep-P pipeline. Therefore, and excision length of 276 was chosen for miRNA discovery in *Avena sativa*.

2.3.2 MiRNA Predictions in Leggett, Jordan, and Betania

Using the modified miRDeep-P pipeline, we were able to annotate candidate miRNAs. Some of these candidate miRNAs belong to a family of known miRNAs (conserved or non-conserved) or have sequence similarity to novel miRNAs previously annotated in other species. Other candidate miRNAs are classified as novel miRNAs, the first miRNAs discovered to be novel in oat species. Candidate miRNA sequences are shown in Tables 6-8. The candidate sequences are named by the first three letters of the cultivar ("leg" for Leggett, "jor" for Jordan, and "bet" for Betania) followed by "cand", a unique number for the cultivar, and an "x" followed by the read count of the sequence in the pooled small RNA library. Allowing for one mismatch between the small RNA and precursor enabled the identification of other small RNA sequences that may be members of the same miRNA family

2.3.2.1 Leggett: Predicted Previously Annotated MiRNAs

Table 6 shows the sequence of candidate miRNA sequences and their respective read counts. Also indicated in Table 6 is if the candidate miRNA has 0 or 1 mismatch to the precursor. For 0 mismatches, if the miR* sequence predicted by the miRDeep-P

pipeline exists in the small RNA library, the miR* sequence is given. For 1 mismatch, the read count with 0 mismatches to the hairpin is given if it exists in the small RNA library.

Of the 20 conserved miRNA families, none were predicted for Leggett. Two members of the non-conserved miRNA family 5049 were predicted (cand8x62 and cand13x11) by the miRDeep-P pipeline; however, a mismatch between the miR5049 member and the respective precursor disqualifies the assembled precursor as the true hairpin for this sequence. In addition to the mismatch, the miRNA sequence that would be produced by the hairpin does not exist in the small RNA library for cand8x62. For cand13x11, the miRNA sequence produced by the hairpin exists in the small RNA library with a read count of 5. Because no normalization was done for miRNA discovery, a read count of 5 may be under the threshold of detection once normalized and considered insignificant. This miRNA family has been annotated in barley and Brachypodium according to miRBase but has also been identified in wheat (Kantar et al., 2012; Kurtoglu, Kantar, Lucas, & Budak, 2013). In barley, miR5049 is down-regulated upon boron stress and was predicted by degradome sequencing to target a tubby-like protein gene; however, the cleavage site could not be identified (Ozhuner et al., 2013).

The candidate miRNAs discussed above passed the plant criteria for the hairpins. Of hairpins that were predicted but did not pass the filtering criteria, a small RNA sequence with sequence similarity to known miRNA cpa-miR8155 was predicted with a mismatch to its precursor. miR8155 is an annotated miRNA from papaya infected with Papaya Ringspot Virus (Aryal et al., 2012). The function or predicted targets of this miRNA are unknown. Allowing for one mismatch enabled the alignment of all small

RNA sequences to the precursor and hairpin for miR8155 (Figure 3). A small RNA sequence with a read count of 159 may be the true small RNA produced from the hairpin (a read count of 171 was scored higher despite a mismatch). Nonetheless, the small RNA sequences aligning within one mismatch to the hairpin for leg_cand7x171 may be members of the same miRNA family.

2.3.2.2 Leggett: Predicted Novel Candidate MiRNAs

Of candidate novel miRNAs with 0 mismatches to the precursor, the presence of a miR* sequence strengthens the confidence of the annotation. The miR* is predicted by the miRDeep-P pipeline as having the 2-nucleotide overlap with the mature miRNA sequence forming the miRNA/miR* duplex produced by the hairpin. The miR* predicted for cand1x9933 exists in the small RNA library with a read count of 116. Figure 4 (A) shows the secondary structure of the miRNA hairpin and the location of the mature miRNA. Figure 4 (B) shows a Geneious alignment of cand1x9933 and its corresponding miR*. The predicted miR* for cand2x418 does not exist in the small RNA library, but the miR* exists in the small RNA library for cand3x213 with a read count of 13. The hairpins for these two candidates are shown in Figures 5-6. The remaining novel candidate miRNAs in the Leggett cultivar that align perfectly to the precursor and pass the plant filtering criteria for miRDeep-P do not have a miR* sequence in the small RNA sequencing library. Figure 7 shows the hairpin and a Geneious alignment of leg_cand4x44 to its hairpin and precursor. Only one other sequence from the small RNA library aligns perfectly to the hairpin and precursor, a sequence with a read count of 43. Because no miR* was found, it is unknown which is the miRNA produced from this precursor hairpin. The candidate miRNA was scored higher due to its higher read count.

Hairpins for candidates leg_cand5x15 and cand6x11 are shown in Figure 8-9. For the hairpin for cand6x11, the 3' end of the mature candidate miRNA sequence is in a large loop, which could interfere with DCL processing. This candidate may be a false positive of the prediction pipeline.

The hairpins for other candidate miRNAs with one mismatch to its precursor are not shown. For some of these precursors, the small RNA sequence predicted to be produced by DCL1 cleavage of the precursor exists at a low read count (Table 6).

2.3.2.3 Jordan: Predicted Known MiRNAs

The modified miRDeep-P pipeline identified 9 candidate miRNAs with perfect match to precursor assemblies that passed the plant filtering criteria in oat cultivar Jordan. Three of these candidates belong to known miRNA families. The most abundant of these is a member of the miR167 family, jor_cand1x12461 with a corresponding miR* with a read count of 781. Figure 10 (A) shows the hairpin and location of the mature miRNA and Figure 10 (B) shows a Geneious alignment of the assembled precursor gene, the hairpin, and the mature miRNA and miR*. The hairpin for a member of the miR528 family, jor_cand3x41, passed the plant filtering criteria (Figure 11). The miR528 hairpin has been used for delivery of artificial miRNAs into plant systems by replacing the mature sequence and the miR* sequence with an artificial miRNA and its complement (Manavella & Rubio-Somoza, 2011). The other known miRNA family discovered by the modified miRDeep-P pipeline in the Jordan oat cultivar is non-conserved family miR827. miR827 has been annotated in many species such as *Populus, Arabidopsis*, apple, tobacco, rice, and maize according to miRBase v.20; however, it is not classified as a conserved miRNA family. In rice, miR827 is involved in a complex mechanism

regulating SPX-MFS protein levels in response to phosphate starvation (Lin et al., 2010). The sequence of miR827 member jor_cand4x29 has a one base difference from the rice miR827 on the 3' end. The hairpin and alignment of the mature miRNA jor_cand4x29, miR*, and hairpin are shown in Figure 12.

Of all candidate miRNAs that were predicted but did not pass the plant filtering criteria of the miRDeep-P pipeline (data not shown), a member of non-conserved miRNA family miR5049 and conserved family miR172 were predicted. The role of miR172 in floral development will be discussed more in Chapter 4. There exists a mismatch between the miRNA sequence and the precursor for both these candidate miRNA sequences. The sequence of the miRNA that would be produced by the precursor does not exist in the small RNA library. In these instances, the hairpins may be false positives or alternatively, an error in the assembly of the hairpins may be reason for the existence of the mismatch.

2.3.2.4 Jordan: Predicted Novel Candidate MiRNAs

The hairpins for other candidate miRNAs are shown in Figures 13-18. Note none have predicted miR* that exist in the Jordan flower small RNA library except for jor_cand2x321. The miR* for this candidate has a read count of 28 (Figure 18). For these candidate miRNAs with 0 mismatches to their respective precursor hairpins, Table 7 (A) shows the predicted miR* sequence as well as the miR* read count if it exists in the small RNA library.

An additional 13 candidates were predicted with one mismatch to the precursor assemblies. Of these candidates, only two have sequence similarity to known miRNAs. With four mismatches to annotated miR7732 from *Brachypodium* and miR6207 from

barley, it is unclear if jor_cand18x13 and jor_cand22x3 are novel to *Avena sativa* or if they may be members of a grass-specific miRNA family. For these, Table 7 (B) gives the read count of the small RNA sequence that would be produced by DCL cleavage of the precursor. Only the hairpins of candidates jor_cand11x365 and jor_cand12x238 would produce miRNAs from the small RNA-seq dataset with a read count above 2. The hairpins for these are shown in Figures 19-20. The candidates and the sequence with 0 mismatches to the precursor may be member of the same novel miRNA family.

2.3.2.5 Betania: Predicted Previously Annotated MiRNAs

Table 8 shows the candidate miRNAs that were predicted within one mismatch to the corresponding hairpin. Of these, Vmatch identified four annotated miRNAs with sequence similarity to the candidate sequences (Table 8). Three of the four annotated miRNAs are not part of any family of miRNAs, but are rather classified as novel miRNAs found in barley, *Brachypodium*, or rice. With 4 mismatches, it's unclear if these candidate miRNAs (bet_cand21x18, bet_cand23x15, and bet_cand26x5) are novel *Avena sativa* miRNAs or are members of previously un-annotated miRNA families. Conservation of the hairpin could give support to a reclassification; however, three of these candidates were predicted with a mismatch to their respective hairpins. This suggests these hairpins are false positive predictions of the modified miRDeep-P pipeline. For the one candidate that was predicted as the DICER product of the hairpin (0 mismatches), no miR* was identified in the small RNA library for bet_cand9x5 but this does not disqualify the candidate as a potential miRNA and member of the miRNA family miR1436.

In addition to these candidates with sequence similarity to known miRNAs

annotated in miRBase, Rfam identified the precursor of bet_cand15x107 to be that of a member of the miR444 family with an e-value of 1.8e-10. An error in the precursor assembly may account for a mismatch between the mature miRNA sequence and its corresponding hairpin (Figure 21). The absence of the sequence predicted to arise from the hairpin (0 mismatches) also supports this interpretation. Bet_cand15x107 shares perfect identity to rice miR444d.3 in a 14 base seed. Outside of the seed, however, there exist more than 4 mismatches. Specifying complete matches with four or fewer mismatches was cause for Vmatch to miss the hit to an annotated miRNA. miR444 is a non-conserved miRNA family, conserved in monocots (Sunkar et al., 2005). Validated targets of miR444 include genes of the MADS-box classification, which have been implicated in floral development (L. Wu et al., 2009).

The miRDeep-P pipeline outputs record of hairpins and candidate miRNA sequences that do not pass the plant filtering criteria. Four previously annotated miRNAs had hairpins in this category. One is bet_cand12x128 with only one mismatch to known miRNA from soybean, miR1511. This miRNA has been annotated in miRBase in the following dicots: soybean, apple, and peach. Although bet_cand12x128 has a perfect match to its hairpin, the predicted miR* does not exist in the library. The prediction did not pass plant filtering criteria due to the asymmetrical bulge opposite the mature sequence (Figure 22). This hairpin may be an example of a non-canonical stem loop. The stem loops of select miR164 family members in *Arabidopsis*, rice, and *Populus* are also non-canonical. In these structures, the mature miRNA sequence is located near or in large loops or bulges but these structural elements do not interfere with DICER processing of the stem loop into the miR164/miR* duplex (Jones-Rhoades et al., 2006).

The other three candidate miRNAs whose hairpins did not pass filtering were bet_cand28x21 with sequence similarity to miR5049, bet_cand29x63 with high sequence similarity to conserved miR398, and bet_cand30x18 with 4 mismatches to a non-conserved miRNA family member from barley. As a member of the conserved miR398 family, bet_cand29x63 likely has functional significance in both abiotic and biotic stress responses and nutrient homeostasis (C. Zhu, Ding, & Liu, 2011). Hvu-miR1130 is a member of a monocot-specific miRNA family miR1130. In wheat, miR1130 is up-regulated during wounding, cold and salt stress, and is also predicted to have an important regulatory role in grain filling of wheat. Targets of miR1130 have been predicted but have not been validated by qRT-PCR, RNA blots, RLM-RACE, or degradome sequencing (F. Meng et al., 2013; B. Wang et al., 2014). miR5049 is also a monocot-specific miRNA family. Members of this family have been annotated in barley and *Brachypodium*, however, they have also been identified in wheat (Kurtoglu et al., 2013). In these three predictions, the hairpins are likely not the true precursors of these candidate miRNAs due to the allowance of one mismatch in the alignment of the small RNA sequence to the precursor. An alternative is to presume there exists a sequencing error in the transcriptome assemblies. Despite not being able to annotate this candidate miRNA sequence with its precursor hairpin, further support for these sequences being miRNAs are their presence in the pooled small RNA libraries across three hexaploid oat cultivars.

2.3.2.6 Betania: Predicted Novel Candidate MiRNAs

Aside from predicting hairpins for known miRNAs discussed above, the modified miRDeep-P pipeline also predicted candidate novel miRNAs. All but three of these

novel candidates are present in small RNA libraries of *Avena sativa* cultivars Jordan and Leggett (denoted by sequence highlighted in Table 8). The sequences of the candidate miRNAs, the miR* sequence predicted from the miRDeep-P pipeline, and the read count of the predicted miR* sequence if it exists in the small RNA library is given in Table 8. The candidates in this table were predicted with 0 mismatches in aligning the small RNA sequence to its precursor. The hairpins for bet_cand3x6 and bet_cand8x9 and are shown in Figures 23-28. Bet_cand4x28 and bet_cand6x14 are 29 bases long. With miRNAs produced from DCL1 being 21 nucleotides long and DCL3 producing 24-nucleotide siRNAs, it is questionable if these are miRNAs. Sequencing errors and errors in adapter trimming may be reasons for the longer length of the small RNA. The hairpin sequence for bet_cand7x11 was searched through the Rfam database and has sequence similarity to mir-207 with an e-value of .0023. This miRNA is annotated in mouse, rat, and dog per miRBase v.20. The mature sequence is highly repetitive of "AG", which would need to be experimentally validated for expression and function.

In allowing 1 mismatch in the alignment of the small RNA sequence to the precursor, 15 candidate miRNAs were predicted, four of which were discussed earlier had sequence similarity to known miRNAs by Vmatch or Rfam search. For these candidates, Table 8 shows the read count of the small RNA sequence with perfect alignment to the precursor, if it exists in the Betania small RNA library. Bet_cand13x391, bet_cand14x369, and bet_cand24x12 are the only of the 15 hairpin predictions passing plant filtering criteria with a small RNA sequence from the library perfectly matching the precursor.

2.3.2.7 Stem-to-loop and Loop-to-base Processing of Bet_cand1x12674

The hairpin for bet_cand1x12674 is 191 bases in length and many small RNA sequences from the Betania small RNA library align perfectly to it. The predicted miR* exists in the small RNA library of Betania with a read count of 6. Figure 29 shows a Geneious alignment of all small RNA sequences (read count greater than 2) aligning perfectly to the hairpin and precursor sequence with the candidate sequence highlighted. The small RNA sequences are clustered in four regions. This represents the potential of two small RNA duplexes being produced from the same gene. This observation has been described as loop-to-base processing in miR319 and miR159 (Bologna, Mateos, Bresso, & Palatnik, 2009). In loop-to-base processing, the loop of the stem loop is cleaved by DCL1 first, rather than the canonical stem-to-loop processing of plant miRNAs. 5' RACE PCR detected the processing intermediates of the miR319a precursor in Arabidopsis. It was hypothesized that the high number of bulges in the upper stem of the miR319 precursor is cause for only one mature miRNA sequence to accumulate at high levels. Comparably, the hairpin for bet_cand1x12674 (Figure 30) has many loops in the upper stem region, which may allow for the accumulation of the miRNA candidate bet_cand1x12674 and not small RNAs from the duplex closer to the loop. Another study has identified these multiple small RNA duplexes in plants other than Arabidopsis and also in miR169 precursors (W. Zhang et al., 2010). The accumulation of other miR-like sequences may be condition or tissue dependent but the validation of function as a true miRNA needs to be investigated.

2.3.2.8 Betania: Novel Candidate MiRNA Families

The hairpin predicted for bet_cand2x1637 has many Betania small RNAs that

align to it within one mismatch. Figure 31 (A) shows the secondary structure of the hairpin and (B) shows small RNA sequences (read count above 2) aligned within one mismatch to the hairpin and precursor. Bet_cand2x1637 is highlighted. Because the predicted miR* sequence was not present in the Betania small RNA library, the high read count sequence bet_3519836_x1016 alternatively may be the miRNA produced from this hairpin. It differs in length by one base on the 3' end to the candidate miRNA sequence, bet_cand2x1637. The candidate sequence was scored higher due to its read count. Sequences aligning to the hairpin with one mismatch may be members of the same miRNA family; however, they would be processed from different precursor genes.

An additional candidate novel miRNA family may include bet_cand13x391. Figure 32 (A) shows the hairpin predicted for the candidate miRNA. The sequence aligning perfectly to the precursor has a read count of 101 and therefore this sequence is more likely the true miRNA produced from DCL1 cleavage of the precursor. The sequences of all small RNA sequences aligning within one mismatch to the precursor are shown in a multiple sequence alignment Figure 32 (B). Sequences with read counts below 3 are excluded. The sequences in this multiple sequence alignment may represent members of a novel miRNA family in *Avena sativa*.

2.4. Conclusions

One of the challenges of a miRNA discovery project in a species with a complex, unfinished genome assembly is that the current assemblies of transcriptome data do not allow for the prediction of all miRNA precursors. An example of this is the absence of precursors for the 20 conserved miRNA families. Of the 20 conserved miRNA families, with most families having multiple members, the modified miRDeep-P pipeline was able

to predict precursors for miR167, miR172, and miR398 in addition to several non-conserved monocot-specific miRNAs. However, with a pipeline for miRNA annotation established, the refined annotation of miRNAs in *Avena sativa* is straightforward once more genomic sequencing becomes available.

Another challenge is the significance of sequences with low read counts. A candidate sequence with a read count of 2 is most likely not significant especially since four small RNA sequencing libraries were pooled for each cultivar. Just as there is no standard pipeline for miRNA annotation, there is no standard normalization of small RNA sequencing libraries. JMP Genomics has an established Basic miRNA-seq Workflow with normalization options. Normalization of read counts will then enable a more comprehensive study looking at the expression profile of not only candidate miRNAs presented here, but also members of the 20 conserved miRNA families and other un-discovered miRNAs. Using JMP Genomics by SAS can further allow a differential expression analysis of these normalized miRNAs to identify isoforms of miRNA families controlling adaptation to cold (19 degrees Celsius) and phenotypic change of the flower developing from 3-4 days post anthesis.

Most of the candidate miRNAs presented here passed the plant filtering criteria for their hairpins. There are many more candidate miRNAs that form non-canonical miRNA hairpins that were screened out by the plant filter. Only candidate miRNA sequences with sequence similarity to known miRNAs in miRBase v.20 are seen in Tables 6-8. For these and multiple instances of the candidate novel miRNAs predictions, the miRNA was predicted with one mismatch to its precursor. The caveat is the potential for numerous false positive predictions. However, we cannot evaluate the false positive

rate due to the discovery nature of the study. If a mismatch exists between the small RNA and the precursor, it does not eliminate the small RNA as a candidate miRNA; the current transcriptome assemblies do not have coverage of the precursor for the small RNA sequence. Alternatively, subgenomic complexity may contribute to errors in the assemblies as well, which could be responsible for numerous false predictions. With the large and complex genome of oat, this is a reasonable assumption.

Allowing for one mismatch in the alignment of the small RNAs to assembled transcripts has enabled the identification of candidate novel miRNA families in both Leggett and Betania cultivars. An additional particular thought-provoking finding is the hypothesis of loop-to-base processing of its pri-bet_cand1x12674. This alternative biogenesis has only been found for conserved miRNA families miR159/319 and miR169. 5' RACE PCR would need to validate the processing intermediates of the precursor. Furthermore, it would be fascinating to detect accumulation of the miR-like sequences from the alternative duplex under different temperatures or stages of floral development.

All but 7 of the candidate miRNAs presented here are present in the pooled small RNA libraries of all three oat cultivars. Cross-library validation adds reinforcement to the predictions that these are true miRNA sequences and that they have functional significance. However, there are many instances of errors in assemblies or false predictions. These small RNAs need to be validated by stem-loop RT-PCR to evaluate the precision of the miRDeep-P pipeline.

From the discovery and annotation of candidate novel miRNAs in *Avena sativa*, we can then begin the investigation into the functional significance of these sequences. In combination with PARE (Parallel Analysis of RNA Ends) in Leggett, Jordan, and

Betania cultivars of the same tissue and growth conditions, we can predict and provide experimental degradome validation of the miRNA sequences and their corresponding targets. Although miRNA target prediction is not required for miRNA annotation, it enhances the annotation. We can then incorporate the miRNAome, transcriptome, and miRNA-cleaved transcriptome into a co-expression network. Through investigating the miRNA-target modules, we can associate miRNAs with desired agronomic traits in *Avena sativa* with the goal of using miRNAs in genomic selection for breeding.

Table 1: Conserved miRNA families and their conserved targets.

Conserved miRNA families	Target Transcription Factor Family	Function
miR156	SBP (SQUAMOSA PROMOTER BINDING PROTEIN)	Vegetative growth
miR157		
miR159/miR319	MYB, TCP	Anther and floral development
miR160	ARF (AUXIN RESPONSE FACTOR)	Hormone signal transduction
miR162		
miR164	NAM/CUC	Floral and shoot development
miR165/miR166	HD-ZIP	Vascular development, meristem formation
miR167	ARF (AUXIN RESPONSE FACTOR)	Hormone signal transduction
miR168		
miR169		
miR170/miR171	SCL (SCARECROW-LIKE)	Flower development
miR172	AP2 (APETALA2)	Flower development
miR390		
miR393		
miR394		
miR395		
miR396	GRF (GROWTH REGULATING FACTOR)	Growth and development
miR397		
miR398		
miR399		

Table 2: Characteristics of four software tools for miRNA detection in small RNA deep sequencing datasets.

Program	Utility	Design (Web, Local)	Alignment Algorithm/Program	Secondary Structure Prediction	Novel miRNA Prediction
DSAP	Comparative miRNA Analysis	Web	BLAST	N/A	N/A
miRExpress	Comparative miRNA Analysis	Local	Smith-Waterman	N/A	N/A
miRDeep	Known and Novel MiRNA Prediction	Local	Bowtie	RNAfold	Bayes Theorem
MIReNA	Known and Novel MiRNA Prediction	Local	MegaBLAST	RNAfold	5 Numerical Criteria

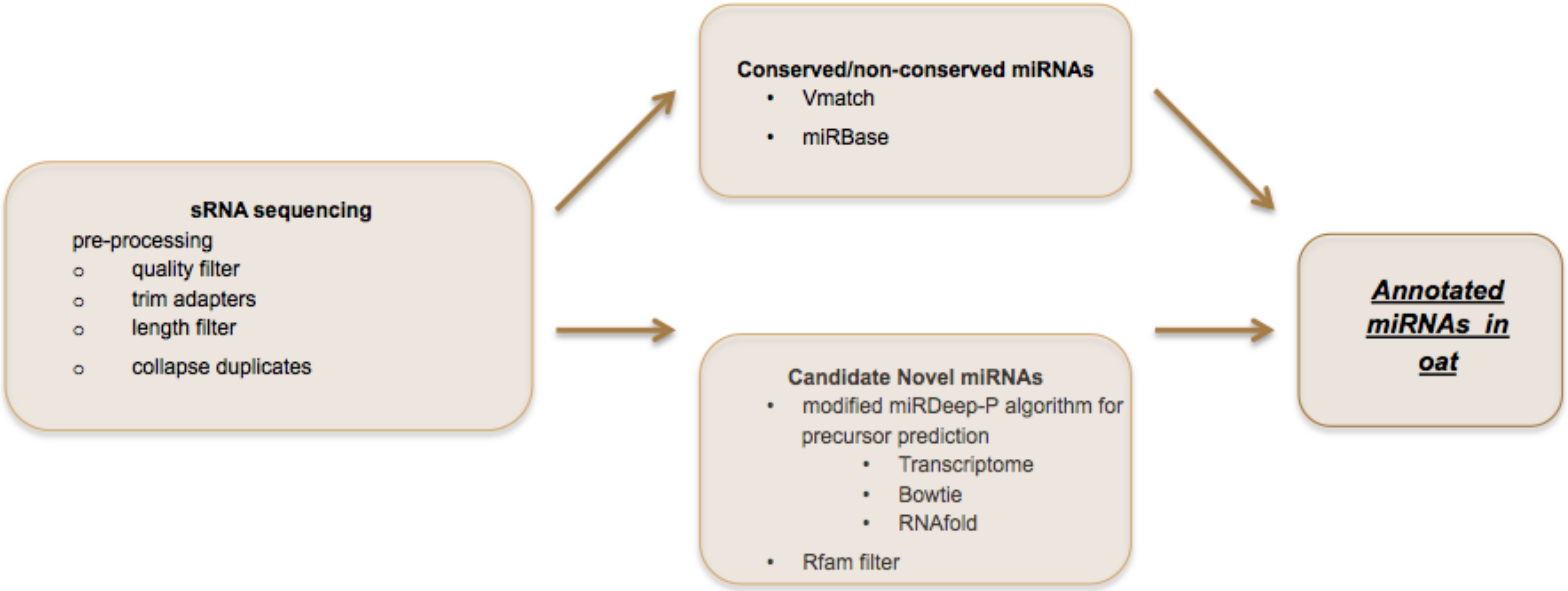


Figure 1: MiRNA annotation pipeline in the absence of a reference genome. Small RNA sequences are pre- processed for quality, adapters, length, and duplicates. Known miRNAs are identified using Vmatch to solve matching tasks between the small RNA sequences and annotated, known miRNAs in miRBase v20. The core miRDeep core probabilistic algorithm is used to score folded hairpins and plant specific parameters filter out hairpins that do not meet the criteria. Finally, the candidates are screened against the Rfam database to eliminate hairpins originating from non-coding RNAs other than miRNAs.

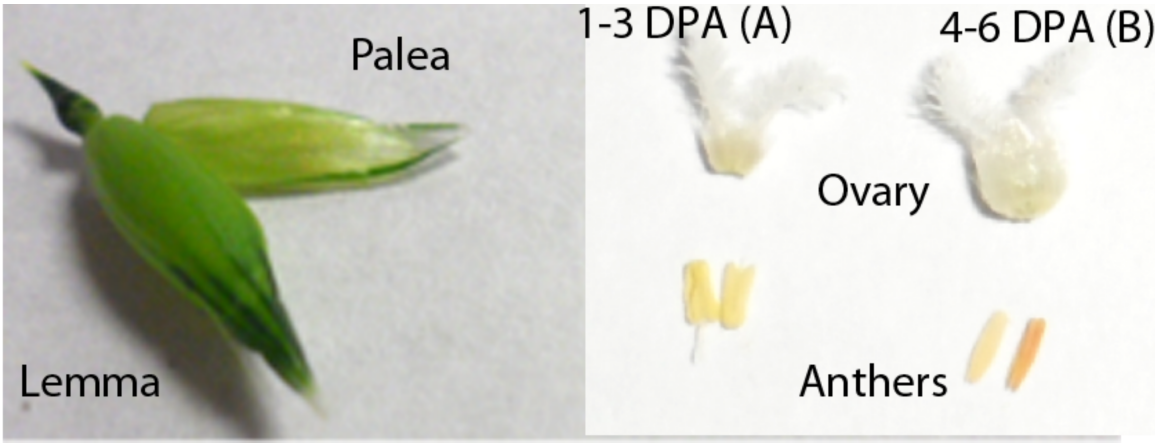


Figure 2: Flower tissue and stage collected from hexaploid oat for small RNA sequencing.

Table 3: Sequencing statistics of transcriptome assemblies of three oat cultivars. All Tissue was collected from plants grown in a 19 degree Celsius growth chamber.

Hexaploid cultivars (*Avena sativa*)	**Stage of Development**	**Number of Sequences**	**N50**
Leggett	1-6 Days Post Anthesis (DPA)	204,738	251
Jordan	1-6 (DPA)	184,040	245
SW Betania	1-6 (DPA)	196,662	254

Table 4: Pre-processing of small RNA reads.

Hexaploid cultivars (*Avena sativa*)	Small RNA Library Stages		Raw reads (pooled stage and growth condition)	Quality Filter	Adapter Trimmed, 18-30	Non-Redundant	Number of Singlets	Filtered Sequences for miRNA Discovery
	19° C	26° C						
Leggett	A (1-3 DPA) B (4-6 DPA)	A, B	10,462,681	9,103,246	6,095,761	1,923,775	1,520,289	403,486
Jordan	A, B	A, B	42,545,304	40,373,000	23,203,002	4,348,133	3,180,952	1,167,181
SW Betania	A, B	A, B	32,263,711	31,100,474	16,023,996	3,534,416	2,653,995	880,421

Table 5: Testing miRDeep-P on known precursors and miRNAs in miRBase.

Number Precursors	Precursor Excision Length	Number Non-Redundant Precursors	Number Precursors Passing Plant Filter	Sensitivity
6189 (miRBase v. 20)	250	2023	1971	31.85%
	276	2025	1974	31.90%
	336	2032	1981	32.01%
2778 (precursors from 20 conserved families	250	1529	1501	54.03%
	276	1532	1503	54.10%
	336	1532	1504	54.14%
568 (high confidence)	250	260	257	45.25%
	276	261	258	45.42%
	336	261	258	45.42%

Table 6: Candidate miRNAs in Leggett predicted by the miRDeep-P pipeline. Candidate miRNAs predicted with 0 mismatches to their hairpin are shown in (A). Candidate miRNAs predicted with 1 mismatch to their precursors are shown in (B). If the candidate sequence has sequence similarity within 4 mismatches to a known miRNA in miRBase v.20, the identity is given. The "*" denoted at leg_cand17x171 indicates that a hairpin was predicted; however, it did not pass the plant filtering criteria of the miRDeep-P pipeline. Highlighted sequences indicate the sequence is present in the pooled small RNA libraries from Leggett, Jordan, and Betania.

A

Candidate miRNA ID	Number of Reads	Predicted Mature miRNA Sequence (0 mismatches)	Predicted miR* Sequence	miR* Read Count
leg_cand1x9933	9933	TGCATTGACTCAAGTTGGACC	TTCAACTTGAGTCGATCCAGT	116
leg_cand2x418	418	GCTTTGGCTGCTCTGTCCCT	GGGCTGGGCGGTTGAAGCGA	0
leg_cand3x213	213	TCATACATCACACATCAGACA	TTTGATGAGTGGTGTACGAGG	13
leg_cand4x44	44	TTAAGCATAGCTGGATTGACT	TCAATCCAATTATGCTTAAAC	0
leg_cand5x15	15	TGTTAACGAAACATGACTGCC	CAGTTATGTTTCATTAACAGA	0
leg_cand6x11	11	AGTACTCTGCTCTGCCTGCT	TGGGTAAAAGCAGAGAACTTT	0

B

Candidate miRNA ID	Number of Reads	Small RNA Sequence (1 mismatch to precursor)	miRBase Hit (<3 mismatches)	miRBase Hit (3 mismatches)	miRBase Hit (4 mismatches)	Read Count of Predicted Sequence (0 mismatch to precursor)
leg_cand7x161	161	GTCCGGCGTGCTCTTCCGTCGAGC				5
leg_cand8x62	62	AATTAATTTGGGTCGGAGGGA	miR5049	mtr-miR5205	mtr-miR5203, bdi-miR5171	0
leg_cand9x34	34	TCCGATCCAAATTAGTTGACT				0
leg_cand10x23	23	TCTGTTCCATAATATAAGCCT				0
leg_cand11x17	17	TAAAAGGAGTCTAGATACACT				0
leg_cand12x14	14	TTGTCTAGATTCGAATGTATC				0
leg_cand13x11	11	AAGTAATATGGATCGGAGGGA	miR5049		bdi-miR5203, bdi-miR5175	5
leg_cand14x9	9	CCGGTGCGCCCCCGGCGGC				0
leg_cand15x8	8	TCCGTTCCATAATATAAGCCT				6
leg_cand16x3	3	TTGTCTAGATTTGGATGTATCT				2
*leg_cand17x171	171	AAACCTGGCTCTGATACCA	cpa-miR8155			10

Table 7: Candidate miRNAs in Jordan predicted by the miRDeep-P pipeline. Candidate miRNAs predicted with 0 mismatches to their hairpin are shown in (A). Candidate miRNAs predicted with 1 mismatch to their precursors are shown in (B). If the candidate sequence has sequence similarity within 4 mismatches to a known miRNA in miRBase v.20, the identity is given. The "*" denoted in front of the miRNA ID indicates that a hairpin was predicted; however, it did not pass the plant filtering criteria of the miRDeep-P pipeline. Highlighted sequences indicate the sequence is present in the pooled small RNA libraries from Leggett, Jordan, and Betania.

A

Candidate miRNA ID	Number of Reads	Predicted Mature miRNA Sequence (0 mismatches)	miRBase Hit (>3 mismatches)	Predicted miR* Sequence	miR* Read Count
jor_cand1x12461	12461	TGAAGCTGCCAGCATGATCTGA	miR167	AGGTCATGCTGGAGTTTCATC	781
jor_cand2x321	321	TCATACATCACACATCAGACA		TTTGATGAGTGGTGTACGAGG	28
jor_cand3x41	41	TGGAAGGGGCATGCAGAGGAG	miR528	CCTGTGCCTGCCTCCTCCATT	2
jor_cand4x29	29	TTAGATGACCATCAGCAAACA	miR827	TTTTGTTGGTTGTCATCTAACC	14
jor_cand5x29	29	TGTTAACGAAACATGACTGCC		CAGTTATGTTTCATTAACAGA	0
jor_cand6x24	24	GACATGTTGGATGTGTTGGCTCCAGTT		CTGGCGCTGACATGTTTTGACAAGTCAG	0
jor_cand7x21	21	CTCCGTCCAGAAATAACTGAC		CAGTTATCTTTGGACGGAGGA	0
jor_cand8x6	6	ACGAGATGCGTCGCTTTGCAGCTGTCCCT		GGACGCTGTAACGCTGCCGTCTCGGTTT	0
jor_cand9x4	4	CTCTGAGCATCCATGATACATGACCT		GTCGTGAATCTGGACTGTAAGAGAT	0

Table 7… (continued): Candidate miRNAs in Jordan predicted by the miRDeep-P pipeline. MiRNAs predicted with 1 mismatch to their precursors are shown in (B). If the candidate sequence has sequence similarity within 4 mismatches to a known miRNA in miRBase v.20, the identity is given. The "*" denoted in front of the miRNA ID indicates that a hairpin was predicted; however, it did not pass the plant filtering criteria of the miRDeep-P pipeline. Highlighted sequences indicate the sequence is present in the pooled small RNA libraries from Leggett, Jordan, and Betania.

B

Candidate miRNA ID	Number of Reads	Small RNA Sequence (1 mismatch to precursor)	miRBase Hit (<3 mismatches)	miRBase Hit (4 mismatches)	Read Count of Predicted Sequence (0 mismatch to precursor)
jor_cand10x3597	3597	TCTTGGATTTATGAAAGACGAACA			2
jor_cand11x365	365	GTCCGGCGTGCTCTTCCGTCGAGC			18
jor_cand12x238	238	TAAAAGGGGTCTAGATACACT			64
jor_cand13x103	103	GCAGGACGGTGGTCATGGA			0
jor_cand14x79	79	CCGGTGCGCCCCCGGCGGC			0
jor_cand15x19	19	TCTGAAATTTGATCATTTGCC			0
jor_cand16x18	18	AGGAGGGCGCGGCGGCCG			0
jor_cand17x15	15	TCCGTTCCACAATATAAGCCT			0
jor_cand18x13	13	CCACCAAGATCTCTGAGGCCC		bdi-miR7732	0
jor_cand19x11	11	TCTGCGACATCTAATTCGGGA			0
jor_cand20x7	7	TATTTTGAAACGGAGGGAGTA			0
jor_cand21x5	5	AATATAAGATGTTTTAGGTTT			0
jor_cand22x3	3	GAGGAGATGGACGCAGAC		hvu-miR6207	0
*jor_cand23x14	14	AATTAATTTGGATTGGAGGGA	miR5049		0
*jor_cand24x13	13	GGAATCTTGATGATGCTG	miR172		0

Table 8: Candidate miRNAs in Betania Predicted by the miRDeep-P pipeline. Candidate miRNAs predicted with 0 mismatches to their hairpin are shown in (A). Candidate miRNAs predicted with 1 mismatch to their precursors are shown in (B). If the candidate sequence has sequence similarity within 4 mismatches to a known miRNA in miRBase v.20, the identity is given. The "*" denoted in front of the candidate miRNA ID indicates that a hairpin was predicted; however, it did not pass the plant filtering criteria of the miRDeep-P pipeline. Highlighted sequences indicate the sequence is present in the pooled small RNA libraries from Leggett, Jordan, and Betania.

A

Candidate miRNA ID	Number of Reads	Predicted Mature miRNA Sequence (0 mismatches)	miRBase Hit (<3 mismatches)	miRBase Hit (4 mismatches)	Predicted miR* Sequence	miR* Read Count
bet_cand1x12674	12674	GCGACAACCGGACCGGCTATTGACCT			GTTAATAGCCCTTCTGGTTGTCGCTA	6
bet_cand2x1637	1637	GCTTTGGCTGCTCTGTCCC			GGCTGGGCGGTTGAAGCGA	0
bet_cand3x54	54	TGTTAACGAAACATGACTGCC			CAGTTATGTTTCATTAACAGA	0
bet_cand4x28	28	GAGGGATGATGCCAGGATCTGGAGGCTCT			AGCTCCTGTTCTTGGTGTTCTTCCTCGA	0
bet_cand5x14	14	ATGTCCCGGGTCGTCGAGGCC			CTTCGACGACGCCAGCGTGG	0
bet_cand6x14	14	GCGCGACGAGGGCTCTCCTGAACGAGTCG			ACTCCTTCAAGAAGGTCCTCGCGCGCAT	0
bet_cand7x11	11	AGAGAGAGAGAGAGAGAGGAGGGC			CACCTCCCTCCCCCTCTCTTCTCC	0
bet_cand8x7	7	CTAGCACCGTATTAGTCGCACGTA			CGTGCGAGTGATACGGTCTAGTT	0
bet_cand9x5	5	ATATTGTGGAACAGAGGGAGT		miR1436	TCCCTCTTGTTCCACAATATAA	0
bet_cand10x2	2	GGCCTCGATGGGTCGAGGGAGGCG			CCTCCCACCGGCCCGTCGACGCCGC	0
bet_cand11x2	2	CGTTGCAGCTCCAGAATCTGGA			CAGAGGCTGGTGCTGTGGCCTC	0
*bet_cand12x128	128	AACCTGGCTCTGATACCA	miR1511		GTAGCTAGGCCAGCACCTCC	0

Table 8 (continued): Candidate miRNAs in Betania Predicted by the miRDeep-P pipeline. Candidate miRNAs predicted with 0 mismatches to their hairpin are shown in (A). Candidate miRNAs predicted with 1 mismatch to their precursors are shown in (B). If the candidate sequence has sequence similarity within 4 mismatches to a known miRNA in miRBase v.20, the identity is given. The "*" denoted in front of the miRNA ID indicates that a hairpin was predicted; however, it did not pass the plant filtering criteria of the miRDeep-P pipeline. Highlighted sequences indicate the sequence is present in the pooled small RNA libraries from Leggett, Jordan, and Betania.

B

Candidate miRNA ID	Number of Reads	Small RNA Sequence (1 mismatch to precursor)	miRBase Hit (<3 mismatches)	miRBase Hit (4 mismatches)	Read Count of Predicted Sequence (0 mismatch to precursor)
bet_cand13x391	391	TAAAAGGGGTCTAGATACACT			101
bet_cand14x369	369	GTCCGGCGTGCTCTTCCGTCGAGC			16
bet_cand15x107	107	TTTTCTTGCAAGTTGTGCAGT			0
bet_cand16x60	60	GGGATAACATCATAGGATT			0
bet_cand17x53	53	GGAGGGCGCGGCGGCCGC			0
bet_cand18x39	39	TTAATGGTCTCAACTTTGTCT			0
bet_cand19x27	27	GACACTTATTATGGATTGGAG			0
bet_cand20x18	18	TCGCAGATTTGTCTAAATTTG			0
bet_cand21x18	18	TCTGTTCACTAATATAAGACT		hvu-miR6197	0
bet_cand22x16	16	TTGTCTAGATTCGTATGTATC			0
bet_cand23x15	15	CCACCAAGATCTCTGAGGCCC		bdi-miR7732	0
bet_cand24x12	12	CAAGTGACTGTATCTCATGAGCCT			7
bet_cand25x8	8	CTCCGTCCAGAAATAAGTGAC			0
bet_cand26x5	5	AAAACATCTTATACTTTGAAACGG		osa-miR5825	0
bet_cand27x2	2	TATCGTTGTGATTTGCTCCAGTTTTGCCC			0
*bet_cand28x21	21	AATTAATTTGGATCGGAGGGA	miR5049	bdi-miR5175	2
*bet_cand29x63	63	TGTGTTCTCAGGTCGCCCCTG	miR398		0
*bet_cand30x18	18	TCTGTTCACTAATATAAGACT		miR1130, hvu-miR6197	0

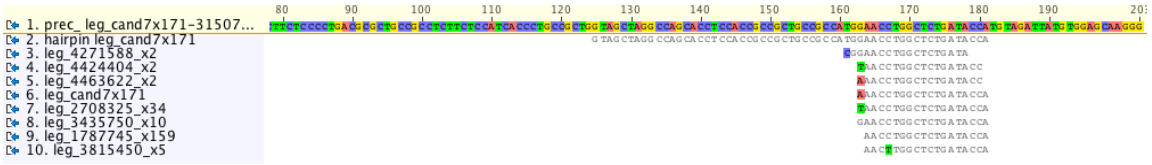
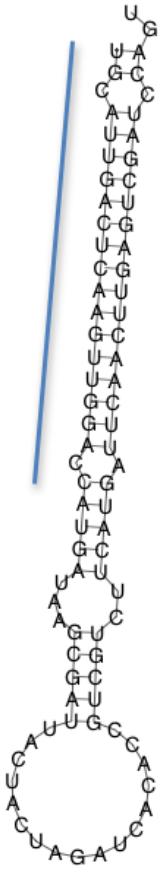


Figure 3: Multiple sequence alignment of all small RNAs within 1 mismatch to hairpin and precursor for leg_cand7x171. This candidate miRNA has sequence similarity to a novel miRNA, miR8155 from papaya.

A



B



Figure 4: Predictions for leg_cand1x9933. The hairpin and location of the candidate sequence is shown in (A) and the multiple sequence alignment of the precursor, hairpin, and the candidate sequence (B). The predicted sequence of the miR* is identical to the sequence 4 from the Leggett small RNA library with a read count of 116.

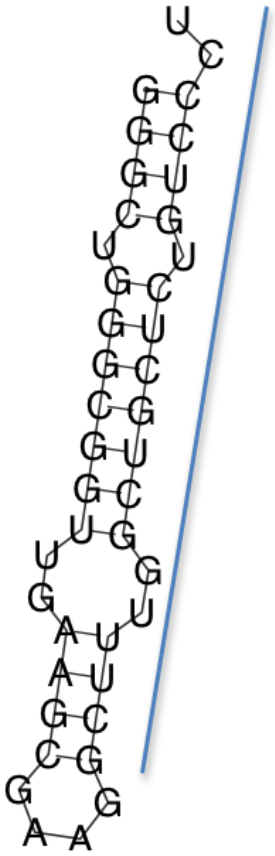


Figure 5: Predicted hairpin for leg_cand2x418. The location of the small RNA sequence to the hairpin is indicated by the blue line. The miR* predicted for this hairpin does not exist in the small RNA library.

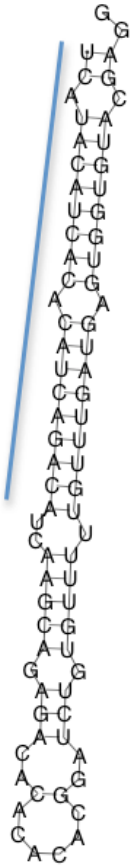
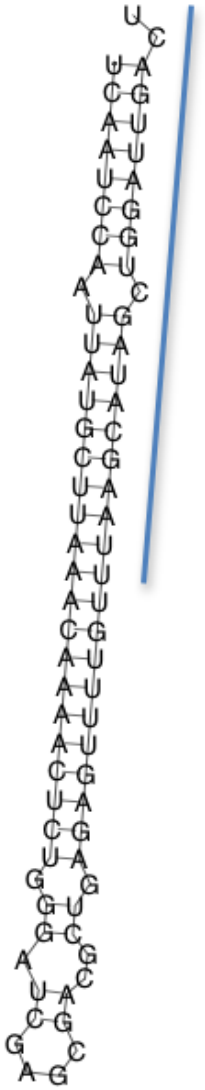


Figure 6: Predicted hairpin for leg_cand3x213. The location of the small RNA sequence to the hairpin is indicated by the blue line. The miR* predicted for this hairpin exists in the small RNA library with a read count of 13.

A



B

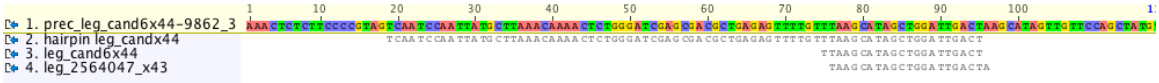


Figure 7: Predictions for leg_cand4x44. The hairpin and location of the candidate sequence indicated by the blue line is shown in (A) and the multiple sequence alignment of the precursor, hairpin, and the candidate sequence (B). The predicted sequence of the miR* does not exist in the small RNA library; however a sequence with a read count of 43 suggests these sequences might be members of the same miRNA family or that the sequence with read count of 43 may be the actual miRNA produced from this hairpin.



Figure 8: Predicted hairpin for leg_cand5x15. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

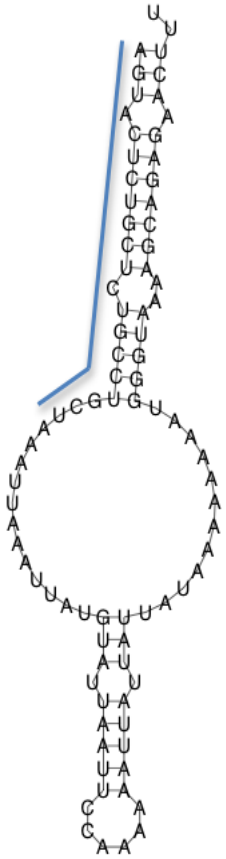
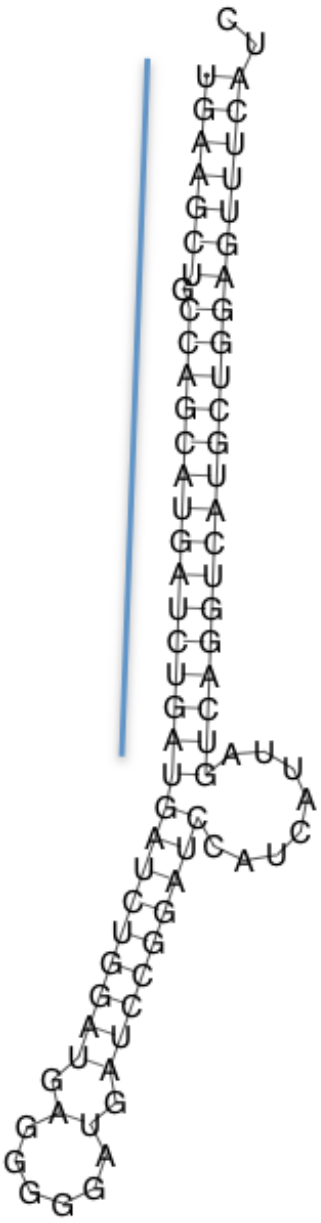


Figure 9: Predicted hairpin for leg_cand6x11. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

A



B



Figure 10: Predictions for jor_cand1x12461. The hairpin and location of the candidate sequence indicated by the blue line is shown in (A) and the multiple sequence alignment of the precursor, hairpin, and the candidate sequence (B). Jor_cand1x12461 is a member of the conserved miR167 family. The predicted sequence of the miR* is identical to the sequence 4 from the Jordan small RNA library with a read count of 781.

A



B

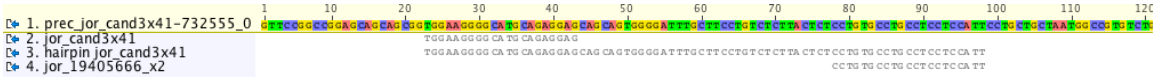
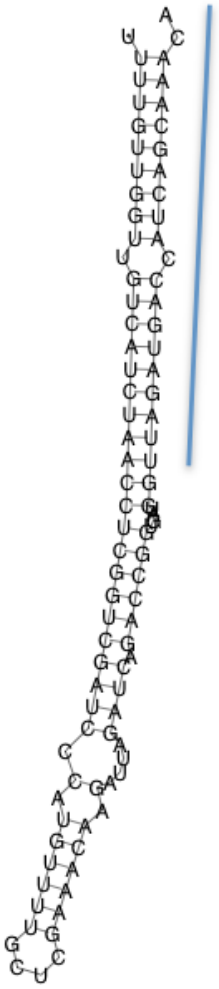


Figure 11: Predictions for jor_cand3x41. The hairpin and location of the candidate sequence indicated by the blue line is shown in (A) and the multiple sequence alignment of the precursor, hairpin, and the candidate sequence (B). Jor_cand3x41 is a member of the non-conserved miR528 family. The predicted sequence of the miR* is identical to the sequence 4 from the Jordan small RNA library with a read count of 2.

A



B

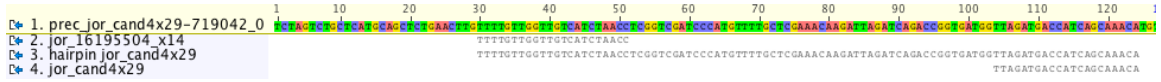


Figure 12: Predictions for jor_cand4x29. The hairpin and location of the candidate sequence indicated by the blue line is shown in (A) and the multiple sequence alignment of the precursor, hairpin, and the candidate sequence (B). Jor_cand4x29 is a member of the non-conserved miR827 family. The predicted sequence of the miR* is identical to sequence 4 from the Jordan small RNA library with a read count of 14.



Figure 13: Predicted hairpin for jor_cand5x29. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

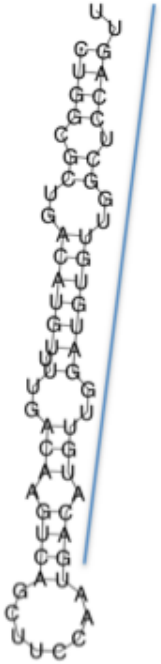


Figure 14: Predicted hairpin for jor_cand6x24. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

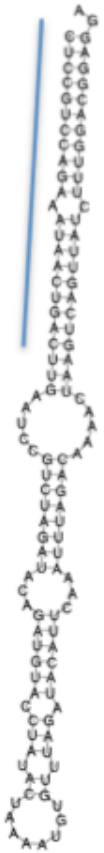


Figure 15: Predicted hairpin for jor_cand7x21. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

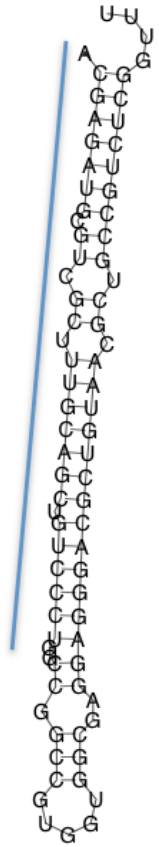


Figure 16: Predicted hairpin for jor_cand8x6. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

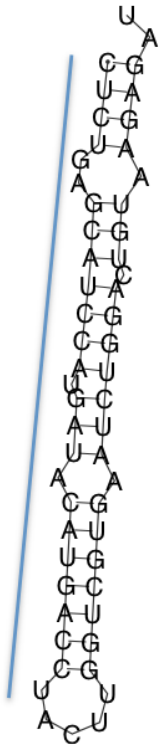
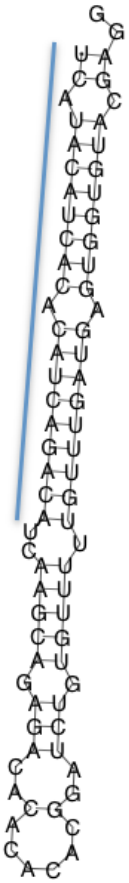


Figure 17: Predicted hairpin for jor_cand9x4. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library

A



B

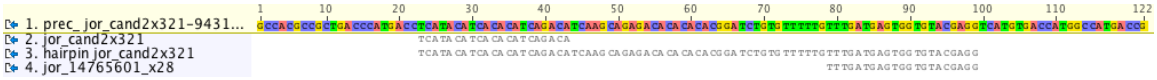


Figure 18: Predictions for jor_cand2x321. The hairpin and location of the candidate sequence indicated by the blue line is shown in (A) and the multiple sequence alignment of the precursor, hairpin, and the candidate sequence (B). The predicted sequence of the miR* is identical to the sequence 4 from the Jordan small RNA library with a read count of 28.

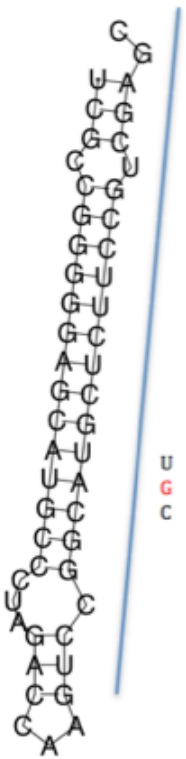


Figure 19: Predicted hairpin for jor_cand11x365. The location of the small RNA sequence to the hairpin is indicated by the blue line. Note that there is a mismatch between the small RNA and the precursor denoted by the red "G". Identical to jor_cand11x365 except for a SNP that matches the precursor perfectly, a small RNA sequence with a read count of 18 is present in the Jordan small RNA library.



Figure 20: Predicted hairpin for jor_cand12x238. The location of the small RNA sequence to the hairpin is indicated by the blue line. Note that there is a mismatch between the small RNA and the precursor denoted by the red "G". Identical to jor_cand12x238 except for a SNP that matches the precursor perfectly, a small RNA sequence with a read count of 64 is present in the Jordan small RNA library.

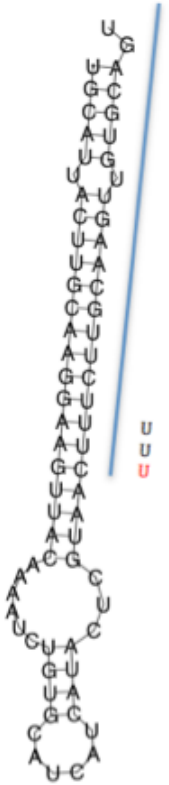
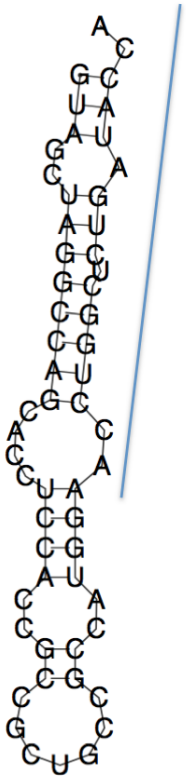


Figure 21: Predicted hairpin for bet_cand15x107. The location of the small RNA sequence to the hairpin is indicated by the blue line. Note that there is a mismatch between the small RNA and the precursor denoted by the red "U". Bet_cand15x107 is predicted to be a member of the non-conserved miR444 family.

A



B

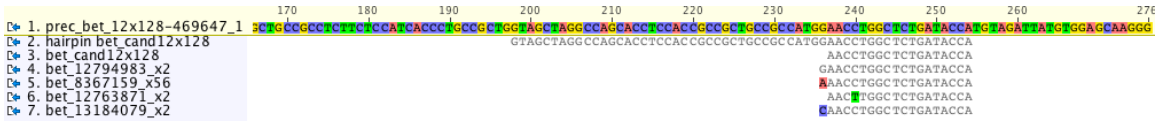


Figure 22: Predictions for bet_cand12x128. The hairpin and location of the candidate sequence indicated by the blue line is shown in (A). Due to the asymmetrical bulge in the miR* region, this hairpin failed the plant prediction criteria. T multiple sequence alignment of the precursor, hairpin, and the candidate sequence is shown in (B). Bet_cand12x128 has sequence similarity to the dicot miRNA family miR1511. The predicted miR* sequence does not exist in the Betania small RNA library. The sequence with a read count of 56 may be a member of the same miRNA family as bet_cand12x128.



Figure 23: Predicted hairpin for bet_cand3x54. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

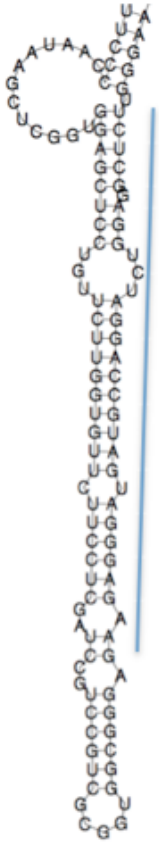


Figure 24: Predicted hairpin for bet_cand4x28. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library

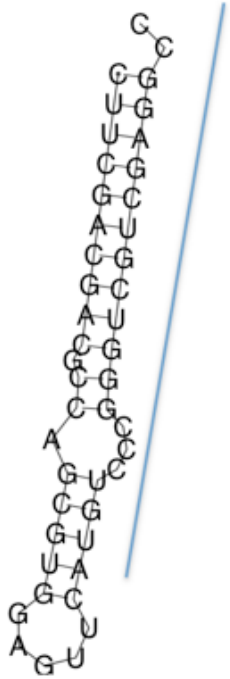


Figure 25: Predicted hairpin for bet_cand5x14. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

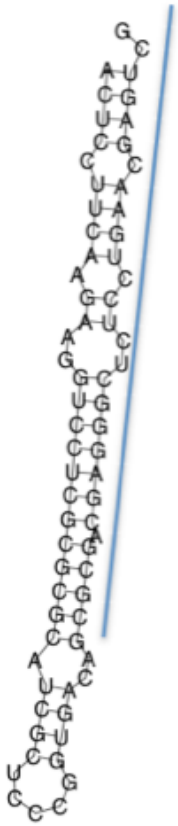


Figure 26: Predicted hairpin for bet_cand6x14. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

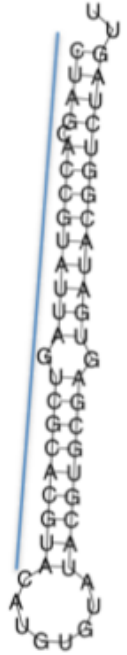


Figure 27: Predicted hairpin for bet_cand8x7. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

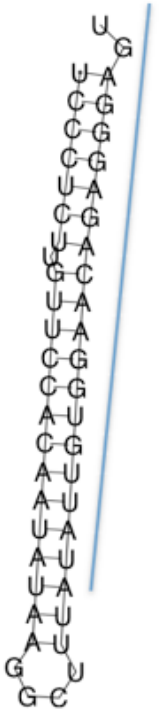


Figure 28: Predicted hairpin for bet_cand9x5. The location of the small RNA sequence to the hairpin is indicated by the blue line. The predicted sequence of the miR* does not exist in the small RNA library.

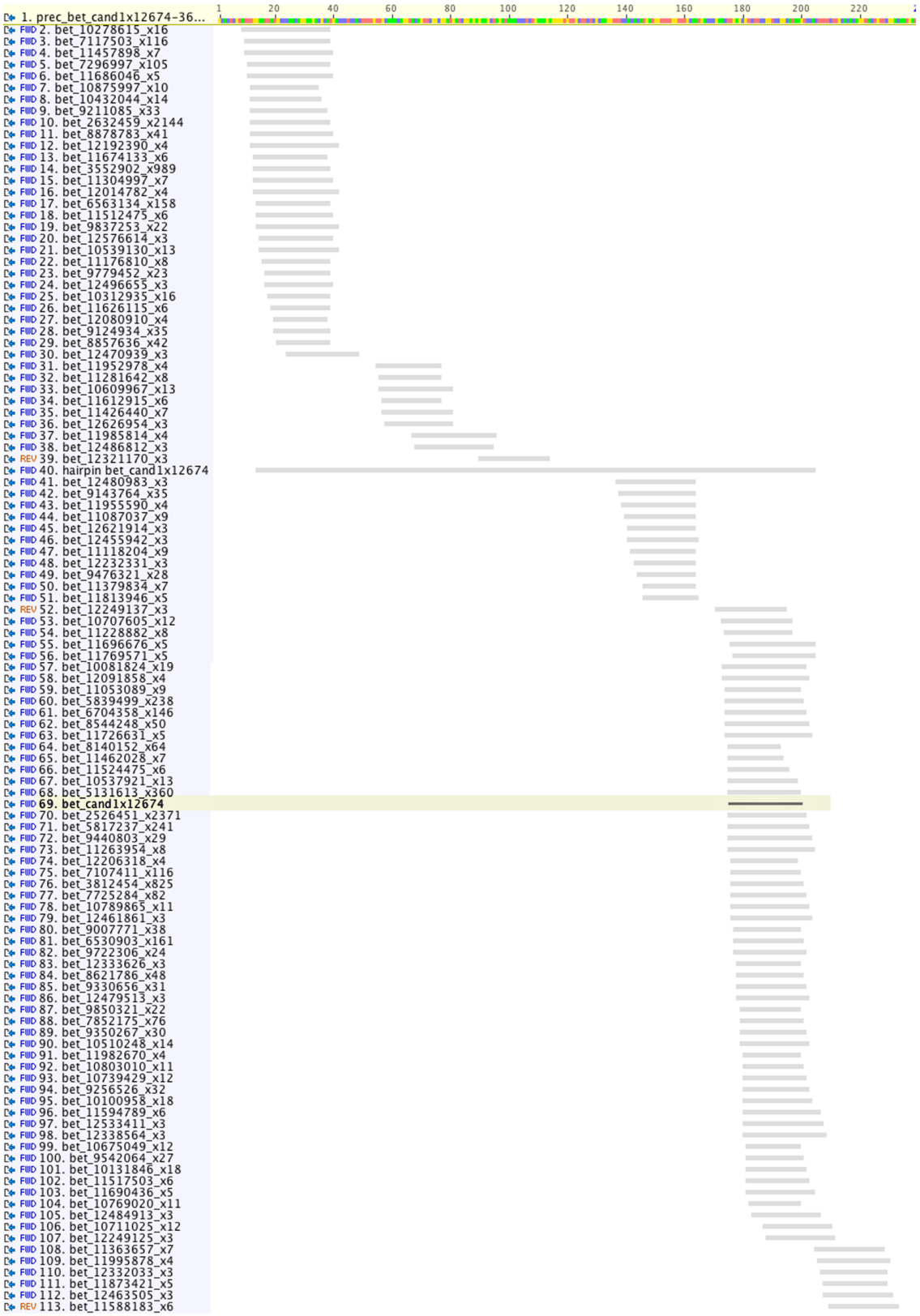


Figure 29: Bet_cand1x12674 as a candidate for loop-to-base processing. Shown here is a multiple sequence alignment of the precursor, hairpin, and all small RNA sequences with 0 mismatches to the hairpin; the candidate sequence is highlighted. The four regions of alignment to the precursor suggest a base-to-loop processing of the hairpin.

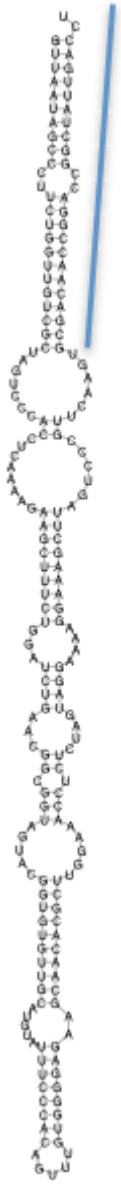
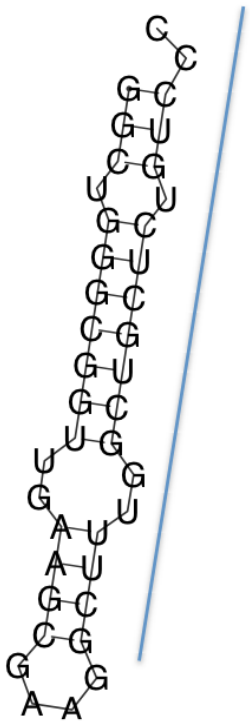


Figure 30: Predicted hairpin of bet_cand1x12674. The blue line indicates the location of the mature candidate sequence.

A



B

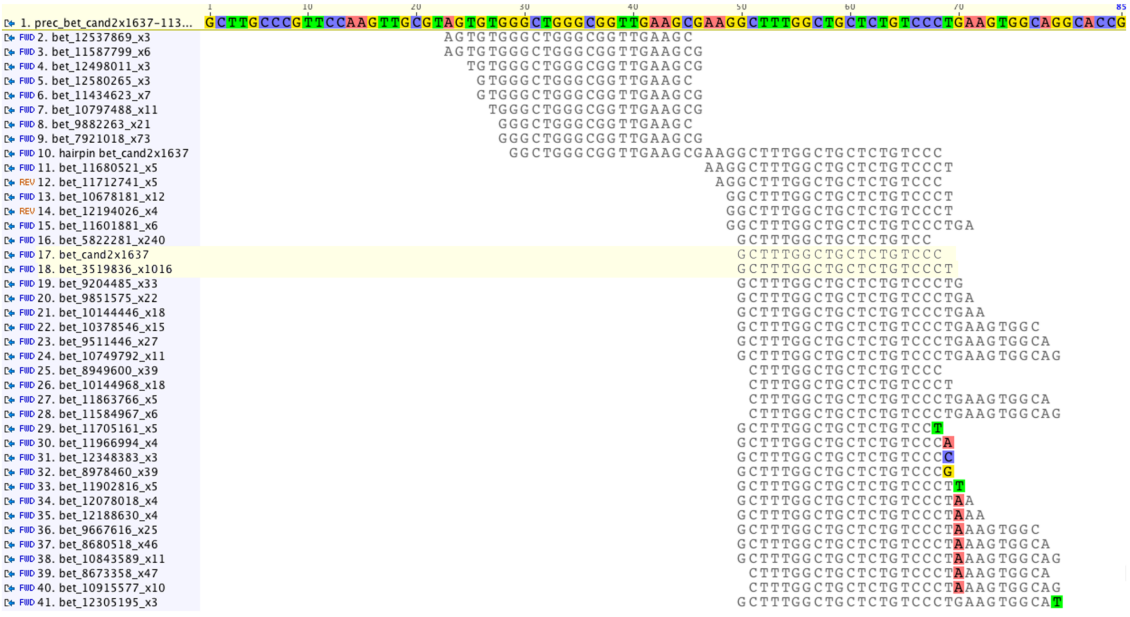
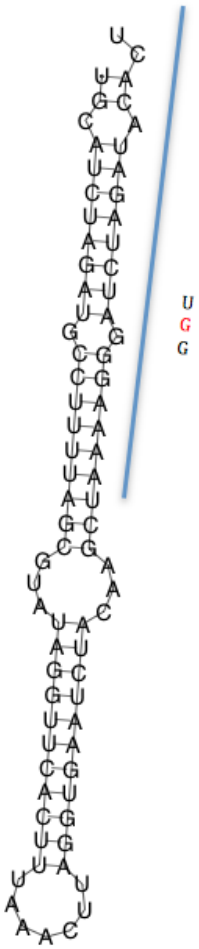


Figure 31: Predictions for bet_cand2x1637. The hairpin and location of the candidate sequence indicated by the blue line is shown in (A). The multiple sequence alignment of the precursor, hairpin, and the candidate sequence is shown in (B) and highlighted. Sequences aligning within on mismatch, especially sequence 18 with the high read count of 1016, may be members of the same miRNA family as bet_cand2x1637 or be the true miRNA cleaved from this hairpin.

A



B

FWD 1. bet_9613025_x26	TAAAAGAGATCTAGATACACT
FWD 2. bet_cand13x391	TAAAAGGGTCTAGATACACT
FWD 3. bet_10704185_x12	AAAAGGGTCTAGATACACT
FWD 4. bet_12179374_x4	CTAAAAGGGTCTAGATACAC
FWD 5. bet_12588692_x3	CTAAAAGGGTCTAGATACACT
FWD 6. bet_12562850_x3	TAAAAGGGTCTAGATACAC
FWD 7. bet_12889153_x2	CTAAAAGGGTCTAGATACACTT
REV 8. bet_12572846_x3	AGGGATCTAGATACACTTCTT
FWD 9. bet_7359980_x101	TAAAAGGGATCTAGATACACT
FWD 10. bet_8432392_x54	AAAAGGGATCTAGATACACTC
FWD 11. bet_12006630_x4	AAAGGGATCTAGATACACTC
FWD 12. bet_12227432_x3	CTAAAAGGGATCTAGATACAC
FWD 13. bet_12405008_x3	CTAAAAGGGATCTAGATACACTC

Figure 32: Predictions for bet_cand13x391. The hairpin and location of the candidate sequence indicated by the blue line is shown in (A). Note that there is a mismatch between the small RNA and the precursor denoted by the red "G". A multiple sequence alignment of all sequences aligning to the hairpin within one mismatch are shown in (B). Identical to bet_cand13x391 except for a SNP that matches the precursor perfectly, the small RNA sequence 9 with a read count of 101 is present in the Betania small RNA library, representing the likely candidate miRNA produced from the precursor. The 13 sequences here may be member of the same novel candidate miRNA family.

CHAPTER 3: MIRNA DISCOVERY IN *AVENA* SPECIES OF THREE PLOIDY LEVELS

3.1 Introduction

The phylogeny of the Aveneae tribe, also called the oat tribe, has long been difficult to decipher. Within this tribe, the *Avena* genus has historically encompassed the genera Helictotrichon, Avenochloa, and Arrhenatherum (Baum, 1977). Traditionally, the Aveneae have been characterized by morphologic traits related to their archetypical spikelet form of long glumes (relative to spikelet length) and a tendency toward a reduced number of flowers per spikelet, commonly 1, 2, or 2–3 per spikelet (Quintanar, Castroviejo, & Catalan, 2007). However, within the tribe, reclassifications have been proposed since the 1800s based on interpretations different classification characters. This is a result of different scoring schemas and variable weights being put on each character for classification. Older classifications have relied on micromorphologic differences and the establishment of clear sterility barriers; however, DNA-based phylogenies have called for reclassifications and have eluded to a complicated phylogeny of the Aveneae tribe (Baum, 1977).

The phylogenetic reconstruction of the Aveneae tribe using plastid DNA and nuclear rRNA sequences has resulted in incongruence among the phylogenies, weak support for placed lineages, and misplaced lineages (Quintanar et al., 2007). Although ITS sequences have challenged the validity of Aveneae phylogenies, these sequences

have been useful in the karyotype evolution of the *Avena* genus (Rodionov, Tyupa, Kim, Machs, & Loskutov, 2005).

Oat, genus *Avena*, exists in three ploidy levels: hexaploid, tetraploid, and diploid. The common, cultivated oat is an allohexaploid (2n = 6x = 42), of which the genome is composed of three subgenomes (AACCDD). Tetraploid oat (2n = 4x = 28) is composed of two subgenomes (AACC, AABB, or CCDD). An auto-tetraploid CCCC has been classified as *Avena* (Rodionov et al., 2005) although previous classifications have included it with Helictotrichon or as a separate genera (Baum, 1977). Diploid oats contain either the AA or CC subgenome combination (2n = 2x = 14). Table 1 shows the genome configurations of the *Avena* genus.

The phylogeny of cultivated hexaploid oat remains elusive. It has been suggested that cultivated hexaploid oat evolved from a divergence of a common ancestral diploid into the A and C genomes, which then converged into the tetraploid. A diploid AA ancestral genome donated the A genome and a CC ancestral genome donated the CC genome to the tetraploid AACC (Thomas, 1992). Of these diploids, variants to each genome have been identified (C: Cv, Cp) and (A: As, Ac, Ad, Al) (Fu & Williams, 2008; Rodionov et al., 2005). The exact identity of the AA and CC genome donors is not definitive. Chloroplast and nuclear gene trees (personal communication, Tim Langdon), ITS phylogenies and karyotype studies (Rodionov et al., 2005), and inferred phylogenies from AFLP studies (Fu & Williams, 2008) propose different progenitors involved in the series of polyploidy events in oat. This may be due to the use of genes with complicated gene histories, gene loss during hybridizations, and translocations between the subgenomes. As a result, deciphering which diploid genome variant contributed to the

tetraploid and which tetraploid is the progenitor of the common hexaploid oat *Avena sativa* is a challenge.

In regards to the D genome, a diploid DD genomic configuration has not yet been discovered. Genomic *in situ* hybridization of AA genome specific markers showed visualization of the probes in 28 of the 42 chromosomes of the hexaploid, pertaining to the AA and DD subgenomes, also supporting the divergence of the AA genome. This was the case until the satellite probe As120 from an AA diploid *Avena strigosa* could be used to distinguish between the A and D genomes (Linares, Ferrer, & Fominaya, 1998). The difficulty in distinguishing the A from D genome using sequence probes supports the hypothesis that the D genome arose from the A genome during a duplication event. Table 1 summarizes the genome configurations of wild and domesticated oat genetic resources and Figure 1 shows the suggested polyploidization pattern of hexaploid common oat.

Although the identity of the wild progenitors of common oat is unknown, wild oats have been sources of improvement of common oat. Breeding common oat with wild oats with desirable agronomic traits such as disease resistance and high yield has resulted in improved common oat characteristics (Valentine, Cowan, & Marshall, 2011). In attempt to improve the wild oat germplasm, Ladizinsky transferred domestication characteristics to the wild tetraploid oat through crossing common hexaploid oat (*Avena sativa*) with a wild tetraploid (*Avena magna*).

An extension of the domestication transfer to wild oat was to create a synthetic hexaploid from the diploid *Avena strigosa* cultivar Saia (AA genome configuration) and the domesticated tetraploid with the CCDD genome configuration (Figure 2). The CCDD

genome configuration of wild tetraploid *Avena magna* was suggested by cytogenetic studies to discriminate between the A and D genomes using a satellite DNA marker (Linares et al., 1998). Using the diploid as the female parent and colchicine to induce chromosome doubling, a synthetic hexaploid was created (Ladizinsky, 2000). Staining fixed anthers of the synthetic hexaploid verified chromosome number (2n=42). Like the tetraploid parent, the synthetic hexaploid is very susceptible to Barley Yellow Dwarf Virus (BYDV) and like the diploid parent, it is also resistant to crown rust.

In trying to understand the gene transfer of *Avena strigosa* to the synthetic hexaploid, transcriptome sequencing of the developing seed has been done. The availability of sequencing resources from the genome configurations *Avena strigosa* (AA), *Avena magna* (CCDD), and the synthetic hexaploid (AACCDD) provides a system for an in depth study of the mechanisms of gene regulation in allopolyploids. As RNA-mediated gene regulation operates at the whole genome scale, there exists a delicate modulation of RNA interference pathways that controls the effect of plant micro RNAs (miRNAs) originating from each subgenome in allopolyploids (Z. J. Chen & Ni, 2006; Hegarty & Hiscock, 2008).

MiRNAs are a class of non-coding RNAs with post-transcriptional gene regulation duties. Plant miRNAs primarily originate from intergenic regions (Rajagopalan et al., 2006) and exhibit trans-acting down-regulation of their targets through utilization of RNA interference machinery (Shabalina & Koonin, 2008). MiRNAs have not been annotated or identified in any configuration of oat. Small RNA sequencing provides the necessary sequencing for annotation of known miRNAs and the prediction of candidate novel, or oat-specific miRNAs. MiRNA identification in the

ploidy series of diploid AA, tetraploid CCDD, and synthetic hexaploid AACCDD genome configurations provides an initial investigation into how miRNAs from different origins interact to affect phenotype and fitness in the course oat allohexaploid evolution through identifying A genome-specific miRNAs.

3.2 Materials and Methods

3.2.1 Total RNA Sequencing Tissue and RNA Extraction

Seed from each line (*A. magna, A. strigosa*, and synthetic hexaploid) was planted in 35.4 cm pots in a 1:1:1 (v:v:v) sand, peat, potting soil mix. They were grown in growth chambers with a 15-hour photoperiod and a 20°/15° Celsius light/dark temperature cycle. Flowers were monitored for pollination and the flowers of the desired stage: 1-3 days post-anthesis (A), 4-6 DPA (B), 7-9 DPA(C) was collected (.5 grams total from multiple plants). Figure 3 shows the morphology of an oat plant and the flower development stages of interest. Collection of tissue and total RNA was extracted by USDA-ARS in Aberdeen, Idaho. The UltraClean Plant RNA Isolation Kit (Mo-Bio, Carlsbad) following kit instructions was used for total RNA extraction.

3.2.2 Total RNA Sequencing and Assemblies

Sequencing library preparations were performed by DHMRI at the North Carolina Research Campus (NCRC) in Kannapolis, North Carolina. Sequencing-by-synthesis using the Illumina HiSeq2000 was done on all three cultivars, stages A, B, and C. 5ug of RNA from each cultivar (pooled seed development stages) was input into the library preparation. Samples were run on separate lanes of a flow cell.

The Bioinformatics Service Division performed the RNA-seq assemblies. Velvet (Zerbino & Birney, 2008) and Oasis (Schulz et al., 2012) were used for the *de novo*

transcriptome assemblies. The Velvet optimizer was used to estimate k-mer size and Oases set the minimum coverage cutoff of 8. Three stage-specific transcriptome assemblies from each cultivar were pooled together for miRNA discovery. Summary statistics of the assembly of the HiSeq transcriptome assemblies from the three cultivars is shown in Table 2.

3.2.3 Small RNA Sequencing Tissue Collection and RNA Extraction

Tissue collection and small RNA extraction for tetraploid *Avena magna* was done by USDA-ARS in Aberdeen. Flower tissue stages A, B, and C were pooled with mature embryos, roots, and shoots for extraction from plants grown in the conditions used for total RNA sequencing. The MiraZol Small RNA Isolation Kit (BiooScientific, Austin) was used per manufacturer's instructions for extraction of RNA and enrichment of small RNA.

For *Avena strigosa* and the synthetic hexaploid, four plants of each cultivar were planted in a 52x18x14 cm pot with Pete's Potting Mix (Daddy Pete's Plant Pleaser, Stony Point) mixed with ¼ cup Tree & Shrub fertilizer/insect protector (Bayer Advanced, Research Triangle Park). Plants were grown in a 19°C and a 26°C Celsius growth chamber with an 18 hr. photoperiod. Three flowers (lemma, palea, ovary, and anthers) from each plant were pooled by stage and cultivar in a tube, collected on ice, and stored in the -80°C freezer. Stages collected were pre-A stage (before flowering or 0 DPA), A (1-3 DPA), B (4-6 DPA), and C (7-9 DPA). The MiraZol Small RNA Isolation Kit (BiooScientific, Austin) was used to extract small RNA from the flower tissue. A modified protocol was used for the extraction (Emir Islamovic, UC Berkeley). The Agilent Bioanalyzer 2100 RNA nano6000 chip and reagents (Agilent, Santa Clara)

ensured lack of phenol contamination and measured the concentration of the small RNA samples.

3.2.4 Small RNA Sequencing and Sequence Pre-processing

Samples from tetraploid *Avena magna* were prepared and sequenced on the HiSeq2000 by DHMRI at the NCRC in Kannapolis, NC. Samples from diploid Avena magna and the synthetic hexaploid were prepared for sequencing using the Illumina TruSeq sRNA Sample Prep Kit (Illumina, San Diego). Following manufacturer's instructions, 3' and 5' adapters were ligated to the short RNA fragments and SuperScriptII (Invitrogen, Carlsbad) facilitated reverse transcription. After ligating unique adapters to each sample, 18 cycles of PCR were used for cDNA enrichment. Samples were pooled equimolarly based on the molarity estimates from the High Sensitivity DNA chip and the Agilent Bioanalyzer2100 instrument (Agilent, Santa Clara). As per the Truseq instructions, the pooled library was run on a 6% pre-cast TBE PAGE gel (Invitrogen, Carlsbad), stained with ethidium bromide and visualized with a UV transilluminator. A custom ladder size 145 and 160 base pairs guided the excision of adapter-ligated cDNA in that size range. Fragments of size 147 bases correspond to miRNAs. Ethanol precipitation of the purified gel product was evaluated for desired size on the Agilent Bioanalyzer DNA1000 chip. Eurofins MWG Operon (Huntsville) performed the small RNA sequencing on the Illumina HiSeq2500.

The fastx_quality_filter tool of the FASTX-Toolkit (Hannon Lab, CSHL) was used for quality filtering of FASTQ raw sequence files. The filter parameters were –q 15 –p 90. In-house python scripts, available upon request, trimmed the 5' and 3' adapters and sequences length 18-30 were retained. Small RNA library sequences from the same

cultivar were pooled for annotation purposes. Reads were collapsed using the collapse_reads_md.pl from miRDeep2 (Friedlander et al., 2012) for input into miRDeep-P. The number of sequences remaining after each of the pre-processing steps is shown in Table 3. Pre-processing of small RNA sequencing and annotation procedure is summarized in Figure 4.

3.2.5 Identification of Known MiRNAs

All mature miRNAs were downloaded from miRBase release 20 (Kozomara & Griffiths-Jones, 2014). Vmatch (Kurtz, 2011) was used to create a suffix tree and perform the matching tasks of the small RNA sequences to known miRNAs in miRBase. The mkvtree program indexed the known miRNAs and the vmatch program solved the matching tasks over the index specifying complete and direct matches between the query and indexed sequences. A hamming distance constrained the number of mismatches between the query and subjects, returning matches with two or fewer mismatches. Vmatch parameters were –h 2 as mentioned above for hamming distance, –complete for complete matches, and –d to specify direct matches.

3.2.6 Identification of Candidate MiRNA Precursors

A modified miRDeep-P (X. Yang & Li, 2011) pipeline was used to identify miRNA precursors. HiSeq data were indexed and aligned to the corresponding pre-processed small RNA libraries by Bowtie-1.0.0 (Langmead et al., 2009) allowing for 1 mismatch. Precursor excision length of 276, as suggested by Thakur, 2011 to correspond to 96% of all plant miRNA precursors, was excised using a miRDeep-P Perl script. Bowtie indexed the filtered precursors and aligned the small RNA sequences to create signatures. Candidate precursors were folded with RNAfold (McCaskill, 1990; Zuker &

Stiegler, 1981) with default parameters. miRDP.pl scored the precursors based on the miRDeep probabilistic model (Friedlander et al., 2008) and candidate precursors that do not meet the plant specific criteria are filtered out of the final prediction by the script rm_redundant_meet_plant.pl. Mismatches between the mature sequence and the complementary arm cannot be greater than 4 and an overhang greater than 2 is not allowed. Using the Rfam web-server (Burge et al., 2012) to search the database of non-coding RNAs with BLAST and covariance models, non-coding RNA hairpins incorrectly identified as miRNA precursors were eliminated for further analysis.

3.2.7 Multiple Sequence Alignment of Small RNAs and Transcript Assemblies

Geneious v5.6.7 (Biomatters, New Zealand) was used for multiple sequence alignments of the selected small RNAs to determine which sequences are likely members of the same miRNA family. To align small RNA sequences to predicted hairpins and the transcript assemblies from which they were predicted, a Geneious Alignment was performed using a 51% similarity cost matrix with a gap open penalty of 12 and a gap extension penalty of 3. These parameters were used for all multiple sequence alignments.

3.3 Results and Discussion

It is challenging to annotate miRNAs from a species without a reference genome. A sequence comparison of known, annotated miRNAs to sequences in a small RNA sequencing library can identify miRNAs. These include both conserved and non-conserved families as well as miRNAs currently annotated in other species as novel. Discovery of these sequences in small RNA libraries of other grasses will require the renaming and classification of the sequence as lineage specific and not novel.

There is a disparity between miRNA identification and annotation. Annotation requires identification of a miRNA precursor gene. In the absence of a reference genome, transcriptome sequencing makes identification of candidate novel miRNAs possible. Utilizing a pipeline designed to identify miRNAs in the absence of a reference genome (Figure 4), this study is the first to identify known and candidate novel miRNAs in a ploidy series- diploid oat *Avena strigosa*, a domesticated tetraploid oat cultivar *Avena magna*, and an oat synthetic hexaploid. In addition to discovering candidate novel miRNAs in these cultivars, predicted candidate miRNAs in common between *Avena strigosa* (AA) and the synthetic hexaploid (AACCDD) were searched in the *Avena magna* (CCDD) small RNA library to identify candidate miRNAs that may be specific to the A genome. The naming scheme for the candidate miRNAs is the cultivar identifier ("str" for *Avena* strigosa, "ama" for *Avena* magna, and "syn" for the synthetic hexaploid) followed by a number specifying the candidate number, and an "x" followed by the read count of the candidate miRNA sequence in the small RNA library of the oat species it was found in.

3.3.1 Identification of Known MiRNAs in *Avena strigosa*

Known miRNAs were identified in the bulk flower small RNA library by comparing the 1.7 million high quality, trimmed, non-redundant sequences against known miRNA sequences in miRBase. Figure 3 shows the morphology of the oat flower and the stages at which flower tissue was collected for small RNA sequencing whereas Table 3 shows the number of small RNAs at each step in the pre-processing. All mature miRNA sequences were downloaded from miRBase v.20. Plant mature miRNAs were filtered for searching. Vmatch (Kurtz, 2011) allowed two mismatches between the query

small RNA sequence and the annotated, known miRNA from miRBase. All 20 conserved miRNA families were identified in the library (Supplementary Table 1). In addition, many non-conserved miRNA families and previously annotated "novel" miRNAs from other species were also identified (Supplementary Table 2). For previously annotated "novel" miRNAs, the miRNA identifier and the corresponding species is also given. Note that numerous sequences from the small RNA library have sequence similarity to multiple "novel" miRNAs from more than one species. To reduce redundancy, only one annotated miRNA is assigned to a sequence present in the small RNA library despite sequence similarity to more than one "novel" miRNA. The total number of reads per miRNA family and the number of unique sequences that have sequence similarity to the annotated miRNAs is shown in Table 4. The most abundant miRNA family is miR166 with over 300,000 reads and 328 members.

The miR165/166 is a family of miRNAs conserved across diverse plant lineages. The family targets HD-ZIPIII (Class III Homeodomain Leucine-Zipper) transcription factors, which have been implicated in leaf polarity and root vasculature (Barik et al., 2014; Kidner & Martienssen, 2004). Although there are 328 unique sequences that can be classified as part of the miR165/166 family, it is doubtful there exist 328 miRNA genes. *Arabidopsis thaliana* contains two miR165 genes and seven mir166 genes; *Oryza sativa* has fourteen miR166 genes, and *Physcomitrella patens* has thirteen. In the sequenced miRNAs from flower tissue of *Avena strigosa*, the members in this conserved miRNA family have slight sequence variations and some occur in the library at only one occurrence. Some of the 328 sequences differ by an additional base on the 3' or 5' end, which may be explained by sequence errors, may originate from the different precursor

genes, or may be alternative DCL1 (DICER1) products from the same precursor gene. Other sequences with one read count may be sequencing errors as well, but were not removed from the Vmatch query due to the correct classification as a member of miR165/166 family. In addition, sequences classified as miR* are included in the classification.

In reference to non-conserved miRNA families described to be specific to the grass lineage, miR444 is the most abundant. miR444 is considered to be a monocot-specific family of miRNAs having been found in rice, sorghum, wheat, barley, maize, and sugarcane (Sunkar et al., 2005). The conserved targets of the miR444 family are MADS-box genes, some of them being involved in flower development (L. Wu et al., 2009) and nitrate signaling (Yan, Wang, Hamera, Chen, & Fang, 2014). miR444 also has a unique self-regulation. In barley, an EST is precursor for miR444a also contains a MADS-box binding site for miR444b (Colaiacovo, Lamontanara, et al., 2012). The absence of reads in the small RNA library for miR441, miR818, miR821, and miR1435 support that these miRNAs are specific to the Ehrhartoideae lineage, which includes sorghum and rice.

3.3.2 Prediction of Previously Annotated MiRNAs in *Avena strigosa*

Of 20 conserved miRNA families, no precursors passed the plant parameter filter by miRDeep-P. However, one hairpin (development stage 1-3 DPA) for a member of the miR396 family was predicted but did not pass the plant parameter filter due to two one-base asymmetrical bulges and 3-base stretch of mismatches (Figure 5). In addition, the candidate miRNA sequence has a mismatch to the transcript, suggesting that the hairpin

is the precursor for a different miR396 member with 0 mismatches to the precursor. This member exists in the *Avena strigosa* small RNA library with a read count of 10.

3.3.3 Predicted Novel Candidate MiRNAs in *Avena strigosa*

The modified miRDeep-P algorithm was used for candidate miRNA prediction of 11 miRNA candidates. If a small RNA sequence with a read count above 1 maps to a contig, the contig consensus sequence is folded to predict secondary structure. The miRDeep core algorithm scores the secondary structure and plant specific parameters filter out hairpins with poor scores. The sequences of the predicted miRNAs and their number of reads are shown in Table 5. Because the HiSeq assemblies from the three development stages were pooled to potentially increase the number of predictions, multiple hairpins predictions could represent multiple genes producing identical DICER products or the same gene, just from two or more stages of development. This is the case for str_cand9x3, str_cand8x6, and str_cand5x8. These three candidate miRNAs have identical hairpins predicted from stages A (1-3 days post anthesis) and B (4-6 days post anthesis). The candidate precursor hairpins are shown for str_cand9x3 in Figure 6, str_cand8x6 in Figure 7, and str_cand5x8 in Figure 8. A multiple sequence alignment of the entire assembled contigs using Geneious 5.6.7 revealed identical contig consensus sequences as well (data not shown).

Other predicted miRNA candidates include two candidates with a read count of only 2, str_cand2x46, and str_cand1x177. It is unlikely a small RNA sequence with a read count of only 2 is a true miRNA despite a precursor hairpin being predicted from the HiSeq assembly of developing oat flower cDNA library sequencing (Figure 9 and 10). In addition to a low read count for str_cand10x2, an asymmetrical bulge of 6 bases on the

complementary arm could influence DCL1 processing of the miRNA from the precursor (Figure 9). The sequence of str_candx46 has very high GC content. A technical issue with the extraction method could explain the high GC content. The Trizol Reagent has been found to enrich for high GC content sequences (Y. K. Kim, Yeo, Ha, Kim, & Kim, 2012). Although the MiraZol Reagent was used in small RNA isolation, it is very similar in composition to the Trizol Reagent. For this high GC-content miRNA candidate, two distinct hairpins were assembled from cDNA reads from stage A (1-3 days post anthesis) of *Avena strigosa* flowers (Figure 11 (A) and (B)). The hairpin for str_cand1x177 (Figure 12) has 6 mismatches between the candidate sequence and complementary strand even though the miRDeep-P plant filter only allows 4 mismatches between the mature miRNA and the opposite arm of the hairpin. 4 of the 6 mismatches are GU wobble pairs, which are not penalized as heavily as other mismatches following Watson-Crick base pairing rules. Hairpins for other candidate miRNAs predicted by the modified miRDeep-P pipeline are shown in Figures 13-16.

In allowing one mismatch between the small RNA and the potential precursor, the sequence of str_cand5x8 (Figure 8), str_cand3x28 (Figure 13), and str_cand2x46 (Figure 11) does not match the precursor identically. This indicates that these candidate miRNAs are not true miRNAs, that there is an error in the assemblies, or that these candidate miRNAs originate from a hairpin other than the one predicted from the HiSeq assemblies.

3.3.4 Identification of Known MiRNAs in *Avena magna*

Of 1.2 million small RNA sequence matches to known miRNAs in miRBase with fewer than 3 mismatches, miR166 was the most abundant known miRNA family with 73,000 reads aligning to known miR166 sequences. The second most abundant miRNA

family is the miR396 family. The conserved miRNA family miR396 targets the GRF (GROWTH-REGULATING FACTOR) family of transcription factors as well as additional non-conserved targets. miR396 spatially-temporally represses GRF to regulate cell growth to ultimately control the number of cells in plant leaves (Rodriguez et al., 2010). Within the miR396 family, there exists a monocot variant of the mature miRNA sequence. This isoform has a one base insertion between base 7 and 8 which when binding to its target GRF, the insertion causes the monocot variant to bind to GRF in a more energetically stable manner. The insertion promotes a more energetically stable binding by eliminating a bulge between the conserved miR396 and GRF, creating a more efficient targeting model in monocots (Debernardi, Rodriguez, Mecchia, & Palatnik, 2012). As a monocot, this variant exists in the oat *Avena magna* small RNA library. Supplementary Table 3 shows all sequences present in the *Avena magna* library within 2 mismatches to known conserved miRNA family members from other plant species, annotated in miRBase v. 20.

miR444 is the most abundant non-conserved family; however, unlike *A. strigosa* miR528 is very abundant as well with 508 total reads and 10 members (Table 4). miR528 targets ascorbate-oxidase, a cell wall enzyme. Extensive studies into how miR528 is involved in tolerance to oxidative stress have not of yet been done (L. Wu et al., 2009). Other non-conserved miRNA families and even previously annotated as "novel" miRNAs are represented in the small RNA library of flower, root, shoot, and mature embryo tissue. The sequences and respective miRNA identification within 2 mismatches is given in Supplementary Table 4.

3.3.5 Predicted Conserved and Non-conserved MiRNAs in *Avena magna*

Using the modified miRDeep-P pipeline, precursor hairpins passing plant filtering criteria were predicted from the HiSeq transcriptome assemblies for 7 candidate miRNA sequences (Table 6). Of known miRNAs annotated in other species and present in miRBase v. 20, hairpins for a non-conserved miR528 family member (Figure 17) and a conserved miR397 family member were predicted. The precursor of miR528 has been used in the vector delivery of artificial miRNAs into plant systems (Manavella & Rubio-Somoza, 2011). The mature sequence of miR528 and miR* are replaced by the artificial miRNA and complement for delivery. A precursor for a member of conserved miRNA family 396 was also predicted but did not pass plant criteria filtering due to two one-base asymmetrical bulges and a 3-base loop of mismatches (Figure 18). There exists a mismatch between the mature miR396 sequence (ama_cand8x9574) and the hairpin. The predicted hairpin may be misassembled or may be a precursor for a different member of the miR396 family.

In the HiSeq assembly of stage A of development, two different hairpins for a member of the miR397 miRNA family were found (Figure 19 (A) and (B)), suggesting two genes for miR397 observed in the assemblies. miR397 is a highly conserved miRNA family and has been found to be involved in abiotic stress response such as drought stress in rice (Zhou et al., 2010) and both drought and salt stress in switchgrass (Sun, Stewart, Xiao, & Zhang, 2012). The predicted target of miR397 is laccase. In *Populus*, the overexpression of miR397 results in reduced levels of laccase, and in turn, a reduced level of lignin (Lu et al., 2013). Although lignin has critical importance in wood formation in *Populus*, lignin is a component of the cell wall in all plants. The down-

regulation of miR397 under abiotic stress would induce lignin production and the strengthening of the cell wall, providing protection from pathogens and other stresses.

A hairpin was also predicted for ama_cand3x20 with 4 mismatches to known *Brachypodium distachyon* miR7732. There are no known or predicted targets for this miRNA; however it is drought responsive (Bertollini et al., 2013). Identical hairpins were predicted for this candidate miRNA from both stage A and stage B of flower development, corresponding to one transcript. There is a three-base asymmetrical bulge in the mature miRNA and the mature sequence produced by the transcript does not match the candidate sequence in the library perfectly (Figure 20).

3.3.6 Predicted Novel Candidate MiRNAs in *Avena magna*

The existence of a mismatch between the small RNA sequence and the precursor is also the situation with ama_cand4x13 (Figure 21), ama_cand5x11 (Figure 22), ama_cand6x2 (Figure 23), and ama_cand7x2 (Figure 24). With a mismatch, these may not be true miRNAs or precursors of mature miRNAs. Just as any read with only one occurrence was eliminated for miRNA discovery, if these read counts were normalized, a count of two would most likely fall beneath a detection threshold. Improved sequencing depth of the transcriptome of the tetraploid oat would increase the discovery rate of miRNAs. 454 sequencing identified 7 additional candidate miRNAs of which 5 are novel miRNA candidates (Appendix).

3.3.7 Identification of Known MiRNAs in the Synthetic Hexaploid

All 20 conserved miRNA families are represented with high sequence similarity to small RNAs in the synthetic hexaploid (Table 4). The annotation and sequences with two or fewer mismatches to known miRNAs annotated in miRBase v.20 is shown in

Supplementary Tables 5 and 6. Like both the *Avena strigosa* and *Avena magna* small RNA library, miR166 family is the most abundant miRNA family.

3.3.8 Predicted Known MiRNAs in the Synthetic Hexaploid.

For prediction of miRNA hairpins in the synthetic hexaploid HiSeq transcriptome library, 32 hairpins were found for 12 candidate miRNAs. Each precursor forms a hairpin loop and passes plant parameter filtering such as location of the mature miRNA on the hairpin and number of mismatches between the miRNA and the opposite arm. Of the hairpins predicted, only syn_candx25 had a match, with 4 mismatches, to a known miRNA in miRBase, miR7722 from *Brachypodium distachyon.* This is the identical small RNA sequence and hairpin sequence predicted from the *Avena magna* small RNA sequence library and transcriptome assembly. Table 7 shows all candidate miRNAs sequences.

The miRDeep-P pipeline predicted a hairpin for a member of the conserved miR398 family; however the hairpin did not pass the plant filtering criteria. miR398 is thought to be involved a complex gene regulatory network involved in both abiotic and biotic stress response (C. Zhu et al., 2011). The hairpin has a 3 mismatches and a 4-base asymmetrical bulge in the middle of the miRNA/miR* duplex (Figure 25). In allowing one mismatch in the alignment of the small RNA sequences to the HiSeq assemblies, a mismatch between the mature miR398 sequence and the hairpin suggests an error in the assembly, that this hairpin is a non-canonical miR398 precursor that produces a different member of the miR398 family, or a false prediction. In addition to a miR398 member, a hairpin was predicted for a small RNA sequence with sequence similarity (4 mismatches)

to known miR2929 from *Oryza sativa* (data not shown). This precursor did not pass plant filtering.

3.3.9 Predicted Novel Candidate MiRNAs in the Synthetic Hexaploid

Four identical hairpins assembled from the cDNA in stage A (1-3 DPA) were predicted for one of the candidate miRNAs with 9 read counts, syn_cand9x9 (Figure 26 (A)). Although the hairpins are identical, a multiple sequence alignment using Geneious 5.6.7 revealed four distinct transcripts, suggesting 4 miRNA genes (Figure 26 (B)); however, it cannot be determined from which genome the transcripts originate.

Additional candidate miRNA precursors passed plant criteria filtering (Table 7). Of candidate novel miRNAs with read counts below 20 not already discussed, only syn_cand8x9 and syn_cand11x5 have a mismatch between it and its precursor sequence (Figure 27 and 28). The hairpins of the other predicted miRNA candidates are shown in s Figures 29-33.

3.3.10 A Novel Candidate MiRNA Family in the Synthetic Hexaploid

The miRDeep-P pipeline also predicted a candidate family of miRNAs. Two hairpins were predicted for syn_cand3x86. The sequences of these two hairpins were found in cDNA assemblies from stage A, B, and C (Figure 34 (A) and (B)). Alignment of the two hairpins and the six cDNA contig consensus sequences from which they were predicted reveals that two transcript assemblies correspond to each hairpin, supporting the presence of two miRNA genes that produce identical miRNA sequences (Figure 34 (C)). The candidate miRNA syn_candx86 has a mismatch between it and the predicted precursor. The small RNA sequence that matches the precursor identically was not predicted for the precursor due to its lower read count, 29, but it does exist in the small

RNA library. This suggests that the two hairpins are precursors for the non-predicted syn_candx29; however, differing by one nucleotide, candidate syn_cand3x86 and the non-predicted syn_candx29 are most likely members of the same family of miRNAs.

Identical hairpins from stage A (1-3 DPA) of flower development were predicted for syn_cand1x336. Despite only one hairpin prediction from the cDNA assemblies for syn_cand1x336 (Figure 35 (A)), a sequence alignment of three cDNA contig consensus sequences from which the hairpin was predicted reveals three different transcripts (Figure 35 (B)). Two are very similar, supporting the A-D genome hypothesis. Alternatively, these transcripts may originate from distinct loci in the same subgenome. Nonetheless, there exists a mismatch between the candidate miRNA and the hairpin suggesting a false positive, a mis-assembly, or that the hairpin produces a different miRNA. Upon evaluating the predictions of every small RNA that aligns within one mismatch to the predicted hairpin for syn_cand1x336, we identify syn_candx298 as the small RNA with 0 mismatches to the predicted precursor for syn_cand1x336. Syn_candx298 was scored lower due to its lower read count, but perfectly matches the hairpin and is more likely the true miRNA produced from that hairpin shown in Figure 35 (A). Also aligning to the predicted precursor for syn_cand1x336 (true pre-syn_cand2x298) with one mismatch is a small RNA sequence with a read count of 29. This is the same sequence of the non-predicted syn_candx29 that aligns perfectly to the hairpin predicted for syn_cand3x86. This discovery adds another potential sequence to the miRNA family: non-predicted syn_candx29, syn_cand3x86, syn_cand2x298, and syn_cand1x336.

The hairpin predicted for syn_cand2x298 (Figure 36), assembly from stage C (7-9 DPA), has one mismatch between it and its corresponding miRNA sequence. The

miRNA sequence that would be produced by the hairpin exists in the library with a read count of 11. Non-predicted syn_candx11 may also be a member of the miRNA family.

To identify other potential members of this candidate miRNA family, we aligned all small RNAs that align within one mismatch to the predicted precursors for syn_cand3x86, syn_cand2x298, and syn_cand1x336 (Figure 37). Sequences with only 2 read counts may not be true miRNAs and may be sequencing errors but are still included in the multiple sequence alignment. These low-read-count sequences need to be validated for classification as members of this newly discovered candidate miRNA family. With such high sequence similarity and the lack of a predicted precursor for every candidate miRNA, deciphering which sequences are mature miRNAs and which are the miR* sequences cannot be done. Figure 38 shows a reference assembly of all candidate miRNA family members assembled toto one of the predicted precursors for syn_cand1x336. Although the precursor is most likely the hairpin that produces syn_cand2x298, a miR* sequence with 0 mismatches to the precursor cannot be found. A precursor was predicted for another member of the candidate miRNA family, syn_cand8x9 (Figure 27).

3.3.11 Predicted A-genome Candidate MiRNAs

Predicted miRNAs in the diploid *Avena strigosa* (AA) by the modified miRDeep-P pipeline were searched for in the synthetic hexaploid (AACCDD) small RNA library and ones predicted in the synthetic hexaploid library were evaluated for presence in the diploid small RNA library. Candidate A-genome-specific miRNAs would be present in the *Avena strigosa* library (AA) and the synthetic hexaploid (AACCDD) but not in the *Avena magna* (CCDD) small non-redundant filtered sequences. 6 sequences are

candidates to be A-genome-specific (Table 8). Of these predicted candidate miRNAs, none have sequence similarity within 4 mismatches to known, annotated miRNAs in miRBase v. 20 (Kozomara & Griffiths-Jones, 2014). These candidates need to be experimentally validated by stem-loop RT-PCR as true miRNAs. Potential function can be predicted using psRNATarget webserver (Dai & Zhao, 2011). Using default parameters and the transcripts of the MSU Rice Genome Annotation version 7 as the pre-loaded transcripts, Table 8 also shows the potential target annotation in rice.

3.4 Conclusions

MiRNAs can be effective annotated in small RNA sequencing libraries. The genome complexity and size of oat has discouraged genome sequencing projects, including small RNAs. Although annotation of miRNAs is ideal with a reference genome, transcriptome assemblies or ESTs have been shown to provide enough evidence for the presence of a miRNA precursor using the miRDeep-P pipeline (K. Wang et al., 2012; Xiaozeng Yang & Li, 2012). Transcriptome assemblies of next-generation high throughput sequencing of the diploid *Avena strigosa*, the tetraploid *Avena magna*, and a synthetic hexaploid were used in a modified miRDeep-P pipeline to predict miRNAs based on the presence of a precursor in the assembly. 11 candidate miRNAs were predicted in the *Avena strigosa* small RNA library, 7 in the *Avena magna* library, and 12 in the synthetic hexaploid library. These candidate miRNAs arise from precursor genes that pass the criteria for miRNA biogenesis. Stem-loop RT-PCR would validate the sequences as true miRNAs and evaluate the precision of the miRDeep-P pipeline.

Due to the discovery nature of this project, transcriptome assemblies from different stages were pooled as well as stage and temperature specific small RNA library

sequences. Missing in the small RNA libraries is the synthetic hexaploid stage C of seed development (7-9 days after anthesis, 26° C), *Avena strigosa* stage C (7-9 days after anthesis, 19° C) and *Avena strigosa* stage pA (prior to anthesis, 26° C). The *Avena magna* small RNA library has miRNAs present from tissues other than only seed development stages and would not be included in a differential expression comparison. To be able to perform a statistical analysis of small RNA expression across the ploidy series, small RNAs from only the pA, A, B, and C stages of development from flower tissue of *Avena magna* under the two different growth temperatures, 19° C and 26° C would need to be done. Inclusion of these tetraploid *Avena magna* small RNA sequences from flower tissue and these missing data would provide a spectrum of small RNA libraries to do temperature and stage comparisons among these oat cultivars of different levels of polyploidy. After normalization, individual miRNA expression profiles of particular miRNAs or miRNA families can be evaluated for differential expression.

Another result of this study may call for a more definitive classification of lineage-specific miRNAs. The absence of miR1435, miR441, miR818, and miR821 in all libraries suggest they are not conserved in the Pooideae lineage. These miRNAs have been annotated in rice and sorghum only, suggesting these miRNAs are conserved in the Ehrhartoideae lineage. Although precursors were not predicted for these non-conserved miRNA families, precursors were predicted for known miRNAs annotated only in one species, such as miR7732 from *Brachypodium*. If stem-loop RT-PCR validates this sequence as a real miRNA in *Avena*, this miRNA may need to be reclassified as non-conserved and not novel. Reclassifications are common and to be expected as more miRNA sequencing projects in plants ensue.

Annotation of miRNAs in these libraries is not complete. Clearly, there are issues with coverage of the RNA-seq assemblies but this is to be expected with the large size of the oat genome. In addition, genome complexity challenges the accuracy of the HiSeq assemblies. In the early alignment step of the miRDeep-P pipeline, allowing for one mismatch in alignments has its advantages in applying it to a large genome with low coverage of transcriptome assemblies. Allowing one mismatch may lead to false positives; however, because miRNA families have high sequence complementarity, members of a miRNA family can be identified even though a precursor cannot be identified for every sequence. Evaluating each precursor for all small RNA sequences aligning within one mismatch in the predictions output file of the miRDeep-P pipeline before applying the plant criteria has been useful in identifying candidate members of a potential miRNA family in the synthetic hexaploid. High-count candidates syn_cand1x336 and syn_candx298 exist in both the *Avena strigosa* and *Avena magna* small RNA libraries, providing support for this candidate novel oat miRNA family. Not allowing a mismatch would reduce the number of total predictions because not all miRNA precursors are represented in the HiSeq assemblies.

With the most abundant conserved miRNA in all three small RNA libraries being miR166, we would expect to find a precursor sequence in all libraries. Of all 20 conserved miRNA families, sequences from all families were present in the small RNA sequencing libraries, however, the modified miRDeep-P pipeline discovered precursors in transcriptomes for only conserved miR397 from the tetraploid. Using available 454 sequencing of total cDNA from developing flowers from these cultivars of oat, a miR159 precursor was also predicted for *Avena strigosa* and a miR166 precursor for *Avena*

magna (see Appendix). With the upcoming release of the oat functional genome (P2EP), we anticipate the refined miRNA annotation to include precursor predictions for all conserved miRNA families and some non-conserved miRNAs. We can then be able to approximate how many miRNA genes correspond to each family. In addition, refined annotation should include the candidate novel miRNA predictions here, as well as new novel candidates. However, challenges still remain without a fully sequenced genome. Annotation of special cases such as miRNAs originating from introns (miRtrons) and alternatively spliced miRNAs due to stress conditions may not be included in the refined annotation.

Despite the need for refined annotation, this is the first miRNA sequencing project in an oat diploid, tetraploid, and synthetic hexaploid. It provides an ideal platform to discover miRNAs that are genome specific. MiRNAs predicted in the diploid and the synthetic hexaploid were evaluated for presence in the opposite small RNA sequencing library. Candidate miRNAs that are present in the diploid (genome configuration AA) also present in the synthetic hexaploid (genome configuration AACCDD) were searched for in the tetraploid small RNA library (CCDD). The candidates absent in the tetraploid are potential A-genome-specific miRNAs. Validation of presence of these sequences by stem-loop RT-PCR in *Avena strigosa* and the synthetic hexaploid total RNA but absence in total RNA from the tetraploid *Avena magna* would confirm that these 6 sequences are indeed miRNAs and that they are A-genome specific. These sequences can then be used as markers to distinguish A from D genomes in oat polyploids.

The small RNA sequencing from this ploidy series in combination with the HiSeq assemblies used for miRNA annotation here and degradome sequencing by PARE (Parallel Analysis of RNA Ends) (Addo-Quaye et al., 2008; German, Luo, Schroth, Meyers, & Green, 2009) will allow an improved annotation of miRNAs and their respective targets. We can then start investigating the effect of different miRNAs on the morphology and fitness of oat. For example we may be able to identify a miRNA or a particular A-genome miRNA target that confers crown rust resistance in the diploid *Avena strigosa* and synthetic hexaploid or a C or D genome miRNA/target that contributes to the susceptibility of both *Avena magna* and the synthetic hexaploid to BYDV. It is these correlations of miRNAs and traits that can be used to introduce improved agronomic traits into breeding common oat.

Table 1**:** Polyploidy in oat. Oat has many different genomic configurations. As a diploid, there are variants of the genomes. Tetraploids exist in combinations of variant diploid genome configurations. Cultivated oat is a hexaploid.

Avena Genome Configurations	Variants
AA	Ac, As, Al, Ad
CC	Cm, Cp, Cv
AABB	
AACC	
CCCC	
CCDD	
AACCDD	

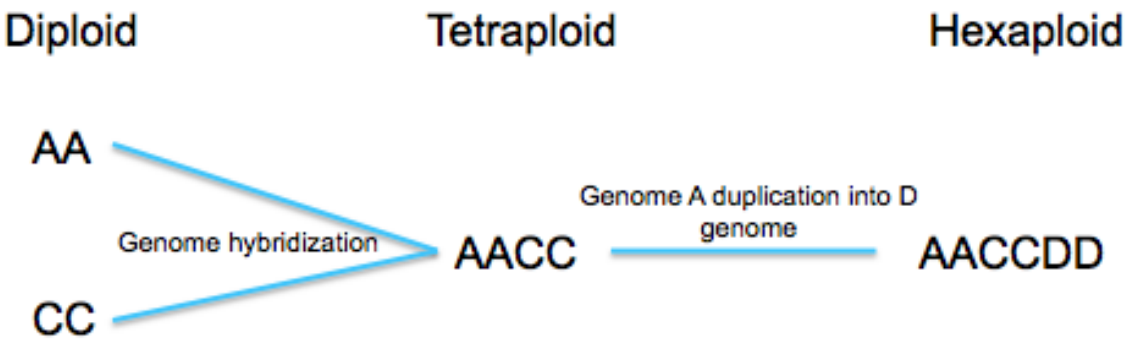


Figure 1: Proposed evolution of common hexaploid oat. Diploid AA and CC genomes hybridized to form a tetraploid. It is proposed that the A genome underwent a whole genome duplication event to form the hexaploid.

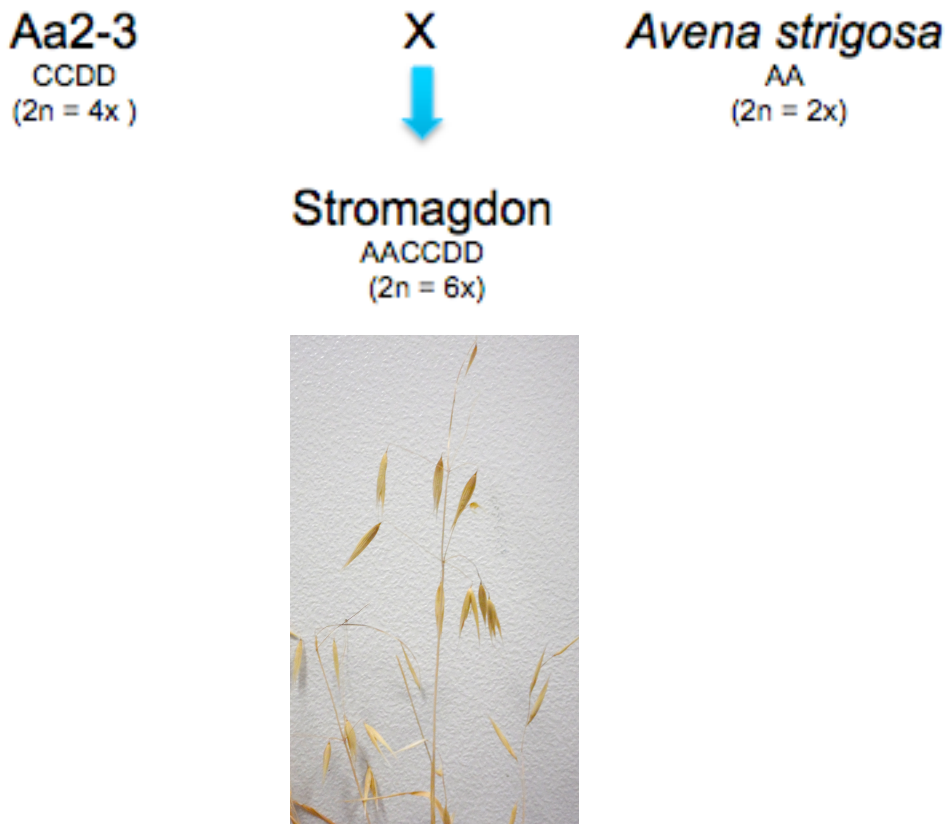


Figure 2: Creation of the synthetic hexaploid. Diploid *Avena strigosa* was crossed with the domesticated tetraploid to create the synthetic hexaploid. Stromagdon is one name proposed for this synthetic oat.

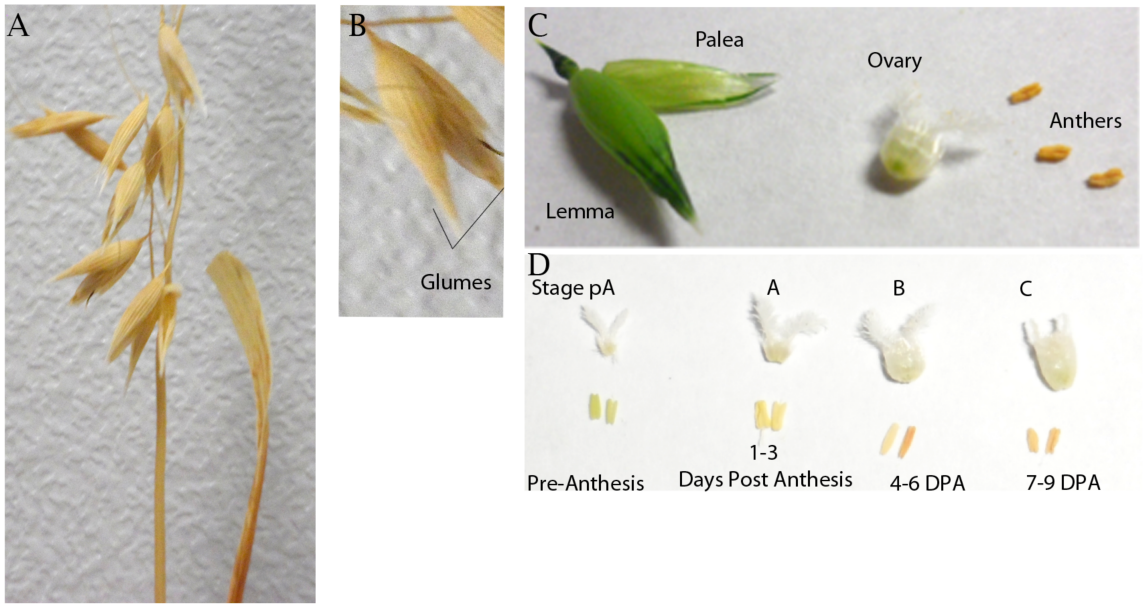


Figure 3: Morphology of oat. The panicle (A) is the branch to which spikelets are attached. A spikelet (B) is composed of individual flowers. Flowers are protected by sterile glumes and reproductive organs lemma and palea (C). The ovary and anthers of four stages in seed/flower development are shown in (D).

Table 2: HiSeq RNA-seq assembly statistics. The table shows the number of RNA-seq assemblies from the three stages of flower development for each cultivar. The N50 sequencing statistic of each assembly is also given.

Cultivar	1-3 Days Post Anthesis (DPA)		4-6 DPA		7-9 DPA	
	Number of Sequences	N50	Number of Sequences	N50	Number of Sequences	N50
Avena strigosa (AA)	14,829	1,677	10,064	1,430	358	753
Avena magna (CCDD)	32,266	1,646	25,903	1,575	223	611
Synthetic hexaploid (AACCDD)	27,209	1,477	29,637	1,456	39,131	1,647

Table 3:Summary read statistics of the small RNA sequencing after each step of pre-processing. The '*' denotes that the tetraploid plants from which flower tissue stages A-C were pooled with mature embryo, root, and shoot tissue for the small RNA sequencing. In addition, these plants were grown with a 20°/15° Celsius light/dark cycle.

Cultivar	Small RNA Library Tissue		Raw reads	Quality Filter	Adapter Trimmed, 18-30	Non-Redundant	Number of Singlets	Filtered Sequences for miRNA Discovery
	19° C	26° C						
Avena strigosa (AA)	pA, A, B	A, B, C	16,002,344	15,111,759	6,620,254	1,759,976	1,390,365	369,611
Avena magna (CCDD)	roots, shoots, mature embryo, flower stage A-C*		15,159,616	8,675,352	6,379,744	1,828,985	1,447,654	381,331
Synthetic hexaploid (AACCDD)	pA, A, B, C	pA, A, B	48,613,969	46,770,485	23,607,454	5,808,930	4,470,305	1,338,625

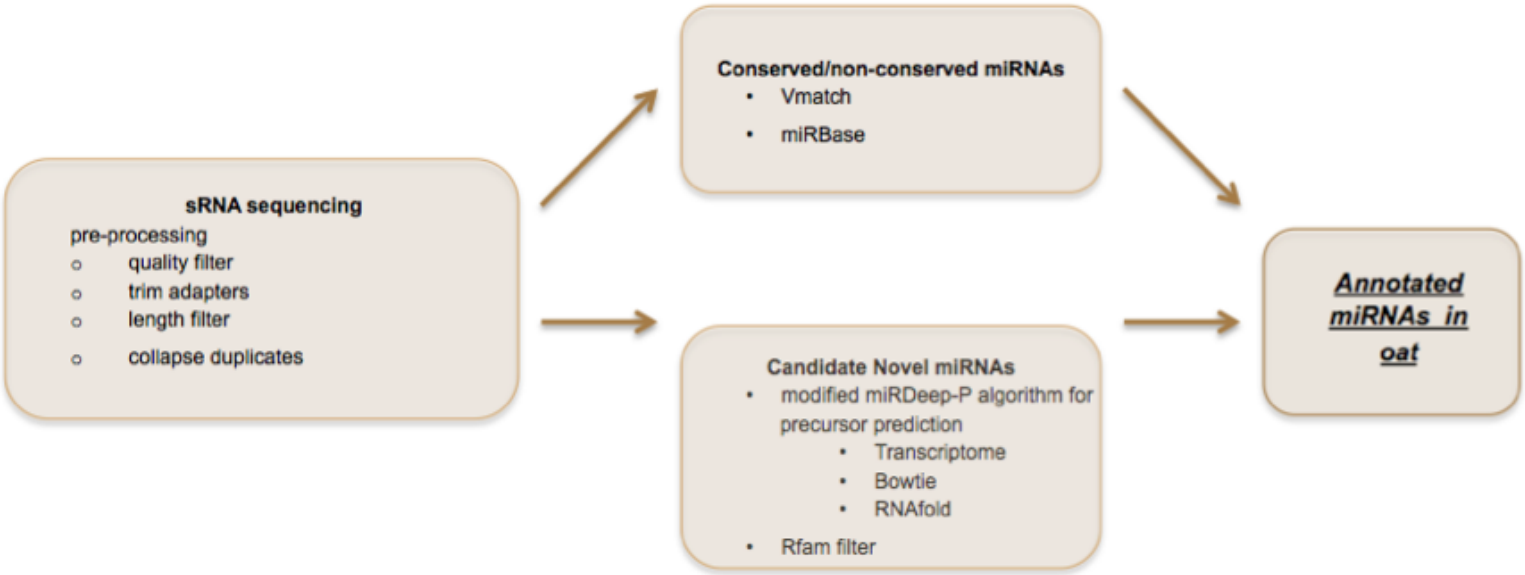


Figure 4: MiRNA annotation pipeline in the absence of a reference genome. Small RNA sequences are pre- processed for quality, adapters, length, and duplicates. Known miRNAs are identified using Vmatch to solve matching tasks between the small RNA sequences and annotated, known miRNAs in miRBase v20. The core miRDeep core probabilistic algorithm is used to score folded hairpins and plant specific parameters filter out hairpins that do not meet the criteria. Finally, the candidates are screened against the Rfam database to eliminate hairpins originating from non-coding RNAs other than miRNAs.

Table 4: Conserved and selected non-conserved miRNA families in *A. strigosa*, *A. magna*, and the synthetic hexaploid. The number of unique reads within 2 mismatches to known conserved or selected non-conserved miRNA families represents the number of potential family members. The number of reads of each classified unique sequence gives a relative abundance of each miRNA family. Over 90% of all miRBase hits are represented by the conserved and non-conserved miRNA families listed here.

	Avena strigosa (350,574 Total miRBase Hits)		Avena magna (119,897 Total miRBase Hits)		Synthetic Hexaploid (1,714,611 Total miRBase Hits)	
Conserved miRNA Family	**% Total Reads**	**Potential Number of Family Members**	**% Total Reads**	**Potential Number of Family Members**	**% Total Reads**	**Potential Number of Family Members**
miR156/157	2.63	115	1.59	53	1.11	169
miR159/319	1.27	110	8.40	169	1.36	191
miR160	< 0.01	3	0.02	8	< 0.01	12
miR162	0.02	4	0.19	5	0.01	8
miR164	< 0.01	5	0.07	7	< 0.01	10
miR165/166	87.33	328	60.90	214	90.52	650
miR167	1.45	67	0.69	31	0.93	107
miR168	0.86	59	2.19	48	0.73	107
miR169	0.04	24	0.30	58	0.11	92
miR170/171	0.06	16	0.50	31	0.05	31
miR172	< 0.01	4	< 0.01	4	< 0.01	9
miR390	0.03	8	0.09	12	0.02	10
miR393	0.05	23	0.04	9	0.08	45
miR394	< 0.01	3	0.02	4	< 0.01	5
miR395	0.03	24	< 0.01	2	0.01	27
miR396	3.08	131	13.41	112	4.04	266
miR397	0.03	2	0.23	5	0.03	8
miR398	0.02	4	0.52	16	0.03	12
miR399	0.01	10	0.04	6	0.02	8
Non-Conserved miRNA Family						
miR408	< 0.01	2	0.01	7	< 0.01	5
miR437	< 0.01	1	0.00	0	< 0.01	3
miR441	0.00	0	0.00	0	0.00	0
miR444	0.39	29	0.83	31	0.13	36
miR528	< 0.01	1	0.42	10	< 0.01	5
miR529	< 0.01	1	< 0.01	1	0.00	0
miR818	0.00	0	0.00	0	0.00	0
miR821	0.00	0	0.00	0	0.00	0
miR827	0.01	2	0.01	2	< 0.01	5
miR1435	0.00	0	0.00	0	0.00	0
miR2118	< 0.01	2	< 0.01	1	< 0.01	4
miR2275	< 0.01	1	0.00	0	< 0.01	11
% Total miRBase Hits:	**97.32**		**90.45**		**99.22**	

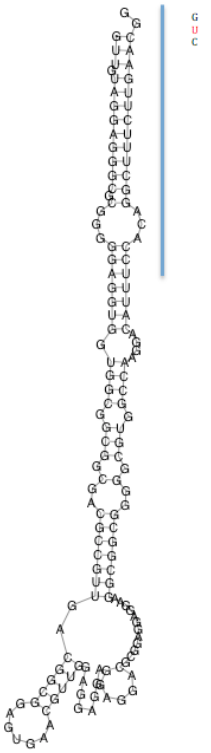


Figure 5: Pre-str_cand12x6047. This precursor was predicted by miRDeep pipeline but did not pass the plant filter criteria. This particular miRNA sequence is a member of the conserved miR396 family. The red "U" indicates the true sequence of the candidate miRNA, which does not match identically the precursor.

Table 5: Predicted candidate miRNAs in *Avena strigosa*. The miRDeep-P algorithm found potential precursors in the RNA-seq assemblies that passed plant criteria. The '*' indicates the hairpin for the candidate miRNA sequence did not pass plant filtering criteria.

Predicted miRNA Candidate ID	Number of Reads	Sequence
str_cand1x177	177	GACATGTTGGATGTGTTGGCTCCAGTT
str_cand2x46	46	GGAGGGCGCGGCGGCCGC
str_cand3x28	28	TTATATTGTGGAACAGAGAGA
str_cand4x13	13	TGTTGCAGATGTGAAGGGCCTGAAGC
str_cand5x8	8	AAAATATAAGACGTTTTAGCT
str_cand6x7	7	GACCCGCTGGCGTCGCAGGACCT
str_cand7x6	6	AACCGGCGAATCCAAGGACTGCAT
str_cand8x6	6	ATTGCGGCGAGTTGCGTCTCTGTT
str_cand9x3	3	ACTATTATCTATGGGTCGACT
str_cand10x2	2	CGCTGGCGCTGCTCGTCGTCTTCTCCGC
str_cand11x2	2	TCAGTTAGAGCCGATGGAACCA
*str_cand12x6047	6047	TCCACAGGCTTTCTTGAACTG



Figure 6: Pre-str_cand9x3



Figure 7: Pre-str_.cand8x6

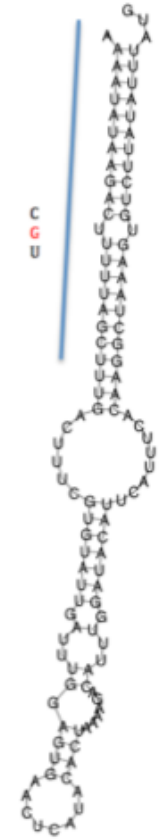


Figure 8: Pre-str_cand5x8. The red "G" indicates the true sequence of the candidate miRNA, which does not identically match the precursor.

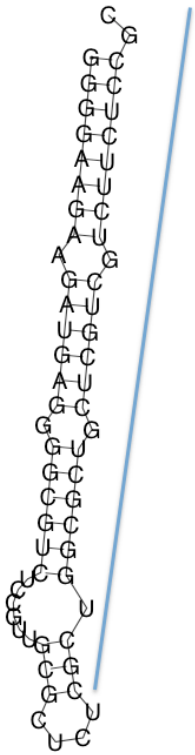


Figure 9: Pre-str_cand10x2



Figure 10: Pre-str_cand11x2

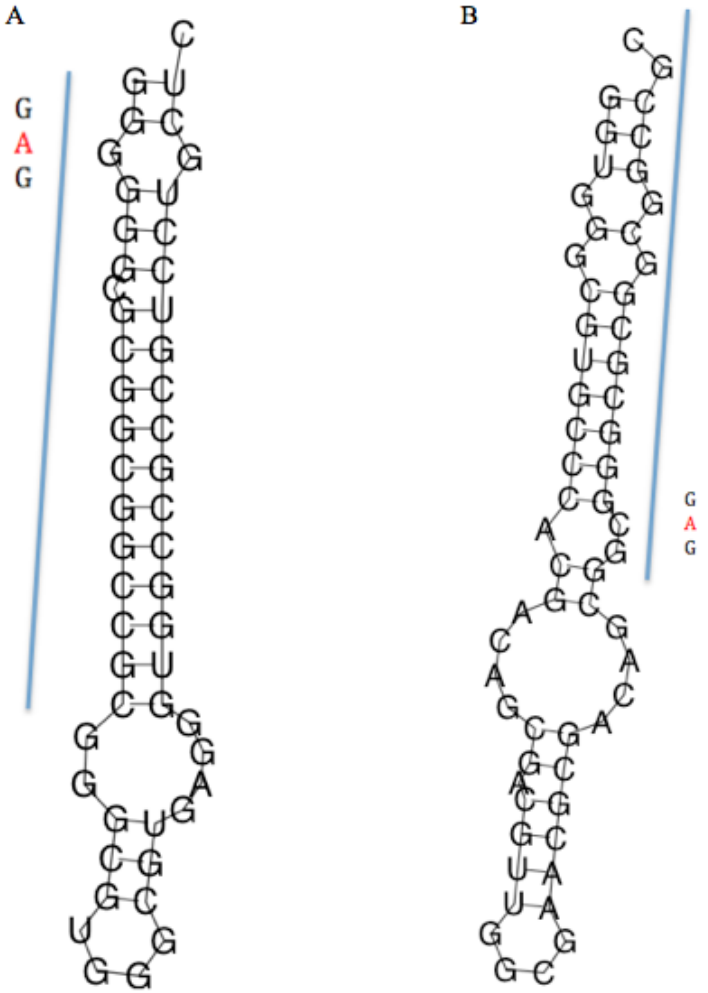


Figure 11: Predicted hairpins for pre-str_can46. For both hairpins 1 (A) and 2 (B), the red "A" indicates the true sequence of the candidate miRNA, which does not match identically the precursor.

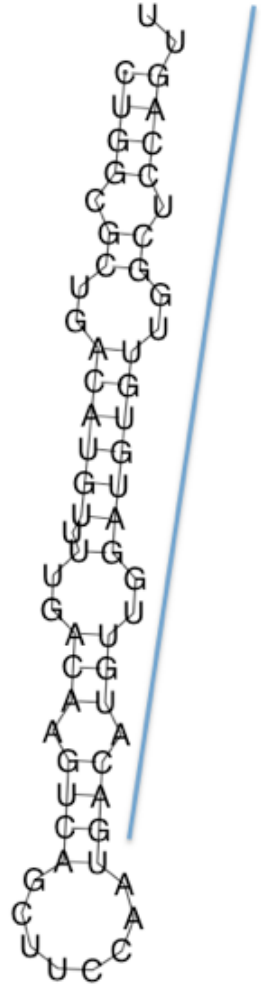


Figure 12: Pre-str_cand1x177



Figure 13: Pre-str_cand3x28. The red "A" indicates the true sequence of the candidate miRNA, which does not match identically the precursor.



Figure 14: Pre-str_cand4x13

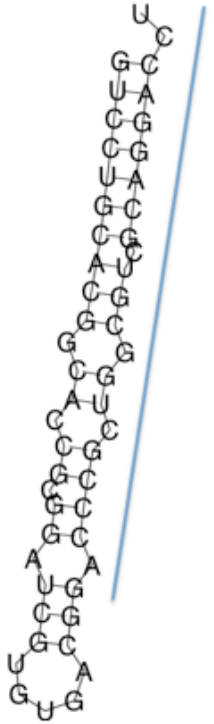


Figure 15: Pre-str_cand6x7



Figure 16: Pre-str_cand7x6

Table 6: Predicted candidate miRNAs in *Avena magna*. The miRDeep-P algorithm found potential precursors in the RNA-seq assemblies that passed plant criteria for the listed candidate miRNA sequences. The annotation of known miRNAs in miRBase v. 20 with less than 3 or 4 mismatches to candidate miRNA sequences is also given. The '*' denotes that the hairpin prediction did not pass plant filtering criteria by miRDeep-P.

Predicted miRNA Candidate ID	Number of Reads	Sequence	miRBase Hit (<3 mismatches)	miRBase Hit (4 mismatches)
ama_cand1x496	496	TGGAAGGGGCATGCAGAGGAG	miR528	
ama_cand2x268	268	TTGAGTGCAGCGTTGATGAAC	miR397	
ama_cand3x20	20	CCACCAAGATCTCTGAGGCCC		miR7732
ama_cand4x13	13	CGCGGCGACGGGGGCGGG		
ama_cand5x11	11	GACGGATCTGGGTAGTACTCC		
ama_cand6x2	2	AGCAGAGCTGGACTGCGGACT		
ama_cand7x2	2	CCTCCTCTCGTCGGCGGCGGCA		
*ama_cand8x9574	9574	TCCACAGGCTTTCTTGAACTG	miR396	

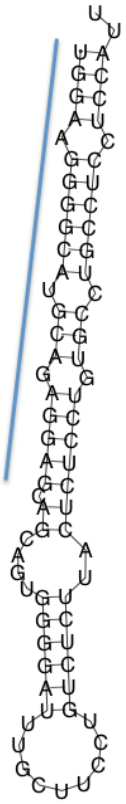


Figure 17: Pre-ama_cand1x496. This candidate miRNA is a member of the non-conserved miR528 family.



Figure 18: Pre-ama_cand8x9574. The red "U" indicates a mismatch between the mature miRNA sequence and the precursor. This precursor did not pass plant criteria by the miRDeep-P prediction pipeline but is a member of the miR396 family.

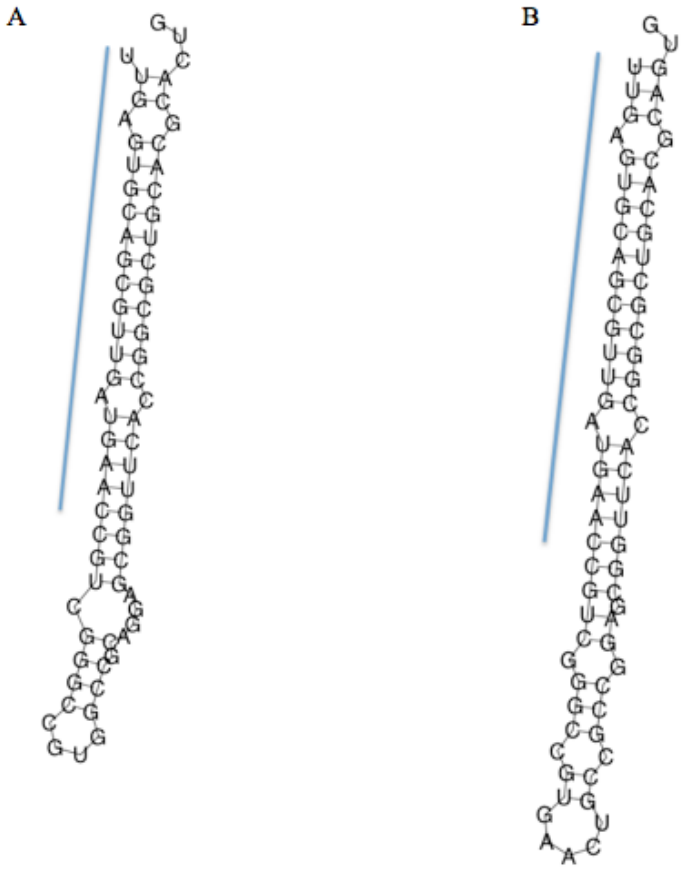


Figure 19: Predicted hairpins for ama_cand2x268. This particular miRNA is a member of the miR397 family. Hairpin 1 is shown in (A) and hairpin 2 is shown in (B).

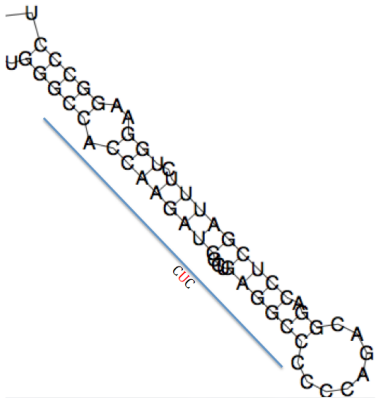


Figure 20: Pre-ama_cand3x20. The red "U" indicates a mismatch between the mature candidate miRNA sequence and the precursor. This candidate miRNA has a read count of 20 and has sequence similarity to miR7732, annotated in *Brachypodium distachyon*.

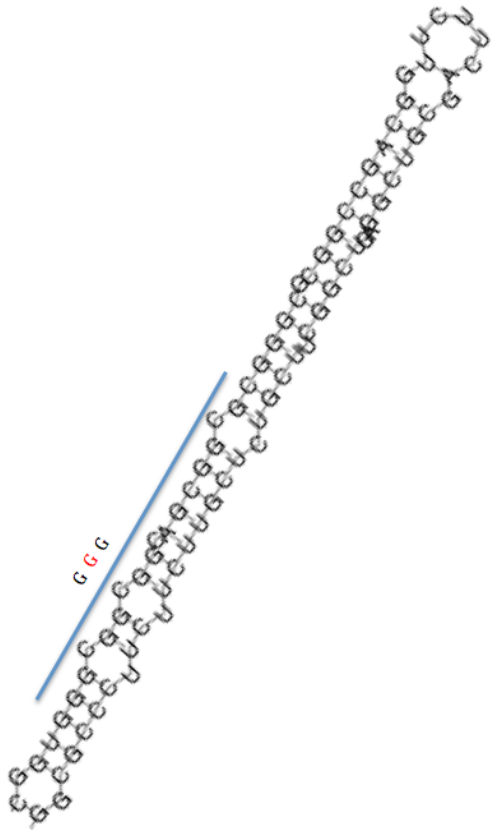


Figure 21: Pre-ama_cand4x13. The red "G" indicates a mismatch between the mature candidate miRNA and the sequence of the predicted precursor.

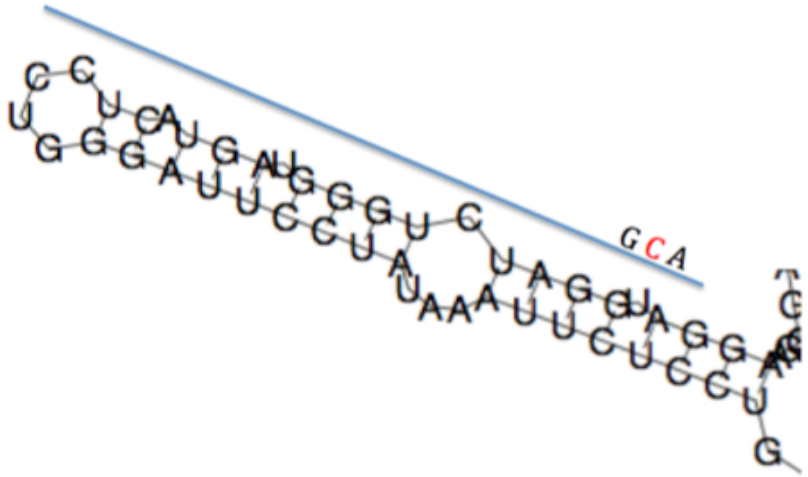


Figure 22: Pre-ama_cand5x11. The red "C" indicates a mismatch between the mature candidate miRNA and the sequence of the predicted precursor.



Figure 23: Pre-ama_cand6x2. The red "U" indicates a mismatch between the mature candidate miRNA and the sequence of the predicted precursor.

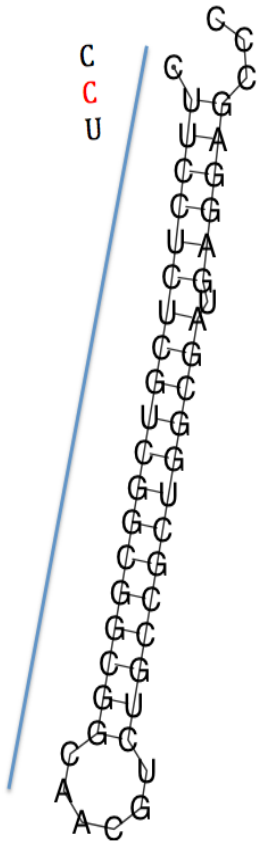


Figure 24: Pre-ama_cand7x2. The red “C” indicates a mismatch between the mature candidate miRNA and the sequence of the predicted precursor.

Table 7. Predicted candidate miRNAs in the synthetic hexaploid. The miRDeep-P algorithm found potential precursors in the RNA-seq assemblies that passed plant criteria for the listed candidate miRNA sequences. The one annotation of known miRNAs in miRBase v. 20 with 4 mismatches to candidate miRNA sequences is also given. The '*' indicates the hairpin predicted for the candidate miRNA sequence did not pass the plant filtering criteria of the miRDeep-P pipeline.

Predicted miRNA Candidate ID	Number of Reads	Sequence	miRBase Hit (<3 mismatches)	miRBase Hit (4 mismatches)
syn_cand1x336	336	TCTGTTCCACAATATAAGCCT		
syn_cand2x298	298	TCTGTTCCATAATATAAGCCT		
syn_cand3x86	86	TTATATTGTGGAACAGAGAGA		
syn_cand4x25	25	CCACCAAGATCTCTGAGGCCC		miR7732
syn_cand5x19	19	TCCTCGACAGGCCTGTGGGCGTCCT		
syn_cand6x15	15	GAAGGGGAGGGCCTGAAGCCCT		
syn_cand7x10	10	GCCAATCTCCTGCTGAGTCAGACT		
syn_cand8x9	9	TCCGTTCCACAATATAAGCCT		
syn_cand9x9	9	TGCCGGGTCGCAGTGGACGTCC		
syn_cand10x5	5	TTACTGGCTGAACTCAAGAATT		
syn_cand11x5	5	CACTTATTTCCGGACGGAGGG		
syn_cand12x2	2	TGAGGTTGAACGTGCTCTCAGGGTGCT		
*syn_cand13x346	346	TGTGTTCTCAGGTCGCCCCTG	miR398	

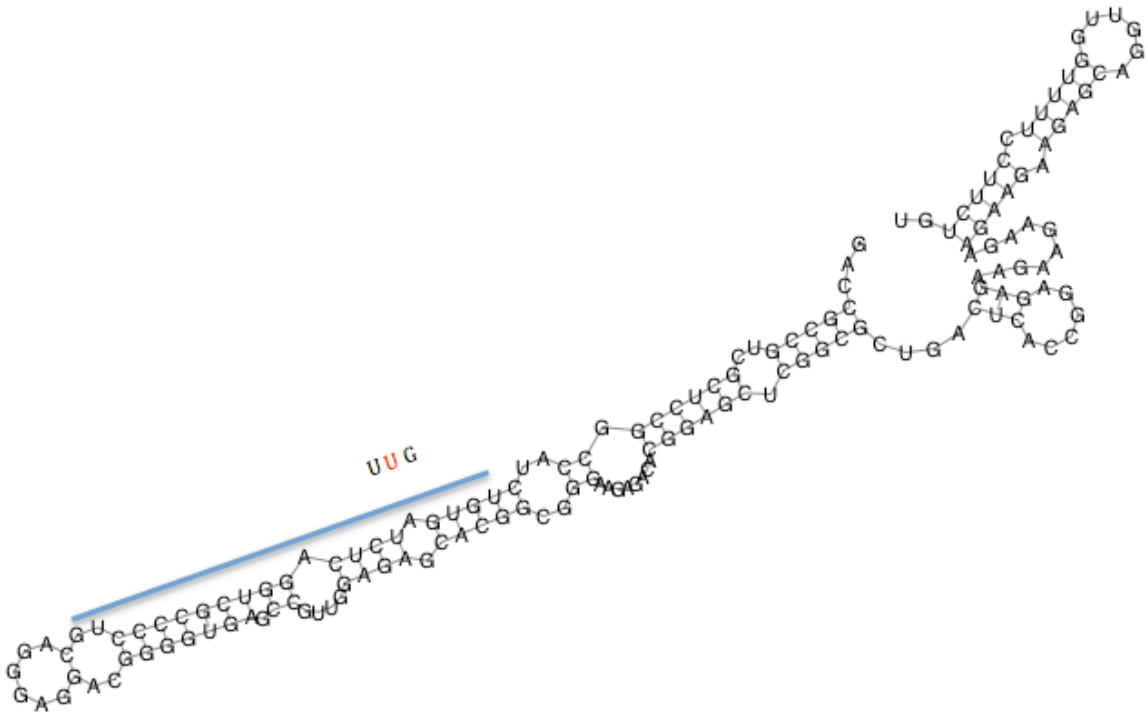
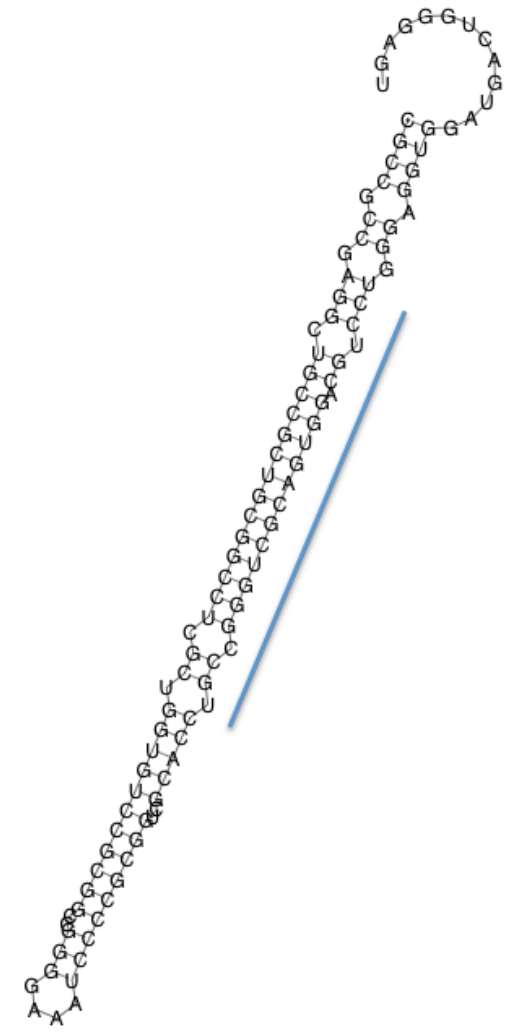


Figure 25: Pre-syn_cand13x346. The red "U" indicates a mismatch between the mature miRNA sequence and the precursor. This precursor did not pass plant criteria by the miRDeep-P prediction pipeline but the candidate miRNA sequence is a member of the miR398 family.

A



B



Figure 26: Predictions for syn_cand9x9. Identical hairpins were predicted for syn_cand9x9 (A) but a Geneious multiple sequence alignment of the candidate miRNA (sequence 1), predicted hairpin (sequence 2), and transcript assemblies (labeled 3-6) suggest that there are four distinct assemblies from stage A of flower development that give rise to the hairpin, and candidate miRNA sequence.



Figure 27: Pre-syn_cand8x9. This small RNA is a candidate of the candidate novel miRNA family. The red "C" in indicates a mismatch between the mature candidate miRNA and the sequence of the predicted precursor.

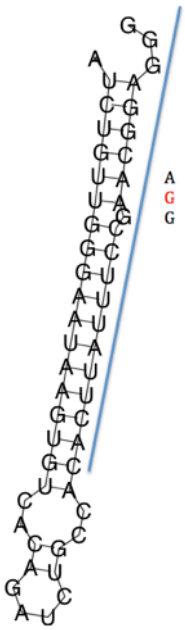


Figure 28: Pre-syn_cand11x5. The red "G" in indicates a mismatch between the mature candidate miRNA and the sequence of the predicted precursor.

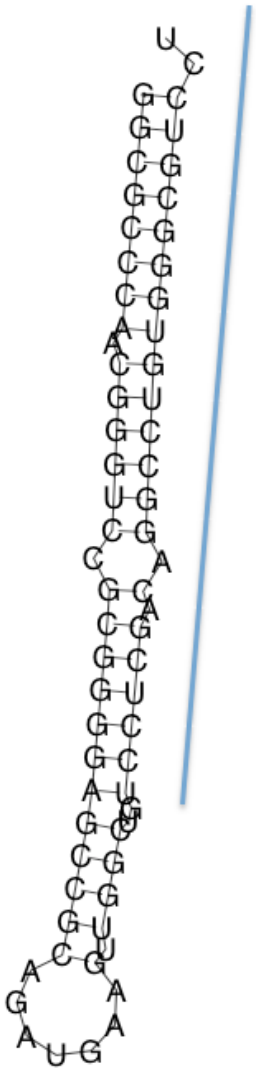


Figure 29: Pre-syn_cand5x19.

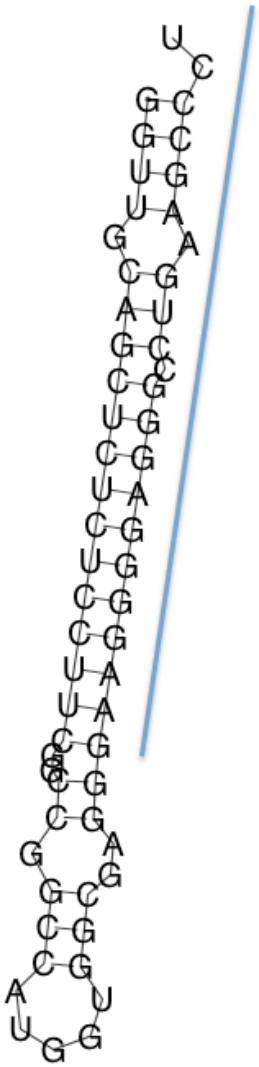


Figure 30: Pre-syn_cand6x15.



Figure 31: Pre-syn_cand7x10.

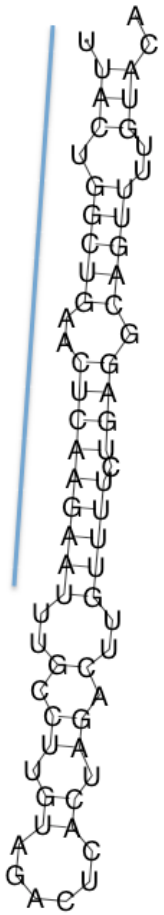


Figure 32: Pre-syn_cand10x5.

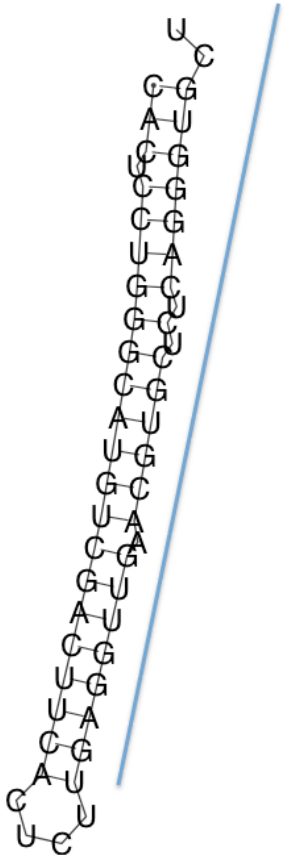


Figure 33: Pre-syn_cand12x2.

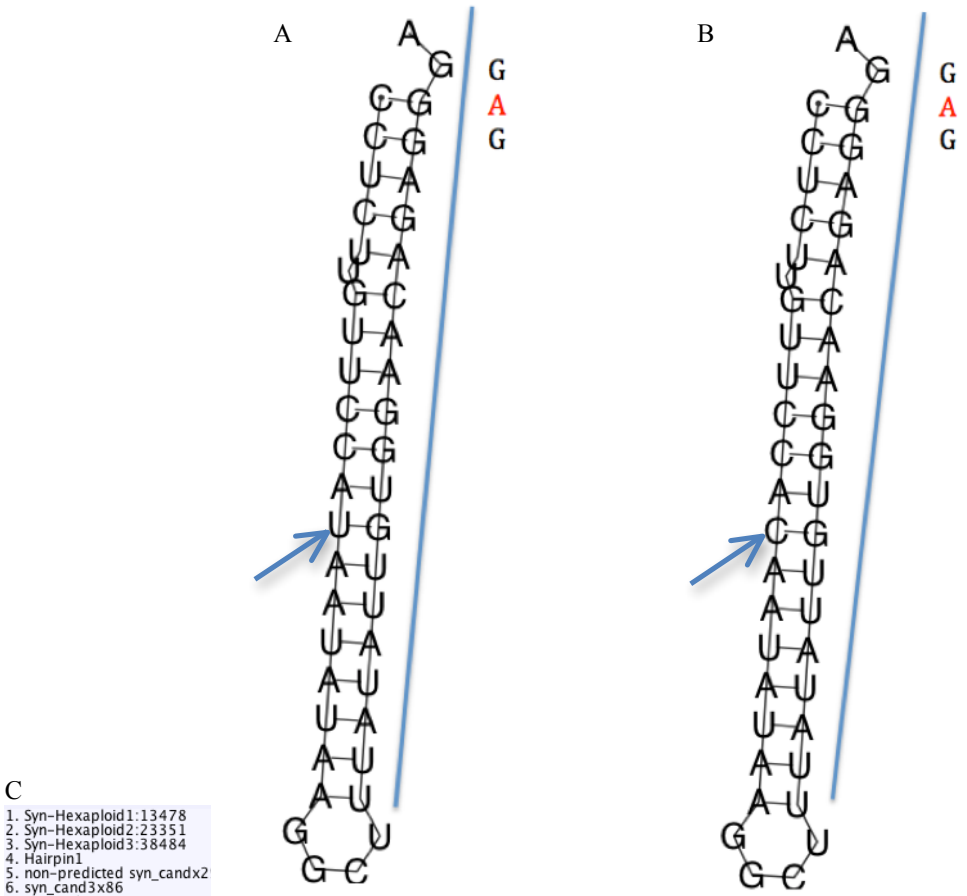
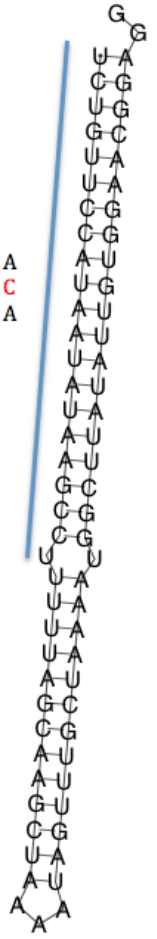


Figure 34: Predictions for syn_cand3x86. Hairpin1 is shown in (A) and Haripin2 is shown in (B). The red "A" in both hairpins indicates a mismatch between the mature candidate miRNA and the sequence of the predicted precursor. The small RNA sequence with 0 mismatches to the precursor was not predicted due to its lower read count but does exist in the small RNA library, non-predicted syn_candx29. The arrow points to the SNP difference between the two predicted hairpins. A Geneious multiple sequence alignment of syn_cand3x86, hairpins and transcript assemblies (C) shows that the transcript assemblies for hairpin 1 (labeled 1-3) reveal at least 2 transcripts that produce the hairpin. The same is for transcript assemblies for hairpin 2 (labeled 8-10). The alignment also shows that the non-predicted syn_candx29 (labeled 5.) aligns perfectly to all transcripts and hairpins whereas syn_cand3x86 has one mismatch in red (labeled 6.).

A



B

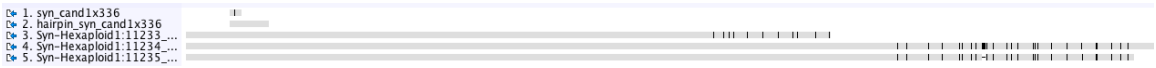


Figure 35: Predictions for syn_cand1x336. The red "C" in indicates a mismatch between the mature candidate miRNA and the sequence of the predicted precursor (A). The small RNA sequence with 0 mismatches to the precursor was not predicted due to its lower read count but does exist in the small RNA library, syn_cand2x298. Syn_cand2x298 was predicted for a different precursor. The multiple sequence alignment of the candidate miRNA sequence, predicted hairpin, and transcript assemblies shows three possible genes giving rise to the hairpin (B).

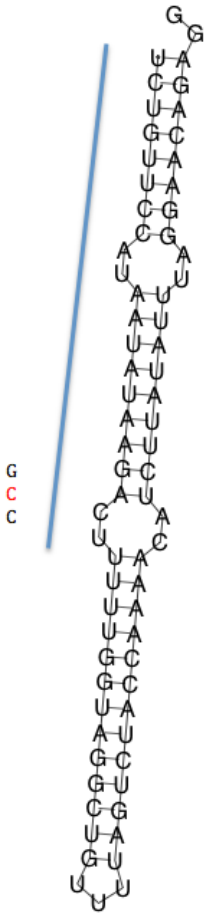


Figure 36: Pre-syn_cand2x298. The red "C" in indicates a mismatch between the mature candidate miRNA and the sequence of the predicted precursor. The small RNA sequence with 0 mismatches to the precursor was not predicted due to its lower read count but does exist in the small RNA library, non-predicted syn_candx11.

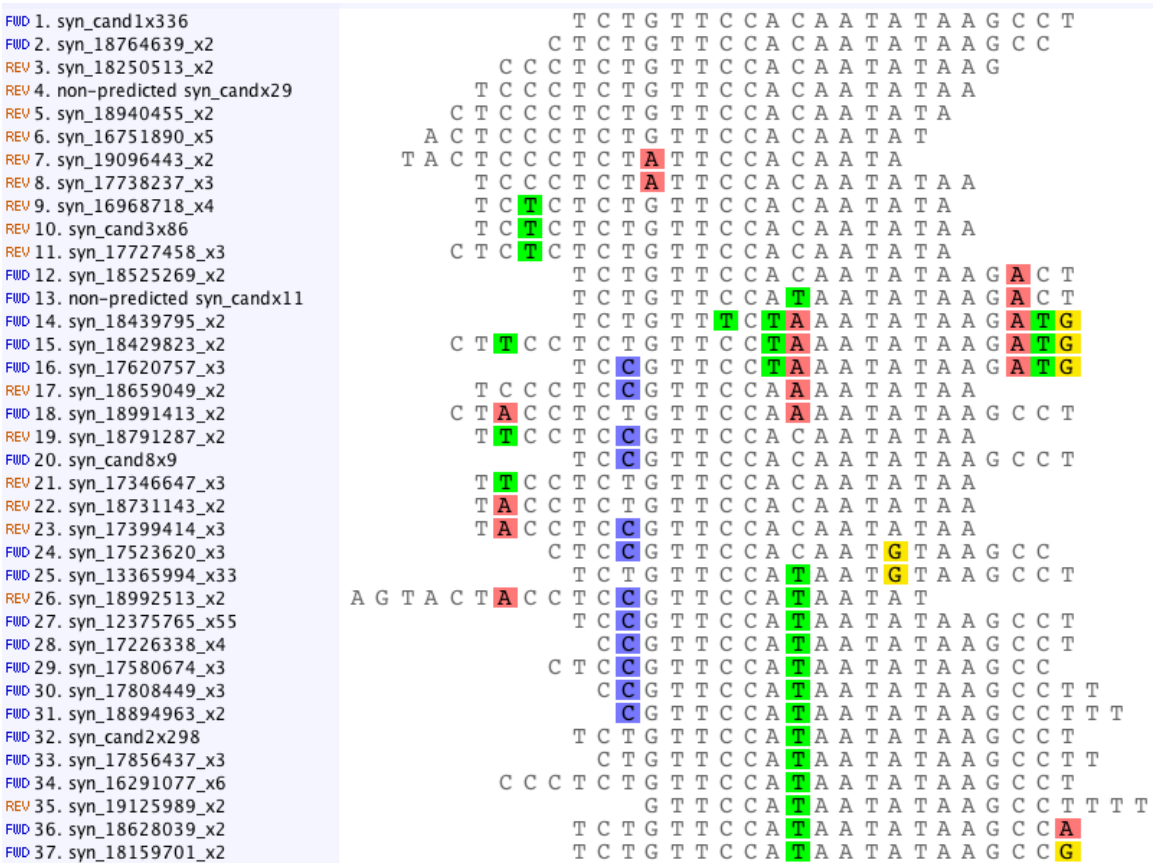


Figure 37: MSA of candidate miRNA family. This multiple sequence alignment shows all small RNAs predicted to have one or fewer mismatches to precursors predicted pre-syn_cand1x336, predicted pre-syn_cand2x298, and predicted pre-syn_cand3x86. Three candidate sequences were excluded due to failing alignment criteria by Geneious 5.6.7.

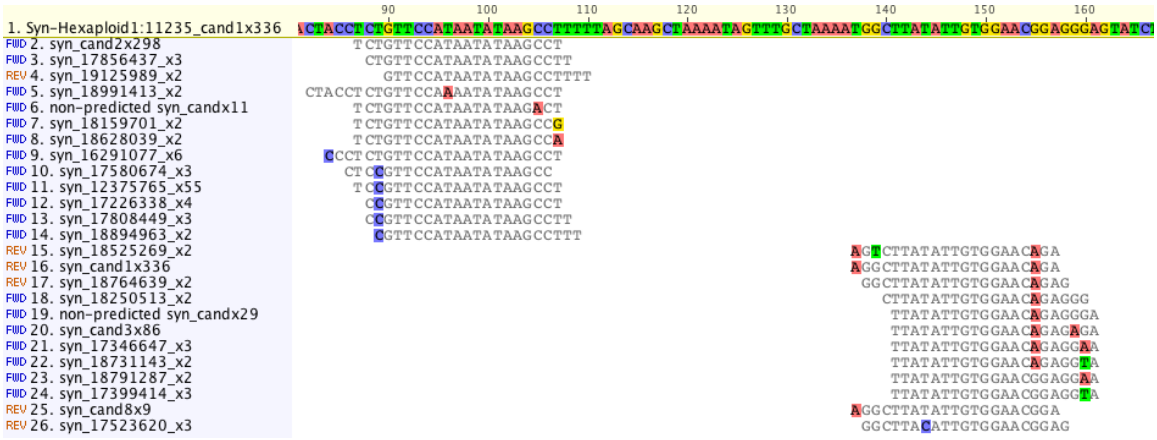


Figure 38: Reference assembly of candidate miRNA family members to one of the precursors for syn_cand1x336. The assembly shows alignment to each arm of the hairpin, corresponding to the mature miRNA sequence and the miR*. Syn_cand2x298 aligns to the precursor with 0 mismatches; however, mismatches do not allow the identification of the miR* produced from the 3' arm of this precursor.

Table 8. Candidate A-genome-specific miRNAs. Predicted miRNA sequences from either *A. strigosa* (AA) or the synthetic hexaploid (AACCDD) that were present in both libraries were searched for in the *A. magna* small RNA library (CCDD). Candidate A-genome-specific miRNAs are present in *A. strigosa* and the synthetic hexaploid but absent in *Avena magna*.

miRNA Accession	Target Accession	Expectation	UPE	Inhibition	Target Description	Read Count in *A. strigosa*	Read Count in Stromagdon
AAAATATAAGACGTTTTAGCT	LOC_Os08g31870.1	2	15.733	Cleavage	cDNA|cell division cycle protein 48, putative, expressed	8	8
AAAATATAAGACGTTTTAGCT	LOC_Os02g38340.1	2.5	18.437	Cleavage	cDNA|actin, putative, expressed	8	8
AAAATATAAGACGTTTTAGCT	LOC_Os02g02540.1	3	13.616	Translation	cDNA|glutamate receptor, putative, expressed	8	8
AAAATATAAGACGTTTTAGCT	LOC_Os03g18030.1	2.5	19.552	Cleavage	cDNA|leucoanthocyanidin dioxygenase, putative, expressed	8	8
AAAATATAAGACGTTTTAGCT	LOC_Os03g18030.2	2.5	19.552	Cleavage	cDNA|leucoanthocyanidin dioxygenase, putative, expressed	8	8
AAAATATAAGACGTTTTAGCT	LOC_Os12g21784.1	2.5	11.965	Cleavage	cDNA|expressed protein	8	8
AAAATATAAGACGTTTTAGCT	LOC_Os06g24490.1	3	15.075	Translation	cDNA|expressed protein	8	8
AAAATATAAGACGTTTTAGCT	LOC_Os03g39610.1	3	14.689	Translation	cDNA|chlorophyll A-B binding protein, putative, expressed	8	8
AAAATATAAGACGTTTTAGCT	LOC_Os10g42100.1	3	16.74	Cleavage	cDNA|pyruvate kinase, putative, expressed	8	8
AAAATATAAGACGTTTTAGCT	LOC_Os03g39610.2	3	14.689	Translation	cDNA|chlorophyll A-B binding protein, putative, expressed	8	8
TGCCGGGTCGCAGTGGACGTCC	LOC_Os05g41450.1	3	16.449	Translation	cDNA|histone-like transcription factor and archaeal histone, putative, expressed	4	9
GACATGTTGGATGTGTTGGCTCCAGTT	LOC_Os02g07870.1	2	21.207	Cleavage	cDNA|ATP synthase, putative, expressed	177	18
GACATGTTGGATGTGTTGGCTCCAGTT	LOC_Os06g45120.2	2	11.245	Cleavage	cDNA|ATP synthase, putative, expressed	177	18
GACATGTTGGATGTGTTGGCTCCAGTT	LOC_Os06g45120.1	2	11.245	Cleavage	cDNA|ATP synthase, putative, expressed	177	18
GACATGTTGGATGTGTTGGCTCCAGTT	LOC_Os10g18550.1	3	13.662	Cleavage	cDNA|retrotransposon protein, putative, Ty3-gypsy subclass, expressed	177	18
TGTTGCAGATGTGAAGGGCCTGAAGC	LOC_Os05g07470.1	3	16.164	Translation	cDNA|D-mannose binding lectin family protein, expressed	13	4
TCCTCGACAGGCCTGTGGGCGTCCT	N/A	N/A	N/A	N/A	N/A	6	19
GCCAATCTCCTGCTGAGTCAGACT	N/A	N/A	N/A	N/A	N/A	10	10

Supplementary Table 1: Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

Annotated miRNA	Sequence in *Avena strigosa* Small RNA Library			
miR156/157	GCTCACCTCTCTCTCTGTCAG	ACAGAAGAGAGTGAGCACT	TGTCAGAAGAGAGTGAGCAC	ATTGACAGAAGAGAGTGAGC
	TTGACAGAAGAGAGTGAACAC	TGACAGAAGAGAGCGAGCACT	TGACAGATGAGAGTGAGCAC	ATTGACAGAAGAGAGTGAGCA
	GCTCACTCTCTCTCTGTCAG	ACAGAAGAGAGTGAGCATC	TGCCAGAAGAGAGTGAGCAC	CTTGACAGAAGAGAGTGAGCAC
	TTGACAGAAGAGAGTGACCAC	TGACAGAAGAGAGTGAGCACA	TGACAGAAGAGAGCGAGCAC	GAGAGAAGATGGAGAACA
	TTAACAGAAGAGAGTGAGCAC	GCTCACTACCCTGTCTGTCACC	TGACAGGAGAGAGTGAGCAC	
	CTGACAGAAGAGAGTGAGCAC	GCTTACTACTCTGTCTGTCACC	TTGACAGAAGAGAGTGAGC	
	TTGACAGAAGAGAGTGAGCTC	GCTCACCACTCTGTCTGTCACC	GCTCACTCCTCTTTCAGTCAGC	
	TCGACAGAAGAGAGTGAGCAC	GCTCATTACTCTGTCTGTCACC	TTGACAGAAGAGAGTGAGCACC	
	CACTACTCTGTCTGTCACC	TGACAGAAGAGAGAGAGCAC	TTGACAGAAGAGAGAGAGCAC	
	TTGACAGAAGAGAGTGAGCAA	GACAGAAGAGAGTGAGCAC	TTGACAGAAGAGAGTGAGCACA	
	GCTCACTACTCTGTCTGTCAC	TGACAGAAGAGAGTGACCAC	TTGACAGAAGAGAGTGAGCAC	
	TTGACAGAAGAGAGTGAGGAC	TGGCAGAAGAGAGTGAGCAC	TTGACAGAAGAGAGTGAGCACT	
	TTGACATAAGAGAGTGAGCAC	TGACAGAAGAGAGTGAGCAC	TTGACAGAAGATAGTGAGCAC	
	TTGACAGAAAAGAGTGAGCAC	TGACAAAAGAGAGTGAGCAC	TTGACAGAAGAGAGTGAGTA	
	TTGACAGAAGAGGGTGAGCAC	TAACAGAAGAGAGTGAGCAC	GCTCACTCCTCTTTCTGTCAGG	
	TTGACAGAAGAGAGTAAGCAC	TGACAGAAGAGATTGAGCAC	TTGACAGAAGAGAGCGAGCAC	
	TTGACAGAAGAGAGTGAGCAT	TGACAGAAGAGAGTGAGCGC	GCTCACTGCTCTATCTGTCAGC	
	TTGACAGAAAAGAGCGAGCAC	TGACAGGTGAGAGTGAGCAC	TTGACAGAAGAGAGGGAGCAC	
	GCTCACTACTCTGTCTGTCACC	TGACATAAGAGAGTGAGCAC	TTGACAGAAGAAAGTGAGCAC	
	ACTACTCTGTCTGTCACC	TGACAGAAGAAAGTGAGCAC	GCTCATTCCTCTTTCTGTCAGC	
	TTGACAGTAGAGAGTGAGCAC	TGACTGAAGAGAGTGAGCAC	GCTCACTCATCTTTCTGTCAG	
	TTCACAGAAGAGAGTGAGCAC	TGACAGAAGAGAGTGAGCTC	TTGACAGAAGAGAGTGAGCA	
	TTGACAGAAGAGAGTGAGAAC	TGACAGAAGAGAGTGAGCT	TGCTCACTACTCTGTCTGTC	
	TTGAGAGAAGAGAGTGAGCAC	TGACAGAGGAGAGTGAGCAC	TGCTCACTTCTCTCTCTGTCACC	
	TTGACAGAAGAGAGTGAGCGC	GACAGAAGAGAGTGAGCA	GCTCACTTCTCTCTCTGTCAGC	
	TTGACAGAAGAGAATGAGCAC	TGACAGAAGAGAGTGAGCA	GCTCACTACTCTTTCTGTCACC	
	TTGACAGAAGAGTGTGAGCAC	TGACAGAAGAGAGTGAGTAC	GCTCACTTCTCTCTCTGTCACC	
	TTGACAGGAGAGAGTGAGCAC	TGACAGAAGAGAGGGAGCAC	TGCTCACTTCTCTCTCTGTCAC	
	TTGACAGAGGAGAGTGAGCAC	TGACAGAAGAGAGTGAGCAT	GCTCACTCCTCTTTCTGTCAG	
	TGACAGAAGTGAGTGAGCACT	TGACAGAAGAGAATGAGCAC	GCTCACTTCTCTCTCTGTCAG	
	TGACAGAAGAGAGTAAGCACC	ACAGAAGAGAGTGAGCAC	TGCTCACTCCTCTTTCTGTCAG	
	TGACAGAAGAGAGTGAGCACC	TGACAGAAGAGAGTGGGCAC	TGCTCACTCCTCTTTCTGTCAGC	
	TGACAGAAGAGAGTGAGCACT	TGACAGAAGGGAGTGAGCAC	GCTCACTCCTCTTTCTGTCAGC	
	TGACAGAAGAGAATGAGCACT	TGACAGAAGAGAGTGAGCAG	GCTCACTCCTCTTTCTGTC	
	GACAGAAGAGAGTGAGCACA	CGACAGAAGAGAGTGAGCAC	GCTCACTCCTCTTTCTGTCA	
	TGACAGAAGAGAGTGAGCATA	TGACAGAAGAGAGTAAGCAC	ATAGACAGAAGAGAGTGAGCAC	
	GACAGAAGAGAGTGAGCACT	TGACAGAAGAGAGTGAGC	ATTGACAGAAGAGAGTGAGCAC	

Supplementary Table 1…(continued): Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR159/319	CTTGGACTGAAGGGTGCTCC	CTGGACTGAAGGGTGCTCCCT	GAGCTCCTATCTTTCCAAT
	CTTGGACTGAAGGGTGCTCCT	TTGGACTGAAGGGCGCTCCCT	TTTGAATTGAAGGGAGCTCTG
	CTTGGACTGAAGGGTGCT	TTGGACTGAAGGGTGCTCCGT	TTTGGATTGAAGGGAGTT
	ATTGGATGACGCGGGAGCTAA	TTGGACTGAAGGGTGCTCGCT	CTTGGATTGAAGGGAGCTCT
	TTGGACGAAAGGGTGCTCCCT	TTGGACTGAAGGGAGCTCCCT	TTTAGATTGAAGGGAGCT
	AGAGCGTCCTTCAGTCCACTC	TTGGACCGAAGGGTGCTCCCT	TTTGGAAAGAAGGGAGCTC
	AGAGCGTCCTTCAGTCCACTT	TTGGACGGAAGGGTGCTCCCT	TTTGGATTGAAGGGAGCT
	AGCTGCCGACTCATTCATCC	TTAGACTGAAGGGTGCTCCCT	TTGGATTGAAGGGAGCTCT
	TTGTACTGAAGCGTGCTCCCT	TTGGACTGAAGGGAGCTCCC	TTTGGATTGAAGGGAGCTTTG
	AGAGCGCCCTTCAGTCCACTC	TTGGACTGAGGGGTGCTCCCT	TTTGGATTGAAGGAAGCTCTG
	TTGGACTGAAGTGTGCTGCCT	TTGGACTGAAGGGTGCTCCCT	TTTGGATTGAAGGGAGCTCTG
	AGCTGCCGACTCATTCATCCT	TTGGACTGAAGGGTGCTCACT	TTTGGATTGAAGGGAGCTCTT
	AGCTGCCGACTCATTCATCCA	TTGGACTGAAGGGTCCTCCC	TTTGGATTGAAGGGAATT
	AGCTGCCGACTCATTCATC	TTGGACTGAAGGGTGCTCCAT	TTTGGATTGAAGGGAGCTT
	ATCTGCCGACTCATTCATCCA	CTTGGACTGAAGGGTGCCCCC	TTTGGATTGAAGGGAGTTTT
	AGAGCGTCCTTCAGTCCACT	CTTGGACTGAAGGGTGCTCCC	TTAGGATTGAAGGGAGCTCT
	AGCTGCCGACTCATTCATCT	CTTGGACTGAAGGGTGCG	TTTGGATTGAAGGGAGCTATT
	TTGGACTGAAGGGCGCTCCC	AGAGCTCTCTTCAGTCCACTC	AAGCTCCTATCTTTCCAATG
	TTGGACTGAAGTGTGCTCCC	TTGGACTGAAGGGAGCTCCCTT	TTTGGATTGAAGAGAGCTCTG
	TTGGACTGAAGGGTGCTCCC	TTTGGATTGAAGGGAGCTTTT	TTTGGATTGAAGGGAGTTTTC
	TTGGACTGAAGGGTGCTCCCG	TTTGGATTAAAGGGAGCTCT	AGCTGCTCGTTCATGGTTCCC
	TTTGACTGAAGGGTGCTCCCT	TTTGGATTGAAGGGAGCTCT	AGCTGCTCGTTCATGGTT
	TTGGACTGAAGGATGCTCCCT	TTTGGATTGAAGTGAGCTCTG	AGCTGCTCGTTCATGGTTCT
	TTGGACTGAATGGTGCTCCCT	TTGGATTGAAGGGAGCTC	CTTGCATGACCGAGGAGCCT
	TTGGAGTGAAGGGTGCTCCCT	TTAGGATTGAAGGGAGCTCTG	AGCTGCTCGTTCATGGTTC
	TTGGACTGAAGGGTGCTCCCC	TTGGGATTGAAGGGAGCTCTG	AGCTGCTCGTTCATGGTTCA
	TTGGACTAAAGGGTGCTCCCT	TTGGATTGAAGGGAGCTCTC	CTTGCATGACCGAGGAGCCGA
	TTAGACTGAAGGGTGCTCCC	TTTGGATTGAAGGGAGCTG	AGCTGCTCGTTCATGGTTCC
	TTGGACTGAAGGGTGCTCCCA	TTGGATTGAAGGGAGCTCTG	AGCTCCCTTCGATCCAAACC
	TTGGATTGAAGGGTGCTCCCT	TTTGGATTGAAGGGGGCTCTG	TTTGGATTGAAGGGAGCTTAT
	TTGGACTGAAGGGGGCTCCCT	TTTGGATTAAAGGGAGCTCTG	GCTCCCTTCGATCCAAAC
	TTGGACTGAAGGGTGCTCC	GAGCTCCTATCTTTCCAATGA	AGCTGCTAAGCTATGGATCCC
	TTGGACTGAAGGGTGTTCCCT	GAGCTTCTATCTTTCCAATG	AGCTGCTAAGCTATGAATCCC
	GGACTGAAGGGTGCTCCCT	TTTGGATTGAAGGGAGCTCTA	ATTGGACTGAAGGGAGCTCCC
	TTGGACTGAACGGTGCTCCCT	TTTGGATTGAACGGAGCTCTG	
	TTGGACTGAAGGGTGCTTCCT	TTGGACTGAAGGGTGCTC	
	TGGACTGAAGGGAGCTCCCT	TTTGGATTGAAGGGAGCTC	
	TTGGACTGAAGGGTGCTCCTT	GAGCTCCTATCTTTCCAATG	

Supplementary Table 1… (continued): Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR160	TGCCTGGCTCCCTGTATGCCT			
	TGCCTGGCTCCCTGTATGCCA			
	GCGTGCAAGGAGCCAAGCATG			
miR162	TCGATAAACCTCTGCATC			
	TCGATAAACCTCTGCATCCGG			
	CGATAAACCTCTGCATCCGG			
	TCGATAAACCTCTGCATCCAG			
miR164	TTAGAGCATTGAGAGGAC			
	TGGAGAAGCAGGGCACGTGCA			
	TGGAGAAGCAGGGCACGTGCT			
	CACGTGTTCTTCTCCTCCATC			
	CATGTGCCCTTCTTCTCCATC			
miR165/166	TCGGACCAGGCTTCATCTTC	TCGGACCAGGCTTCAGTCCCC	TCAGACCAGGCTTCACTCCCC	TCGGACCAGGCTTCATTT
	TCGGACCAGGCTTATTCCCC	TCGGACCAGGCTTCATTCTC	TCGGACCAAGCTTCATTCCCC	TCAGACCAGGCTTCATTTC
	GGGTTGTTGTCTGGTTCA	TCGGACCAGGCTTCAATCCT	TCGGACCAGACTTCAATCCCT	TCCGACCGGGCTTCATTCCCC
	GGGTTGTTGTCTGGTTCAA	TCGGACCAGGGTTCAATCCCT	TCGGACCAGGCTTCATTATC	TCGGACCAGGCTTCAATCCC
	GAATGACGCCGGGTCTGAAAG	TCGGACCACGCTTCATTCCCC	TCGGACAAGGCTTCATTCCCC	TCGGACCAAGATTCATTCCCC
	GGGTTGTTGTCTGGTTCAAGG	TCGGACCATGCTTCATTCCCC	TCGGACCAGACTCCATTCCCC	TCGGACCAGGCTTCATTCGTC
	TCGGACCAGGCTTCATTGTTT	TCGGAACAGGCTTCATTCCCC	TCGGCCCAGGCTTCATTTCCC	TCGGACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCATTTGTT	TCGGACCAGGCATCCTTCCCC	GGTATGTTGGCTGGCTCGA	TCGGACCAGGCTTAATTCCCC
	TCGGACCAGGCTTCATTCCTTT	TCGTACCAGGCTTTATTCCCC	CGGACCAGGCTTCAATCCCT	CGGACCAGGCTTCATTCCCA
	GGTATGTTGTCTGGCCCGG	TCGGACCAGGCTACATTCCCT	TCGGACCAGGCTACATTCCCC	TCGGACCAGTCGTCATTCCCC
	TCGGACCAGGCTTCATTTATT	TTGGACAAGGCTTCATTCCCC	TCGGACCGGGCTTCATTCCC	TCGGACCAGGCTTCATTTCT
	TCGGACCAGGCTTCATTCTAA	TCGGACCAGGCTTCATTCCAC	TCGGACCAGGCTCCATTCCCG	TTGGACCAGGCTTTATTCCCC
	TCGGACCAGGCTTCATTCCATT	TCGGACCAGGCTGCATTCCCC	CGAACCAGGCTTCATTCCCC	TCGGACCAGGGTTTATTCCCC
	TCGGACCAGGCTTCATTATTT	TCCGACCAGGCTTCAATCCCT	TCGGACCAGGCTTCATTCCGT	TCGGACCAGGCTTCATCCCCC
	TCGGACCAGGCTTCATTTTTT	TCGGACCAGGCTTCATCTC	TCGGACCAGGCTTCGATCCC	TCGGACCAGGCTTCAATC
	TCGGACCAGGCTTCATTCTTTT	TCAGACCAGGCTTCATTCCCT	TCATACCAGGCTTCATTCCCC	TCGGACCAGGCTTTATTCT
	TCGGACCAGGCTTCATTTTTA	TCGGACCAGGCTTTATTCCCC	TCGGACCAGGCTTCATTCCT	TCGGATTAGGCTTCATTCCCC
	TCGGACCAGGCTTCAATCTTT	TCGGACGGGGCTTCATTC	GGAATGTTGTCTGGTTCAAG	TCGGACCAGGCTTCATTCCC
	TCGGATTAGGCTTCATTCCTT	TCCGACTAGGCTTCATTCCCC	TCGGACCAGGCCTCATTCCC	GCGAACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCATTCTCC	TCCGACCAGGCTTCATTCCCC	TCGGGCCAGGTTTCATTCCCC	TCGGACCAGGCTTTAATCCC
	GGAATGTTGTCTGGTTCAAGG	TCGGACCAGGCTTCATTATCC	TCGGACCAGGCTTCAATA	TTGGACCAGGCTTCATTCTCC
	TCGGACCAGGCTTTATTCCC	TCGGACCAGGCTTCATCT	TCGGACCAGGCTTCTTTCT	TCGGACCAGGCCTCAATCCCT
	ACCGACCAGGCTTCATTCCCC	TCGGACTAGGCTTCAATCCCT	TCGGACTAGGATTCATTCCCC	TCGGACCAGGCTTCATTCAT
	TCGGACCAGGCTTCATTCCGC	TCGGACCAGGCTTCATTTCCC	TCGGACCAGTCTTCATTCCCC	TCGGACCAGGCGTCATTCCCC
	TCGGACCAGGCTTCATACCCC	TAGGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTCC	TCGGACCAAGCCTCATTCCCC
	TCGGACCAGGCTTCATTGCCC	TCGGACCAGGCTTCAATCC	TCGGACCAGGCTTCAATCT	TCGGACTAGGCTTCATTCCCC

Supplementary Table 1… (continued): Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR165/166	TCAGACCAGGCTTCAATCCC	TCGGACGTGGCTTCATTCCCC	TCGGACCTAGCTTCATTCCCC	TCGGACCAGACTTCACTCCCC
	TCGGACCAGGCTTCATTCCTC	TCGGATCAGGCTTCATTCCCC	TCGCAACAGGCTTCATTCCCC	TCACACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCATTCT	TCAGACCGGGCTTCATTCCCC	TCGGACCAGGCTTCATTCCCG	TCGGACCGGGCTTCAATCCC
	GGGGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTC	TCGGACCAGGCTTCATTTTC	TTGGACCAGGCTTCAATCCCT
	TCAGACCAGGCTTCGTTCCCC	TCGGACCCGGCTTCATTCCCC	TCAGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTTTCT
	CCGGACCAGGCTTCAATCCCT	TCGAACCAGGCTTCATTCCCC	CCGGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTCACC
	TCGGACCAGGCTTCGATCCCT	TGGGACCAGGCTTCATTCCCC	TCGGACTGGGCTTCATTCCCC	CGGACCAGGCTTCATTCCC
	TCGGACGAGGCTTCATTCCCC	TCGGACCAGGGGTCATTCCCC	TCGGACCAGGCTTCGTTCCAC	TCGGACCAGGCTTCATTGTCC
	TCGGACCAGGCTTCACTCCCC	TCGGACCGGGCTTCATTCCCC	GGAATGTTGTCTGGTTGGAGA	GCGTACCAGGCTTCATTCCCC
	TCAAACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTCCG	TCGGACCAGGCTTCATTACT	TCGGACCAGGCTTCGTTCCCC
	TCGGACCAGGCTTTATTCCAC	TCGGACCAGGCTTCATTCTT	TCGGACCAGGATTCATTCCCC	GCGGACCAGGCTTCATTCCCC
	TCGGACCAGTCTTCAATCCCT	TCGGACCAGGATTCATTCCCT	TTGGACCAGGCTTCATTCCCT	TCGGACAAGGCTTCATTCCC
	TCGGACCAGGTTTCATTCCCC	TCGGACCAGGCTTTGTTCCCC	TCGGACCAGACTTCATTCCCC	CCGGACCAGGCTTCATTCCC
	TCGGACCAGGCTTCATTTTCC	TCTGACTAGGCTTCATTCCCC	TCGGACCAGGCTTCTTCCCC	TCGGACCAGGCTTAAATCCCT
	TCGTACCAGGCTTCATTCCCC	TCGGGCCAGGCTTCATTCCCC	TCGGACCAGGCTTCCTTCCCC	TCCGACAAGGCTTCATTCCCC
	TTGGACCAGGCTTCATCCCCC	TCGGACACGGCTTCATTCCCC	TCAGACCAGGCTTCCTTCCCC	TCGTACCAGGCTTCATTCCAC
	TCGGACCAGGCTTTATCCCCC	TCGGACCAGGCTTGATTCCCC	TCGACCAGGCTTCATTCCCC	TCGGATCAGGCTTAATTCCCC
	TCGGACTATGCTTCATTCCCC	TCGGACGAGGCTTCAATCCCT	TCGGCCCAGGCTTCATTCCCC	TCGGACCAGGCTTGAATCCCT
	TCGGACCAGGCTTCACCCCC	TCGGACCAGGCTTCATTCCAT	TCAGACCAGGCTTCATTTCCC	TTCGGACTAGGCTTCATTCCC
	TCGGACCAGGCTTCATTTT	TCGGACCAGGCTTCATTCGCC	TCAGACCAGGCTTCATTCCC	TTCGGACCGGGCTTCATTCCC
	TCGGACCAGGCTTTATTCGCC	TCAGACCAGGCTTCAATCCCT	TCGGACCAGGGTTCATTCCAC	TTTGGACCAGGCTTCATTCCC
	TCGGACTAGGCTTCATTCCCT	TCCGACCCGGCTTCATTCCCC	TCGGACCAGCCTTCAATCCCT	TTCGGACCAGACTTCATTCCC
	TCGGACATGGCTTCATTCCCC	TCGAACCAGGCTTCATTCCCT	TCGGACCAGGCTTCATTCGC	CTCGGACCAGGCATCATTCCC
	TCGGACCAGGCTTCATTACCT	TCGGACCAGGCTCCATTCCCC	TCGGACCAGGCTTCATTCCTA	GGAATGTTGTCTGGCTCGG
	CGGACCAGGCTTCATTCCCC	TCGGACCAGACTTCATTCCCT	TCGGCCCAGGCTTCATTCCCT	AATTTTGTCTGGCTCGAGG
	TCGGACCAGTTTTCATTCCCC	TCGGACCAGGCTTCATTTC	TCGGACCAGGCTTCATTCCCT	TTCGGACCAGGCTTCATTGCC
	TCGGACTAGGCTTTATTCCCC	TCTGACCAGGCTTCATTCCCC	TCGGACCAGGCCTTATTCCCC	GGAATGTTGTCTGGCTCGAGG
	TCGGACCTGGCTTCATTCCCC	TCGGACCAGGCTTCAATCCCG	TCGGACCAGGCTTCATTCCCA	TTCGGACCAGGCTTCATTCGC
	TCGGACCAGGCTACAATCCCT	TCGGACGGGGCTTCATTCCCC	TCGGAGCAGGCTTCATTCCCC	CTCGGACCAGGCTTCATTCCC
	TCGGACCAGGCTTAGTTCCCC	TCGGACCAGGCTTCAATCCCC	TCGGACCAGGCTTCATCCCC	TTCGGACCAGGCCTCATTCCC
	TCGCACCAGGCTTCATTCCCC	TCGGACAGGGCTTCATTCCCC	TCGGACCAGGCATCATTCCCC	GGAATGTTGTCTGGCTCG
	TCGGTCCAGGCTTCATTCCCC	TCGGACCAGGCTTCTTTCCCC	ACGGACCAGGCTTCATTCCCC	TTCGGACCAGGCTTCATTCCT
	TCGGACCAGGCTTCAATCCAT	TCGGACCAGGCTTCATTCCTT	TCGGACCAGGCCTCATTCCCC	TTCGGACCAGTCTTCATTCCC
	TCGGACCAGGCTTTAATCCCT	TCGGACCAGGGTTCATTCCCC	TCGGACCAGGCTTCATTTCC	TTGGGACCAGGCTTCATTCCC
	GGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTCTTC	TCGGACCAGGCTTTATTTCCC	CTCGGACCAGGCTTCAATACC
	TCGGACCAGGCTTCAATCCCT	TCGGACCAGGCTTTATTCCCT	TCAGACCAGGCTTCATTCCT	TTCGGACCAGGCTTCATTTCC
	TCGGACCAGGCTTCAATCCCA	TTGGACCAGGCTTCATTCCCC	TCGGACCAGCCTTCATTCCCC	TTCGGACCAGGCTTCATTCCA
	TCGGACCAGGCTTCATTACCC	TTGGACCAGGCTTCATTCC	TCGGCCCAGGCTTCAATCCCT	TTCGGACCAGGTTTCATTCCC

Supplementary Table 1… (continued): Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR165/166	TTCGGACCAGGCTACATTCCC	GGAATGTTGTCTGGCTCGA	TCGGACCAGGCTTCATTTTA	TTCGGACCAGGCTTCATTCC
	TTCGGACCAGGCTGCATTCCC	TTCGGACCAGGCTTCATTCCC	TCGGACCAGGCTTCATTCCTAC	CCTCGGACCAGGCTTCATTCC
	TTCGGACCAGGCTTCATCCCC	TTCGAACCAGGCTTCATTCCC	TCGGACCAGGCTTCAATTCCCC	TCGGACATGGCTTCAATCCCT
	GAATGTTGTCTGGCTCGGGG	TTAGGACCAGGCTTCATTCCC	CGGACCAGGCTTCATTCCCCC	TCGGACCAGGCTTCATTTTTC
	GGAATGTTGTCTGGCTCGGGG	TTCGGACCAGGCTTCATTCCG	TCGGACCAGGCTTCATTCCCCC	TCGGACCAGGCTTCAATCCTT
	TTCGGACCAGGCTCCATTCCC	CTCGGACCAGGCTTCATTACC	TCGGACCAGGCTTCATTCCCTC	TCGGACCAGGCTTCATTTATC
	CTCGGACGAGGCTTCATTCCC	GGAATGTTGTCTGGCTCGGG	TCGGACCAGGCTTCATTCCACC	TCGGACCAGGCTTCATTCTTT
	TTCGGACCAAGCTTCATTCCC	TCCGGACCAGGCTTCATTCCC	TTTCGGACCAGGCTACATTCC	TCGGACCAGGCTTCAATTTC
	TTCGGACCATGCTTCATTCCC	TTCGGACCAGGCTTGATTCCC	TCTCGGACCAGGCTTCAATC	TCGGACCAGGCTTCATTTAT
	GGAATGTTGTCTGGCTCGAG	GGAATGCTGTCTGGCTCGAGG	TTTCGGACCAGGCTTCATTCC	TCGGACCAAGCTTCATTCCTT
	CTCGGACCAGGCTTCAATCCC	TTCGGAGCAGGCTTCATTCCC	TTTCGGACCAGGCTTCATACC	GGGTTGTTGTCTGGTTCAAGT
	TTCGGACCAGGCTTCACTCCC	TCGGACCAGGCTTCATGTT	TCTCGGACCAGGCTTCATTCC	TCGGACCAGGCTTCATTTTT
	TCCAACCAGGCTTCATTTCCC	TCGGACCAGGCTTCATTTCCCC	TCTCAGACCAGGCTTCATTCC	TCGGACCAGGCTTCATTATT
	CTCGGACCAGGCTTTATTCCC	TCGGACCAGGCTTCATTCCCCG	TCTCGGACCAGGCTTCAATCC	GGTATGTTGGCTGGCTCGATG
	TTCGGACCAGGCTTTATTCCC	TCGGACCAGGCTTCATTCCCTT	CTCGGACCAGGCTTCATTCC	TCGGACCAGGCTCAATCCCT
	TTCGGACCAGGCTTAATTCCC	TCGGACCAGGCTTCATTCCCCT	TTCGGACCAGGCTTCATTT	
	CTCGGACCAGGCTACATTCCC	TCGGACCAGGCTTCATTCCTTC	TTTCGGACCAGGCTTCGTTCC	
	GAATGTTGTCTGGCTCGAGG	CGGACCAGGCTTCATTCCCCT	CTCGGACCAGGCTTCATT	
	GTCGGACCAGGCTTCATTCCC	TCGGACCAGGCTTCATTCCCCA	TTCGGACCAGGCTTCATTCT	
miR167	GATCATGACTGACAGCCT	TGAATCTGCCAGCATGATCTG	TGAAGCTGCGAGCATGATCTGA	TGAAGCTGCCAGCATGATCGGG
	ATCATGACTGACAGCCTCATT	TGAAGCTGCCAGCGTGATCTGA	TGAAGCCGCCAGCATGATCTGA	TGAAGCTGCCTGCATGATCTGA
	ATCATGACTGACAGCCTCA	TGAAGCTGCTAGCATGATCTGA	AGATCATGTTGCAGCTTCATT	TGAAGCTGCCAGCATGATCTGG
	GATCATGACTGACAGCCTCATT	AGATCATGTTGCAGCTTCAT	ATCATGCTGTGCAGTTTCACC	TGAAGCTGCCAGCATGTTCTGA
	GATCATGACTGACAGCCTCA	TGAAGCTGCCGGCATGATCTGA	TGAGGCTGCCAGCATGATCTGA	AGGTCATGCTGGAGTTTCATC
	TGAAGCTGCCAGCATGATC	TGAAGCTGCCAGCATGATCTGA	TGAAACTGCCAGCATGATCTGA	AGGTCATGCTCGAGTTTCATC
	TGAAGCTGCCAGCATGATCT	TTAAGCTGCCAGCATGATCTGA	TGAAGCTGCCAGCATAATCTGA	TCATGCTGGAGTTTCATC
	ATCGTGCTGTGACAGCTTCACT	TGAAGCTACCAGCATGATCTGA	TAAAGCTGCCAGCATGATCTGT	AGGTCATGCTGGAGTTTCAT
	GCTGCCAGCATGATCTGA	TGAATCTGCCAGCATGATCTGG	TGAAGCGTCCAGCATGATCTGA	AGGTCATGCTGGAGTTTCA
	AAGCTGCCAGCATGATCTGA	TGAAGCTGCCAGCATGACCTGA	GATCATGCTGTGCAGTTTCATC	AGGTCATGCTGGAGTTCCATC
	TGAAGCCGCCAGCATGATCTG	TAAAGCTGCCAGCATGATCTGA	ATCATGCTGTGCAGTTTCATC	
	TGAAGCTGCCAGCATGATCTA	TGAAGCTGCCAGCATGATCTGT	TGAAGCTGCAAGCATGATCTGA	
	TGAAGCTGCCAGCATGATCTG	TGAAGCAACCAGCATGATCTGA	TGAAGCTGCCAGCATTATCTGA	
	TGAAGCTGCGAGCATGATCTG	TGAAGCTTCCAGCATGATCTGA	TGAAGCTGCCATCATGATCTGA	
	TGAAGCTGCCAGCATGTTCTG	TGAAGCTGCCAGCATGATATGA	TGAAGCAGCCAGCATGATCTGA	
	TAAAGCTGCCAGCATGATCTG	TGAAGCTGCCAGCATGATTTGC	TGAAGCTGCCAGCATGATTTGA	
	TGAAGCTGCCAGCATAATCTG	TGAAGCTGCCAGCATGATCAGA	GAAGCTGCCAGCATGATCTGA	
	TGAAGCGGCCAGCATGATCTG	TGAAGCTGTCAGCATGATCTGA	CGAAGCTGCCAGCATGATCTGA	
	TGAAGCTGCCAGCTTGATCTG	TGAAGCTGCCAGCATGATCTAA	TGAAGCTGCCAGCATGATCTGC	

Supplementary Table 1… (continued): Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR168	CCCGTGTTGCACCAAGTGAAT	TCGCTTGGTGCAGATCGGGCA	TCGCTTGGTGCAGACCGGGAC	TCGCTCGGTGCAGATCGGGAC
	CCCGCGTTGCGCCAAGTGAAT	CCCCGCGTTGCACCAAGTGAAT	TCGCTTGGTGAAGATCGGGA	ACGCTTGGTGCAGATCGGGAC
	CCCGCGTTGCACCAAGTAAAT	CCCGCGTTGCACCAAGTGA	TCGCTTGGTGCAGATCGGGG	
	CCCGCGTTGCACCAAGTGAT	CCCGCGTTGCACCAAGTGAAT	TCGCTTGGTGCAGATCGGGAA	
	CCCGCGTTGAACCAAGTGAAT	CCCGCGTTGCACCAAGTGAAG	TCGCTTGGTGCGGATCGGGAC	
	CCGCGTTGCACCAAGTGAA	CCCGCGTTGCACCAATTGAAT	CGCTTGGTGCAGATCGGGA	
	CCCGCGTTGCACCAAGTGATT	TCGCTTGGTGCACATCGGGAC	TTGCTTGGTGCAGATCGGGAC	
	CCTGCGTTGCACCAAGTGAAT	TCGCTTGGTGCATATCGGGAC	TCGCTTGGTGCAGGTCGGGAA	
	CCCGCGTTGCACCAAGTGAA	TCACTTGGTGCAGATCGGGAC	TCGCTTGGCGCAGATCGGGAC	
	CCCGCGTTGCACCAGGTGAAT	TCTCTTGGTGCAGATCGGGAC	TCGCTTGGTGCAGATCGTGAC	
	CCCGCGTTGCACCAAGTG	TGCTTGGTGCAGATCGGGAC	TCGCTTGGTACAGATCGGGA	
	CCGCGTTGCACCAAGTGAAT	TCGCTTGGTGCAGATTGGGAC	TCGCTTGGTGCAGATCGGCAC	
	CCCGCGTTGCACCAAGTGAAA	TCGCTTGGTGCAGATCGGGAG	TCGCATGGTGCAGATCGGGAC	
	TCGCTTGGTGCAGATCGGGCT	CGCTTGGTGCAGATCGGGAC	TCGCTTGTTGCAGATCGGGAC	
	CCCGCGTTGCACCAAGGGAAT	TCGCTTGGTGCAGATCGGGAT	TCGCTTGGTGCAGATCGGTAC	
	CCCGCGTTGCACCAAGTGG	CGCTTGGTGCAGATCGGG	TCGCTTGGTGCAGCTCGGGAC	
	CCCGCGTTGCACCAAGCGAAT	TCGCTTGGTGCAGATCGGGA	TCGCTTGGTGCAGATCGGGT	
	CCCGCGGTGCACCAAGTGAAT	TCGCTTGGTGCAGATCGGG	TCGCTTGGTGCAGTTCGGGAC	
	TCCGCGTTGCACCAAGTGA	TCGCTTGTGCAGATCGGGA	TCGCTTGGTGCAGATCGGGAC	
miR169	TAGCCAAGGATGACTTGCCG	GGCGGTCACCTTGGCTAG		
	GGCAGGTTGTCCTTGGCTAC	GGGCAAGTCATCCTGGTTACC		
	GGCAGGTTGTCCTTGGCTA	GGGCAAGTCAGCCTGGTTACC		
	TAGCCAAGGATGACTTGCCT	GGTGGGTCTTCTTGGCTAGC		
	CAGCCAAGGATGACTTGCC	GGCAAATCATCTGGGACT		
	GGCAGGTTGTCCTTGGCTACA			
	AGGCGGTCACCTTGGCTAGT			
	AGCCAAGGATGACTTGCCGGC			
	GAGAAGACCGTTGTTCGG			
	AGAGAAGACCGTTGTTCGG			
	CAGCCAAGGATGACTTGCCGGC			
	GGCAGGTTGTCCTTGGCT			
	GGCAGTCACCTTGGCTAG			
	GGCAGTCACCTTGGCTAGC			
	TAGCCAAGGATGACTTGCCTA			
	GGCAGTCTCCTCGGCTAGC			
	GGCGGTCACCTTGGCTAGC			
	GGCGGTCACCTTGGCTAGA			
	TAGCCAAGAATGACTTGCCT			

Supplementary Table 1… (continued): Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR170/171	TGAGCCGAACCAATATCACT		
	TGAGCCGAACCAATATCA		
	TGATTGAGCCGCGCCAATATC		
	TGATTGAGCCGTGCCAAT		
	TGATTGAGCCGTGCCAATA		
	TGATTGAGCCGTGCCAATATC		
	TGATTGAGCCGCGCAAATATC		
	TTGAGCCGTGCCAATATCTCG		
	TATTGGCTCGGCTCACTCAG		
	TGTTGGCTCGACTCACTCAGA		
	TTGAGCCGTGCCAATATCACG		
	TTGAGCCGTGCCAATATCTCT		
	TATTGGCTCGGCTCACTCAGG		
	ATTGGCTCGGCTCACTCAGG		
	TGTTGGCTCGACTCACTCAG		
	TGATTGAGCCGTGCCAATATCT		
miR172	GTGGCATCATCAAGATTCAC		
	GTGGCATCATCAAGATTCA		
	GGAATCTTGATGATGCTG		
	ATCTTGATGATGCTGCATCGG		
miR390	AAGCTCAGGAGTGATAGCGCC		
	AAGCTCAGGAGGGGTAGCGCC		
	AAGCTCAGGAGGGATAGCGCC		
	AGCTCAGGAGGGATAGCGCC		
	GCTATCTATCCTGAGCTCT		
	GCTATCTATCCTGAGCTCC		
	GCTATCTATCCTGAGCTCCT		
	CGCTATCTATCCTGAGCTCC		
miR393	ATTCCAAAGGGATCGCATT	TCAGTGCAATCCCTCTGGAAT	TTCCAAAGGGATCGCATTTTT
	TCCAAAGGGATCGCATTGATCT	TTCCAAAGGGATCGCATTG	TTCCAAAGGGATCGCATTTT
	TCCATAGGGATCGCATTGATC	TTCCAAAGGGATCGCATTAATT	TTCCAAAGGGATCGCATTGAT
	CATGCGTTCCTTTTGGAG	TTCCAAAGGGATCGCATT	
	TCCAAAGGGATCGCATTGATT	TTCCAAAGGGATCGCATTGTT	
	TCCAAAGGGATCGCATTGATC	TTCCAAAGGGATCGCATTGATT	
	TCCAAAGGGATCGCATTGATTT	TTCCAAAGGGATCGCATTGATC	
	TCCAAAGGGATCGCATTGTTT	TTCCAAAGGGATCGCATTT	
	TCCAAAGGGATCGCATTGATCC	TTCCAAAGGGATAGCATTGAT	
	TCAGTGCAATCCCTCTGGA	TTCCAAAGGGATCGCATTGA	

Supplementary Table 1… (continued): Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR394	TTGGCATTCTGTCCATCTCCT			
	TTGGCATTCTGTCCACCTCC			
	TTGGCATTCTGTCCATCTCC			
miR395	TGAAGTGTTTGGAGGATATT			
	TGAAGTGTTTGGAGGAACT			
	GAAGTGTTTGGGGGAACT			
	TGAAGTGTTTGGAGGAACTC			
	CTGAAGTGTTTGGGGGAACTC			
	GTTCTCCTCAAATCACTTCAGT			
	TGAAGTGTTTGGGGGAACTC			
	CTGAAGTGTTTGGGGGAACT			
	GAAGTGTTTGGAGGAACT			
	TGAAGTGTTTGGGTGAACTC			
	TGAAGTGTTTGGGGGAAATC			
	TGAAGTGTTTGGGGGAACT			
	TGAAGTGTTTGGGTGAACT			
	TGAAGTGTTTGGGGGAACTT			
	TGAAGTGTTTGGAGGAAC			
	GAAGTGTTTGGGTGAACT			
	AAGTGTTTGGAGGAACTTT			
	TGAAGTGTTTGGAGGAACTCT			
	AAGTGTTTGGAGGAACTCT			
	GTTCCTTTCAAGCACTTCATG			
	GTTCCCTGCAAGCACTTCACG			
	GTTTCCTACAAGCACTTCACG			
	GTTCCTTACAAACACTTCG			
	GTTCCCTACAAGCACTTCACG			
miR396	TCCCACAGCGTTCTTGAACTT	TTCCACAGCTTTCTTGAACTA	TTCCACAGCTTTCCTGAACTT	TTTCACAGCTTTCTTGAACTT
	TTCCAAAGCTTTCTTGAACTT	TTCCACAGCTTTCTTGAACT	TTACACAGCTTTCTTGAACTT	TTCCACAGCTTTCTTAAACTT
	TTCCACGGCTTTCTTGAACTG	CACAGCTTTCTTGAACTT	TTCCACAGCTTTCTTGAACCT	TTCCACAGCTGTCTTGAACTT
	TCCACAGCTTTCTTGAACTG	TTTCACAGCTTTCTTGAACTG	TTCCACAGCCTTCTTGAACTT	ATCCACAGCTTTCTTGAACTT
	TTCCACAGCTTTCTTGGACTG	TTCCACAGCTTTCTTGAAGTT	TTCCACAGCTTTCTAGAACTT	CCACAGGCTTTCTTGAACTG
	TTCCACAGCTTTCTTGAATTC	TCCACAGCTTTCTTGAACTT	TTCCGCAGCTTTCTTGAACTT	TCCACAGGCCTTCTTGAACT
	CACAGCTTTCTTGAACTG	TTCCACAGCTTTCTTGAACGT	TTCCACAGCTATCTTGAACTT	TCCACAGGCTTTCTTGAACTT
	TTCCACAGCTTTCTTGAACTC	TTCCACAGCTTTCTTGAACTT	TCCCACAGCTTTCTTGAACTG	ACAGGCTTTCTTAAACTG
	CCCCACAGCTTTCTTGAACTT	TTCGACAGCTTTCTTGAACTT	CCACAGCTTTCTTGAACTT	TCCACAGGCTTTCTTGAAGTG
	TTCCACAGCTTTCTTGAAC	TTCCACAGCTTTCTTGAACAG	TTCCACAGCTTTCGTGAACTT	TCCACAGGCTTTCTTGAACTC
	TTCCACAGCTTTCTTGAACTG	TTCCACGGCTTTCTTGAACTT	TTCCACAGCTCTCTTGAACTT	TCCACAGGCATTCTTGAACTG

Supplementary Table 1… (continued): Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR396	TCCACAGGCTTTCTTGAACC	TCCACAGGCTTTCTTGAACGG	TCCACAGGCTTTCTTAAACT	AAGAAAGCTGTGGGAAACT
	TCCACAGGCTTTCTTGAACTG	TCCACAGGCTTTCCTGAACTG	TACACAGGCTTTCTTGAACTG	CTCCACAGGCTTTCTTGAACTG
	TGGACAGGCTTTCTTGAACTG	TCCACAGGCTTTCTTGAATTG	TCCACAGACTTTCTTGAACTG	TTCCACAGCTTTCTTGAACTCA
	TCCACAGGCCTTCTTGAACTG	TCCACAGGCTTTCATGAACT	TCCACGGGCTTTCTTGAACTG	TTCCACAGCTTTCTTGAACTTC
	TCCACAAGCTTTCTTGAACTG	TCCACAGGCGTTCTTGAACTG	TCCACAGGCTTTCGTGAACTG	TTCCACAGCTTTCTTGAACTGT
	TCCACAGGCTTTCTTAAACTG	TCCAGAGGCTTTCTTGAACTG	TCCACAGGCTTTCTTTAACTG	TTCCACAGCTTTCTTGAACTTA
	TCCACAGGCTTTCTTGAACT	TCCGCAGGCTTTCTTGAACTG	TCCACAGGATTTCTTGAACTG	TTCCACAGCTTTCTTGAACTTT
	TCCACAGGCTTTCTGGAACTG	CCACAGGCTTTCTTGAAC	CCACAGGCTTTCTTGAACT	GGTCAAGAAAGCTGTGGGAAG
	TCCACAGGCTTTCTTCAACTG	TCCACAGGCTCTCTTGAACTG	TCCCCAGGCTTTCTTGAACT	ACAGTTCAAGAAAGTCCTTG
	TCCACAGGCTTTCTTGAAATG	TCCACAGGCTTTCTTGATCTG	TCCACAGGCTTTTTTGAACTG	AGTTCAAGAAAGTCCTTG
	TCCACAGGCTTTCATGAACTG	TCCACAGGCTTCTTGAACTG	TCCACATGCTTTCTTGAACTG	AGCTTTCTTGAACTTCTCT
	TCCATAGGCTTTCTTGAACTT	TCCAGAGGCTTTCTTGAACT	TCCACAGGCTTTCTTGAACTA	TCTCCACAGGCTTTCTTGAAC
	TTCACAGGCTTTCTTGAACTG	TCCACAGGCTTTCTTGAACCG	TCCACAGGCTTTCGTAAACTG	
	TCCACAGGCTTTCTTGAACAG	ACAGGCTTTCTTGAGCTG	TCCACAGTCTTTCTTGAACTG	
	TCCACAAACTTTCTTGAACTG	TCCACAGACTTTCTTGAACT	TCCACAGGCTTTCTTGATCT	
	TCCTCAGGCTTTCTTGAACTG	ACCACAGGCTTTCTTGAATTG	TTCAATAAAGCTGTGGGA	
	TCCACAGGCTTTCTTGGACTG	TCCACAGGCTTACTTGAACTG	TTCAATAAAGCTGTGGGAAA	
	TCCACCGGCTTTCTTGAACTG	TCCATAGGCTTTCTTGAACTG	TTCAATAAAGCTGTGGGAAAT	
	CCCACAGGCTTTCTTGAACTG	TCTACAGGCTTTCTTGAACTG	GTTCAATAAAGCTGTGGG	
	TCCACAGGCTTTCTTGAACA	TCCACAGGCGTTCTTGAACT	GTTCAATAAAGCTGTGGGAAA	
	ACCGGCTTTCTTGAACTG	TCCCCAGGCTTTCTTGAACTG	GTCAAGAAAGCTGTGGGAA	
	TCCACAGGCTTTCTTTCACTG	TCCACAGGCTTTCTCGAACTG	GTTCAATAAAGCTGTGGGAA	
	TCCAGGCTTTCTTGAACTG	ACAGGCTTTCCTGAACTG	GTTCAATAAAGCTGTGGGAAT	
	TCCACAGGCTTTCTTGAAC	ACAGGCTTTCTTGAACTG	GTTCAATAAAGCTGTGGGA	
	TCCACAGGTTTTCTTGAACTG	CACAGGCTTTCTTGAACTG	CTCCACAGGCTTTCTTGAACT	
miR397	TTGAGTGCAGCGTTGATGAAC			
	TTGAGTGCAGCGTTAATGAAC			
miR398	ATGTGTTCTCAGGTCGCCCCT			
	TGTGTTCTCAGGTCGCCCCTG			
	GTGTTCTCAGGTCGCCCCTG			
	TGTGTTCTCAGGTCACCCCTG			

Supplementary Table 1… (continued): Conserved miRNAs in *Avena strigosa*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR399	TGCCAAAGGAGAGTTGCCCTG
	TGCCAAAGGAGAATTGCCCTG
	TGCCAAAGGAGAGATGCCCTG
	GTGCGGTTCTCCTCTGGCATG
	GCGTACGTCTCCTTCGGC
	TGCCAAAGGAGATTTGCCCAG
	TGCCAAAGGAGAATTGTCCTG
	GCGCGGTTCTCCTCTGGCATG
	GTGCGTCTCTTCTCTGGCA
	CGCGGTTCTCCTCTGGCATG

Supplementary Table 2: Small RNA sequences in *Avena strigosa* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

Annotated miRNA	Sequence in *Avena strigosa* Small RNA Library		
miR158 (A. thaliana, A. lyrata)	TTCCAAATGTAGACAAAGAA		
	TTCCAAATGTAGACAAAGCA		
	TCCCAAATGTAGACAAAGCA		
miR401 (A. thaliana)	CGAAACTGGTGTCGACCGACA		
miR403	TTAGATTCACGCACAAACTCG		
miR404 (A. thaliana)	GCTGTCGGTTGCGGCTGC		
miR408	CTGCACTGCCTCTTCCCTGG		
	CTGCACTGCCTCTTCCCTGGC		
miR437	TTGTCCTAAGTCAAACTTT		
miR444	CGGGGCTCTCCGGCGCACC	TGTTGTCTCAAGCTTGCTGCC	TGCAGTTGTTGTCTTAAGCTT
	TGTTATCTCAAGCTTGCTGCC	TGCAGTTGTTGTCTCAATG	TGCAGTTGCTGCCTCAAGCTT
	GTTGTCTCAAGCTTGCTGCC	TGCAGTTGTTGTCTCAAG	GCAGTTGTTGTCTCAAGCTT
	TGTTGTCTCAAGCTTTCTGCC	CAGTTGTTGTCTCAAGCTT	TGCAGTTGTTGTCTCAAGCTT
	TGTTGTCTCAGGCTTGCTGCC	TGCAGTTGTTGTCTCAAGCT	TTCAGTTGTTGTCTCAAGCTT
	TGTCGTCTCAAGCTTGCTGCC	TGCAGTTGTTGTCTCAAGCCT	TGCAGTTGTTGTCTCAAGCTG
	TGTTGTCTCTAGCTTGCTGCC	GCAGTTGTTGTCTCAAGCT	TGTTGTCTCAAGCTTGCTGC
	TGTTGTCTCAAGCGTGCTGCC	CGCAGTTGTTGTCTCAAGCTT	TGTTGTCTCAAGCTTGCT
	TGTTGTCTCAAGCTTGCTGCT	TGCAGTTGTTGTGTCAAGCT	CTGTGACTTTCTTGCAAGTTG
	TGTTGTCTTAAGCTTGCTGCC	TGCAGTTGTTGTCTCAAGC	
miR479	TGAGCTGAACCAATATCACTC		
	TGAGCCGAACCAATATCACTC		
	TGAGCCGGACCAATATCACTC		
	TGAGCCGAACCAATATCAATC		
	TGAGCCAAACCAATATCACTC		
miR482	GGAATGGGCTGATTGGGAAGC		
	GGACTGGGCTGTTTGGGCT		
miR528	TGGAAGGGGCATGCAGAGGAG		
miR530	TGCATTTGCACCTGCACCTT		
miR538 (P. patens)	TGCAAGGAGTCTATGTCAG		
miR820 (O. sativa)	CGGCCTCGTGGAGGGACT		
miR824	CCTTCTCATCGATGGTCTAGA		
miR827	TTTTGTTGGTTGTCATCTAACC		
	TTAGATGACCATCAGCAAACA		
miR845	AGCTCTGATACCAATTGTTT		
	CGGCTCTGATACCAATTG		
	CGGCTCTGATACCAATTT		
miR894 (P. patens)	CGATTCACGTAGGGTTCACC	ATTCACGTCGGGTTCACC	CGTTTCACGTAGGGTTCACC
	TTTCACGTAGGGTTCACC	TTTCACGTCGGGTTCACC	GTTTCACGTAGGGTTCACC

Supplementary Table 2 …(continued): Small RNA sequences in *Avena strigosa* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR1030 (P. patens)	TGCATTTGCACCTGCACCT
miR1128	TTCTATTCCCTCCGTCCCAAA
miR1133 (T. aestivum)	ATACTCCCTCCGTCCAAG TACTCCCTTCGTCCTAAA
miR1310	ATGGGGGGCGCAACGCCT AGGCATGGGGGGCGCAACGCC AGGCATGGGGGGCGCAACGCCT CATGGGGGGCGCAACGCCT
miR1432	TCAGGAGAGATGACACCAAC ATCAGGAGAGATGACACCAAC
miR1847 (O. sativa)	ATTTTGCAGTTGTGGAAC
miR1863	CGCTCTGATACCATGTTAG
miR1878	ATTTGTAGTGTTCAGATTAAGTT ATTTGTAGTGTTCAGATTAAGT ATTTGTAGTGTTCAGATT ATTTGTAGTGTTCAGATTAAGTTT
miR2088 (M. truncatula)	CCAATGTAATCTCGGACT
miR2111	ATCCTCGGGACACAGATTACC
miR2118	TTCCCGATGCCTCTCATTCC TTCCTGATGTCTCCCATTCCTA
miR2275	TTCAGTTTCCTCTAACAT
miR2673 (M. truncatula)	CCGCTTCCTCTTCCTCTTCCTC
miR3438 (A. lyrata)	CATAGCTCAATCTCTGAC
miR3630	GGGAATCTCTCTGATGTT
miR5048 (H. vulgare)	TGTTTGCAGGTTTTAGGTCCAA TGTTTGCAGGTTTTAGGTTT TATTTGCAGGTTTGAGATCTAA GTTTGCAGGTTTAAGGTC TATTTGCAGGTTTGAGATCTA GTTTGCAGGTTTTAGGTCC GTTTGCAGGTTTTAGGTCCA ATTTGCAGGTTTGAGATCTA ATTTGCAGGTTTGAGATC ATTTGCAGGTTTGAGATCT TGTTTGCAGGTTTTAGGTCCA GTTTGCAGGTTTAAGGTCT TGTTTGCAGGTTTTAGGTCT TATTTGCAGGTTTGAGATC TTTGCAGGTTTTAGGTCCA ATTTGCAGGTTTGAGATCTAA TGTTTGCAGGTTTTAGGTC TTGTAGGTTTAAGGTCTA TGTTTGCAGGTTTTAGGTCC TATTTGCAGGTTTGAGATCT ATTTGCAGGTTTTAGGTC GTTTGCAGGTTTTAGGTC

Supplementary Table 2 …(continued): Small RNA sequences in *Avena strigosa* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR5049	AAGTAATATGGATCGGAGGAGT	AAGTAATTTGGATCGGAGGGA	ACTAATATGGGTCGGAGG
	GTAATATGGGTCGGAGGA	ACTAATATGGGTCGGAGGAAG	AGACATCTATTTTGGGACGGA
	AAATAATATGGATCGGAGGGAG	CAATTAATATGGATCGAAGGAAGT	AATTAATATGGGTTGGAGGGA
	CTCCGATCCATATTACTTG	AGTAATATGGGTCGGAGGGAG	AATTAATATGGATCAAAGGGA
	CAACTAATATGGATCGGAGGGAGT	ACTAATATGGATCGGAGGGAG	AATTAATATGGATCGGAGGGG
	AACTAATATGGATCGGAGGGA	AATTAATATGGATCGGAGGGAGT	ATTAATTTGGATCGGAGGGA
	CAATTAATATGGATCGGAGGGAG	CAACTAATATGGATCGAAGGAAGT	CATTAATATGGATTGGAGGGA
	AATTAATATGGATCGGAGGGA	CAAGTAATATGGGTCGGAGGGAGT	AATTAATATGGGTCGGAGGGA
	AGTAATATGAATCGGAGGGAGT	CAAGTAATATGGATCGGAGAGAGT	AATTAATATGGACTGGAGGG
	CCTCCGATCCATATTACTTGT	CAACTAATATGAATCGGAGGAAGT	CATTAATATGGACCGGAGGGA
	AAGTAATATGGATCGGAGGGA	CAAGTAATATGGATCAGAGGGAGT	AAATAATATGGGTCGGAGGGA
	AGTAATTTGGATCGGAGGGAG	TTCCGATCCATATTACTTG	AATTAATATTGGTCGGAGGGA
	AATATGGGTCGGAGGGAGT	CAACTAATATGGATCGGAAGAAGT	ATTAATATGGGTCGGAGGGA
	AATATGGGTCGGAGGGAG	CAATTAATATGGGTCGGAGG	ATTAATTTGGATCAGAGGGA
	AAGTAATATGGATCGGAGGGAGT	CAACTAATATGGATCGGAGGAAGT	AATTAATATGGGCCGGAGGGA
	GCAATATGGATCGGAGGGA	AACTAATATGGATCGGAGGGAG	AACTAATATGGGTCGGAGGGA
	CAATTAATATGGATCGGAGGG	ATTAATATGGACCGGAGGAAG	AATTAATATGGACTGGAGGGA
	CAAGTAATATGGATCGGAGGGAG	CCTCCGACCCATATTACTTG	ATAATATGGATCGGAGAGA
	CAAATAATATGGATCGGAGGGAGT	AGTAATATGGATCGGAGAGA	
	CAACTAATATGGATCGGAGAAAGT	CAAGTAATATGGATCGGAGGGAGT	
	CAAATAATATGGATCGGAGGAAGT	CAATTAATATGGATCGGAGGGAGT	
	CAACTAATATGGGTCGGAGGAAGT	AAGTAATATGGGTCGGAGGGA	
miR5053 (H. vulgare)	GCAGCCGTCGTCGCCGGCG		
miR5054 (B.distachyon)	TCCCCACAGTCGGCGCCA		
	TCCCCACGGTCGGCGCCA		
miR5059 (B.distachyon)	TCCTGGGCAGCAGCACCA		
miR5062 (B.distachyon)	TGAATCTTGGTGAAAAGC		
miR5067 (B.distachyon)	TCTGCGACAATTAATATGGAT		
	TCAGCGACAATTAATATGGAT		
	TGAGCGACAATTAATATGGAT		
miR5071 (O. sativa)	TCAAGTATCATATCATGGACA		
	TCAAGCACCATATCATGGACA		
	TCAAGTATCATGTCGTGGACA		
	TCAAGCATCATGTCATGGACA		
	TCAAGCATCATATCATGTAC		
miR5072 (O. sativa)	CGTCCCCAGCGGAGTCGCCA	ATCCCCAGCGGAGTCGCC	GTCCCCAGCGGAGTCGCCA
	TCCCCAGCGGAGTCGCCA	TTCCCCAGCGGAGTCGCC	TTCCCCAGCGGAGTCGCCA
	CGTTCCCCAGCGGAGTCGCC	ATCCCCAGCGGAGTCGCCA	CTCCCCAGCGGAGTCGCCA

Supplementary Table 2 …(continued): Small RNA sequences in *Avena strigosa* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR5077 (O. sativa)	TTCACGTCGGGTTTACCA
	TTCACGTCGGGTTCACCA
	ATTCACGTCGGGTTCACCA
	TTTCACGTCGGGTTCACCA
	TTCACGTAGGGTTCACCA
miR5083 (O. sativa)	GACTACAATTATCTGATC
	ACTACAATTATCTGATCA
miR5139 (R. glutinosa)	AAACCTGGCTCTGATATA
miR5164 (B.distachyon)	CTCAATTTTGTCTAGATACGC
	CAACTTTGTCTAGATACCT
	CTCAACTTTGTCTAGATACGC
miR5174 (B.distachyon)	ATTATGGATCGGAGGGAG
	TTTTGGGACGGAGGGAGT
miR5175 (B.distachyon)	AAGAATTATGGAACGGAGGGA
miR5180 (B.distachyon)	AGTGTTGCAGTTTTGAACT
miR5181 (B.distachyon)	TCCGATCCAAAATAAGTGTCT
	CCTCCGATCCATAATAAGTGT
miR5185 (B.distachyon)	TGAGAAATGAACTAGAAGT
	TTGAGAAATGAACTAGAAGT
miR5203 (B.distachyon)	ACTTATTATGGATCGGAGGGA
	ACTTATTTTGGATCGGAGGGA
	CTTATTATGGATCGGAGGGA
	ACTTATTATGGATCCGAGGGA
	ACTTATTATGGATCAGAGGGA
	ATTATTATGGACCAGAGGG
miR5267 (M. truncatula)	CATTTGCTAGAACACACCG
miR5299 (M. truncatula)	TTGGTTTTGGAAAGCGAC
miR5387 (S. bicolor)	CACGAACCGGTGTTAAAGGAC
miR5568 (S. bicolor)	TCCGTTCCAAATTATAAGATG
	AAAACATCTTATATTTTGGA
	GATGATGATGATGATGATC
	ATGAAGATGAGGATGATGAA
	ATGATGACGAGGATGATGA
	ATGAAGATGAGGATGATGAAA
miR5711 (B. rapa)	TGTTCTGTGGGTTTTTACCGA
miR5783 (G. max)	ACGAGGACGAGGAGGACG
	ACGACGACGAGGAGGACT
miR5803 (O. sativa)	ACCACTGTAGGCGCAACG

Supplementary Table 2 …(continued): Small RNA sequences in *Avena strigosa* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR6173 (H. brasiliensis)	CCGTAAACGATGGATACT		
	AGCCGTAAACGATGGATACT		
miR6191 (H. vulgare)	GATTTGTCTAGATTTGGA		
miR6197 (H. vulgare)	TCTGTTCCTAAATGTGAGA		
	TCTGTTCCTAAATATAAGA		
miR6219 (S. bicolor)	AACCGGGACTAAAGGTTC		
	CGGGACTAAAGGTGAGACTT		
	AACCGGGACTAAAGGTAGGTC		
miR6248 (O. sativa)	ATTTGCGGATGGAGGGAGTA		
miR6250 (O. sativa)	GGGGATAGATCGACGCGTC		
	GGGATAGATCGACGCGTC		
	GGGGATAGATCGACGCGT		
miR6300 (G. max)	GTCGTTGTAGTATAGTGA	GTCGGTGTAGTATAGTGG	GTCGCTGTAGTTTAGTGG
	GTTGTTGTAGTATAGTGG	GTCGTTAAAGTATAGTGG	GTCGTTGTAGAATAGTGG
	GTCGTTGTAGTATAGTGG	GTCGTTGTAGTATAGCGG	
	GTCGATGTAGTATAGTGG	GTCGTTGTAGTATTGTGG	
	GTCGTTGTAGTTTAGTGG	GTCGTTGTGGTATAGTGG	
	GTCTTTGTAGTATAGTGG	GTCGTTGTAGTATAATGG	
	GTCGTTGTAGTATAGGGG	GTCGTTGTAGGATAGTGG	
miR6478 (P. trichocarpa)	CAGACCTTAGCTCAGTTGGT	ACGACCTTAGCTCAGTTGGT	CCGACCTTACCTCAGTTGGT
	CCGACCATAGCTCAGTTGGT	CTGACCTTAGCTCAGTTGGG	CCGACCTTAGCTCGGTTGGT
	TCGACCTTAGCTCAGTTGGT	CCGACCGTAGCTCAGTTGG	CCAACCTTAGCTCAGTTGG
	CCGACCTTATCTCAGTTGGT	CCGACCTTAGCTCAGTTGGT	CCGACCTTAGCTCAATTGGT
	CCGAGCTTAGCTCAGTTGG	CCGACTTTAGGTCAGTTGGT	CCGACCTTAGCTCAGTTTGT
	CCGACCTTAGGTCAGTTGGT	CCGATCTTAGCTCAGTTGGT	CCGACCTTAGCTCAGTGGT
	CCGACCTTAGCTCAGTTGGTA	CCGACCTTAACTCAGTTGGTA	CCGACCTTAGCTCAGTTGGA
	CCGACCTTAGCTCAGTTGCT	CGACCTTAGCTCAGTTGGT	CCGACCTTAGCTCAGTTGG
	CTGACCTTAGCTCAGTTGG	CCAACCTTAGCTCAGTTGGT	CCGACCTTAGCTCGGTTGG
	CCGACCTTAACTCAGTTGG	CCGACCTTAGCTCAGTTAGT	CCGACCTTAGCTCAGATGGT
	CCGACCTTAGCTCAGTTGGC	CCGACCTTAGCTCAGTGGG	CCGACCTTAGCTCTGTTGGT
	CCGACCTTAGCTTAGTTGGT	CCGACCTTAGATCAGTTGGT	CCGACCTTAGCTCAGATGG
	GACCTTAGCTCAGTTGGT	CCGACCTTAGCCCAGTTGGT	CCGACCTTAGCTCAGGTGGT
	CGACCTTAGCTCAGTTGG	CGACCTTAACTCAGTTGGT	CCGCCCTTAACTCAGTTGGT
	CCGACCGTAGCTCAGTTGGT	CCGACCTTAGCTCAGTTAA	CCGACCTTAGCTCAGTTGGG
	CCGACCTTAGCTCAGTTAG	CCGACCCTAGCTCAGTTGGT	
	CCGACCTTAGCTCAGTTGTT	CCGACCTTAACTCAGTTGGT	
	CCGACCTTAGCTCAGTTGAT	CCGACCTTAGCTCAGCTGGT	
	GCGACCTTAGCTCAGTTGGT	CCGACCTTAGCTTAGTTGG	

Supplementary Table 2 …(continued): Small RNA sequences in *Avena strigosa* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR7714 (B.distachyon)	AATATGTATCGGAGGGAGTAT
	TAATATGGATCGGAGGGAGTA
	AATATGGATCGGAGGGAGT
	ATGGATCGGAGGGAGTAGC
	AATATGGATCGAAGGGAGTAC
miR7716 (B.distachyon)	GTCGTTATTCTCCAGTAA
miR7717 (B.distachyon)	CATTGAGATTCGTGGATA
miR7732 (B.distachyon)	CCAAGATCTCTGAGGCTC
miR7754 (B.distachyon)	TCTCTCGGTTGAGGAACT
miR7757 (B.distachyon)	TACAAAACCTTCAGCTATC
	CACAAAACCTTCAGCTATC
	CACAAAACCTTCAGCTATCCC
	CACAAAACCTTCAGCTATCCA
	TACAAAACCTTCAGCTATCC
	GATAGCTGAATGTTTTGTTT
	GATAGTTGAATGTTTTGTTT
	CACAAAACCTTCAGCTATCC
	TACAAAACCTTCAGCTATCCA
	CACAAAACCTTCACCTATCCA
	CACAAAACCTTCAGCTTT
	CCCCAAGATGAGTGCTCTC
	CCAAGATGAGAGCTCTCTC
miR7782 (B.distachyon)	CCGCTCTGATACCATGTG
	CTGCTCTGATACCATGTTGA
miR7783 (B.distachyon)	TGCTCTGATACCATGTTGA
	GCTCTGATACCATGTGGACT
miR8005 (S. tuberosum)	TTTAGGGTTTAGGGTTTAGAG
miR8024 (S. tuberosum)	GAGGATTTGAAGAATGCA
miR8155 (C. papaya)	AACTTGGCTCTGATACCA
	AAACCTGGCTCTGATACCA
	AACCTGGCTCTGATACCA
	AAACTTGGCTCTGATACCA
	AACACGGCTCTGATACCA

Supplementary Table 3: Conserved miRNAs in *Avena magna*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

Annotated miRNA	Sequence in *Avena magna* Small RNA Library			
miR156/157	TGACAGAAGAGAGTGACTTC	TGACAGAAGAGAGTGAACAC	TTGACAGAAGAGAGCGAGCAC	
	GCTCACTACTCTGTCTGTCAC	ACAGAAGAGAGTGAGCAC	TTGACAGAAGAGAGGGAGCAC	
	ACTACTCTGTCTGTCACC	TGACAGAAGAGAGTGATCAC	TTGACAGAAGAGAGTGAGCACT	
	GCTCACTACTCTGTCTGTCACC	TGACAGAAGAGGGTGAGCAC	TTGACAGAAGAGAGTGAGC	
	GCTCACTGCTCTGTCTGTCACC	CGACAGAAGAGAGTGAGCAC	TTGACAGAAGAGAGAGAGCAC	
	TCACTACTCTGTCTGTCACC	TGACAGAAGAGAGAGAGCAC	GCTCACTGCTCTATCTGTCAGC	
	TTGACAGAAGAGAGTGAGAAC	TGACAGACGAGAGTGAGCAC	TTGACAGAAGAGAGTGAGCAC	
	TTGAAAGAAGAGAGTGAGCAC	GACAGAAGAGAGTGAGCAC	GCTCACTCTCTCTCTGTCAGC	
	TTGACAGAAGAGAGTCAGCAC	TGACAGAAGGGAGTGAGCAC	GCTCACTTCTCTCTCTGTCAGC	
	CTGACAGAAGAGAGTGAGCAC	TGACAGAAGAGAGTGAGCAC	GCTCACTACTCTGTCTGTC	
	TTGACAGAAGAGAGTGACCAC	TGACAGAAGAGAGGGAGCAC	GCTCACTCCTCTTTCTGTCAGC	
	ACAGAAGAGAGTGAGCACA	TGACAGAAGAGAGTGAGAAC	TGCTCACTCCTCTTTCTGTCAGC	
	TGACAGAAGAGAGTGAGCACC	TGACAGAAGAGAGTGAGCTT	TGCTCACTACTCTGTCTGTCATC	
	TGACAGAAGAGAGTGAGCACT	TGACAGAAAAGAGTGAGCAC	TGCTCACTTCTCTCTCTGTCA	
	TGACAGAAGAGAGTGAGCACA	TGAGAGAAGAGAGTGAGCAC	TGCTCACTACTCTGTCTGTC	
	GCTCACTATTCTGTCTGTCACC	TGACAGAAGAGAGTGAGCAT	TGCTCACTACTCTGTCTG	
	GCTCACCACTCTGTCTGTCACC	TGACAGAAGAGAGTGAGCAA	ATTGACAGAAGAGAGTGAGCAC	
	GCTCACTACTCTGTCTGTCACA	TGACAGAAGAGAGTGAGCA		
miR159/319	CTTGGACTGAAGGGTGCTCC	TTGGACTGAAGGGCGCTCCCT	TTGGACTGAAGGGTGCTCCCA	CTTGGACTGAAGGGTGCTCCC
	CTTGGACTGAAGGGAGCTCCC	TTGGATTGAAGGGTGCTCCCT	TTGGACTGAAGGGTGCTCCC	TTGGACTGAAGGGTGCTCCCTC
	CTTGGACTGAAGGGTGCT	TTGGACTGAAGGGTGCCCCCT	TTGGGCTGAAGGGTGCTCCCT	TGGACTGAAGGGAGCTCCCTC
	ATTGGATGACGCGGGAGCTAA	TGGACTGAAGGGTGCTCCC	TTGGACTGAAGGGCGGTCCCT	TTGGACTGAAGGGAGCTCCCTT
	ACTGGATGACGCGAGAGCTAA	TTGGACTGTAGGGGGCTCCCT	TTAGACTGAAGGGTGCTCCCT	CATTCGAGAACTCTGATG
	ATTGGATGACGCGGGAGCT	TTGGACTGAAGGGTTCTCCCT	GGACTGAAGGGAGCTCCCT	TTTGGATTGGAGGGAGCTCTG
	AGAGAGTCCTTCAGTCCACTC	TTGGACTGAAGGGTGCTCGC	TGGACTGAAGGGTGCTCCCT	TTGGACTGAAGGGTGCTC
	AGAGCGTCCTTCAGTCCACT	TTGGACTGAAGGGTGCTCCT	TTGGACTGAAGGGAGCTCCCT	TTTGGATTGAAGGGATTT
	AGCTGCCGACTCATTCATCC	TTGGACTGAAGGGTGCTCCTT	TTGGACTGAAGGGGGCTCCCT	TTTGGATTGAAGGGAGCCA
	TTGGACTGAAGGGTGCTCCAG	TTGGACTGAAGGGTGTTCCCT	GGACTGAAGGGTGCTCCCT	GTTGGATTGAAGGGAGCTCTG
	TTGGACTGAAGGGTGCTTT	TTGGACTGAAGAGTGCTCCCT	TTGGACTGAAGGGGGCTCCC	TTTGGATTGACGGGAGCT
	AGCTGCCGACTCATTCATCCA	TTGGACCGAAGGGTGCTCCCT	GTGGACTGAAGGGTGCTCCCT	TTTGGATTGAAGGGAGCTCTT
	ATAGCGTCCTTCAGTCCACTC	TTGGACTGAAGGCTGCTCCCT	TGGACTGAAGGGAGCTCCCT	TTTGGATTGAAGGGAGCTTTG
	AGAGCGTCCTTCAGTCCACTC	TTGGACTGAAGGGTGCTCCCC	TTGGACTGAAGGGTGATCCCT	TTTGGATTGAAGGGAGCA
	TTGGACTGAAGGGTGCTCCCG	TTGGACTGAAGTGTGCTCCCT	TTGGACTGAAGGGAGATCCCT	TTTGGGTTGAAGGGAGCTCTG
	TTGGACTGAAGGGTGCTCC	TTGGACTGAAGGGTGCTCGCT	TTGGACTGAAGGGTACTCCCT	TTTGGATTGAAGGGAGCC
	TTGGACTGAAGGGCGCTCCC	TTGGACTGAAGGGAGCTCCC	TTGGACTGAGGGGTGCTCCCT	TTTGGATTGAAAGGAGCT
	TTGGACTGAAGGGCGCTCC	TTGGACTGAAGGGTGCTCCCT	CTTGGACTGAAGGGTGCCCCC	TTTGGATTGAAGGGAGCTC
	TTGGACTGAAGGGTGCTTCCT	TTGGACTGAAGGGTGATCCC	AGAGCTCTCTTCAGTCCACT	TTTGGATTGAAGGGAGCATTA

Supplementary Table 3 … (continued): Conserved miRNAs in *Avena magna*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR159/319	TTTGGATTGAAGGTAGCTCTG	TTTGGATTGAAGGGAGCTCGG	CTTGCATGACCGAGGAGCCGA
	TTTTGATTGAAGGGAGCTCTG	CTTGGATTGAAGGGAGCT	AGCTGCTTGTTCATGGTTCC
	TTTGGATTGAAGGGAGCTCT	TTTGGATTGAAGGGAGCTTT	AGCTGCTCGTTCATGGTTC
	TTTAGATTGAAGGGAGCTCTG	TTGGACTGAAGGGAGCTCC	TTGCATGACCGAGGAGCCGA
	TTTGGATTGAAGGGAGATCTG	TTTGGATTGAAGGGAGTA	AGCTGCTTGTTCATGGTTCCC
	TTTGGCTTGAAGGGAGCTCT	GAGCTCCTATCTTTCCAATGA	TTTGCATGACCGAGGAGCTGG
	TCTGGATTGAAGGGAGCTCT	TTTGGATTGAAGGGAGTTC	TTTGCATGACCGAGGAGC
	TTTGGATTGAAGGGAATT	TTTGGCTTGAAGGGAGCTCTG	CTTGCATGACCGAGGAGCCG
	TTTGGATTGAAGGGAGGTC	TTTGGATTGAAGGGCGCTCT	TTGCATGACCGAGGAGCTGA
	TTTGGATTGAAGGGGGCTCTG	TTTGGATTGAAGGGAGTTCT	TTTGCATGACCGAGGAGCT
	TTTAGATTGAAGGGAGCT	TCTGGATTGAAGGGAGCTCTG	AGCTGCTTGTTCATGGTTC
	TTTGGATTGAAGGGAGCTG	TTCGGATTGAAGGGAGCTCTG	TTTGGATTGAAGGGAGCTTAT
	TTTGGATTGAAGGGATTTC	TGGATTGAAGGGAGCTCT	TTTGGATTGAAGGGATTTCC
	TTTGGATTGAAGGGAGCTCC	TTGGATTGAAGGGAGCTC	AGCTCCCTTTGATCCAAACC
	TTTGGATTGAAGGGAGCTCTA	TTTGGCTTGAAGGGAGCTC	AGCTCCCTTTGATCCAAAC
	TTTGGATTGAAGGGAGTTTT	TTGGATTGAAGGGAGCTCTG	TTTGGATTGAAGGGAGTTTCT
	TTTGGATTGAAGGGAGCTTTT	TTTGGATTGAAGGGAGCG	CTCCCTTTGATCCAAACC
	TTTGGATTGAAGAGAGCTCT	TTTGGAATGAAGGGAGCT	CTTGGACTGAAGGGAGCTCC
	TTTGGATTGAAGGGAGTTT	CTTGGATTGAAGGGAGCTCTG	
	TTTGGATTGAAGGGTGCTCTG	TTTGGATTGAAGGGAGCCCTG	
	CTTGGATTGAAGGGAGCTCT	TTTGGATTGAAGGGCGCTCTG	
	TTTGGATTGAAGGGGGCTC	TTGGATTGAAGGGAGCTCTC	
	TTTGGATTGAAGGGAGCTCTC	TTTGGATTGGAGGGAGCTCT	
	TTTGGTTTGAAGGGAGCTCTG	TTTGGATTGAAGGGAGCTCA	
	TTTGGATTGAAGGGAGAT	TTTGGATTGAAGGGAGCTCTG	
	TTTGGATTGAAGGGAGCTATC	TTTGGATTGAAGGGAACTCTG	
	CTTGGACTGAAGGGAGCT	TTTGGATTGAAGGGAGCTTTA	
	GAGCTCCTATCTTTCCAA	TGGACTGAAGGGAGCTCC	
	TTTGGATTGAAGGGATCT	GAGCTCCTATCTTTCCAATG	
	TTTGGATTGAAGGGATCTCT	TTTGGATTGAAGGGAGCTT	
	CTTGGATTGAAGGGAGCTC	TTGGATTGAAGGGAGCTCT	
	TTTGGATTGAAGGGAGTTCA	TTTGGATTGAAGGGAGCTCG	
	TTTGGATTGAAGGGAGTT	TTTGCATGACCGAGGAGCTGA	
	TTTGGATTGAAGGGAGCT	AGCTGCTCGTTCATGGTTCCC	
	TTTGGATTGAAGGGAGCTA	AGCTGCTCGTTCATGGTTCC	
	TTTGGATTGAATGGAGCTCTG	AGCTGCTCGTTCATGGTT	
	TTCGGATTGAAGGGAGCTCT	AGCTGCTCGTTCATGGTTCT	
	GAGCTCCTATCTTTCCAATGG	TTTGCATGACCGAGGAGCTG	

Supplementary Table 3 … (continued): Conserved miRNAs in *Avena magna*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR160	CCTGGCTCCCTGAATGCC			
	TGCCTGGCTCCCTGTATTTC			
	TGCCTGGCTCCCTGTATGCCA			
	TGCCTGGCTCCCTGTATG			
	TGCCTGGCTCCCTGTATGCTC			
	TGCCTGGCTCTCTGTATGCCA			
	GCGTGCAAGGAGCCAAGCATG			
	CTCGCTCCCTGAATGCCG			
miR162	GGGCGCAGTGGTTTATCGATC			
	TCGATAAACCTCTGCATCCGG			
	CGATAAACCTCTGCATCCGG			
	TCGATAAACCTCTGCATC			
	TCGATAAACCTCTGCGTCCGG			
miR163	TGAAGAGGTCATGGAACT			
miR164	TGGAGAAGCAGGGCACGTGCA			
	TGGAGAAGCAGGGCGCGTGCA			
	TGGAGAAGCAGGGCACGTGCT			
	TGGAGAAGGAGCGCACATGCT			
	TCTTTGGTTATCTAGCTGTATGA			
	CATGTGCGCTCCTTCTCCAGC			
	CATGTGCCCTTCTTCTCCATC			
miR165/166	TCGGACCAGGCTTCCCCC	TCGGACCAGGCTTCATTTACT	TCGGACCAGGATTCAATCCCT	TCGGACCAGGATTCATTCACC
	GGGTTGTTGTCTGGTTCAAG	TCGGATCAGGCTTCATTCCCC	TCGGCCCAGGCTTCATTCCCC	TCGTACCAGGGTTCATTCCCC
	GGGTTGTTGTCTGGTTCA	TCGGACCAGGCTTCATTCCAA	TCGGACCAGGCTTAAATCCCT	TCGGACCAGCCTTAATTCCCC
	GGAATGTTGTCTGGTTCAAGA	TCGGACCAGGCTTCATTCCCT	CCGGACCAGGCTTCATTCCCC	TCGGACCAGGCTACAATCCCT
	GGGTTGTTGTCTGGTTCAAGG	TCGGACAAGGCTTCAATCCCT	TCGGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTACCC
	GAATGACGCCGGGTCTGAAAG	TCGTACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTTCG	TCGGATCAGGCTTCATTTCCC
	TCGGACCAGGCTTCAATCCCTT	TCGGACCAGGCTTCAATC	TCGGACCAGGCTTCATTCC	TCGGACCAGGCATCATTCCCC
	GGTATGTTGTCTGGCTCGGTG	TCGGACCAGACTTCATTCCCC	TCCGACCAGGCTTCATTCCCC	ACGGACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCATTTTTA	TCGGACCAGGCTTCCTTCCCC	TCGGACCAGGCTTCATTGCCC	TTGGACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCAATCCCTA	TCGGACCAGGCTTCAATCCCC	CGGACCAGGCTTCATTCC	TCGGACCATGCTTCAATCCCT
	TCGGACCAGGCTTCATTATTT	TCGGACCAGACTTCAATCCCT	TCGGACCACGCTTCATTCCCC	TCGGACCAGGCTTCATTCCTA
	TCGGACCAGGCTTCATTCCTTT	GGTATGTTGGCTGGCTCGA	TCGGACCAGGCTTCAGTCCCC	TCGGACCAGGCTTCATTTCT
	TCGGACCAGGCTTCAATCTTT	TCGGACCAGGCTTCATTCTC	TCGGAACAGGCTTCATTCCCC	GGAATGTTGTCTGGTTCA
	GGTATGTTGTCTGGCTCGGT	TCGGACCAGGCTTCAATTC	TCGGACCAGGATTCATTCCCT	TCGGACCAGGCTTCATTCCTC
	TCGGACCAGGCTTCATTTATA	GCGGACCAGGCTTCATTCCCC	TCGGACCAGGCCTCATTCCCC	TCGGACCAGGATTCATTCCCC
	GGTATGTTGTCTGGCCCGG	TCGGACCAGGCTTCATTACT	TCGGACCAGGATTAATTCCCC	TCGGACCAGGCTTCACCCCCC
	TCGGACCAGGCTTCATTTTTT	TCGGACCAGGCTTCAATCCCG	TCGACCAGGCTTCATTCCCC	TCGCACCAGGCTTCATTCCCC

Supplementary Table 3… (continued): Conserved miRNAs in *Avena magna*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR165/166	TCGGACCAGGCTTCAATT	TCGGGCCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTCCT	TCGGACCAGGCTTCATTCCCCG
	TCGGACCAAGCTTCATTACCC	TCGGACCAGGCTTCATTTCCC	TCGGACCAGGCTTCATTCACC	TCGGACCAGGCTTCATTCCCCC
	TCGGACCAGGCTTCATTTTC	TCGGACCAGGCTTCATTCCAC	TCGGACCAGGCTTTAATCCCT	TCGGACCAGGCTTCATTCCCCT
	TCGGACCAGGCGTCATTCCCC	TCGGACCAGGCTTCATCCCCC	TCGGACGAGGCTTCATTCCCC	TCGGACCAGGCTTCATTTCCCC
	TGGGACCAGGCTTCATTCCCC	TCGGACCAGGCTTAATTCCCT	TCGGACCAAGCTTCAATCCCT	CGGACCAGGCTTCATTCCCCT
	TCGGACCAGGCTTCATTTC	TCTGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTCTCT	TTCGGACCAGGCTTCATTCC
	GGAATGTTGTCTGGTTCAAGG	TCGGACCAGGCTACATTCCCC	TCGGACCAGGCTTTATTCCCC	TCTCGGACAAGGCTTCATTCC
	TCGGACCAGGCTTCATTACCA	TCGGACCAGGCTTCATTA	TCGGAGCAGGCTTCATTCCCC	CTCGGACCAGGCTTCATTCC
	GGACCAGGCTTCAATCCCT	GGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCAATCACT	TCTCGGACCAGGCTTCATTCC
	TCGGACCAGGGTTCATTCCCC	TCGGACCAGGCTTCATACCCC	TCGGACCAGGCTTCATTCCTT	TTTCGGACCAGGCTTCATTCC
	TCAGACCAGGCTTCATTCCCC	TCGGACCAGCCTTCATTCCCC	TCGGGCCAGGCTTCAATCCCT	TTTCGGACCAGGCTTCATTAC
	GGAATGTTGTCTGGCTCTGGG	TCGGACCAGGCTTCATCCCC	TCGGACCAGGCTTAATTCCCA	TTTCAGACCAGGCTTCATTCC
	TCGGACCAGGCTTCATTCTCC	TCGGACTAGGCTTCATTCCCC	CGGACCAGGCTTCATTCCC	CCTCGGACCAGGCTTCATTCC
	TCGGTCCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTTTCC	TCGGACCAGGCTTCATTCCCG	TCTCGGACCAGGCTTCAATCC
	TCGGAGCAGGCTTCAATCCCT	TCGGCCCAGGCTTAATTCCCC	TCGGACCAGGCTTCATTTTCT	TCGGACCAGGCTTCATTTATC
	TCGGACCAGGCTTCACTCCCC	TCGGACAGGGCTTCATTCCCC	TCGGACCAGGCTTCAATTCCT	TCGGACCAGGCTTAAATCCAT
	TCGGACCAGGCTTCAACCCCT	TCGGACCAGGCTTCAATCCCT	TCGGACCTGGCTTCATTCCCC	TCGGACCAGGCTTCATTACTT
	TGGACCAGGCTTCATTCCCC	TCGGACCAGACATCATTCCCC	TCGGACCATGCTTCATTCCCC	TCGGACCAGGCTTCATTCTTT
	TCGGACCGGGCTTCATTCCCC	TCGGACAAGGCTTAATTCCCC	TCGGACCTGGCTTCATTCCCT	TCGGACCAGGCTTCAATTTCT
	TCGGACCAGGCTTCTTTCCCC	TCGGACCAGGCTTGATTCCCC	TCGGACCAGGCTTCATTC	TCGGACCAGGCTTCATTATT
	TCGGACCAGGCTTCATTCA	TCGGACCAGGCTTCGTTCCCC	TCGGCCAAGGCTTCATTCCCC	TCGGACCAGGCTTCATTTTT
	TCGAACCAGGCTTCATTCCCC	TCGGACCAGTCTTCATTCCCC	TCGGACCAGGCTTCATTCCGC	TTGGACCAGGCTTCAACCC
	TCGGGCCAGGCTTCATTTCCC	TCGGACCAGGCTTCATTCTT	GGAATGTTGTCTGGCTCGAGG	TCGGACCAGGCTTCATTTTTC
	TCGGACCAGGCTTCAATCT	TCGGACCAGGCTTAATTCCAC	GGAATGTTGTCTGGCTCG	TCGGACCAGGCTTCAATTTC
	TCGGACCAGGCTTCAATCCC	TCGGAACAGGCTTCATTCCAC	GGAATGTTGTCTGGCTCGGG	GGGTTGTTGTCTGGGTCAAGG
	TCGGACCAGGCTTCATTCGCC	GGAATGTTGTCTGGTTGGAGA	GGAATGTTGTCTGGCTCGGGG	TCGGACCAGGCTTCAGTCCTT
	TCGGACCAGGCTTCATTCCC	TCGGACCAGGCTTCATTT	GAATGTTGTCTGGCTCGG	TCGGACCAGGCTTCATTCTTA
	TCGGACCAGGTTTCATTCCCC	TCAGACCAGGCTTCAATCCCT	TTCGGACCAGGCGTCATTCCC	TCGGACCAGGCTTCATTATTC
	TCGGACCCGGCTTCATTCCCC	GGAATGTTGTCTGGCTCGGGT	GTCGGACCAGGCTTCATTCCC	TCGGACCAGGCTTCATTTAT
	TAGGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTTT	TTCGGACCAGGCTTCATTCCC	TCGGACCAGGCTTCAATATC
	TCGGACCAGGCTTCACCCCC	TCGGACCAGGCTCCATTCCCC	GGAATGTTGTCTGGCTCGAG	GGTATGTTGGCTGGCTCGAT
	TCGGACCAGGCTTCAATCCCA	TCGGACCAAGCTTCATTCCCC	GGAATGTTGTCTGGCTCGG	GGTATGTTGGCTGGCTCGATG
	TCGGACCAGGCTTAATTCCCC	CGGACCAGGCTTCATTCCCC	GGAATGTTGTCTGGCTCAAGG	
	TCGGACCAGGCTTCATTCCCA	TCGGACCAGGATTCATTCCC	CTCGGACCAGGCTTCATTCCC	
	TCGGACAAGGCTTCATTCCCC	TCGGATCAGGCTTCAATCCCT	GGAATGTTGTCTGGCTCGA	
	TCGGACCAGGCTTCATTCTTC	TCGGACCAGGCTGCATTCCCC	GGTATGTTGTCTGGCCCG	
	GGAATGTTGTCTGGTTCAAG	TCGGACCAGGCTTCAATCC	CTCGGACCAGGCTTCAATCCC	
	TCGGACCAGGCTTCATTCT	TCGGACCAGGCTTCTTTCCC	TCGGACCAGGCTTCATTCCCCA	

Supplementary Table 3… (continued): Conserved miRNAs in *Avena magna*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR167	TCATGACTGACAGCCTCAT	ATCGTGCTGTGACAGCTTCACT	TGAAGCTGCTAGCATGATCTGA	AGATCATGCTGGAGTTTCATC
	ATCATGACTGACAGCCTCATT	GAAGCTGCCAGCATGATCTG	TGAAGCTGCCAGCATGATCTTT	
	GATCATGACTGACAGCCTCATT	GGTCATGCTGCGGCAGCCTCACT	ATCATGCTGTGCAGTTTCATC	
	TGAAGCTGCCAGCATGATCT	TGAAGCTGCCAGCATGATCTA	TGAAGCTGCCAGCATGATCTGA	
	TGAAGCTGTCAGCATGATC	TGAAGCTGTCAGCATGATCTG	TGAAGCTGCCAGCATGATCTGG	
	TGAAGCTGCCAGCATGATC	TGAAGCTGCCAGCATGATCTG	TGAAGCTGCCAGCATGATCTGC	
	AAGCTGACAGCATGGACT	TGAAGCTGTCAGCATGATCTA	TGAAGCTGCCAGCATGATCTGT	
	GATCGTGCTGTGACAGCTTCACT	TGAAGCTGTCAGCATGATCTGC	AGGTCATGCTGGAGTTTCATC	
	TGAAGCTGTCAGCATGAT	TCATGCTGTGCAGTTTCA	AGGTCATGCTGGAGTTTCAT	
	TGAAGCTGTCAGCATGATCT	GATCATGCTGTGCAGTTTCATC	GGGTCATGCTGGAGTTTCATC	
miR168	CCCGCGGTGCACCAAGTGAAT	CTTGGTGCAGATCGGGAC		
	ACGCGTTGCACCAAGTGAAT	TCGCTTGGTGCAGATCGTGAC		
	CCCGCGTTGCACCAAGTG	CGCTTGGTGCAGATCGGGA		
	CCGCGTTGCACCAAGTGAAA	TCGCTTGGAGCAGATCGGGAC		
	CCCGCGTTGCACCAAGTGG	TCGCTTGGTGCAGATCGGGAG		
	TCGCTTGGTGAAAATCGGGAC	TCGCTTGGTGCAGATCGGGAT		
	CCCGCGTTGCACCAAGTGAAT	TCGCTTGGTGCAGGTCGGGAC		
	CCCGCGTTGCACCAAGTGA	TCGCTTGGCGCAGATCGGGAC		
	CCCGCGTTGCACTAAGTGAAT	TCGCTCGGTGCAGATCGGGAC		
	TCGCTTGGTGCAGATCGTC	TCGCTAGGTGCAGATCGGGAC		
	CCCGCGTTGCACCAAGTGAA	TCGCTTGGTGCAGATCGGGGC		
	TCGCTTGGTGCAGATCGGTTC	TCGCTTGGTGCAGATCGGGAA		
	CCGCGTTGCACCAAGTGAA	TCGCTTGGTGCAGATCGGGG		
	CCCGCCTTGCACCAAGTGAAT	CGCTTGGTGCAGATCGGGC		
	CCGCGTTGCACCAAGTGAAT	TCGCTTGGGGCAGATCGGGAC		
	CCCGCGTTGCCCCAAGTGAAT	CGCTTGGTGCAGATCGGGAC		
	TCGCTTTGTGCAGATCGGGA	TCGCTTGGTGCAGATCGAGAC		
	TCGCTTGGTGCAGGTCGGGAA	TCGCTTGGTGCAGATCGGG		
	TCGCTTGATGCAGATCGGGAC	CCGCTTGGTGCAGATCGGGAC		
	TCGCTTGGAGCAGATCGGGA	TCGCTTGGTGTAGATCGGGAC		
	TCACTTGGTGCAGATCGGGAC			
	TCGCTTGGTGCAGATCGGGAC			
	TCGCTTGGTGCAGATCGG			
	GCTTGGTGCAGATCGGGAC			
	TCGCTTGGTGCAGATCGGGA			
	CGCTTGGTGCAGATCGGG			
	TCGCTTGGTGCAGATCGGAAC			
	TCGCTTGGTGCAGATCCGGAC			

Supplementary Table 3… (continued): Conserved miRNAs in *Avena magna*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR169	TAGCCAAGGATGACTTGCCT	GGCAGGTTGTCCTTGGCTAC	TAGCCAAGGATGACTTGCCGGC	TAGCCAGGGATGATTTGCCTC
	GGCAGGTTGTCCTTGGCTACC	CAAGTTGTCCTTGGCTAAC	AGAGAAGACCGTTGTTCGG	
	GGCAGGGTGTCCTTGGCTAC	CAGCCAAGGATGACTTGC	CAGCCAAGGATGACTTGCCGGC	
	GCAGGTTGTCCTTGGCTACA	GCAGGTTGTCCTTGGCTAC	GGCAGGTTGTCCTTGGCT	
	CAGCCAAGGATGACTTGCCGG	CAGGTTGTCCTTGGCTAC	GGCAAGTTGTCTTTGGCT	
	GGCAGGTTGTCCTTGGCTACA	GGCGAGTTGTTCTTGGCTAC	CCGGCAGGTTGTCCTTGGC	
	GGCAGGTTGTCGTTGGCTACA	GCAGGTTGTCCTTGGCTACT	CGGCAGGTTGTCCTTGGC	
	GGCGAGTTGTTCTTGGCT	GGCAGGTTGTCCTTGGCTA	AGCGGTCACCTTGGCTAGC	
	CAGGTTGTCCTTGGCTACA	AGCCAAGGATGACTTGCCGA	GGCAGTCTCCTCGGCTAGC	
	CGAGTTGTTCTTGGCTACT	CGAGTTGTTCTTGGCTACA	GGCAGTCACCTTGGCTAGC	
	AGCCAAGGATGACTTGCCGG	GCAGGTTGTCCTTGGCTAT	GAGACATTTTGGCTAATT	
	GGCAGGTTGTCCTTGGCG	GGCAGGTTGTCCTTGGATAC	GGCAGTCACCTTGGCTAGA	
	GGCAGGTTGTCCTTGGGTAC	GCAGGTTGTCCTTGGCTA	GGCGGTCTCCTTGGCTAGC	
	GGCAGGTTGTCCTTGGCTAT	CAAGTTGTCCTTGGCTAGC	GGCGGTCACCTTGGCTAGC	
	TAGCCAAGGATGACTTGCC	CAGCCAAGGATGACTTGCCG	GGCGGTCACCTTGGCTAGA	
	CAGCCAAGGATGACTTGCCGA	GCAAGTCCGTCCTTGGCT	ATCGGCGAGTTGTTCTTGGC	
	CGAGTTGTTCTTGGCTAC	GGCAAGTCCGTCCTTGGCTA	GGGCAAGTCATCCTGGTTACC	
	GGCGAGTTGTTCTTGGCTA	GGCAAGTCCGTCCTTGGCTAC	GGCAAGTCATCCTGGTTACC	
	GGCAGGTTGTCCTTGGCTACT	AGCCAAGGATGACTTGCCGGC	GGTGGGTCTTCTTGGCTAGC	
miR170/171	TATTGGCTTGGCTCACTC	TTGAGCCGTGCCAATATCACG		
	CGGTATTGGTGCGGTTCAATC	TTGAGCCGTGCCAATATCA		
	AGGTATTGGTGCGGTTCAATC	TATTGGCTCGGCTCACTCAG		
	TGATTGAGCCGTGCCAATATC	TGTTGGCTCGACTCACTCAGT		
	TGATTGAGCCGAGCCAATATC	TGTTGGCTCGACTCGCTCAGA		
	TGATTGAGCCGTGCCAATA	TGTTGGCTCGACTCACTC		
	TGATTGAGCCGCGCCAATTTC	TTGAGCCGTGCCAATATCTCT		
	TGATTGAGCCGCGCCAATATC	TGATTGAGCCGTGCCAATATCT		
	TGATTGATCCGTGCCAATATC	TTGGCATGGTTCAATCAA		
	TGATTGAGCCGGGCCAATATC	TGTTGGCATGGTTCAATCAGA		
	TGATTGAGCCGTGCCAAT	TGTGATGTTGGCATGGTTCAA		
	TGATTGCGCCGTGCCAATATC	CGAGCCGAACCAATATCACTC		
	TGTTGGCTCGACTCACTCAGG			
	TATTGGCTCGGCTCACTCA			
	ATTGGCTCGGCTCACTCAGG			
	TGTTGGCTCGACTCACTCAGA			
	TGTTGGCTCGACTCCCTCAGA			
	TATTGGCTCGGCTCACTCAGG			
	TGTTGGCTCGACTCACTCCGA			

Supplementary Table 3… (continued): Conserved miRNAs in *Avena magna*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR172	GGAATCTTGATGATGCTGCAT			
	GTGGCATCATCAAGATTCA			
	GGAATCTTGATGATGCTG			
	GGAATCTTGATGATGCTGCA			
miR390	AAGCTCAGGAGGGATAGCGCC			
	AAGCTCAGGAGGGATAGAGCC			
	AAGCTCAGGAGGGATAGCGCA			
	AAGCTCAGGAGGGATAGC			
	CGCTATCTATCCTGAGCTC			
	GCTATCTATCCTGAGCTCT			
	GCTATCTATCCTGAGCTCC			
	GCTATCTATCCTGAGCTC			
	GCTATCTATCGTGAGCTC			
	GCTATCTATCCTGAGCTCTT			
	CGCTATCTATCCTGAGCTCCA			
	CGCTATCTATCCTGAGCT			
miR393	TCCAAAGGGATCGCATTGATCC			
	TCCAAAGGGATCGCATTGATC			
	TCCAAAGGGATCGCATTGACC			
	TCCAAAGGGATCGCATTGATTT			
	TCAGTGCAATCCCTCTGGA			
	TTCCAAAGGGATTGCACTGAT			
	TTCCAAAGGGATCGCATTGATC			
	TTCCAAAGGGATCGCATTGAT			
	TTCCAAAGGGATCGCATTGATT			
miR394	TTGGCATTCTGTCCATCTCCT			
	AGGTGGGCATATTGCCAATG			
	TTGGCATTCTGTCCATCTCC			
	TTGGCATTCTGTCCATCTC			
miR395	GCAGGGTTTTGAGGAACT			
	TGAAGTGTTTGGGGGAACTC			
miR396	TTCCACGGCTTTCTTGAACTT	CCACAGCTTTCTTGAACTG	TTCCACAGCTTTCTTGAACTC	CCACGGCTTTCTTGAACT
	TTCCGCAGCTTTCTTGAACTG	TTCCACGGCTTTCTTGAACT	TTCCATAGCTTTCTTGAACTG	CACAGCTTTCTTGAGCTG
	TTCAACGGCTTTCTTGAACTT	TTCCACAGCTTTCTTGAACGG	TCCACAGCTTTCTTGAAC	CCACAGCTTTCTTGAACT
	TTCCACAGCTTTCTTGAACTT	TTCCACAGCTTTCTTGAACTG	TACCACAGCTTTCTTGAACTT	TTCCACAGATTTCTTGAACTG
	TTCCACAGCTTTCTTGAAC	TTCACAGCTTTCTTGAACTG	CACAGCTTTCTTGAACTG	ACCACAGTCTTTCTTGAACTG
	TCCACAGCTTTCTTGAACTT	CACAGTTTTCTTGAACTG	TACCACAGCTTTCTTGAACT	TCGACAGGCTTTCTTGAACTG
	TTCCACATCTTTCTTGAACTG	TCCACAGCTTTCTTGAACTG	TTCCACAGCTTTCTTGAACT	TCCACAGGCCTTCTTGAACTG

Supplementary Table 3… (continued): Conserved miRNAs in *Avena magna*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR396	TCCACAGGCTTCCTTGAACT	TCCATAGGCTTTCTTGAACTG	GTTCAATAAAGCTGTGGGA
	GCCACAGGCTTTCTTGGACTG	TCCACGGGCTTTCTTGAACTG	ATCCACAGGCTTTCTTGAACTG
	ACAGGATTTCTTGAACTG	TCCACAGGATTTCTTGAACT	CTCCACAGGCTTTCTTGAACTG
	TCCACAGGCTTCCTTGAACTG	ACAGGTTTTCTTGAACTG	GGTCAAGAAAGCTGTGGGAAG
	TCCACAGGTTTTCTTGAACTG	TCCACAGGCTTTCATGAACTG	TTCCACAGCTTTCTTGAACTGT
	CCCACAGGCTTTCTTGAACTG	TCCGCAGGCTTTCTTGAACTG	TTCCACAGCTTTCTTGAACTCT
	CACAGGCTTTCTTGAACTT	CCACAGGCTTTCTTGAACTG	TTCCACAGCTTTCTTGAACTTT
	TCCACAGGCTTTCCTGAACTG	CACAGGCTTTCTTGAACT	TCTCCACAGGCTTTCTTGAACT
	TCCACAGGCTTTCTTGAATTG	TCCAAAGGCTTTCTTGAACTG	
	GTTCAAGAAAGTCCTTGGAA	TCCACAGGCTTTCTTGAACTG	
	TCCCAGGCTTTCTTGAACTG	TCCACAGGCTTTCTCGAACTG	
	TCCACAGGCTTTCTTGAATTC	CATAGGCTTTCTTGAACTG	
	TCCACAGGCTCTCTTGAACTG	ACAGGCTTTCTTGAACTG	
	ACAGGGTTTCTTGAACTG	TCCACAGGCTTTCTTGATCTG	
	ACAGGCTTTCTTGAACGG	TCCACAGGCTTTGTTGAACTG	
	TCCTCAGGCTTTCTTGAACTG	TCCACAGACTCTCTTGAACTG	
	ACAGGCTTACTTGAACTG	TCCACAGGCTTTCTTGAACG	
	CCAGGCTTTCTTGAACTG	TCCAGAGGCTTTCTTGAACTG	
	TCCACAGGCTTTTTTGAACTG	GTTCAAGAAAGCCCATGGAAA	
	TCCACAGACTTTCTTGAACTG	TCCACAGGCTTTCTTAAACTG	
	CACAGGCTTTCTTGAACGG	TCCACAGGCTTTCTTGAAC	
	TCCACAGGCTTTCTTGAACA	CACAGGCTTTCTTGAATTG	
	TCCACAGGCTTTCTTGAACT	CCACAGGCTTTCTTGAACT	
	ACAGGCTTTATTGAACTG	TCCACAGGCTTTCTTGAACTC	
	CACAGGCTTTCTTGAACCG	CACAGGCTTTCTTGAACTG	
	TCCACAGGCTTTCTTGAA	TCCACAGGCTTTCTTGAACTA	
	TCCACAGGCTTTCTTGGACTG	TCCACAGGCTTTCTTGAACGG	
	ACCACAGGCTTTCTTGAACTG	CACAGGCTTTCTTGGACTG	
	TCCACAGGCTTTCTTGAACCG	TCCACCGGCTTTCTTGAACTG	
	TCCAGGCTTTCTTGAACTG	TCCACTGGCTTTCTTGAACTG	
	CTACAGGCTTTCTTGAACTG	TTCACAGGCTTTCTTGAACTG	
	TCCACATGCTTTCTTGAACTG	TTCAATAAAGCTGTGGGAAA	
	ACAGGCTTTCTTGAACTA	TTCAATAAAGCTGTGGGAAT	
	TCCACAGGATTTCTTGAACTG	TTCAATAAAGCTGTGGGA	
	TCCACAGGCTTTCTTGAACTT	TTCAATAAAGCTGTGGGAAAT	
	TGCACAGGCTTTCTTGAACTG	GTCAAGAAAGCTGTGGGAAGA	
	TCCACAGGCTTTATTGAACTG	GTTCAATAAAGCTGTGGGAAA	
	TCTACAGGCTTTCTTGAACTG	GTTCAATAAAGCTGTGGGAA	

Supplementary Table 3… (continued): Conserved miRNAs in *Avena magna*. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR397	TCACCGGCGCTGCACGCACT
	TTGAGTGCAGCGCTGATGAAC
	TTGAGTGCAGCGTTGATGACC
	TGAGTGCAGCGTTGATGAAC
	TTGAGTGCAGCGTTGATGAAC
miR398	CGACGACATGAGATCACG
	TGTGTTCTCAGGTCGCCCCTA
	GTGTTCTCAGGTCGCCCC
	TGTGTTCTCAGGTCGCCCC
	TGTGTTCTCAGGGCGCCCCTG
	TGTGTTCTTAGGTCGCCCCT
	TGTGTTCTCAGGTCGCCT
	TGTTCTCAGGTCGCCCCTG
	TGTGTTCTCAGGTCGCCCCCC
	TGTGTTCTCAGGTCGCCCCTG
	TGTGTTCTCAGGTCGCCCCT
	GTTCTCAGGTCGCCCCTG
	GTGTTCTCAGGTCGCCCCT
	GTGTTCTCAGGTCGCCCCTG
	TGTGTTCTCAGGTCGCCC
	TGTGTTCTCAGGTCGCCCCGG
miR399	TGCCAAAGGAGAGTTGCCCTG
	TGCCAAAGGAGAATTGCCCTG
	CGTGCCAAAGGAGAATTGCCC
	CAACTTCTCCATTGCCAGA
	TGCCAAAGGAGAGTTGCCC
	TGCCAAAGGAGAATTGTCCTG

Supplementary Table 4: Small RNA sequences in *Avena magna* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

Annotated miRNA	Sequence in *Avena magna* Small RNA Library		
miR403	GTGCGTGCATCTGAGGGC		
miR408	TGCACTGCCTCTTCCCTGGCTT		
	ATGGACTGCCCCTTCCCTGG		
	CAGGGGTGGAGCAGAGCAAGG		
	TGCACTGCCTCTTCCCTGGCTC		
	TGCACTGCCTCTTCCCTGGCT		
	CTGCACTGCCTCTTCCCTGG		
	CTGCACTGCCTCTTCCCTGGC		
miR444	GTTGTCTCAAGCTTGCTGCC	TGCAGTTGCTGCCTCAAGCTT	GCAGTTGTTGTCTCAAGCT
	TGTTGTCTCAAGCTTGATGCC	CAGTTGTTGTCTCAAGCT	TGTTGTCTCAAGCTTGCTGC
	TGTTGTCTCAAGCTTGCCGCC	TGCAGTTGCTGCCTCAAGCT	GTTGTCTCAAGCTTGCTG
	TGTTGTCTCAAGCTTGCTGCC	TGCAGTTGTTGTCTCAAGCTT	TGTTGTCTCAAGCTTGCT
	TGTTGTCTCAAGCTTTCTGCC	TGCAGTTGTTGTATCAAGCTT	TGTTGTCTCAAGCTTGCTG
	TATTGTCTCAAGCTTGCTGCC	TGCAGTTGTTGTCTCAAGTTT	CTGTGACTTTCTTGCAAGTTG
	TGTTGCCTCAAGCTTGCTGCC	AGTTGTTGTCTCAAGCTT	CTGTGACTTTCTTGCAAGT
	TCTTGTCTCAAGCTTGCTGCC	TGCAGTTGTTGTCTCAAGCCT	CTGTGACTTTCTTGCAAG
	TGTTGTCTCAAGCTTGCTGAC	TGCAGTTGTTGTCTCAAGC	CTGTGACTTTCTTGCAAGTT
	TGCAGTTGCTGCCTCAAGC	TGCAGTTGTTGTCTCAAGCT	
	TGCAGTTGTTGTCTCAAGTTC	TGCAGTTGTTGTCTCAAGCC	
miR479	TGAGCCGAACCAATATCACTC		
	TGAGCCGAACCAAAATCACTC		
miR528	TGGAAGGGGCATGCAGAG		
	TGGGAGGGGCATGCAGAGGAG		
	TGGAAGGGGCATGCAGAGGAG		
	TGGAAGGGGCATGCAGGGGAG		
	TGGAAGGGGCATGCAGAGA		
	TGGAAGGGGCATGCAGAGG		
	GGAAGGGGCATGCAGAGGAG		
	GGAAGGGGCATGCAGAGGA		
	CCTGTGCCTGCCCCCTCCATT		
	CCTGTGCCTGCCTCCTCCATT		
miR827	TTAGATGACCATCAGCAAACA		
	TTTTGTTGGTTGTCATCTAACC		
miR831 (A. thaliana, A. lyrata)	TCTCTTTGTACTGTTCTTG		
miR845	CGCTCTGATACCAATTGT		
miR854	TGATCGATCATGTAAGCT		
miR858	TTCGTTGTCTGTTCGACCT		
	TTCGTTGTCTGTTCGACCTCG		

Supplementary Table 4… (continued): Small RNA sequences in *Avena magna* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR894 (P. patens)	ATTCACGTCGGGTTCACC
	CGTTTCACGTAGGGTTCA
	TTTCACGTAGGGTTCACC
	TTTCACGTCGGGTTCACC
	ATTCACGTAGGGTTCACC
miR1117 (T. aestivum)	CCCGGTTCTTGGCACGAACC
	CCGGTTTGTGACACGAACC
	CCGGTTCGTGGCTTGAAC
	CCGGTTCGTGGCTTGAACC
	TTCCGGTTCGTGGCATGAACC
	CACCGATTCGTGGCACGAAC
miR1120	TCTTATATTTTGAAACGGA
	TTATTATATTATGGGACT
	CTTATATTATAAAACGGA
miR1128	CTACTCCCTCCGTTCGGAA
miR1135	GACAAGTATTTCCGGACG
miR1310	CATGGGGGGCGCAACGCC
	CATGGGGGGCGCAACGCCA
	CATGGGGGGCGCAACGCCT
	GCATGGGGGGCGCAACGCCT
	ATGGGGGGCGCAACGCCA
	ATGGGGGGCGCAACGCCT
miR1432	TCAGGAGAGATGACACCGAC
	ATCAGGAGAGATGACACCGA
	ATCAGGAGAGATGACACCAAC
	CAGGTGTCATCTCCCCTGAAC
	TCAGGAGAGATGACACCAAC
miR1436	ATATTATGGGACGGAGGTAGT
miR1511	CTTGGCTCTGATACCATAAA
miR1514	TTCCGTTTTAAAATAGAC
miR1863	TCTGATACCATGTAAGATTG
	GGCTCTGATACCATGTTAGTTTA
	GGCTCTGATACCATGTTAGTTTAG
	TCAACATGGTATCAGAGCGGC
miR1878	ATTTGTAGTGTTCAGATTGAGTTT ATTTGTAGTGTTCAGATTGAGTT
	TTGTAGTGTTCAGATTAAGTTT ATTTGTAGTGTTCAGATTGAGT
	ATTTGTAGTGTTCAGATTA ATTTGTAGTGTTCAGATTGA
	ATTTGTAGTGTTCAGATTAAGTTT TTTGTAGTGTTCAGATTGAGTTT

Supplementary Table 4… (continued): Small RNA sequences in *Avena magna* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR1882 (O. sativa)	ATTGTTTTCAAGGTGATT		
miR2111	AATCTGACTCCTGAGGTC		
miR2118	TATTGCACTAGTCCCGGG		
miR2120 (O. sativa)	ATACCTTTAGTCCCGGTTGT		
miR2655 (M. truncatula)	CGTTTTGGTCCGTTAACTTG		
miR3630	ATGGGAATCTCTCTGATGT		
	TGGGAATCTCTCTGATGTT		
	GGGAATCTCTCTGATGTG		
	GGGAATCTCTCTGATGTT		
miR3711 (*P. abies*)	GCGCTAGAAGGAGGGCAT		
miR4364 (G. max)	ACAGCGGAAGAAACTGCTT		
	CCAACAGCGGAAGAAACT		
	CAACAGCGGAAGAAACTGCT		
miR5048 (H. vulgare)	TGTTTGCAGGTTTTAGGTCCAA	TTTTGCAGGTTTTAGGTC	TGTTTGCAGGTTTTAGGTCTT
	ATTTGCAGGTTTTAGGTCC	TATTTGCAGGTTTGAGATC	GTTTGCAGGTTTAAGGTCTAA
	TGTTTGCAGGTTTTAGGTC	TTTGCAGGTTTTAGGTCTAA	TGTTTGCAGGTTTAAGGTC
	ATTTGCAGGTTTGAGATCT	TGTTTGCAGGTTTAAGGTCTAA	TGTTTGCAGGTTTTAGGTCCA
	TTTGCAGGTTTAAGGTCT	TTTGCAGGTTTTAGGTCT	TTTGTAGGTTTAAGGTCT
	TGTTTGCAGGTTTTAGGTCC	TATTTGCAGGTTTTAGGTCTA	TGTTTGCAGGTTTAAGGTCT
	TTTGCAGGTTTGAGATCT	TATTTGCAGGTTTTAGGTCT	TATTTTCAGGTTTCAGGTC
	TGTTTGCAGGTTTAAGGTCTA	ATTTGCAGGTTTGAGATCTAA	TATTTGCAGGTTTGAGATCTA
	TTTGTAGGTTTAAGGTCTA	ATTTGCAGGTTTGAGATCTA	
	TGTTTGCAGGTTTTAGGTCT	GTTTGCAGGTTTTAGGTCCAA	
miR5049	CAAGTAATTTGGATCGGAGGGAG	CAACTAATATGGATCGGAGGG	ATTAATATGGATCGGAGGGAGT
	AAGTAATATGGATCGGAGGTAGT	CAATTAATATGGATCGGAGGGAGT	ACTAATATGGACCGGAGG
	CAACTAATATGGATCGGAGAAAGT	ATATGGGTCGGAGGGAGT	AATTAATATGGGTCGGAGGAAGT
	CAATTAATTTGGATCGGAG	CAAGTAATGTGGATCGGAGGGAGT	CAACTAATATGGATCGGAGGTAGT
	CAAGTATTATGGATCGGAGGGAGT	ACTAATATGGATCGGAGGGAGT	CAATTAATTTGGATCGGA
	AGTAATATGGATCGGAGGGAGT	ACTAATATGGGTCGGAGG	CAAGTAATATGGGTCGGAGAAAGT
	CAATTAATATGAATCGGAGGAAGT	CAAGTAATATGGATCAGAGGGAGT	AATTAATATGGATCGGAGGGAGT
	CAAGTAATATGGATTGGAGGGAG	CAAGTAATATGGATCGGAGGAAGT	CCCTCCGATCCATATTACTTG
	CAATTATTATGGATCGGAGGA	TTAATATGGATCGGAGGGAGT	TAATATGGGTCGAAGGAAGT
	CAATTAATATGGATCGAAGGAAGT	CAAGTAATATGAATCGGAGGGAGT	CAAGTAATTTGGATCGGAGGGAGT
	CAAGTAATATGGATCGGAGGGAG	CAACTAATATGGATCGGAGGGAGT	CAACTAATATGAATCGGAGGAAGT
	ATTAATTTGGATCGGAGG	TAATATGGGTCGGAGGGAG	AGTAATATGAATCGGAGGGAG
	AACTAATATGGGTCGGAGGAAGT	CAACTAATATGGGTCGGAGGAAGT	TTAATATAGATCGGAGGAAGT
	CAAGTAATATGGGTCGGAGGGAGT	TCCCTCCGATCCATATTACTTGT	AATTTGGATCGGAGGGAGT
	AACTAATATGGGTCGGAGGA	AATTAATTTGGATCGGAGGAA	CAACTAATATGGATCGGAGGA

Supplementary Table 4… (continued): Small RNA sequences in *Avena magna* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR5049	AACTAATATGGATCGGAGGGA	AATTAATTTGGATCGGAGAGA	CATTAATATGGACCGGAGGGA
	ATTAATATGGATCGGAGGGAG	AATTAATATGGGTCGGAA	AACTAATTTGGATCGGAGGGA
	CAAGTAATATGGATCGAAGGGAGT	ACTAATTTGGATCGGAGGGA	AATTAATATGGGTTGGAGGGA
	CAAGTAATATGGATCGGAGAGAGT	TTAATATGGATCGGAGAGA	AACTAATATGGATTGGAGGGA
	CAACTAATATGGATCGGAGGAAGT	ATTAATATGGACCGGAGT	AATTAATATGGATCGGAGAGA
	CAATTAATTTGGATCGGAGGA	AATTAATATGGGTCGGAGGGA	CAACTAATATGGGTCGGAGG
	GTAATATGGATCGGAGGGAGT	AATTAATATGGATCGGAAGGA	AAGTAATATGGATCAGAGGGAGT
	ACAATTAATTTGGGACTGAGG	AATTAATTTGGATTGGAGGGA	TAAGTAATATGGATCGGAGGGAGT
	ACACTTATTTAGGATCGGAG	AATTAATTTGGATCGGAGGGA	CAAGTAATATGGATCGGAGGGAGT
	ACACTTATTTAGGATCGGA	AATTAATTTGGATCGAAGGGA	
	AACTAATATGGATCAGAGGGA	AACTAATATGGGTCGGAGGGA	
miR5054 (B. distachyon)	TCCCCACGGACGGCGCCA		
	GCCCCACGGTGGGCGCCA		
	TCCCCACGGTCGGCGCCA		
miR5059 (B. distachyon)	TCCTGGGCAGCAACACCA		
	TCCTGGGCAGCAGCACCA		
miR5062 (B. distachyon)	TGAATCTTGGTGAAAAGCCG	TGAACCTTGGATAAAAGCCGCCT	
	TGAATCTTGGGTAAAAGCCGCCT	TGAATCTTGGTGAAAAGC	
	CTTGGATAAAAGCCGCCT		
miR5067 (B. distachyon)	TCAGCGACAATTAATCTGGAT	CAGCGACAATTAATATGGAT	
	TCAGCGAGAATTAATATGGAT	TCAGCGACAATTAATATGGAT	
	TAGCGACAATTAATATGGAT	CAGCGACAAAGAATATGGAT	
miR5071 (O. sativa)	TCAAGCATCATATCATGT		
	TCAAGCATCATATCATGTACA		
	TCAAGCATCATGTCATGGACA		
miR5072 (O. sativa)	ATCCCCAGCGGAGTCGCC	ATCCCCAGCGGAGTCGCCA	
	TTCCCCAGCGGAGTCGCCA	TTCCCCAGCGGAGTCGCC	
	TCCCCAGCGGAGTCGCCA	GTTCCCCAGCGGAGTCGCCA	
miR5073 (O. sativa)	ATTTGGTGAATCGGAAACC		
miR5075 (O. sativa)	CTCCGTCACCGTCGTCCG		
miR5077 (O. sativa)	TTTCACGTCGGGTTCACCA	TTCACGTAGGGTTCACCA	
	TTCACGTCGGGTTCACCA		
	TTCACGTCGTGTTCACCA		
miR5082 (O. sativa)	CGCGATGATGGCCGCGCGGGCTCA		
miR5083 (O. sativa)	AGACTACAATTATCTGATCA		
miR5160 (O. sativa)	CGAGATCGATGGTGCATT		

Supplementary Table 4… (continued): Small RNA sequences in *Avena magna* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR5161 (O. sativa)	TATGAATCAGAGGGAGTA
miR5164 (B. distachyon)	CAACCTTGTCTAGATACG
	CTCAACTTTGTCTAGATACG
	CTCAACTTTGTCTAGATACGT
	TAACTTTGTCTAGATACGA
miR5174 (B. distachyon)	TTATGGATCGGAGGGAGT
	ATATGGATCGGAGGGAGTA
	TTTGGATCGGAGGGAGTA
	TTTATGTGACGGAGGGAGTA
miR5181 (B. distachyon)	CGACACTTATTATGGATCGGA
	ACACTTATTATGGATCGG
	TCCGATCCGTCATAAGTGTCG
	TCCGATCCATAATAAGTGTCG
	TCTGATCCATAATAAGTGTTG
miR5185 (B. distachyon)	TTTGAGAAATGAACTAGAAGT
	TTGACAAATGAACTAGAAG
miR5203 (B. distachyon)	ACTTATTTTGGATCGGAGGGA
	ACTTATTTAGGACCGGAGGGA
	ACTTATTATGGATCAGAGGGA
	ACTTATTATGGACCGGAGGGA
	ACTTATTATGGATCGGAGGGG
	ACTTATTATGGATTGGAGGGA
	ACTTATTATGGATCGGAGGGA
miR5205 (M. truncatula)	TATTTTGGGACGGAGGGAGT
	ATTTTGGGACGGAGGGAGT
miR5227 (M. truncatula)	AAGAGAAGCAGGTTGATG
miR5265 (M. truncatula)	GACATGTTGGAATGGTTA
miR5299 (M. truncatula)	TTGGTTTTGGAAAGCGAC
miR5503 (O. sativa)	TTCGGATCTTTATAGATGC
miR5568 (S. bicolor)	AAGGAGTGTATCTAGAAA
	CAAATTATAAGATGTTGTAGC
	ATGAAGATGATGAGGATG
miR5742 (M. truncatula)	AGAGCAATGGTCGTTGGA
miR5770 (G. max)	TAGGACGGTGGTTTGGAC
miR5778 (G. max)	ACGACCTCTTCGTCGGCT
miR5786 (G. max)	CCGCAGGATAGAGGGCAC
miR6031 (B. napus)	AGGTTCGTAGAGGTTTGAAG
miR6151 (*N. tabacum*)	AATGTGAGGCATTTCATT

Supplementary Table 4… (continued): Small RNA sequences in *Avena magna* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR6173 (*H. brasiliensis*)	GCCGTAAACGATGGATACT		
	CCGTAAACGATGGATACT		
	AGCCGTAAACGATGGATACT		
miR6181 (*H. vulgare*)	GCTCTTCTTGGACTGTGGCG		
	CTTCTTGGACTGTGGCGC		
	CTCTTCTTGGACTGTGGCG		
miR6191 (H. vulgare)	TAGATTTGTCTAGATTTGG		
miR6197 (H. vulgare)	TCTGTTCCTAAATATAAGACG		
	TCTGTTCCTAAATGTAAGCCT		
	TCTGTTCCTAAATATAAGATG		
miR6219 (S. bicolor)	GAACCGGGACTAAAGGTCGGACCT		
miR6224 (S. bicolor)	TTCGTCCTAAAATATAAGGC		
miR6300 (G. max)	GTCGTTGTAGTATAATGA		
	GTCGTTGTAGTATAGTGG		
	GTGGTTGTAGTTTAGTGG		
	GTCGTTGTAGTATAGTGC		
	GTCGCTGTAGTTTAGTGG		
	GTCGTTGTAGTATAGTGT		
	GTCGTTTTAGTATAGTGG		
	CTCGTTGTAGTATAGTGG		
	GGCGTTGTAGTATAGTGG		
miR6424 (P. trichocarpa)	CGTTCGCCAGCTTTGCACC		
miR6425 (P. trichocarpa)	CCATGGAAGACAGTGACT		
miR6455 (P. trichocarpa)	AATAGCATCCTCAACTAT		
miR6478 (P. trichocarpa)	CCGACCTTCGCTCAGTTGGT	CCGACCTTAACTCAGCTGGT	CGACCTTAGCTCAGTGGC
	CCGACCTTAACTCAGTTGGC	CCGACCTTAGCTCAATTGGT	CCGACCTTAGCTCAGTTGG
	CCGACCTTAGCTAAGTTGGT	CGACCTTAGCTCAGTTGAT	TCGACCTTAGCTCAGTTGGT
	GACCTTAACTCAGCTGGT	CCGACCTTAGATCAGTTGGG	CGACCTTAGCTCAGCTGG
	CCGACCTTAACTCAGTTGG	CCGACCTTAGCTCAGTTCG	CCACCTTAGCTCAGTTGGT
	CGACCTTAACTCAGTTGG	CTGACCTTAGCTCAGTTGG	CCGACCTTAGCTCAGTCGGT
	CCGACCTTAACTCAGCTGG	CCGACCTTAGCTCAGCTGC	CCGACCTTACCTCAGTTGG
	CCGAGCTTAGCTCAGTTGGT	CCGATCTTAACTCAGTTGGT	CCGACATTAGCTCAGTTGGT
	CGACCTTAACTCAGTTGGT	CCGACCATAGCTCAGTTGGT	CGATCTTAGCTCAGTTGGTA
	CCGATCTTAGCTCAGTTGGT	CCGACCTTAGCTCAGCTGGT	CCGACCCTAGCTCAGTTGGT
	CCGACCTTAGCTCAGCTGGG	CCGACCTTACCTCAGTTGGT	CCGACCTTAGCTCAGATGGT
	CCGACCTTAGCTCAGCTGG	CCGACCTTAGCTCAGTTG	CCGACCTTAGCTCAGTGGT
	CGACTTTAGCTCAGTTGG	CCGACCTTAGCTCAGTTGGT	CCGACCTTAGCTCATTTGGT
	CCGACCTTAGCTCAGCTG	CCGACCTTAGCCCAGTTGGT	CGACCTTAGCTCAGTTGGT

Supplementary Table 4… (continued): Small RNA sequences in *Avena magna* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR6478 (P. trichocarpa)	CGACCTCAGCTCAGTTGGT	CCGACCTTGGCTCAGTTGGT	CCGACCTTAGCTCAGTTGGA
	CCAACCTTAGCTCAGTTGGT	CCGATCTTAGCTCAGTTGG	CCGACCTTAGCTTAGTTGGT
	CCGACCTTAGCTCTGTTGGT	CCGACTTTAGCTCAGTTGGT	GGACCTTAGCTCAGTTGGT
	CCGATCTTACCTCAGTTGGT	ACGACCTTAGCTCAGTTGGT	CCGACCTGAGCTCAGTTGGA
	CCGACCTGAGCTCAGTTGG	CCGACCTTAACTCAGTTGGT	CCCCCTTAGCTCAGTTGG
	CCGACCTTAGCTCAGTGGG	CCGACCTTAACTCAGTTGGG	CGACCTTAGCTCAGTTGGC
	CGACCTTAGCTCAGTTGGG	CCGACCTTAGCTCAGTTGCT	CCGACCGTAGCTCAGTTGGT
	CCGACCTTAGGTCAGTTGGT	CCGACCTTAGCTCAGGTGGT	CCGACCTTAGCTCAACTGG
	GCGACCTTAACTCAGTTGGT	CGACCTTAGCTCAGTTGGA	CCGACCTCAGCTCAGTTGGT
	CCGACCTTAGCTCAGTTGGTA	GACCTTAGCTCAGTTGGC	CCGACCTTAGCTCAGTTAGT
	CCGACCTTAGCTCAGTTGGCA	CCGACCTGAGCTCAGTTGGT	CCGACCTGAGCTCAGTTG
	CTGACCTTAGCTCAGTTGGT	CCGACCTGAGCTCAGTTGGTA	CCGACCTTCGCTCAGTTGGC
	CCGACCTTAGCTCAGTTGGC	CGACCTTATCTCAGTTGGT	CCGAACTTAGCTCAGTTGGT
	GACCTTAGCTCAGTTGGT	CCGACTTTAGCTCAGTTGG	CGACCTTAGTTCAGTTGGT
	CGACCTTAGCTCAGTTGG	CCGACCTTAAATCAGTTGGT	CCGACCTTAGCTCAGTTGGG
	GCGACCTTAGCTCAGTTGGT	CGACCTTAACTCAGTTGGC	CCTACCTTAGCTCAGTTGGT
	CGACCTTAGCTCAGCTGGT	CCGACCTTAGCTCGGTTGGT	CCGACCTTAGCTCAGTGG
	CCGACCTTAGCTCAGTTGA	CGACCTTAACTCAGTTGGG	CCGACATTAGCTCAGTTGG
	CCGACCTTAGATCAGTTGGT	CGATCTTAGCTCAGTTGGT	
	TCGACCTTAGCTCAGTTGG	CCGACCTTAACTTAGTTGGT	
	CCGACCTTAGCACAGTTGGT	CCGACATTAGCTCAGGTGGT	
miR7714 (B. distachyon)	TAATATGGACCGAAGGGA		
	AATATGGATCGGAGGGAGT		
	AATTTGGATCGAAGGGAGTA		
	AATATGAATCGGAGGGAGT		
	CTAATATTGATCGAAGGGA		
miR7717 (B. distachyon)	AGGCGACTGAGAATTAACT		
	AGATGACTGAGAAATAACT		
	TGCTATTTCTCGGTCGACT		
miR7720 (B. distachyon)	TTACACATGACTTGGGTA		
miR7725 (B. distachyon)	CAAGGGACATGGTTTTCG		
miR7732 (B. distachyon)	CCAAGATCTCTGAGGCTC		
miR7737 (B. distachyon)	TTTGGTCACTGTATTTCA		
miR7757 (B. distachyon)	GATAGCTGAATGTTTTGTT	CACAAAATCTTCAGCTATCCA	CACAAAACCTTCAGCTATCCA
	CACAAAACCTTCAGCTATCCG	GATAGCTGAATGTTTTGTTTA	GATAGTTGAATGTTTTGTTTA
	CATAAAACCTTCAGCTATCC	CACAAAACCTTCAGCTATCCC	ATAGCTGAATGTTTTGTTTA
	TACAAAACCTTCAGCTATCCA	TAAAACCTTCAGCTATCCA	CACAAAACCTTCAGCTATCC
	CACAAAACCTTCAGCTATC	ACAAAACCTTCAGCTATCC	

Supplementary Table 4… (continued): Small RNA sequences in *Avena magna* with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR7767 (B. distachyon)	CCCCAAGATGAGTGCTCTC
	CCCCAAGATGAGTGCTCT
miR7782 (B. distachyon)	AACCGCTCTGATACCATG
	CCGCTCTGATACCATGTTGA
	ACCTGCTCTGATACCATGTTGTGT
	GCCGCTCTGATACCATGTT
	ACCTGCTCTGATACCATGTTAGGA
miR7998 (S. tuberosum)	ACGGACCGTAGATGAACC
miR8005 (S. tuberosum)	TTTAGGGTTTAGGGTTTAGA
	TTTAGGGTTTAGGGTTTAG
miR8021 (S. tuberosum)	TGACTCAAACTCGAGACCT
miR8028 (S. tuberosum)	TATTTATAGTACAAGGAT
miR8030 (S. tuberosum)	CGGTTGGTTTGGTCTTGG
	GGTTGGTTTGGTCTTGGA
miR8047 (S. tuberosum)	ATATTTTCGAAATTGGAC
miR8122 (P. persica)	GGGAAGATTTGTTGAAAG
miR8155 (*C. papaya*)	AACTTGGCTCTGATACCA
	AACCTGGCTCTGATACCA
	AAACCTGGCTCTGATACCA
	AACCTGGATCTGATACCA
	AACCTGGCTCTAATACCA

Supplementary Table 5: Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

Annotated miRNA	Sequence in Synthetic Hexaploid Small RNA Library			
miR156/157	TTTGACAGAAGAGAGTGAGCAC	TTGACAGAAGAGAGTGAGGAC	TGACAGAAGAGAGTTAGCAC	CGACAGAAGAGAGTGAGCAC
	GCTCACTGCTCTATCTGTCAG	TTGACAGAAGAGAGTGGGCAC	TGACAGAAGAAAGTGAGCAC	TGACAGAAGAGAGTGAGCAT
	GCTCACCTCTCTCTCTGTCAGC	CTCACTACTCTGTCTGTCACC	TGACCGAAGAGAAATGAGCAC	TGACAGAAGAGAGTGATCAC
	GCTCACCTCTCTCTCTGTCAG	TTGACATAAGAGAGTGAGCAC	TGACAGAAGAGAGTAAGCAC	TGTCAGAAGAGAATGAGCAC
	GCTCACTTCTCTCTCTGTCACA	TTGACAGAAGGGAGTGAGCAC	TGACAGAAGTGAGTGAGCAC	GACAGAAGAGAGTGAGCAC
	TTGACAGAAGAGAGCGACCAC	CACTACTCTGTCTGTCACC	TGACAGAAGAGAGTGATCCC	TGACAGAAGAGAGTGAGCGC
	TTGACAGAAAAGAGTGAGCAC	TTGACAGAAGAGCGTGAGCAC	TGACAGAAGAGAGAGAGCAC	TTACAGAAGAGAGTGAGCAC
	TTGACAGAAGAGAGTGAGCTC	TTGACAGAAGAGAGTGAACAC	TGACAGAAGAGAGTGACCAC	TGACAGAAGAGAGTGAGCT
	ACTACTCTGTCTGTCACC	TTGACAGAAGAGAGTGAGCAA	AGACAGAAGAGAGTGAGCAC	TGACAGAAGAGAGTGAGCAA
	TTGACAAAAGAGAGTGAGCAC	TTGACAGAAGAGAGTAAGCAC	TGGCAGAAGAGAGTGAGCAT	TGTCAGAAGAGAGTGAGCAC
	TTGTCAGAAGAGAGTGAGCAC	TTGACAGGAGAGAGTGAGCAC	TGACAGAAGAGAGCGAGCAC	TTTACAGAAGAGAGAGAGCAC
	TTGACAGAAGAGATTGAGCAC	TGACAGAAGAGAGTGAGCACT	TGACAGAAGAGACTGAGCAC	TTGACAGAAGAGAGTGAGCACC
	TTGATAGAAGAGAGTGAGCAC	TGACAGAAGAGAGGGAGCACA	TGACAGAAGAGAGGGAGCAC	TTGACAGAAGAAAGTGAGCAC
	TTGACAGAATAGAGTGAGCAC	TGACAGAAGAGAGTGAGCACG	TGACAGAGGAGAGTGAGCAC	TTGACAGAAGAGAGTGAGCA
	TTGACAGAAGAGAGTGACCAC	GACAGAAGAGAGTGAGCACT	TGACAGAAGATAGTGAGCAC	CTCACTCCTCTTTCTGTCAGTC
	TTGACAGATGAGAGTGAGCAC	TGACAGAAGAGAGTAAGCACC	TGGCAGAATAGAGTGAGCAC	TTGACAGAAGAGAGTGAGCAC
	TCACTACTCTGTCTGTCACC	TGACAGAAGAGAGTGAGCACC	TGACAGAAAAGAGTGAGCAC	TTGACAGAAGAGAGCGAGCAC
	GCTCACTACTCTGTCTGTCACC	TGACAGAAGAGAGTGAGCATT	TGACAGAAGAGAGTGAGCTT	GCTCACTACTCTGTCTGTCAGC
	TTGACAGAAGAGAGTGATCAC	GACAGAAGAGAGTGAGCACA	TGACAGAAGAGAATGAGCAC	GCTCACTGCTCTATCTGTCAGC
	GCTCACTTCTCTCTCTGTTACC	TGACAGAAGAGAGAGAGCACA	TGACAGAAGAGAATGAGCA	TTGACAGAAGAGAGAGAGCAC
	TTGACAGAAGAGAGTTAGCAC	TGACAGAAGAGAGTGAGCACA	TACAGAAGAGAGTGAGCAC	GCTTACTCCTCTTTCTGTCAGC
	TTGGCAGAAGAGAGTGAGCAC	ACAGAAGAGAGTGAGCACA	TGACAAAAGAGAGTGAGCAC	TTGACAGAAGAGAGTGAGCATA
	TTGACAGAGGAGAGTGAGCAC	TGACAGAAGAGAGCGAGCACT	TGACAGGAGAGAGTGAGCAC	TTGACAGAAGAGAGTGAGC
	TTAACAGAAGAGAGTGAGCAC	ACAGAAGAGAGTGAGCACT	TGACAGAAGAGAGTGCGCAC	TTGACAGAAGAGAGTGAGCACT
	TTGACAGAAGAGGGTGAGCAC	TGACAGAAGAGAGTGAGCGCT	TAACAGAAGAGAGTGAGCAC	TTGACAGAAGAGAGTGAGCACA
	GCTCACTACTCTGTCTGTCAC	TGACAGAAGAGAGTGAGCATC	TAACAGAAGAGAATGAGCAC	TTGACAGAAGAGAGGGAGCAC
	TCGACAGAAGAGAGTGAGCAC	GCTCACTACTCTGTCTGTCGCC	TGACAGAAGAGAGTGAGCA	GCTCACTCTCTCTCTGTCAGC
	TTGACAGAAGAGAGTGAGCAT	GCTCACTACTCTGTCTGCCACC	TGGCAGAAGAGAGTGAGCAC	GCTCACTGCTCTATCTGTAAGC
	TTGACAGAAGAGAGTGAGAAC	GCTTACTACTCTGTCTGTCACC	TGACAGAAGAGAGTGGGCAC	GCTTACTGCTCTATCTGTCAGC
	GCTCACTCTCTCTCTGTCAG	GCTCACTACTCTGTCTGTTACC	TGACAGAAGAGAGTGAGCAC	TGCTCACTCCTCTTTCTGTCA
	CTGACAGAAGAGAGTGAGCAC	GCTCACTACTCTGTCTATCACC	ACAGAAGAGAGTGAGCAC	TGCTCACTTCTCTGTCTGTCACC
	TTGACAGAAGAGAGTGAGCGC	CCTCACTACTCTGTCTGTCACC	TGCCTGAAGAGAGTGAGCAC	TGCTCACTTCTCTCTCTGTCACC
	TTGACGGAAGAGAGTGAGCAC	TCTCACTACTCTGTCTGTCACC	TGACAGAAGAGAGTGAACAC	GCTCACTCCTCTTTCTGTCA
	TTGACAGAAGAGAGTCAGCAC	GCTCACTACTCTGTCTGTGACC	TGACAGAAGAGAGTCAGCAC	TGCTCACTCCTCTTTCTGTC
	TTGACAGAAGCGAGTGAGCAC	GCTCGCTACTCTGTCTGTCACC	TGACAGAAGAGAGTGAGC	GCTCACTGCTCTTCCTGTCATC
	TTTACAGAAGAGAGTGAGCAC	GCTCAATACTCTGTCTGTCACC	TGACAGAAGAGGGTGAGCAC	GCTCACTTCTCTCTCTGTCACC
	TTGACAGAAGAGAGTGTGCAC	TGACAGAAGAGAGTGAGCCC	TGACAGAAGAGAGTGAGCTC	TGCTCACTACTCTGTCTGTCATC

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR156/157	GCTCACTCCTCTTTCTGTC			
	GCTCACTCCTCTTTCTGTCAGC			
	TGCTCACTCCTCTTTCTGTCAGC			
	TGCTCACTCCTCTTTCTGTCAG			
	TGCTCACTACTCTGTCTGTC			
	TGCTCACTTCTCTCTCTGTCAGC			
	TGCTCACTACTCTTTCTGTCACC			
	GCTCACTTCTCTCTCTGTCAG			
	GCTCACTCCTCTTTCTGTCAG			
	GCTCACTTCTCTCTCTGTCAGC			
	TGCTCACTACTCTGTCTGTCA			
	GCTCACCTCTCTCTCTGTCA			
	ATTGACAGAAGAGAGTGAGCA			
	ATTGACAGAAGAGAGTGAGCAC			
	ATTGACAGAAGAGAGTGAGC			
	ATTAACAGAAGAGAGTGAGCAC			
	TCTGACAGAAGAGAGTGAGCAC			
	CTTGACAGAAGAGAGTGAGCAC			
	TTGACAGAAGAGAGTGAGCCT			
	TTGACATAAGAGAGTAAGCAC			
	GAGAGAAGATGGAGAACA			
miR159/319	CTTGGACTGAAGGGTGCTC	AGCGTCCTTCAGTCCACT	TTGGGCTGAAGGGTGCTCCCT	TTGGACTGAAGGGTGCTCTCT
	CTTGGACTGAAGGGTGCTCCT	AGAGCGTCCTTCAGTCGACTC	TTGGACTGAAAGGTGCTCCCT	TTGGACTGAAGGGTGCACCCT
	CTTGGACTGAAGGGAGCTCCC	AGAGCGTCCTTCAGTCTACTC	TTGCACTGAAGGGTGCTCCCT	TTGGACTGAAGGGTGCCCCCT
	CTTGGACTGAAGGGTGCT	AGAGCGTCCTTCAGTCCACT	TTGGACTGAAGAGTGCTCCCT	TTGGACTGAAGGGGGCTCCCT
	CTTGGACTGAAGGGTGCTCC	AGCTGCCGACTCATTCATCC	TTGGACTGAAGGGTGCTTCCT	TTGGACTGAAGGGTGCTA
	ACTGGATGACGCGAGAGCTAA	AGAGCGTCCTTCAGTCCT	TTGGACTGAAGGGTGCTCCC	TTGGACTGAAGGGCCCTCCCT
	GAGCTTTCTTCGGTCCACT	AGAGCGTCCTTCAGTCCACTT	TTGGACTGAATGGTGCTCCCT	CTGGACTGAAGGGTGCTCCCT
	ATTGGATGACGCGGGAGCTAA	AGAGCGTTCTTCAGTCCACTC	TTGGAATGAAGGGTGCTCCCT	ATGGACTGAAGGGTGCTCCCT
	AGAGCGTCCTTCAGTCCACAC	TTGGACTGAAGGGAGCTCCC	GACTGAAGGGTGCTCCCT	TTAGACTGAAGGGTGCTCCCT
	AGCTGCCAACTCATTCATCCA	TTGGACTGAAGGGGGCTCCCC	TTGGACTGAAGGGTGCTCCCT	GTGGACTGAAGGGTGCTCCCT
	AGAGCGTCCTTCAGTCCACTC	TTGGACTGAAGGGAGCTCCCT	TTGGACTGAAGGGTCCTCCCT	TTGGATTGAAGGGTGCTCCCT
	AGAGCGTCCTTCAGTCTAGTC	TTGGGCTGAAGGGTGCTCCC	TTGGACTGAAGGGCGCTCCCT	TTGGACTGAAGTGTGCTCCCT
	AGACCGTCCTTCAGTCCACTC	TTGGACTGAAGGGTGCTCGCT	TGGACTGAAGGGTGCTCCCC	TTGGACTGAAGGGTGCTCCCG
	AGAGCGTCCTTCAGTCCAC	TTGGACTGAAGGGTACTCCCT	TTGGACTGAAGGGTGCTCC	TTTGACTGAAGGGTGCTCCCT
	AGCTGCCGACTCATTCATCCA	TTGGACTGTAGGGTGCTCCCT	TTGGACTGAAGGGTACTCCC	TTCGACTGAAGGGTGCTCCCT
	GAGCGTCCTTCAGTCCACT	TTGAACTGAAGGGTGCTCCCT	TCGGACTGAAGGGTGCTCCCT	TTGGACTGAAGGGCGCTCCC
	AGCTGCCGATTCATTCATTCA	TTGGACTGAGGGGTGCTCCCT	TTGGACTGAAGGGAACTCCCT	TTGGACTTAAGGGTGCTCCCT

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR159/319	TGGACTGAAGGGTGCTCCC	TTGGATTGAAGGGAGCTCTG	TTTGGATTGAAGGGAGCTC	TTTGGATTGAAGGGAGTTTTT
	TTGGACCGAAGGGTGCTCCCT	TTTGGATTGAAGGGAGTT	GAGCTCCTATCTTTCCAATGA	AGCTCCCTTCGATCCAAACC
	TTGGACTGATGGGTGCTCCCT	TTTGGATTAAAGGGAGCTC	TTTGAATTGAAGGGAGCTCTG	TTTGGATTGAAGGGAGTTTCT
	TTGGACTGAAGGGTGCTCCTT	TTGGATTGAAGGGAGCTCT	GCTCCTATCTTTCCAATG	AGCTCCCTTTGATCCAAACC
	TTGGACTGAAGGGTGCTACCT	TTTGGATTGAACGGAGCTCTG	TTTGGATTGAAGGGAGCTCCG	TTTGGATTGAAGGGATCTTCT
	TTGGACTGAAGGGCGCTCACT	TTTGGATTGAAGGGGGCTCTG	TTTGTATTGAAGGGAGCTCTG	TTTGGATTGAAGGGATATCTT
	TGGACTGAAGGGTGCTCCCT	TTTGGATTGAAGGGAGCTATT	GAGCTCCTATCTTTCCAATGG	ATTGGATTGAAGGGACCTG
	TTGGACTGAAGGGAGCCCCCT	CTTGGATTGAAGGGAGCTCTG	TTTGGATTGAAGGGAGCTTTC	ATTGGACTGAAGGGAGCTCCC
	TTGGACTGAAGGGTGCCCCC	TTTGGATTGAAGGGAGCTCTC	TTTGGATTGAAGGGAGCTTTG	TTTGGACTGAAGGGAGCTCCT
	TTGTACTGAAGGGTGCTCCCT	TTTGGATTGAAGGGAGCG	GAGCTCCTATCTTTCCAAAG	
	TTGGACTGAAGGGTGCTCCCC	TTTGGATTGAAGGGAGTTTT	TTTGGATTGAAGGGAGCT	
	TTGGACTGAGGGGCGCTCCCT	TTTGGATTGAAGGGAGCTTT	TTTGGATTGAAGGGAGCTCT	
	TTGGACTAAAGGGTGCTCCCT	TTTGGATTGAAGGCAGCTCTG	TTTGGATTGAAGGGGGCTCT	
	TTGACTGAAGGGTGCTCCCT	TTTGGATTGAAGGGAGCTCGG	GAGCTCCTATCTTTCCAATG	
	TTGGACTGAAGGGTGCTCCCA	TTTGGATTGAAGGGAGCTGTG	TTTGGGTTGAAGGGAGCTCTG	
	TTGGACTGAAGGTTGCTCCCT	CCTGGATTGAAGGGAGCTCT	CTTGGATTGAAGGGAGCTCT	
	TTGGACTGAAGGGGGCTCCC	TGTGGATTGAAGGGAGCTCTG	TTTGGATTGAAGGGAGATCTG	
	TTGGACTGAAGGGTGCTCCT	TTTGGATTGAAGGGAGCTTTT	GAGCTCCTATCTTTCCAAT	
	TTGGACTGAAGGGTGTTCCCT	TTTGGATTGAAGGGAGCTCTG	AGCTCCTATCTTTCCAATG	
	TTGGACTGAAGGGTTCTCCCT	TTGGATTGAAGGGAGCTC	TTTGGATTGAAGGGAACTCTG	
	TTGGACTGAAGGGTGCTCACT	TTTGGATTGAACGGAGCT	TTTGCATGACCGAGGAGCTGA	
	TTGGACTGAAGGGTGCGCCCT	TTTGGAATGAAGGGAGCTC	AGCTGCTCGTTCATGGTTCC	
	TTGGACTGAAGGGTGCCCCTT	TTGGACTGAAGGGTGCTCT	AGCTGCTCGTTCATGGTTCCC	
	CTTGGACTGAAGGGTGCTCCA	TTGGATTGAAGGGAGCTCTC	TTTGCATGACCGAGGAGCTGT	
	CTTGGACTGAAGGGTGCCCCC	TTTGGCTTGAAGGGAGCTCTG	AGCTGCTTGTTCATGGTTCC	
	TTGGACTGAAGGGTGCCCCT	TTTGGATCGAAGGGAGCTCTG	AGCTGCTTGTTCATGGTT	
	AGAGCTCTCTTCAGTCCACT	AGCTCCTATCTTTCCAATGA	TTTGCATGACCGAGGAGCTGG	
	TTGGACTGAAGGGTGCTCTT	TTTGGATTGAAGGGAGCTCG	TTTGGATTGAAGGGCTCTCTG	
	CTTGGACTGAAGGTTGCTCCC	TTGGACTGAAGGGAGCTCC	TTTGGATTGAAGGGAGTTTTC	
	TTGGACTGAAGGGTGCTCTTT	TTTGGATTGAAGGGAGCTCTA	AGCTGCTCGTTCATGGTT	
	AGAGCTCTCTTCAGTCCACTC	GAGCTCCTATCTTTCCAATGT	TTGCATGACCGAGGAGCTGA	
	TTGGACTGAAGGGTGCTTCT	TTTGGATAGAAGGGAGCTCTG	CTTGCATGACCGAGGAGCCG	
	CTTGGACTGAAGGGTGCTCCC	TTTGGATTGAAGGGAGTTCTG	AGCTGCTCGTTCATGGTTCT	
	TGGACTGAAGGGAGCTCCCTC	TTTGGATTGAAGGGAGCCT	AGCTGCTCGTTCATGGTTC	
	TTGGACTGAAGGGTGCTCCCTC	TTTGGATTGAAGGGAGCATT	TTTAAATTGAAGGGAGCTCTG	
	TTGGACTGAAGGGAGCTCCTTC	TTTGGATTGACGGGAGCTCTG	CTTGCATGACCGAGGAGCC	
	TTGGACTGAAGGGAGCTCCCTT	TTTGGATTGAAGGGAGCTCTT	AGCTGCTTGTTCATGGTTCCC	
	TTGGACTGAAGGGTGCTCCTTT	TTTGGATTGAAGGGAGCTA	TTTGGATTGAAGGGAGTTACT	

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR160	TGCCTGGCTCTCTGTATGCCA			
	TGCCTGGCTCCCCGTATGCCA			
	TGCCTGGCTCCCTGTATGCCT			
	TGCCTGGCTCCCTGAATGCCT			
	TGCCTGGCTCCCTGCATGCCA			
	TGCCTGGCTCCCTGTATGCCA			
	TGCCTGGCTCCCTGTATGCC			
	TGCCTGGCTCCCTGTATTTC			
	TGCCTGGCTCCCTGAATGCCA			
	GCGTGCAAGGAGCCAAGCATG			
	GCGTTGAGGGAGTCATGCAGG			
	TGCCTGGCTCCCTGTATGCTTTT			
miR162	TCGATGAACCCCTGCATCCGG			
	GGAGGCAGCGGTTGATCGATC			
	TTGATAAACCTCTGCATCCGG			
	TCGATAAACCTCTGCATCCAG			
	TCGATAAACCTCTGCATCGGG			
	TCGATAAACCTCTGCATCCGC			
	TCGATAAACCTCTGCATCCGG			
	TCGATAAACCTCTGCATCC			
miR164	TGGAGAAGCAGGGCACGTGCA			
	GGAGAAGCAGGGCACGTG			
	TGGACAAGCAGGGCACGTGAA			
	TGGAGAAGCAGGGCACGTGCT			
	CACGTGTTCTTCTCCTCCATC			
	CATGTGCCCGCCTTCTCCACC			
	CATGTGCGCTCCTTCTCCAGC			
	ATGTGCCCTTCTTCTCCATC			
	CATGTGCCCTTCTTCTCCACC			
	CATGTGCCCTTCTTCTCCATC			
miR165/166	TCGGACCAGGCTTCTCCCC	GAATGACGCCGGGTCTGAAAG	TCGGACCAGGCTTCATTAATT	TCGGACCAGGCTTCATTATTT
	TCGGACCAGGCTTCCCCC	GGGTTGTTGTCTGGTTCAA	TCGGACCAGGCTTCATTTTTA	GGTATGTTGTCTGGCTCGGTG
	TCGGACCAGGCTTATTCCCC	GGAACGTTGTCTGGTTCAAGG	TCGGACCAGGCTTCATTCCTAT	TCGGACCAGGCTTCATTATTA
	TCGGACCAGGCTTCATCTTC	TCGGACCAGGCTTCAATTCTT	TCGGACCAGGCTTCATTTGTT	TCGGACCAGGCTTCATTTATT
	TCGGACCAGGCTTCCCCCC	TCGGACCAGGCTTCAATCTTT	TCGGACCAGGCTTCAATCCCTA	TCGGACCAGGCTTCATTTCTTT
	GGAATGTTGTCTGGTCCAAG	TCGGACCAGGCTTCATTTTCTT	TCGGACCAGGCTTCATTTACTT	TCGGACCAGGCTTCAATCCTTT
	GGGTTGTTGTCTGGTTCAAG	TCGGACCAGGCTTCAATCCCTT	TCGGACCAGGCTTCATTCATTT	TCGGACCAGGCTTCATTCTATT
	GGGTTGTTGTCTGGTTCAAGG	GGTATGTTGTCTGGCCCGG	TCGGACCAGGCTTCATTCCTTT	TCGGACCAGGCTTCATTTTTT

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR165/166	TCGGACCAGGCTTCATTTATA	TCGGACCAGGCTTCATTC	TCGGACCAGCCTTCATTCCCC	TCGTACCAGACTTCATTCCCC
	TCGGACCAGGCTTCATTCTTTT	TCGGACCAGGCTTCAATCCGT	TCGGACCAGGCTTCATTCCGT	TCGAACCAGGCTTCATTTCCC
	TCGAACCAGGCTTCAATCCC	TCGGATCAGGCTTCATTCCCC	TCGGACCAGGCTACATTCCC	TCGGACCAGGCTTCATTCCC
	TCGGACCAAGCTTCATTCCCA	TCGGACCAGGCTTCACTCCCC	TCGGGCCAGGCTTCAATCCC	GGAATGTTGTCTGGTTCAAGG
	TAGGACCAGGTTTCATTCCCC	TCGGACCAGGCTTCACGCC	GGAATGTCGTCTGGCTCGG	GGAATGTTGTCTGGTTCA
	TCGGACCAGGCTTCATTCT	TCGGACCAGGCTTCATTCCTT	TCGTACCAGGCCTCATTCCCC	TCGGATCAGGCATCATTCCCC
	TCGAACCATGCTTCATTCCCC	TCGGACCAGGCTGCATTCCCC	TCTGACCAGGCTTCATTCCCC	TCGGACCAGGCTCCAATCCC
	TCGGACCAAGCTTCATTCC	GGAATGTTGTCTGGTTGGAGA	TCGGACCAGGCTTCACTCGCC	TCGCACCAGGCTTCATTCCTC
	TCGGACCAGTCTTCAATCCCT	TCGGACCAGGCATCATTCCTC	TCGGACCAGGTTTCGTTCCCC	TCGGACCTGGCTTCATTCCCC
	TCGGACCAGGCTTCTTTCCC	TCGGACCCGGCTTTATTCCCC	TCGGAGTAGGCTTCATTCCCC	TCGGACCAGGCTTCACCCCCC
	TCCGACCAGTCTTCATTCCCC	TCGGACCAGGCTTCATTCAA	TCGGACCAGGTTTGATTCCCC	GGGGACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCGTTCCCT	TCGGATTAGGCTTCATTCCCC	TCGGCCCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTTCT
	TCGGACCTGGCTTCAATCCC	TCGGACCAGGCTTCATTATCC	TCGGACCAGGCTTCATTCTT	TCGGACCAGGCTTCACCCCC
	TTGGACCAGGCTTTATTCCCC	TCGGACCAGGCTTCAACCCCC	TCGGACCAGGCTTCAATCCCT	GGACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCACTCCCT	TCGAACCGGGCTTCATTCCCC	TCGGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCTTTTC
	TCGGACCAGGATTCATTCCCG	TCGGATCAGGCTTTATTCCCC	GCGAACCAGGCTTCATTCCCC	GGAATGTTGTCTGGCTCGGC
	TCGGACCTGGCTTTATTCCCC	TCGGACCAGGCTTCATTTT	TCGGGCCAGGCTTCATTTCCC	TCGTACCAGGCTTCAATCCC
	TCGGACCAGGCTTCATCCCCC	TCGGACCAGGCTTGAATCCCT	AATGTTGTCTGGTTGGAGA	TCGAACCAGGCTTCATTCCCC
	TCGGACCAGGCTTTAATCCC	CGGACCAGGCTTTATTCCCC	GAATGTTGTCTGGTTGGAGA	TCGGACCAGGCTTCATATC
	TCGGAACCGGCTTCATTCCCC	TCGGACCAGGCTTCATTCTC	TCTGACCAGTCTTCATTCCCC	TTGGACCAGGCTTCAATCCCT
	TCGGATCAGGCTTCATTCCC	TTGGACCAGGCTTCAATCCC	TCGAACCAGGCTTCATTCCC	TCGGACCAGGCTTTATCCCCC
	TCGGACCAGGCCTCATTCCC	TCGGACCGGGCTTCATTCC	CCGGACCACGCTTCATTCCCC	TCGTACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCATCCCT	TCGGACCAGGCTTCGTTCTCC	TTGGACCAGGCTTCATTCGCC	TCCGACCAGGCTTCATTCCCC
	TCAGACCAGGCTTCATCCCCC	TCGGACCAGTCTTCATTCCCT	TCGGACCAGGCTTGATTCCCC	TCGGACCAGGCTTCATGCCCC
	ACGGATCAGGCTTCATTCCCC	TCGACCAGGCTTCATTCCCC	TTGGACCAGGCTTCATTCCC	TCTGACCAGGCTTCATTCCGC
	TCGGACCAGGCTTCATTTA	TCGGACCGGTCTTCATTCCCC	TCGGACCAGGCTTCATTCACC	TCATACCAGGCTTCATTCCCC
	CGGACCAGGCTTCATTCCT	TCGGACCAGGCTTCAACCCC	TCGGACCAGGCTTAATTCCCC	TCGGACCAGGCTTCATTTTCC
	TCTCACCAGGCTTCATTCCCC	TCGGATAAGGCTTCATTCCCC	CGGACCAGGCTTCTTCCCCT	GCGGACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCAATTCCT	TCGGACCAGGCCTTATTCCCC	CCGGACCAGCCTTCATTCCCC	TCTGACCAAGCTTCATTCCCC
	ACGGACTAGGCTTCATTCCCC	TCACCAGGCTTCATTCCCC	TCGAACCAGGCTTCATTCACC	TCGGCCCAGGCTTAATTCCCC
	ACCGACCAGGCTTCATTCCCC	TCGGACCAGGCTTTATTCACC	TCGGACCAGGCTTCATTCGC	CGGACCAGGCTTCATTCCCT
	TCGGACCGGGCTTCATTCCCC	TCGGACCAGGCTTCTTTCGCC	GCGGACCAGGCTTCAATCCC	TCGGAACAGGCTTCATTCCCC
	TCGGACCAGGCTTCATTCC	TCGGACCATGCTCCATTCCCC	TGAACCAGGCTTCATTCCCC	TCGGACCAGGCTTTAATCCCT
	TCGGACCAGGCTTCAATCAC	TCGGACCAGGCTTCATTTACC	TCGGACAAGGCTTCATTCCCC	TCGGACCAGGCTTCCTCCCC
	TCGGACCAGGGTTCATTCCCC	TCGGACCATGCTTCATTCTCC	TCGGACCATCCTTCATTCCCC	TCGGACCAGGATTCAATCCCT
	TCAGACCAGGCTTCAATCCCT	TCGGACCAGTTTTCATTCCCC	TCAGACCAGGCTTCATTCCCT	TCGGATGAGGCTTCATTCCCC
	TCGGATCAGGCTTCAATCCCT	ACGAACCAGGCTTCATTCCCC	GGTATGTTGGCTGGCTCGA	TCGGACCAGGCCTCCTTCCCC
	TTTGACCAGGCTTCATTCCCC	CGGACCGGGCTTCATTCCCC	TCCGCCCAGGCTTCATTCCCC	TCGCAACAGGCTTCATTCCCC

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR165/166	TCGGACCTGGCTTCAATCCCT	TCGGACCAGGCTTCATTGTCC	TCGGACTAGGCTTCAATCCCT	TCGGGCCAGGGTTCATTCCCC
	TCGGACCAGTCTTCAATCCC	TCGGACCAGGCTTCATTCGCT	TCGAACCAGGCTTTATTCCCC	TCGCACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCATTTCCC	TCGGAACAGGCTTCAATCCCT	TCGGACCAGGCTTCAGTCCC	TCTAACCAGGCTTCATTCCCC
	TCGCACCAGGCTTTATTCCCC	TCGGCCCAGGCTTCCTTCCCC	TCGGACCAGGCTTCTATCCCT	TCGGACCAGGCTTCATTCA
	TCGGTCCAGGCTTCATTCCCC	TCGGACCATGCTTCATTCCCC	TGGACCAGGCTTCGTTCCCC	TCGGACCACGCATCATTCCCC
	TCGGAAAAGGCTTCATTCCCC	TCGGACCAGGCTTCATCACCC	TCGGACCAGGCTTTATTCCTC	TCGGACCAGGCTTCATCCCCT
	TCGCACCAGGATTCATTCCCC	CCGGACCAGGCTTCAATCCCT	TCAGACCAGGCTTCGTTCCCC	TCGGACCAGGCTTCATTT
	TCGGACCAGGCTTCATTACT	TCGGACCAGTCTTCATTCCC	TCGGACCAGGCTTCATCCTCC	GCGGACCAAGCTTCATTCCCC
	TCGGCCCAGGCTTTATTCCCC	ACGGACCAGGCTTCATTCCCC	GGAATGTTGTCTGGTTCAAG	AGGGACCAGGCTTCATTCCCC
	TCGGACCAGGCTTTATTCCCT	TCGGACCAGGCTAAATTCCCC	TCGGACAGGGCTTCATTCCCC	TCGGACCAGGCTTCATTCCT
	TCGCATCAGGCTTCATTCCCC	GGCCCAGGCTTCATTCCCC	TCGCACGAGGCTTCATTCCCC	TCGTACCAGGCTTCATTCCCT
	TGGGACCATGCTTCATTCCCC	CCGGACCAGTCTTCATTCCCC	TCAGACCAGTCTTCATTCCCC	CGGACCAGGCTTCATTCCCC
	CCGGACCAGGCTTCATTCCCT	TCGGACCAGGCTTCATTCATC	TCGGACCAGGCTTCCTTCTCC	TCGGACCAGGCTTCAATTC
	TCGGACCAGGCTTCCTTCCCC	TCGGACCAGGCTTCATACT	TCGGACCAGGCTTCAATACC	GGACCAGGCTTCATTCCCT
	TCAGACCAGCCTTCATTCCCC	TGGGGCCAGGCTTCATTCCCC	TCGGACCAGGCTTAATCCC	TCGGACCAGGCTTCAATCA
	TCGGACCAGCCTTCAATCCCT	TCGGACCAGGCTTCATTCTGC	TCGGACCACGCTTCATTCCCC	TCGGACCAGGCTTCGATCCC
	TCTGACCAGGCTTCATTTCCC	TCGGACCAGGCTTCAGTCCCC	TCGGGCCAGGCTTCATTCCCC	TCGGACCAGACTTCATTTCCC
	TCGGACCAGGCTTCAATCCT	TCGGACTAGGCTTCATTCCCC	CGAACCAGGCTTCATTCCCC	TTGGACCAGGCTTCATTCCCC
	TCGGACCAGGCTTCATTACCC	CCGGACCAGGCTTCAATCCC	TCGGACCAGGCTTCATTCTCT	TCGGACGAGGCTTCATTCCCC
	TCGGACCGGGCTTCAATCCCT	TCGACCCAGGCTTCATTCCCC	TCGGACCAGGCTACATTCCCT	TCAGACCAGGCTTAATTCCCC
	TCGGACCAGGCTTCATTTTC	TCGGACCATGCATCATTCCCC	TCGGACCATGCTTCATTCCCT	TCGGACCAGGCTTCAAACCCC
	TCGGACCAGGCTTTATTCCCC	TCGGACCAGCCTTCATACCCC	TCGAACCAGGCTTCAATCCCT	AAGGACCAGGCTTCATTCCCC
	CGGACGAGGCTTCATTCCCC	TGGGACCAGGCTTCATTCCCC	TCGGTCCAGGCTTCAATCCCT	TCGGACCAGGCTTCTTCCCCC
	TCGGACCAGGGTTTATTCCCC	TTGGACCAGGCTACATTCCCC	TCGGACCAGGCTTCAATCC	TCGCACCAGTCTTCATTCCCC
	TCGGACCAGGCTTCGTTCCCC	TCGGACCAGGCTTAGTTCCCC	TCGTACCAGGCTTCAATCCCT	TCGGACCAGGCTTCTTTCCCT
	TCAGACCAGGCTTCATTCCC	TTGGACCAGGCTTCATTCTCC	TCGGACCAAGCTTCAATCCC	TCGGCACAGGCTTCATTCCCC
	TAGGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCAGTCCCT	TCGGACCAGGCTTCATTCCCA	TCGGAACAGGCTTCAATCCC
	TCAAACCAGGCTTCATTCCCC	TCGGACCGGCCTTCATTCCCC	TCGGACCAGGCTTCAATCGC	TCGGACCAGGCTTCATTTTCT
	TCGGACCACCCTTCATTCCCC	TCGGACCAGCCTTGATTCCCC	TCGGACCAGTCTTTATTCCCC	TCGGACCAGGCTTCATTCCCG
	TCGGACCAGGGTTCATTCCCT	TCGGACCAGTCTTCATTCCCC	CAGGACCAGGCTTCATTCCCC	TTGGACCAGCCTTCATTCCCC
	TCGAACCAGGCTTCATTCCAC	TCGGACCAGGCTTCATTATCT	TCGGACCAGGCTTCATTCCAA	TCGGACCAGGCTTCATTTCTC
	TCGGACCAGGCCTCAATCCCT	CGGACCAGGCTTCAATCCC	TCGGACTAGGCTTCATTCGCC	TCGGACCAGGCTTTATTCCC
	TCGGACCAGGCTTCATCT	TCGGACCAGGCTTCATTCGAC	TCGGACCAGGCTTCAATCACT	TCTGACCAGGCTTCAATCCCT
	TCTGACCAGGCTGCATTCCCC	TCGGACCAGACTTCATTCCCC	TCGGACCAGGCTTCATTTC	GGACCAGGCTTCATTCCCG
	TCGGACCAGGCTTCCATCCC	TCGGACCAGGCTTCATTCAC	TCGGGCCAGTCTTCATTCCCC	CCGCACCAGGCTTCATTCCCC
	TAGGACCAGGCTTCAATCCC	GGAATGTTGTCTGGCTCGGGT	TCGGACCAGGCGTCATTCCCC	TCGGACCAGGCTTCATTTACT
	TCGGACCAGGGTTCATCCCCC	TCTGACCAGGCTTCAATCCC	CGGACCAGGCTTCCTTCCCC	TCGGACCAGACTTCAATCCCT
	TCGGACCAGGCGTCAATCCCT	TCGGACCAGGCTTCATTCTCC	TCGGACCAGGCTTTATTCTCC	TCGGAGCAGGCTTCAATCCCT

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR165/166	TCGGACCAGGCCTCAATCCC	ACGGACCAGGCGTCATTCCCC	TCGGACCAGGTTTCCTTCCCC	TCGGACTAGGCTTTATTCCCC
	TCGGACCAGGCCTCATTCCCT	TCAAACCAGGCTTCATTCCC	TCGGAATAGGCTTCATTCCCC	TCGGACCAGGCTTCAATCCCA
	TCGGACCAGGCTTCAATCTC	TCCGACCAGGCTTTATTCCCC	TCGGACCAGGCTTCAATCCCG	TCGGACCAGGCTTCAATC
	TCGGACCAGGCTTCTTCCCC	TCGGACCAGTCTTCATTACCC	TCGGACCAGGCCTCATTCCCC	TCGGACCAGGCTTCTTTCCCC
	TCGCACCAGGCATCATTCCCC	TCGGACCAGGCTTCATTAT	CCGGACTAGGCTTCATTCCCC	TCGGCCCAGGCTTCATGCCCC
	GGAATGTTGTCTGGTTCAA	TCGGACCAGGCTTCATTGCCT	TCGGACCAGCCTTCAGTCCCC	TCGGACCAGGCTACATTCCCC
	TCGGAACAGGCTTCGTTCCCC	TCGGACCAGGCTTCATTCCTC	CCGGACCAGGCTTCATTCCCC	TTGGACCAGGCTTCATTCACC
	TCGGACCAGGCTTCATTCCA	TCGGACCAGGCTTCATACCCC	TCGGACCAGGCTTCATTTCCT	TCGGACCCGGCTTCATTCCCC
	TCGGACCATACTTCATTCCCC	TCGCACCATGCTTCATTCCCC	CCGGACCAGGCTTCATCCCCC	TCGGACTAGGCTTCAATCCC
	TCGGACTAGGGTTCATTCCCC	TCGGATCAGGCTTCAATCCC	TCGGACCAAGCTTCATCCCCC	TCGGACCAAGCTTCATTACCC
	TCGGAGCAGGCTTCATTCCCC	TCCGACCAGGCCTCATTCCCC	TCTGACGAGGCTTCATTCCCC	TCGGCTCAGGCTTCATTCCCC
	TCGGACCAGGCTTCGATCCCT	TCGGACCAGGTTTCAATCCCT	TCGGACCAGTCCTCATTCCCC	TCGGACCAGGCTTCAATCTCT
	TCGGACCAGGCTTCAACCCCT	TCGAACCAGGCTTCATTCGCC	TCGGACCAGGCTTCATCCCC	TCGGATCAGACTTCATTCCCC
	TCGGACCAGGCTTCAATCGCT	CGGATCAGGCTTCATTCCCC	TCGGACCAGACCTCATTCCCC	CGGACCAGGCTTCATTCCC
	TCGGACCAGGCTTCAAGCCC	TCGCACCAGCCTTCATTCCCC	TCGGACCAGGCTTCATTATC	CTCGGACCAGCCTTCATTCCC
	ACGGACCAGGCTTGATTCCCC	TCGGACCAGGCCACATTCCCC	TCGGACCAGGATTTATTCCCC	TTCGGACCAGGCTCCATTCCC
	TCGGGCCAGGCTTCAATCCCT	TCGGACCAGGCTCCATTCCTC	TCGGACCAAGCTTCATTCCCT	TTCGGACCAGCCTTCATTCCC
	TCGGACCAGGCTTCAATT	CGGACCAGGCTTCAATCCCT	TCGGACCAGGTTTCATTCCCC	TTCGGACCAAGCTTCATTCCC
	TCGGACCAGGCTTCATTCGCC	TGGACCAGGCTTCATTCCCC	TCGGACCAGGCCCCATTCCCC	CTCGGACCAGGCTTCATTCCT
	TCTGACCAGGCTTCATCCCCC	TCGGACCAGGCTTCATAT	TCGGACCAGGCTTCAATTCC	CTCGGACCAGGCTTCATTCTT
	TCGGACCAGCCGTCATTCCCC	ACGGACCAGGCTTCATTCCC	TCGGACCAGGCTTCATTCCCT	CTCGGACCAGGCTTTAATCCC
	TCGGACCAGGCTTTATTCGCC	TCGGACCATTCTTCATTCCCC	TCGGACCAGGCTTCATTCCAT	TTCGGACCAGGCTTCATTTCC
	TCGGACCAGGCTTCATTCCGC	TCGCACCAGGCTTCATCCCCC	TCGGACCAGGATTCATTCACC	GGAATGTTGTCTGGCTCGGGG
	TCGGACCCGGCTCCATTCCCC	TCGGACCAAGCTTCATTCCCC	TCACACCAGGCTTCATTCCCC	GGAATGTTGTCTGGCTCGAT
	TCGAACCAGGCTTCATTCCCT	TCGGACCAGGCTTCAATACCT	TCAGACCAGGCTTCAATCCC	TTCCGACCAGGCTTCATTCCC
	TCGGACCAGGCTTCACGC	TCGGACCAGGCATCATTCCCC	GCGGACCAGACTTCATTCCCC	CTCGGACCAGGCTTCATTTCC
	TCGGCCCAGGCCTCATTCCCC	TCGGACCAGGCTTCACTCTCC	TCGGACCAGGCTTCAATCT	AATGCTGTCTGGCTCGAGG
	TCGGACCAGACTTCATTCGCC	TCGGACCAGCCTTCATTCACC	TCTGACCAGGCTTCATTCCCT	TTCGGACCAGGCTACATTCCC
	TCGGACCAGGCTTCATTCCAC	TCGGACCAGGCTTCTTTCTCC	TCGGACCAGGCTTCATTGCCC	TTCGGACCAGGCTTCATTCCA
	TCAGACCAGGCTTCATTCCCC	TCGGACCAGGATTCATTCCCC	TCGGACCAGGCTTCAATCCAT	TTCGGACCAGGCTTAATTCCC
	TTGGACCAGGCTTCATTCCCT	TCGGACCAGACTTCATCCCCC	TCGGACCAGGCTTCATTCTTC	TTCGGACCCGGCTTCATTCCC
	TCGGTCCAGGCTCCATTCCCC	TCGGACAAGGCTTCAATCCCT	TCGGACCAGGCTTGATTCCCT	GGAATGTTGTCTGGCTCGGG
	TCGGCCCAGGTTTCATTCCCC	TCGGACCACGCTTCCTTCCCC	TCGGACCAGGTTTCATTCTCC	TTCGGACCAGGATTCATTCCC
	TCGGCCCAGGCTTCAATCCCT	TGGGACCAGGCTTTATTCCCC	TCGGATCAGGCTTCATTCCAC	TTCGGACCAGGCTTCATTCTC
	TCGGACCAGGCTTCATTCACT	TCGGACTAGGCTTCATTCCCT	TGGACCAGGCTTCAATCCCT	TTCGGACAAGGCTTCATTCCC
	TCGGACCATGCTTCAATCCCT	CGGACCAGGCTTCATTCC	TCGGACCAGGCTTCATTCGTC	TTCGAACCAGGCTTCATTCCC
	TCGGACCAGGCTCCATTCCCT	TCGGACCAGGCTTCAATCCCC	TCGGACCAGGCTCCATTCCCC	TTCGGACCAGGCTTCTTTCCC
	TCGGACCAGGGTTCAATCCCT	TCGCACCAAGCTTCATTCCCC	TCGGACCAGGCTTCAATCCC	TTCAGACCAGGCTTCATTCCC

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR165/166	GGAATGTTGTCTGGCTCG	TTCGGATCAGGCTTCATTCCC	TCGGACCAGGCTTCATTCCTAC	TTTCGGACCAGGCTTCATTC
	CTCGGACCAAGCTTCATTCCC	TTCGGACCATGCTTCATTCCC	TCGGACCAGGCTTCATTCCCTC	TTCGGACCAGGCTTCATTT
	TTCGGACTAGGCTTCATTCCC	AATGTTGTCTGGCTCGGGG	TCGGCCAGGCTTCATTCCCC	CCTCGGACCAGGCTTCATTCC
	TTCGTACCAGGCTTCATTCCC	AGCCGGGCTTCATTCCCC	TCGGACCAGGCTTCATTTTA	TTCGGACCAGGCTTCATTTC
	TTCGGACCAGGCTTTATTCCC	TTAGGACCAGGCTTCATTCCC	CGGACCAGGCTTCCTTCCCCC	TCTCGGACCAGGCTTCATACC
	TTCGGGCCAGGCTTCATTCCC	GGAATGTTGTCTGGCTCGG	TCGGACCAGGCTTCATTCCCCC	CCTCGGACCAGGCTTCAATCC
	TTCGGACCAGGCTTCATCCCC	TTCGGACCAGGCTTCATTGCC	TCGGACCAGGCTTCAATCCCTC	TCGGACCAGGCTTCATATCT
	TCCGGACCAGGCTTCATTCCC	GGTATGTTGTCTGGCTCGA	TCGGACCAGGCTTTATTCCCTC	TTGGACCAGGCTTCAATCCCG
	TTCGGACCAGGCTTCAATCCC	CTCGGACCAGGCTTCATTACC	TCAACCAGGCTTCATTCCCC	TCGGACCAGGCTTCAATTT
	CTCGGACCAGGCTTTATTCCC	CTCGGACCAGGCTTCATTGCC	TCGGACCAGGCTTCATTCCCTT	TCGGACCAGGCTTCAATCTTC
	TTCGGACCAGGCTTGATTCCC	TTCGGACCAGGCTTCATTACC	TCGGACCAGGCTTCATTCCCTA	TCGGCCCAGACTTCATTCCTC
	TTCGGACCAGACTTCATTCCC	GAATGTTGTCTGGCTCGGGG	TCGGACCAGGCTTTATTCCCCT	TCGGACCAGGCTTCAATCAT
	GGAATTTTGTCTGGCTCGA	TTCGGACCAGGCCTCATTCCC	TTCGGCCAGGCTTCATTCCC	TCCTGACCAGGCTTCATTCCC
	CTCGGACCAAGCTTCATTGCC	CTCGGACCAGGCTTCGTTCCC	TCGGACCAGGCTTCATTCCCCA	TCGGATTAGGCTTCAATCCCT
	TTCGGACCGGGCTTCATTCCC	TTCGGACCAGGCTTCATTCCC	TCGGACCAGGCTTCATTCCGCC	TCGGACCAGGCTTTAATCCCG
	TTGGGACCAGGCTTCATTCCC	CTCGGACCATGCTTCATTCCC	TCGGACCAGGCTTCATTCCTCT	TCGGACCAGGCTTCAATTTCT
	GGAATGTTGTCTGGCTCGA	CTCGGACCAGGCTTCATTCAC	TCTCGGTCCAGGCTTCATTCT	TCCGACCCAGGCTTCATTCCC
	TCGACCCACGCTTCATTCCCC	TTCGGACCACGCTTCATTCCC	TCTCGGACCAGGCTTCGTTCC	GCGAACCAGGCTTCATTCCTC
	TTCGGACCAGGCTTCGTTCCC	CTCGGACCAGGCTTCATTCCC	TCTCGGACCAGGCTTCAATC	TCAAACCAGGCTTCAATCCC
	CTCGGACCAGGCTTCATCCCC	TTCGGACCAGGCTTCAGTCCC	TTTCGGACCAGGCTTCATTCC	TCGGACCAGGCTTCATTTGT
	ATCGGACCAGGCTTCATTCCC	TTCGGTCCAGGCTTCATTCCC	TTTCGGACCAGGCTTCATGCC	TCCGGACCAGCCTTCATTCCC
	GTCGGACCAGGCTTCATTCCC	TTCGGACCAGGCTTCATACCC	TTTCGGACCAGGCTACATTCC	TCGGACCAGGCTTCAATTA
	GGAATGTTGTCTGGCTCGTGG	CTCGGACCAGGCTTCAATCCC	TCTCGGACCAGGCTTCATTC	TCGGACCAGGCTTCAATCCTG
	TTCGGCCCAGGCTTCATTCCC	TTCTGACCAGGCTTCATTCCC	CTCGGACCAGGCTTCATTCC	TCGGACCAGGCTTCATTTTTC
	GGTATGTTGTCTGGCTCG	TTCGGACCAGGCTTCATTCCT	TTCGGACCAGGCTTCATTCT	TTGCACCAGGCTTCAATCCCT
	GAATGTTGTCTGGCTCGAG	TTCGGACCAGGCTTCACTCCC	TTTCGGACCTGGCTTCATTCC	TCGGACCAGGCTTCCTTCCTT
	GGAATGTTGTCTGGCTCGAGT	TTCGCACCAGGCTTCATTCCC	TTCGGACCAGGCTTCATTC	TCGGACCAGGCTTCAATCTT
	CCCGGACCAGGCTTCATTCCC	TCGGACCAGGCTTCATTCCCCG	TCTCGGACCAGGCTTCAATCC	TCGGACCAGGCTTCATTTCTA
	TTCGGACCAGTCTTCATTCCC	TCGGACCAGGCTTCATATT	TTTCGGACCAGGCTTCAATCC	TCGGACCAGGCTTAATCCCT
	TTCGGACCAGGTTTCATTCCC	TCGGACCAGGCTTCATGTTC	TTTCGGACCAGACTTCATTCC	TCGGACCAGGCTTCTATTC
	GGAATGTTGTCTGGCTCGAGG	TCGGACCAGGCTTCATATTC	TCTCGGACCAGGCTTTATTCC	TCGGACCAGGCTTCATTATT
	AATTTTGTCTGGCTCGAGG	TCGGACCAGGCTTCATTTCCCC	TATCTGACCAGGCTTCATTCC	TCGGACCAGGCTCCAATCCCG
	CTCGGTCCAGGCTTCATTCCC	CGGACCAGGCTTCATTCCCCC	TTTCGGACCAGGCTTTATTCC	TCGGACCAGGCTTCGATCCCG
	TTCGGACCAGGCTTCATTCAC	TCGGACCAGGCTTCATTCGCCC	TTTCAGACCAGGCTTCATTCC	TCGGACCAGGCTTCAACCCCG
	CTTGGACCAGGCTTCATTCCC	TCGGACCAGGCTTCATTCCCCT	TTCGGACCAGGCTTCATTCC	TCGGACCAGGCTTCAATTTC
	TTCGGACCTGGCTTCATTCCC	TCAGACCAGGCTTCATTCCCCT	TTTCGGACCAGGCTTCCTTCC	TCGGACCAGGCTTCATTCTTA
	GGAATGTTGTCTGGCTCGAG	TCGGACCAGGCTTCATTCCTTC	TTTTGGACCAGGCTTCATTCC	TCGGACCAGGCTTCATTTTT
	TTTGGACCAGGCTTCATTCCC	CGGACCAGGCTTCATTCCCCT	TCTCGGACCAGGCTTCATTCC	TCGGACCAGGCTTCAATTCT

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR165/166	TCGGACCAGGCTTTATTCTTC			
	TCGGATTAGGCTTCAATCCC			
	TCGGACCAGGCTTCATTATTC			
	TCGGACCAGGCTTCATTTATC			
	TCGGACCAGGCTTCATTCTTT			
	TCGGACCAGGCTTCATTTAT			
	TCGGACCAGGCTTCATTGTT			
	GGTATGTTGGCTGGCTCGATG			
	GGTATGTTGGCTGGCTCGAT			
	TCGGACCAGGCTCAATCCCT			
miR167	ATCATGACTGACAGCCTCA	TGAATCTGCCAGCATGATCTG	TGAAGCTGCCAGCTTGATCTGA	TGAAGCTGCCAGCGTGATCTGA
	GATCATGACTGACAGCCTCAT	TGAAGCTGCTAGCATGATCTG	AGATCATGCTGGAGTCTCATC	TGAAGCTGTCAGCATGATCTGC
	GATCATGACTGACAGCCTCATT	TGAAGCTGCCAGCATGATCTG	TGAAGCTTCCAGCATGATCTGA	TGAAGCGGCCAGCATGATCTGA
	GATCATGACTGACAGCCTCA	TGAAGCTGCCAGCATGATGTG	TGTAGCTGCCAGCATGATCTGA	TGAAGCTGCCAGCATGATCTAG
	ATCATGACTGACAGCCTCATT	TGAAGCTGTCAGCATGATCTG	GAAGCTGCCAGCATGATCTGA	GATCATGCTGTGCAGTTTCATC
	GATCATGACTGACAGCCT	TGAGGCTGCCAGCATGATCTG	TGGAGCTGCCAGCATGATCTGA	TGAAGCTGCCTGCATGATCTGC
	GATCATGACTGACAGCCTC	TGAAGCTGCCAGCGTGATCTG	TGAAGCTGCCAGCATGATCAGA	TGAAGCTGACAGCATGATCTGA
	ATCATGACTGACAGCCTCAT	TGTAGCTGCCAGCATGATCTG	TGAAGCTGCTAGCATGATCTGA	TGAAGCTGCCAGCATGATCTTA
	ATCGTGCTGTGACAGCTTCACT	TGAAGCTGCCAGCATGATCTA	TGATCATGCTGGAGTTTCATC	AGATCATGCTGGAGATTCATC
	TGAAGCTGCCAGCATGATTT	TGAAGTTGCCAGCATGATCTG	TGAAGCTGCCAGCATTATCTGA	TGAAGCTGCCAGCATGATCTGT
	TGAAGCTGCCAGCATGATC	TGAAGCTGCCAGCATGATCTC	TGAAGCTGTCAGCATGATCTGA	AAGCTGCCACCATGATCTGA
	TGAAGCTGCCAGCATGATCT	TGAAGCTGCCAACATGATCTG	TGAAGCTGCCAGCATGAACTGA	TGCAGCTGCCAGCATGATCTGA
	GAAGCTGCCAGCATGATCTG	TGAAGCTGCCAGCATGATCGG	AGATCATGCTGGAGGTTCATC	TGAAGCTGCCAGCATGGTCTGA
	GGTCATGCTGCGGCAGCCTCACT	TGAAGCTCCCAGCATGATCTG	AGATCATGCTGGAGTTTCGTC	TGAAGCTGCGAGCATGATCTGA
	AGCTGCCAGCATGATCTGA	TGAAGCTGCAAGCATGATCTG	AGATCATGCTGGAGTTTCAGC	TGAAGCTGCCAGCACGATCTGA
	GAAGCTGCCAGCATGATCT	TGAAGCTGCCAGCATGATCCGA	TGAAGCTGCCAGTATGATCTGA	TGAAGCTGCCAACATGATCTGA
	TGAAGCTGCCGGCATGATCTG	AGATCATGCTGGAGTTTTATC	TGAAGCTGCCAGCATGATCTGC	TGAAGCTGCCAGCATGATATGA

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR167	TAAAGCTGCCAGCATGATCTGA	TTAAGCTGCCAGCATGATCTGA	AGATCATGCTGGAGTTTCATC	
	TGAAGCTGCCATCATGATCTGA	TGACGCTGCCAGCATGATCTGA		
	TGAAGCTGCCAGCAAGATCTGA	AGATCATGTTGCAGCTTCATT		
	TGAAGCTGCCAGCATGATCTGA	ATCATGCTGTGCAGTTTCATC		
	CGAAGCTGCCAGCATGATCTGA	TGAAGCTGCCAGCATAATCTGA		
	TGAAGCTGCCAGCATGCTCTGA	TGAAGCTGCCTGCATGATCTGA		
	TGAAGCTGCCAGCATGATTTGA	TGAAGCTGCAAGCATGATCTGC		
	TGAAGTTGCCAGCATGATCTGA	TGAAGCTGCTAGCATGATCTGC		
	TGAAGCTGCCAGCATGATGTGA	TGAAGCTGCCAGCATGATCTGG		
	TGAAGCTGCCAGCATGATCTAA	AGGTCATGCTGGAGTTTCATC		
	TGAAGCTGCCAGAATGATCTGA	AGATCATGCTGGAGTTTCAT		
	TGAAGCTGCCAGCATGTTCTGA	AGGGCATGCTGGAGTTTCATC		
	TGAAGCTGCCAGCATGATAGG	GTCATGCTGGAGTTTCATC		
	TGAAGCAGCCAGCATGATCTGA	AGATCATGCTGAAGTTTCATC		
	GAAGCTGCCAGCATGATCTGC	TCATGCTGGAGTTTCATC		
	TGAAGCTGCCAGCATGACCTGA	GATCATGCTGGAGTTTCATC		
	TGAATCTGCCAGCATGATCTGA	AGGTCATGCTGGAGTTTTATC		
	TGAAGCTACCAGCATGATCTGA	GGTCATGCTGGAGTTTCATC		
	AGATCATGCTGGAGCTTCATC	ATCATGCTGGAGTTTCATC		
miR169	TAGCCAAGGATGACTTGCCA	CAAGTTGTCCTTGGCTAGC	AGGCGGTCACCTTGGCTAGC	GGCAGTCTCCTCGGCTAGC
	GGCAGGTTGTCCTTGGCTAC	TAGCCAAGGATGACTTGCCT	GGCGGGCCCCTTGGCTAGC	GGCGGTCACCTTGGCTAGA
	GGCAGGTTGTCCTTGGCTACT	GCAGGTTGTCCTTGGCTAC	AGCCAAGGATGACTTGCCGGC	TAGCCAAGGATGACTTCCCT
	CAGCCAAGGATGACTTGCC	GGCGAGTTGTTCTTGGCT	GGCAGTCTCCTCCGCTAGC	GGCAGTCACCTTGGCTAG
	TAGCCAAGGATGACTTGC	TAGCCAAGGATGACTTGCCTG	CAGCCAAGGATGACTTGCCGGC	TAGCCAAGGATGACTTTT
	CAGCCAAGGATGACTTGCCGA	TAGCCAAGGATGACTTGCC	TAGCCAAGGATGATTTGCCTGT	GGCGGTCTCCTTGGCTAG
	TAGCCAAGGATGACTTGCCGG	CAGCCAAGGATGACTTGCCGC	CAGCCAAGGATGACTTGCCAGC	GGCAGTCACCTTGGCTAGA
	GGCAGGTTGTCCTTGGCTA	GGCGAGTTGTTCTTGGCTACA	AGAGAAGACCGTTGTTCGG	GGCGGTCACCTTGGCTAGT
	GGCAGGTTGTCCTTGGCTACA	TAGCCAAGGATGACTCTCT	AGCCAAGGATGATTTGCCTGT	GGCGGTCACCTTGGTTAGC
	AGCCAAGGATGACTTGCCT	GGCAAGTCCGTCCTTGGCTACA	GAGAAGACCGTTGTTCGG	GGCGGTCACCTTGGCTAGC
	GGCAGGTTGTCCTTGGCTTCA	GGCAAGTCCGTCCTTGGCTA	GCAAGTTGTCCTTGGCTA	GGCGATCACCTTGGCTAGC
	AGCCAAGGATGATTTGCCTG	GGCAAGTCCGTCCTTGGCTAC	GGCAGGTTGTCCTTGGCT	GTCGGTCACCTTGGCTAGC
	CAGCCAAGGATGACTTGCCGG	GGCAAGTCCGTCCTTGGCTACT	GGTGGTCACCTTGGCTAGC	TAGCCAAGAATGACTTGCCT
	GCAAGTTGTCCTTGGCTAGC	CAAGTCCGTCCTTGGCTACT	TAGCCAAGGATGACTTTCCT	TAGCCAAGGATGAATTGCCTT
	CAAGTTGTCCTTGGCTAAC	GGCAAGTCCGTCCTTGGCTAT	TAGCCAAGGATGATTTGCCTG	TAGCCAAGGATGACTTGCCTT
	GGCGAGTTGTTCTTGGCTA	GGGCGGTCACCTTGGCTAG	GGCGGTCTCCTTGGCTAGC	TAGCCAAGGATGACTTGCCTA
	GGCAGGTTGTCCTTGGCTACC	AGGCGGTCACCTTGGCTAGT	TAGCCAAGGGTGACTTGCCT	TAGCCAAGGATGACTTGCCTC
	GCAGGTTGTCCTTGGCTA	AGGCGGTCACCTTGGCTAG	GGCGGGCACCTTGGCTAGC	CCAGCCAAGAATGGCTTGCCTA
	TAGCCAAGGATGACTTGCCC	AGCGGTCACCTTGGCTAGT	GGCAGTCACCTTGGCTAGC	GGCGGTCACCTTGGCTAG

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR169	ATCGGCGAGTTGTTCTTGGC	
	TAGCCAGGGATGATTTGCCTCT	
	GGCAAGTCAGCCTGGTTACC	
	GGGCAAGTCATCCTGGTTACC	
	GGGCAAGTCAGCCTGGTTATC	
	GGCAAGTCATCCTGGTTACC	
	GGGCAAGTCAGCCTGGTTACC	
	TAGCCAGGAATGATTTGCCTC	
	TATCCAGGGATGATTTGCCTC	
	TATCCAGGGATGATTTGCCTT	
	TAGCCAGGGAGGATTTGCCTC	
	GGCAGTCTCCTTGGCTATC	
	AGCCAGGGATGATTTGCCTC	
	GGTGGGTCTTCTTGGCTAGC	
	TAGCCAGGGATGATTTGCCTC	
	TAGCCAAGGATGAATTGCCGC	
miR170/171	TGAGCCGAACCAATATCACT	TATTGGCTCGACTCACTCAGA
	TGAGCCGTGCCAATATCAC	TGTTGGCTTGACTCACTCAG
	AGGTATTGGTGCGGTTCAATC	TTTAGCCGTGCCAATATCACC
	CGGTATTGGTGCGGTTCAATC	TATTGGCTCGGCTCACTCAG
	TGATTGAGCCGTGCCAATATC	TGTTGGCTCGACTCACTCAGA
	TGATTGAGCCGTGCCAAA	TTGAGCCGTGCCAATATCTCT
	TGATTGAGCCGTGCCAAAATC	TTGAGCCGTGCCAATATCA
	TGATTGAGCCGTGCTAATATC	TGATTGAGCCGTGCCAATATCT
	TGATTGAGCCGTGCCAATA	TGATTGAGCCGTGCCAATATTT
	TGATTGAGCCGCGTCAATATC	
	GATTGAGCCGTGCCAATATC	
	TGATTGAGCCGTGCCAATC	
	TGATTGAGCCGTGCCAATAT	
	TGATTGAGCCGTGCCAAT	
	TGATTGAGCCGTGCCAATATT	
	TGATTGAGCCGCGCCAATATC	
	TGATTGAGCCGTGCCAATACC	
	TGTTGGCTCGACTCACTCAGT	
	TATTGGCTCGGCTCACTCAGG	
	TGTTGGCTCGACTCACTCAGG	
	TGTTGGCTCGACTCACTCAG	
	TTGAGCCGTGCCAATATCACG	

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR172	GGAATCTTGATGATGCTGCAT		
	AGAATCTTGATGATGCTGTT		
	GGAATCTTGATGATGCCGCAT		
	AGAATCTTGATGATGCTGCAT		
	GTGGCATCATCAAGATTCA		
	GGGAATCTTGATGATGCTGCA		
	GGAATCTTGATGATGCTGCA		
	GGAATCTTGATGATGCTG		
	GAATCTTGATGATGCTGCA		
miR390	AAGCTCAGGAGGGATAGCGCT		
	AAGCTCAAGAGGGATAGCGCC		
	AAGCTCAGGAGGGATAGCGCC		
	AAGCTCAGGAGGGATAGCGC		
	AGCTCAGGAGGGATAGCGCC		
	GCTATCTATCCTGAGCTCCT		
	GCTATCTATCCTGAGCTC		
	GCTATCTATCCTGAGCTCT		
	GCTATCTATCCTGAGCTCC		
	CGCTATCTATCCTGAGCTCCA		
miR393	TCCAAAGGGATCGCATTGATCC	TCAGTGCAATCCCTCTGGAAT	TTCCAAAGGGATTGCACTGAT
	TCCAAAGGGATCGCATTGAC	TCAGTGCAATCCCTTTGGAAT	TTCCAAAGGGATCGCATTTCT
	TTCCAAGGGATCGCATTGAT	TTCTAAAGGGATCGCATTGAT	TTCCAAAGGGATCGCATTGAC
	TGCAAAGGGATCGCATTGATC	TTCCAAAGGGATCGCATTGTT	TTCCAAAGGGATCGCATTGAT
	TCCAAAGGGATCGCTTTGATC	TTCCAAAGGGATCGCATTTTT	TTCCAAAGTGATCGCATTGAT
	TCCAAAGGGATCGCATTGATT	TTCCAAAGGGAACTCATTGAT	TTCCAAAGGGATCGCATTGATC
	TCCAAAGGGATTGCATTGACC	TTCCAAAGGGATCGCATTGTTT	TTCCAAAGGGATCGCATTGATT
	TCCAAAGGGACCGCATTGATC	TTCCAAAGGGATCGCATTGACC	
	TCCAAAGGGATCGCATTG	TTCCAAAGGGATCGCATTC	
	TCCAAAGGGATCGCATTGATCT	CTCCAAAGGGATCGCATTGACC	
	TCCAAAGGGATCGCATTGATTT	TTCCAAAGGGATCGCATTGA	
	CAATGCGATCCCTTTGGATG	TTCCAAAGGGATCGCATT	
	TCCAAAGGGATCGCATTGATC	TTCCAAAGGGATCGCATTTT	
	TCCAAAGGGATCGCATTGATTC	TTCCAAAGGGATCGCATTCT	
	CATGCGTTCCTTTTGGAGG	TTCCAAAGGGATCGCATTGT	
	TCCAAAGGGATCGCATTGACCA	TTCCAAAGGGATCGCATTGACT	
	TCCAAAGGGATCGCATTGACCT	TTCCAAAGGTATCGCATTGAT	
	TCCAAAGGGATCGCATTGACC	TTCCAAAGGGACCGCATTGAT	
	TTAGTGCAATCCCTCTGGAAT	TTCCAAAGGGATCGCATTT	

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR394	TTGGCATTCTGTCCATCTCCT			
	TTGGCATTCTGTCCATCTCC			
	TTGGCATTCTGTCCATCTC			
	TTGGCATTCTGTCCATCTCT			
	AGGTGGGCATATTGCCAATG			
miR395	GTTCACTACAAGCACTTCGCA	TGAAGTGTTTGGAGGAACTC		
	GTTCCCTACAAGCACTTCACGG	ATTCTCCTCAAATCACTTCAGT		
	TTTCTCTGCAAACACTTC	CTGAAGTGTTTGGGGGAACT		
	CGAAGTGTTTGGGGGAACTC	GAAGTGTTTGGAGGAACT		
	TGAAGTGTTTGGGTGAACT	TGAGGTGTTTGGGGAAACTC		
	TGAAGTGTTTGGGTGAAC	TGAAGTGTTTGGGGGAACTC		
	GAAGTGTTTGGGGGAACT	CGAAGTGTTTGGGGGAACT		
	TGAAGTGTTTGGGAGAACTC	TGAAGTGTTTGGGGGAACG		
	GAAGTGTTTGGGGGAACTC	AAGTGTTTGGAGGAACTCT		
	TGAAGTGTTTGGGGGAACT	TGAAGTGTTTGGGTGAACTCT		
	GTTCTCCTCAAATCACTTCAGT	TGAAGTGTTTGGAGGAACTCT		
	TGAAGTGTTTGGCGGAACTC	GTTCCCTACAAGCACTTCTCGT		
	TGAAGTGTTTGGGTGAACTC	GTTCCCTACAAGCACTTCACG		
	TGAAGTGTTTGGAGGAACT	TTCCACGGCTCTCTTGAACTT		
miR396	TTCCATAGGTTTCTTGAACTT	TTCCACGGTTTTCTTGAACTT	TTCCACAGCTTTCTTGAAC	TTCCCCAGCTTTCTTGAACTT
	TTCCACAGCTTTCTTGAGCTT	TTCCAAAGCTTTCTTGAACTT	TTCCACAGCTTTCTTGAACT	TCCACAGCTTTCTTGAACT
	TTCTACAGCTTTCTTGAACTG	TTCCACAGCTTTATTGAACTT	TTCCACAGGTTTCTTGAACTG	TTCCACAGCTTTCTTGAACGG
	TTCCACAGCTTTCTTGAACTC	TTCCACGGCTTTCTTGAAC	TTCCACAGCATTCTTGAACTG	TTCCACAGCCTTCTTGAACTG
	TTCCACAGCTTTCTTGAAATT	TTCCACGGCTTTCTTAAACTT	TTCCACGGCTTTCTTGAACC	TTCGACAGCTTTCTTGAACTT
	TTCCACAGCTTTTTTGAACTT	TTCCACAGCTTTCTTAAACTG	TTCCATAGCTTTCTTGAACTG	TTCCACAGGTTTCTTGAACTT
	TTCCACGGCTTTCTTGAAATT	TTCCACAACTTTCTTGAACTG	TTCCACAGCTTTCTTGAACCT	TTCCACGGCTTCCTTGAACTT
	TTCCACAGCTTTCTTGAACTG	TTCCACAGCTTCCTTGAACTT	TTCCACAGCTTTCTTGAATTC	TTCCACAGCTTTCTTGAAGTG
	CACAGCTTTCTTGAACTG	TTACACAGCTTTCTTGAACTG	TTCCACAGCCTTCTTGAACTT	TTCCACCGCTTTCTTGAACTT
	TTCCACAGCTTTCATGAACTT	TTCCACAGCTTTCTTGAATTT	TTCCACAGCTTTCTTGAACTA	CCCACAGCTTTCTTGACCAG
	TTCCACAGCTCTCTTGAACTT	TTCCACGGCTTTCTTGAACTT	TTCCACATCTTTCTTGAACTG	ACCTCTCTTGAACTTTCT
	TTCCACAGCTTTCTTGAACTT	TTCCACAGCTCTCTTGAACTG	ATCCACAGCTTTCTTGAAC	TCCACACGCTTTCTTGAACT
	CCACAGCTTTCTTGAACTG	TTCAACAGCTTTCTTGAACTT	TTGCACAGCTTTCTTGAACTG	TCCATGGGCTTTCTTGAACT
	TCCCACAGCTTTCTTGAACTT	TTCCACTGCTTTCTTGAACTT	TCCACAGCTTTCTTGAACTT	TTTTACAGCTTTCTTGAACTG
	TTCCGCAGCTTTCTTGAACTT	TTCCACAGCTTTCTTGATT	TCCACAGCTTTCTTGAACTG	ACAGGCTTTCTTGAACTG
	CTCCACAGCTTTCTTGAACTT	TTCTACAGCTTTCTTGAACTT	TTCCACAGCTTTCTTGAAGTT	TCCACAGGCTTTCTTGAAATC
	TCCACGGCTTTCTTGAACT	TTCCACAGCTTTCTTAAACTT	TTCCCCAGCTTTCTTGAACTG	TTCACAGGCTTTCTTGAACTG
	TTCCACAGCTTTCTTGTACTT	TTCCACGGCTTTCTTGAACT	TTCCACAGCTTTCTTGGACTT	TCCACAGGCTTTCCTGAACT
	CACAGCTTTCTTGAACTT	TTTCACAGCTTTCTTGAACTG	TTCCACAGCTTTCCTGAACTT	TCCCCAGGCTTTCTTGAACTG

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR396	TCCACAGGCTTTCTTTAACTG	TCCACAGGCTTTCTTTAACT	TCCACACGCTTTCTTGAACTG	TCCGCAGGCTTTCTTGAACTG
	GTTCAAGAAAGTCCTTGGAAA	TCCACAGGATTTCTTGAACTG	TCCATAGGCTTTCTTGAACTT	TCCACAGGCTTTCTTGACCGG
	TCCACAGGCTTTCTTGAATT	TCAACAGGCTTTCTTGAACTG	TCCACAGGCTTCTTGAACTG	TCCACAAGCTTTCTTGAACTG
	CCAGGCTTTCTTGAACTG	TCCACGGGCTTTCTTGAACT	TCCACAGGCTTTCTTGAAG	TCCACAGGCTTTCTCGAACTG
	TCCACAGACTTTCTTTAACTG	TTATACAGCTTTCTTGAACTG	TCCATAGGCTTTCATGAACTG	CACAGGCTTTCTTGAACGG
	TCCACAGGCTTTCTTGAAATG	TCCACAGGCTTTCTTGACTT	TCCACAGGCTTTCTTGCACTG	TCCACAGGCTTTCTTGAACCA
	TCCCACAGCTTTATTGAACTG	TCCACAGGCTTTGTTGAACTG	TCCACAGGCTTTCTTGAATTC	CACAGGCTTTCTTGAACCG
	TCCACAGGCTTTCTTGAAAT	TCCCAGGCTTTCTTGAACTG	TCCACAGGCTTTCATGAACTG	TCCACAGGCTCTCTTGAACT
	TCCACAGGTTTTCTTGAACT	TCCACAGGCTTTCTTGAACTC	TCCATAGGCTTTCTCGAACTG	TCAACAGGCTTTCTTGAACTA
	CCACAGGCTTTCTTGAACTT	TCCACAGGCTTACTTGAACTG	TCACAGGCTTTCTTGAACT	TCCACAGGCTTTGTTGAACTT
	TCCACAGGCCTTCTTGAACCG	TCCACAGGCTTCCTTGAACTG	TCCACCGGCTTTCTTGAACTG	GTTCAAGAAAGTCCTTGGA
	TCACAGGCTTTCTTGAACTG	CACAGGCTTTCTTAAACTG	TCCACAAGCTTTCTTGAACT	TCCACAGACTTTCTTGAACT
	TCCACAGGCTTTCTTCAACTG	ACAGGCTTTCTTGAATTG	TCCACAGGCTTTCTTGAACC	GTTCAAGAAAGCCCATGGAAT
	ACAGGCTTTCTTGAACTC	TCAAGAAAGCCCATGGAA	TCCAGGCTTTCTTGAACTG	TCCACAGGGTTTCTTGAACTG
	TTTACAGGCTTTCTTGAACT	TCCACAGGCTATCTTGAACTG	TCCAAAGGCTTTCTTGAACTG	TCCACAGGCTTTCTTGATC
	TCCACAGGCGTTCTTGAACTG	TCCACAGGCTTTCTTGGACTG	TCCACAGCCTTTCTTGGACTG	TCCACAGGCTTTCTAGAACTG
	TCTACAGGCTTTCTTGAACT	TCCACAGCCTTTCTTGAACTG	TCCACAGGCTTTCTTGAACTA	ACCACAGGCTTTCTTGAACTG
	GTTCAAGAAAGCCCATGGAAA	TCCACAGGCTTTCTTGAACTT	TCGACAGGCTTTCTTGAACTG	TCCACAGGCTTTCTTGATCTG
	ACAGGCTTTCTTGAACGG	TCCACAGGCTTTCTTGATCT	TCCACAGACTTTCTTGAACTG	CCACAGGCTTTCTTGAACTG
	TCCACAGGCTTTTTTGAACTG	TCCACAGGCTTTCGTGAACTG	TACACAGGCTTTCTTGAACTG	TCCACAGGCTTTTTTGAGCTG
	TCCACAGGCTGTCTTGAACTG	TCCACAGGCTTTCTTGAACT	GTTCAAGAAAGTCCTTGGAA	TCCACAGGCCTTCTTGAACTG
	TCCACAGGCTCTCTTGAACTG	CACAGGCTTTCTTGAACT	TCCACAGGCTTTCTTGATT	TCCACAGGCTTTCTTGAAGTT
	TCCACAGGCTTTCTTGAACA	TCCACAGTCTTTCTTGAACTG	TCCACGGGCTTTCTTGAACTG	TCTCCAGGCTTTCTTGAACTG
	TCCACAGGCTTTCTTGAAC	TCCTCAGGCTTTCTTGAACTG	TCCACAGGCTTTCTTGAACCG	TCTACAGGCTTTCTTGAACTG
	TCCACAGGCTTTCTTGACCTG	TTCCACAGCTTTCTTGAGGTG	TCCACTGGCTTTCTTGAACTG	TGCACAGGCTTTCTTGAACTG
	CCACAGGCTTTCTTGAACT	CCCACAGGCTTTCTTGAACTG	TCCACGGTCTTTCTTGAACTG	ACAGGCCTTCTTGAACTG
	TCCACAGGCTTTATTGAACTG	TCCACAGGCTTTCTTAAACTG	CCACAGGCTTTCTTGAAC	GTTCAATAAACCTGTGGGAA
	TCCACAGGCTTTCTTGTACTG	TCCACAGGTTTTCTTGAACTG	TCCACAGGCTTTCTTGAACG	GTTCAATAAAGCTGTGGGAA
	TCCACAGGCATTCTTGAAC	TCCACAGGCTTTCTTGAAGTG	TACACAGGCTTTCTTGAACT	TTCAATAAAGCTGTGGGATT
	TCCAGAGGCTTTCTTGAACTG	TCCACAGGCTTTCCTGAACTG	TCCACAGGCTTTCTTGATCTT	GTTCAATAAAGCTGTGGGAAA
	CACATGCTTTCTTGAACTG	TCCACAGGCTTTCTTGCACTT	TCCACAGGCTTTCTTGAACGG	GTTCAATAAAGCTGTGGGT
	TCCACCGACTTTCTTGAACTG	TCCACAGGCTTTCTTGAGCTG	CCCCCAGGCTTTCTTGAACT	GTCAAGAAAGCTGTGGGA
	TCCACAGGCTTTCTTGAACAG	ACAGGCTTTCTTGAACCG	TCCACAGGCATTTTTGAACTG	GTTCAATAAAGCTGTGGGA
	CACAGTCTTTCTTGAACTG	CCACAGGCTTTCTTGAACGG	TCCACAGGCTTTTTTGTACTG	CAAGAAAGCTGTGGAAAT
	TCCACAGGCTTTCTTGAATTG	TCCACAGGCATTCTTGAACTG	CCCACAGGCTTTCTTGAACTT	TTCAATAAAGCTGTGGGAAA
	TCCACATGCTTTCTTGAACTG	TCCACAGGCTTTCTTGTACCG	TCCATAGGCTTTCTTGAACTG	GTTCAATAAAGCTGTGGGAAT
	TCGACAGGCTTTCTTGAACT	TCCACAGGCTTTCTTAAACT	TCCACAGGCTTTCTTGAACTG	GGTCAAGAAAGCTGTGGGAA
	CACAGGCTTTCTTGAACTG	TCCACAGGCTTTCTTGAACCT	TCCAAAGGCTTTCTTTAACTG	GTTCAATAAAGCTGTGGG

Supplementary Table 5 … (continued): Conserved miRNAs in synthetic hexaploid. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20.

miR396	TTCAATAAAGCTGTGGGA	CTCCACAGGCTTTCTTGAACTG	GGTCAAGAAAGCTTTGGGAAG	ACGGTTCAAGAAAGTCCTTG
	TCAAGAAAGCTGTGGGAA	CTCCACAGGCTTTCTTGAACTA	GGTCAAGAAGGCTGTGGGAAG	ACAGTTCAAGAAAGTCCTTGC
	TCAATAAAGCTGTGGGAAA	CTCCACAGGCTTTCTTGAACT	TTCCACAGCTTTCTTGAACTCT	ACGGTTCAAGAAAGTCCTTGGAAA
	GTCAAGAAAGCTGTGGGAAG	TTCCACAGCTTTCTTGAACTTT	GGTCAAGAAAGCTGTGGGAAT	TCTCCACAGGCTTTCTTGAAC
	TTCAATAAAGCTGTGGGAAAT	ATGGTTCAAGAAAGCCCATGGAAA	GGTCAAGAAAGCTGTGGGAAG	GGTTCAAGAAAGTCCTTGGAAA
	TTCAATAAAGCTGTGGGATC	TTCCACAGCTTTCTTGAACTGC	TTCCACAGCTTTCTTGAACTTA	ACAGTTCAAGAAAGTCCTTG
	GGTCAAGAAAGCTGTGGGAAA	ACGGTTCAAGAAAGCCCA	TCTCCACAGGCTTTCTTGAACT	ATGGTTCAAGAAAGTCCT
	CAAGAAAGCTGTGGAAAC	TTCCACAGCTTTCTTGAACTTC	ACAGTTCAAGAAAGTCCTTGT	
	GTCAAGAAAGCTGTGGGAA	TTCCACAGCTTTCTTGAACTGA	ACTTTTCTTGAACTTCTTATGCCT	
	TTCAATAAAGCTGTGGGAAAA	TTCCACAGCTTTCTTGAACTGT	ACAGTTCAAGAAAGTCCTTGG	
miR397	TAATACTGTCAGGTAAAGATGTC			
	TCACCGGCGCTGCACGCACT			
	TCACCGGCGCTGCACGCA			
	TCCTTGAGAGCAGCGTTGATGT			
	TGAGTGCAGCGTTGATGAAC			
	TTGAGTGCAGCGTTGATGAAC			
	TTGAGTGCAGCGTTGATTAAC			
	TTGAGAGCAGCGTTGATGAAC			
miR398	TGTGTTCTCGGGTCGCCCCTG			
	TGTGTTCTCAGGTCGCCCCTG			
	TGTGTTCTCAGGTCGCCCCT			
	TGTGTTCTCAGGTCGCCCC			
	TGTGTTCTCAGGTCGCCCATG			
	GTGTTCTCAGGTCGCCCC			
	TGTTCTCAGGTCGCCCCT			
	GTGTTCTCAGGTCGCCCCTG			
	TGTGTTCTCAGGTCACCCCTG			
	TGTGTTCTCAGGTCGCCCCGG			
	GTGTTCTCAGGTCGCCCCT			
	GTTCTCAGGTCGCCCCTG			
miR399	TGCCAAAGGAGAATTGCCCTG			
	TGCCGAAAGAGAGCTGCCCTG			
	GTGCGGTTCTCCTCTGGCATG			
	TGCCAAAGGAGAGTTGCCCTG			
	TGCCAAAGGAGAGTTGCCC			
	TGCCAAAGGAGAATTGTCCTG			
	TGCCAAAGGAGAATTACCCTG			
	GCGCGGTTCTCCTCTGGCATG			

Supplementary Table 6: Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

Annotated miRNA	Sequence in Synthetic Hexaploid Small RNA Library		
miR158 (A. thaliana, A. lyrata)	TTCCAAATGTAGACAAAACA		
	TTCCAAATGTAGACAGAGCA		
	TTCCAAATGTAGACAAAGCA		
	TCCCAAATGTAGACAAAGCA		
miR401 (A. thaliana)	CGAAACTGTTGTCGACCGACA		
	CGAAACTGGTGTCGACCGACA		
	CGAAACTGGTGTCGACCGCA		
	GAAACTGTTGTCGACCGACA		
	TGAAACTGGTGTCGACCGACA		
	CGAAACTGATGTCGACCGACA		
	AACTGGTGTCGACCGACA		
	CGACACTGGTGTCGACCGACA		
	CGAAATTGGTGTCGACCGACA		
	CGAAACTGGTGTCGACTGACA		
miR403	TTAGATTCACACACAAACTCG		
	TTAGATTCACGCACGAACTCG		
	TTAGATTCACGCACAAACTCG		
miR408	TGCACTGCCTCTTCCCTGGCTT		
	CAGGGATGGAGCAGAGCAAGG		
	TGCACTGCCTCTTCCCTGGCT		
	CTGCACTGCCTCTTCCCTGGC		
	ATGCACTGCCTCTTCCCTGGC		
miR437	TTATTCTAAGTCAAACTT		
	TTGTTCTAAGTCAAAGTT		
	TTGTCCTAAGTCAAACTCG		
miR444	TGTTGTCTCAAGCTTGCTGCC	GCAGTTGTTGTCTCAAGCT	GCAGTTGTTGTCTCAAGC
	TTTTGTCTCAAGCTTGCTGCC	CGCAGTTGTTGTCTCAAGCTT	TGCAGTTGTTGTCCCAAGCTT
	TGTTGTCTTAAGCTTGCTGCC	TGCAGTTGCTGCCTCAAGCT	GCAGTTGTTGTCTCAAGCTT
	TGTTGTCTCAAGCTTGCTGCA	TGCAGTTGTTGTCTCAAGCTA	TGCAGTTGTTGTCTCAAGCTG
	TGTTGTTTCAAGCTTGCTGCC	CAGTTGTTGTCTCAAGCTT	TGCAGTTGTTGTCTCAAGCT
	TGTTGTCTCAAGCTTGCCGCC	TGCAGTTGTTGTCTCAAGCCG	TGCAGTTGCTGCCTCAAGCTT
	TGTTGTCTCAAGCTTACTGCC	TGCAGTTGTTGTCTGAAGCTT	TTCAGTTGTTGTCTCAAGCTT
	TGTTGTCTCAAGCTTCCTGCC	TGCTGTTGTTGTCTCAAGCTT	TGCAGTTGTTGTCTTAAGCTT
	TGCATTTGTTGTCTCAAGCTT	TGCAGTTGTTGTCTCAAGCTT	TGTTGTCTCAAGCTTGCT
	CAGTTGTTGTCTCAAGCT	TGCAGTTGTCGTCTCAAGCTT	TGTTGTCTCAAGCTTGCTGC
	TGTAGTTGTTGTCTCAAGCTT	TGCAGTTGATGCCTCAAGCTT	CTGTGACTTTCTTGCAAGTT
	TGCAGTTGTTGTCTCAAGC	TGCAGTTGTTGTCTCAAGCCT	CTGTGACTTTCTTGCAAGTTG

Supplementary Table 6 … (continued): Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR477	TCTCCCTCAAAGGTCTCT
	ATCTTTCTCGAGGGCTTC
	TCTCCCTCAAAGGTCTCC
miR478 (P. trichocarpa)	CACCTCTTCTATTTTTAGT
miR479	TGAGCCGAACCAATATTACTC
	TGAGCCGAACCAATATCACTC
	TGAGCCAAACCAATATCACTC
miR528	TGGAAGGGGCATGCAGAGGAG
	AAGGGGCATGCAGAGGAG
	CCTGTGCCTGCCTCCTCCATT
	TGGAAGGGGCATGCAGAGG
miR812 (O. sativa)	GGTGAAATGGTAGACACG
miR824	CCTTCTCATCGATGGTCTAGA
miR827	TTAGATGACCATCAGCAAACA
	TTAGATGACCATCAGCAAAC
	TTAGATGACCATCAGCAAACT
	TTTTGTTGGTTGTCATCTAACC
	TTAGATGACCATCAGCAAAAG
miR829 (A. thaliana, A. lyrata)	GCTCTGATACCAATTGTTGGA
miR845	AGCTACATTGTCTGCTGGGT
miR858	TTTCGTTGTCTGTTCGACCT
miR894 (P. patens)	ATTCACGTAGGGTTCACC
	TTTCACGTAGGGTTCACC
	CGTTTCACGTAGGGTTCACC
	TTTCACGTCGGGTTCACC
	CGATTCACGTAGGGTTCACC
	TTTCACGTCGGGTTCACT
	CGTTTCACGTCGGGTTCA
miR904 (P. patens)	TGTCAATGTTTTGGAGCA
miR1098 (S. moellendorffii)	GATGACGTTTGTGTTGGA
miR1117 (T. aestivum)	TAACACCGGTTCGTGGCACGAACC
	TAGTCCCGGTTCGTAGCACGAAC
	TAATACCGGTTCGTGGCTCGAACC
miR1119	CACGGCCTGATGCTTAGT
miR1120	CTTATATTTTGAAACGGA
	ATCTTATATTATGAAACGGA
miR1122	AGATACGTCCGAATCTAGA
	TAGATACGTCCGAATCTAGA

Supplementary Table 6 … (continued): Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR1130	CCTCCGTCTAGAAATGTAAGA		
miR1133 (T. aestivum)	ATACTCCCTCCGTCCTAAT		
	TACTCCCTCCGTCCAAGA		
	TATACTCCCTCCGTCCAGAA		
miR1310	GCATGGGGGGCGCAACGCCT		
	CATGGGGGGCGCAACGCCT		
	AGGCATGGGGGGCGCAACGCCT		
	CATGGGGGGCGCAACGCC		
	ATGGGGGGCGCAACGCCT		
miR1319 (O. sativa)	CGGCATCTGTAGTATTTT		
miR1432	TCAGGAGAGATGACACCAAC	TAGGTGTCATCTCCCCTGAAC	
	ATCAGGAGAGATGACACCGA	ATCAGGAGAGATGACACCAA	
	TCAGGAGAGATGACACCGAC	ATCAGGAGAGATGACACCAAC	
	ATCAGGAGAGATGACACCGAC		
miR1436	ACATTAGTGGACGGAGGGAG		
miR1439 (O. sativa)	TTTTGGAACGGAGAGAGTACT		
	TTTAGAACGGAGTAAGTATT		
	TTTTGAAACGGAGGGAGTATT		
	TTTGAAACGGAGGGAGTATT		
miR1520 (G. max)	ACATCGTGGTTGGCCATG		
miR1863	GGCTCTGATACCATGTTAGTTTAG		
miR1871	TTGGCTCTGATACCATGTTGGTTT		
	TGGCTCTGATACCATGTTGGA		
miR1878	ATTTGTAGTGTTCAGATT	ATTTGTAGTGTTGAGATTAAGTTT	
	ATTTGTAGTGTTCAGATTAAGT	TTTGTAGTGTTCAGATTAA	
	ATTTGTAGTGTTCAGATTGAGTTT	ATTTCTAGTGTTCAGATTAAGTTT	
	ATTTGTAGTGTTCAGATTAAGTTT	ATTTGTAGTGTTCAGATTAAGTT	
miR2111	AATGCACCAATTTTCTGAGA		
miR2118	TTCCGGATGCCTTCCATGCCT		
	TTCCTGATGCCTCTCATTCCTA		
	TTTCTGATGCCTCCCATTCCTA		
	TTCCTGATGTCTCCCATTCCTA		
miR2275	TTTGGTTTCCTCCAATATCTCA	TTTGTTTTCCTCCAATATTTC	AGAATTGGAGGGAAACAAACT
	CTTGTTTTCCTCCAATATCTCA	AGGATTAGAGGGCACTGAACC	AGAGTTGGAGGAAAACAAACT
	TTTGTTTTCCTCCAATATCTT	AGAGTTGGAGGAGAACAAACT	AGAATTGGAGGATATCAAACT
	TTTGTTTTCCTCCAATATCTTA	AGAATTGGAGGAAATCAAACT	
miR2592 (M. truncatula)	AAACATCAAGGTCGGGTG		
miR2678 (M. truncatula)	AATTTGTTGCGAGTGTTT		

Supplementary Table 6 … (continued): Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR3513 (A. hypogaea)	TGATCTGATAGAAATTGT		
miR3630	TGGGAATCTCTCTGATGTT		
	GGGAATCTCTCTGATGTT		
	GGGAATCTCTCTGATGTA		
	ATGGGAATCTCTCTGATGTA		
miR4393 (G. max)	AAAAGACGGCAGAGAAGC		
miR4408 (G. max)	CATTGGATGAGGGTCTGA		
miR4995 (G. max)	GCAGTGGCTTGGTTAAGG		
miR5048 (H. vulgare)	GTTTGCAGGTTTTAGGTCC	TGTTTGCAGGTTTTAGGTCCAA	ATTTGCAGGTTTGAGATCT
	TGTTTGCAGGTTTTAGGTCC	TATTTGCAGGTTTGAGATCTA	TATTTGCAGGTTTGAGATC
	TTTGCAGGTTTTAGGTCC	GTTTGCAGGTTTAAGGTCT	TGTTTGCAGGTTTAAGGTC
	ATTTGCAGGTTTTAGGTCCA	TTTGTAGGTTTAAGGTCTA	TGTTTGCAGGTTTAAGGTCTA
	GTTTGCAGGTTTAAGGTCTA	TATTTGCAGGTTTGAGATCT	TGTTTGCAGGTTTAAGGTCTAA
	TGTTTGCAGGTTTAAGGTCT	ATTTGCAGGTTTGAGATCTA	GTTTGCAGGTTTTAGGTC
	ATTTGCAGGTTTGAGATC	TTTGCAGGTTTGAGATCT	
	TGTTTGCAGGTTTTAGGTC	TGTTTGCAGGTTTTAGGTCT	
	TATTTTCAGGCTTTAGGTC	TTTGCAGGTTTAAGGTCT	
	TGTTTGCAGGTTTTAGGTCCA	TTTGTAGGTTTAAGGTCT	
	TATTTGCAGGTTTGAGATCTAA	ATTTGCAGGTTTGAGATCTAA	
miR5049	CAATTAATATGGGTCGGAGGA	CAAGTAATATGGATCGGAGGG	AATTAATTTGGATCGGAGGAA
	CAAGTAATATGGATCGGAGGGAGT	AAATAATATGGATCGGAGGGA	AAGTAATATGGATCGGAGGA
	CAAGTAATATGAATCGGAGGGAGT	ATTAATATGGACCGGAGGAAG	CAAGTAATATGGATTGGAGGGAGT
	AACTAATATGGATCGGAGGGAG	CAAGTAATATGGGTCGGAGGGA	TCCGATCCATATTACTTG
	CAAGTAATATGTATCGGAGGGA	CAATTAATATGGATCGGAGGGA	TTAATATGGATCGGAGGGAGT
	AATAATATGGATCGGAGGGAGT	AAGTAATATGGATCGGAGAGAGT	CAACTAATTTGGATCGGAGGAAGT
	CAATTAATTTGGATCGGAGG	CCTCCGATCCATATTACTTGT	AAGTAACATGGATCGGAGGAA
	AATTTGGATCGGAGGGAGT	AATATGGACCGGAGGGAGT	CAAGTAATATGGATCGGAGGGAG
	AAGTAATATGAATCGGAGGGAG	CAATTAATATGGATCGGAGG	CAAGTAATATGGATCGGACA
	GAGTAATATGGATCGGAGC	CAACTAATATGGATCGGAGGGAGT	CAACTAATATGGGTCGGAGGAAGT
	ATAATATGGACCGGAGGAA	CAACTAATATGGATCGGAGGAAGT	CAAGTAATATGGATCAGAGGGAGT
	CAATTAATTTGGATCGGAG	CAATTAATATGGATCGGAGGAAAT	CAAGTAATATGGATCGGAGGAAG
	CAATTAATATGGATCGGAGGA	ATTAATATGGATCGGAGGGAG	AATATGGGTCGGAGGGAG
	AAGTAATATGGATCGGAGGGAA	CAATTAATTTGGATCGGAGGA	AAGTAATATGGATCGGAGGGGG
	CAACTAATACGGATCGGAGGAAGT	AAGTAATATGAATCGGAGGGAGT	AATTAATATGGATCGGAGGGAGT
	CAACTAATATGGATCGTAGGAAGT	CAAGTAATATGGATCGGAGGA	ATTAATATGGATCGGAGGGAGT
	AAGTAATATGGATCAGAGGGAGT	CAACTAATATGGATCGGATGAAGT	GTAATATGGATCGGAGAGA
	CAAGTAATTTGGATCGGAGGGAGT	CAAGTAATATGGGTCGGAGGG	CAACTAATTTGGATCGGAGGAAG
	CAACTAATATGGATCGGAAGAAGT	TCCCTCCGATCCATATTACTTGT	AATAATATGGTTCGGAGGA

Supplementary Table 6 … (continued): Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR5049	AGTAATATGGGTCGGAGGGAG	AACTAATATGGATCGGAGGAAGT	AACTAATATGGATCGGAGAGA
	AAGTAATTTGGATCGGAGGGA	CAAGTAATATGGATCGGAGAGAGT	AATTAATTTGGATTGGAGGGA
	AGTAATATGGATCGGAGGGAG	CAATTAATATGGATCGGAGGGAGT	TTAATTTGGATCGGAGAGA
	CAACTAATATGGATCGAAGGAAGT	AATATGGATCGGAGGGAAT	AATTAATTTGGGTCGGAGGGA
	AATTAATATGGGTCGGAGGAA	AAGTAATATGGATCGGAGGGAGT	AATTAATTTGGATCGGAT
	CAAGTAATATGGGTCGGAGGGAGT	CAAGTAATATGGATCGGAGGGA	AATTAATTTGGATCAGAGGGA
	CAATTATTATGGATCGGAGGA	TAATATGGGTCGGAGGGA	AATTAATATGGACTGGAGGGA
	AATTAATATGGGTCGGAGGA	CAACTAATATGGATCGAAGGAA	ATTAATTTGGATCGGAGGGA
	CAAGTAATATGGATCGGAGGGACT	CAAGTAATATGGATCGGAAGGAGT	AATTAATTTGGATCGGAGGGT
	AGTAATATGGATCGGAGGGAGT	GAGTAATATGGATCGGAGGGAGT	AATTAGTTTGGATCGGAGGGA
	ATAATATGGGTCGGAGGAAGT	CCCTCCGACCCATATTAGTTG	ATTAATTTGGATCGGAGAGA
	CAACTAATATGTATCGGAGGAAGT	ATAATATGGATCGGAGGGAGT	ACTAATTTGGATCGGAGGG
	AACTAATATGGATCGGAGGGA	AGACATCTATTTTGGGACGGA	AATTAATATGGATCGAAGGGA
	AACTAATATGGATCGGAG	AGTCACTTATTTTGGGAC	CATTAATATGGATTGGAGGGA
	AGTAATATGGATCAGAGGGAG	ACACTTATTTAGGATCGGAG	ACTAATTTGGATCGGAGGGA
	CAACTAATATAGATCGGAGGAAGT	AACTAATATGGGTCGGAGGGA	AATTAATATGGGTCGGAGGGA
	AATTAATATGGATCGGAGGGA	ATTAATATGGGTCGGAGGGA	AACCAATATGGATCGGAGGGA
	CAACTAATATGGATCGGAGGGA	ATTAATATGGATCAGAGGGA	
	AATAATATGGATCGGAGGGA	AATTAATTTGGATCGGAGGGA	
	AAGTAATATGGGTCGGAGGGA	AAATAATATGGGTCGGAGGGA	
	AAGTAATATGGATCGGAGGAAGT	TTAATTTGGATCGGAGGGA	
	CAAGCAATATGGATCGGAGGGAGT	ACTAATATGGGTCGGAGGGA	
	TCCGATCCATATTACTTGT	ATTAATTTGGATCGGAGT	
	TATTATGGATCGGAGGGAGT	AATTAATTTGGATCGAAGGGA	
	CTAATATGGGTCGGAGGAAGT	ATTAATTTGGATCAGAGGGA	
	AATTAATATGGACCGGAGGAAGT	CATTAATATGGACCGGAGGGA	
	CAACTAATATGGATCGGAGAAAGT	AACTAATATGGATCAGAGGGA	
	CAAGTAATATAGATCGGAGGGAGT	AATTAATATGAGTCGGAGGGA	
	ACTAATATGGATCGGAGGGAGT	AATTAATATGGATCAGAGGGA	
	AAGTAATATGGATCGGAGGGA	AACTAATTTGGATCGGAGGGA	
	CAAGTAATATGGATCGGAGAGA	CATTAATTTGGATCGGAGGGA	
	AATTAATATGGATCGGAGGG	AATTAATATGGATCGGAAGGA	
	GTAATATGGATCGGAGGGA	AATTAATATGGGCCGGAGGGA	
	CTCCGGCCCATATTACTTG	AATTAATATGGGTTGGAGGGA	
	CCCTCCGATCCATATTACTTG	TAATATGGGTCAGAGGGA	
	TAATATGGACCGGAGGGAGT	ATTAATATGGATCGGAGAGA	
	CAAGTAATATGGATCGGAGG	AGTTAATATGGATCGGAGGGA	
	AAGTAATATGAATCGGAGGGA	AACTAATATGGATCGAAGGGA	

Supplementary Table 6 … (continued): Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR5054 (B. distachyon)	TCCCCACGGTCGGCGCCA		
	TCCCCACAGTCGGCGCCA		
	TCCCCACGGACGGCGCCA		
miR5059 (B. distachyon)	TCCTGGGCAGCAACACCA		
	TCCTGGGCAGCAGCACCA		
miR5062 (B. distachyon)	TGAATCTTGGTGAAAAGCC		
	TGAATCTTGGTGAAAAGC		
	TGAATCTTGGGTAAAAGCCGCCT		
miR5067 (B. distachyon)	TCAGCGACAAGTAATATGGAT	TCAGCGACAAGTAATATGGAC	TCAGCGAGAACTAATATGGAT
	CAGCAACAATTAATATGGAT	TGAGCGACAATTAATATGGAT	TCAGCGACAATTAATATGGAT
	TCAACGACAATTAATATGGAT	TCAGAGACAATTAATATGGAT	GCGACAACTAATTTGGAT
miR5071 (O. sativa)	TCAAGCATCATTTCATGGACA	TCAAGCATCATGTCATGGACA	
	TCAAGCACCATATCATGGAC	TCAAGCACCATATCATGGACA	
	TCAAGTATCATATCATGGACA	TCAAGTATCATGTCGTGGACA	
miR5072 (O. sativa)	CGTTCCCCAGCGGAGTCGCCA	ATCCCCAGCGGAGTCGCCA	TCCCCAGCGTAGTCGCCA
	TCCCCAGCGGGGTCGCCA	TCCCCAGTGGAGTCGCCA	TCCCCAGCGGAGTCGCCA
	TCCCCAGCGGAGTCGCCG	GTCCCCAGCGGAGTCGCC	ATCCCCAGCGGAGTCGCC
	TCCCCAGCGTAGTCGCCT	TCTCCCCAGCGGAGTCGCCA	
	TCTCCAGCGGAGTCGCCA	GTTCCCCAGCGGAGTCGCCA	
	TTCCCCAGCGGAGTCCCCA	TTCCCCAGCGGAGTCGCCA	
	TTCCCCAGCGGAGTCGCC	ATCCCCAGCGTAGTCGCC	
miR5073 (O. sativa)	TTTGGTGAATCGGAAACCCT		
	ATTTGGTGAATCGGAAACC		
miR5077 (O. sativa)	TTTCACGTCGGGTTCACCA		
	TTCACGTCGGGTTCACCT		
	ATTCACGTCGGGTTCACCA		
	TTCACGTAGGGTTCACCA		
	TTCACGTCGGGCTCACCA		
	TTCACGTCGGGTTCACCA		
miR5082 (O. sativa)	CGCGATGATGGCCGCGCGGGCTCA		
miR5083 (O. sativa)	ATACTACAATTATCTGATC		
miR5139 (R. glutinosa)	AAACCTGGCTCTGATATA		
miR5161 (O. sativa)	TTTGGATCAGAGGGAGTACA		
	CTGGATCAGATGGAGTCT		
miR5164 (B. distachyon)	TCAACTTTGTCTAGATACG	CTCAACTTTGTCTAGATACGA	
	CGCAGATTTGTCTAGATACGC	TCAACTTTGTCTAGATACGG	
	CTCAACTTTGTCTAGATACG	TCAACTTTGTCTAGATATG	
	CTCAACTTTGTTTAGATACG		

Supplementary Table 6 … (continued): Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR5174 (B. distachyon)	TATGGATCGGAGGGAGTA		
	TTTTGGGACGGAGGGAGTA		
	ACTCCCTCTGATCCATAAT		
miR5178 (B. distachyon)	TCTGACCTGTGGGCCAGAGC		
miR5180 (B. distachyon)	GTGTTGCAGTTTTGAACT	TATGTGTCACAGTTTTGAACT	
	AAGTGTCACAGTTTTGAACT		
	TAAGTGTCACGGTTTTGAAC		
miR5181 (B. distachyon)	ACACTTATTATGGATCGGA	TCCGATCCAAAATAAGTGCC	CTCCGATCCATAATAAGTG
	CAACACTTATTATGGATC	TCCGATCCTAAATAAGTGTC	CTTCCGGTCCATAATAAGTGT
	CGACACTTATTATGGATCGGA	TCTGATCCATAATAAGTGTTG	ACTTTTTATGGATTGGAGGGA
	CACTTATTTAGGATCAGA	TCCGATCCATAATAAGTGACG	CCTCCGATCCATAATAAGTGT
	GACACTTATTATGGATCG	TCCGATCCGTCATAAGTGTCG	
	CACTTGTTATGGATCGGA	TCCGATCCATAATAAGTATTG	
	TCCAATCCATAATAAGTGTCG	CCTCCGATCCATAATAAGTG	
miR5183 (B. distachyon)	ATTTTGACAAATTTGAGACA		
miR5185 (B. distachyon)	TTCTAGTTCATTTGCCAAA	TTTGAGAAATGAACTAGAAG	
	TTCTAGTTCATTTGTCAAAAC	TTGAGAAATGAACTAGAAGT	
	TTGGCAATTGAACTAGAAG	TTGAAAAATGAACTAGAAGT	
miR5203 (B. distachyon)	ACTTATTATGGATCAGAGGGA	CTTAATATGGATCGGAGGGA	ACTTATTATGCATCGGAGGG
	CTTATTATGGATCGGAGGGA	ACTTATTTTGGATCGGAGGGA	ACTTATTATGGATCGGAGGGA
	ACTTATTATGAATCGGAGGGA	ACTTATTTTGGATCGGAGGG	ACTTATTATGCATCGGAGGGA
	ATTAATATGGACCGGAGGGA	ACTTATTATGGATTGGAGGGA	ACTTATTATGGATCGGAGGAA
	ACTTATTATGGATCGGAGAGA	AATTAATATGGACCGGAGGGA	
miR5205 (M. truncatula)	AATTTGGGTCGGAGGGAGT		
	TAATTTGGGTCGGAGGGAGT		
	CTTATATTTTGGGACGGAGGG		
miR5225	TCAGGAGAGATGACACCG		
miR5227 (M. truncatula)	GAAGAGAAGAATATTGAA		
miR5276 (M. truncatula)	GGAGCGGCTTGCTGGGGC		
miR5281 (M. truncatula)	GTAATTTGGATCGGAGGGA		
miR5387 (S. bicolor)	AACACGAACCGGTACTAAAGGAAC	ACGAACCGGTGATAAAGGC	
	ACGAACCGGTGTTAAAGGACC		
	ACACGAACCGGTGTTAAAGG		
miR5389 (S. bicolor)	ACAGGATATGAGAATCTCTT		
miR5528 (O. sativa)	ACGGATTTAGATGGTGCC		
miR5568 (S. bicolor)	TCTATTCCAAATTATAAGATG	AATATAAGATGTTTTAGC	
	TCCATTTCAAAATGTAAGATG	AAATATAAGATGTTTTAGC	
	TCCGTTCCAAATTATAAGATG		

Supplementary Table 6 … (continued): Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR5658 (A. thaliana)	AGGATGATGATGATGATGT ATGATGAGGATGATGATAA GATGATGATGATGATGAT
miR5711 (B. rapa)	TGTTCTGTGGGTTTTTACCGA
miR5762 (G. max)	TAGGAGGTATCAAATGGC
miR5809 (O. sativa)	CGTCGCCGGCGACCACCT
miR5831 (O. sativa)	TGGTCAAACTTTGAATAGTT
miR6030 (B. napus)	TCCACCCATACCATACAGACC
miR6173 (H. brasiliensis)	CCGTAAACGATGGATACT
miR6174 (H. brasiliensis)	CTGCTAAACTGTCAGACTT ATGCGAAACTGTCAGACT
miR6191 (H. vulgare)	TAAATTTGTCTAGATACGA AAATTTGTCTAGATTTGA AAATTTGTCTAGATATGGA
miR6196 (H. vulgare)	AGGACGAGGAGCTGGACA
miR6197 (H. vulgare)	TCTGTTCCTAAATGTAAGCCT TCTGTTCCTAAATATAAGACG
miR6207 (H. vulgare)	ACGACCTGGACGCCGGCG
miR6219 (S. bicolor)	GAATCGGGACTAAAGGTC GAACCGGGACTAAAGGTCGGACCT AGTCCCGATACGTTAGTCCCG GAACCGGGACTAAAGGTGTGACCT ACCGGGACTAAAGGTCGGACC AACCGGGACTAAAGGTCGGACCT AACCGGGACTAAAGGTCGGACC
miR6224 (S. bicolor)	CGTTCTAAAATATAAGGC CTCCGTCCCAAAATATAAGGC
miR6250 (O. sativa)	GGGATAGATCGATGCGTCT GGGATAGATCGATGCGTCA
miR6253 (O. sativa)	GGAAGTGGGCAGTTGGAT
miR6273 (P. persica)	AATGGAGGATGATTTTTTTT
miR6300 (G. max)	GTGGTTGTAGTTTAGTGG GTAGTTGTAGTATAGTGG GTCGTTGTAGTGTAGTGG GTCGTTGTAATATAGTGG GTGGCTGTAGTATAGTGG GTCGTTGTAGTATAGTGA GTCGTTGTAGTATAGGGG GTCGTTGTAGTATAGTGG GTCGTTGCAGTATAGTGG GTCGTTGTAGTATGGTGG GTCGTTGTAGTATAATGG GTCGTTGTAGTATTGTGG

Supplementary Table 6 … (continued): Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR6478 (P. trichocarpa)	CCGACCTTAGCTCAGTTG	CCGACCTCAGCTCAGTTGGT	CCGACCTTAGCTCAGTTGGC
	CCGACCTTAGCTCAGTTGG	CGACCTTAGCTCAGTGGT	CCGACCTTAGCTCAGTGG
	CCGACCTTAGCTCAGTTGGTG	CCGACCTTAGCCCAGTTGG	CCGAGCTTAGCTCAGTTGG
	CCGACCTTAGCTCAGTTGGTA	CCGACCTTAGCTCAGGTGGT	CCGACCTTAACTCAGTTGG
	GCGACCTTAGCTCAGTTGGT	CCGACCTTAGCCCAGTTGGT	CCGACCTTAGCTCAGTTGT
	CCGAGCTTAGCTCAGTTGGT	CCGACCTTAGCTCAGTGGT	
	CGACCTTAGCTCAGTTGGG	CCGACCTTAGCTCCGTTGGT	
	CCGTCCTTAGCTCAGTTGGT	CCGACCTTAGCTCAGTTAGT	
	CCGACCTTAGCTCAGTTT	CCGACCTTAGCTCAGTTGCT	
	CCGACCTTAACTCAGTTGGT	CCGACCTTAGCTCAGTTGGTT	
	CCGACCCTAGCTCAGTTGG	CCGACCTTAGCTCATTTGGT	
	TCGACCTTAGCTCAGTTGGT	CGACCTTAGCTCAGTTGG	
	CCGACCTTAGCTCAGTAGG	CCGACCTTAGCTCAGTTGGCA	
	ACGACCTTAGCTCAGTTGG	CCGACCTTAGCTCAGTAA	
	CCGACCTTAGCTAAGTTGGT	CGACCTTAGCTCAGTTGGT	
	CCGATCTTAGCTCAGTTGGT	CCGACCTTAGCTCAGTTCGT	
	CCAACCTTAGCTCAGTTGGT	CCGACCTTAGCTCGGTTGG	
	CCGACCTTAGTTCAGTTGGT	CGACCCTAGCTCAGTTGGT	
	CAGACCTTAGCTCAGTTGGT	CCGACCTTAGCTCAATTGGT	
	CTGACCTTAGCTCAGTTGGT	CCGACCTTAGCTCAGTTAG	
	CCGACTTTAGCTCAGTTGGT	CCGACCCTAGCTCAGTTGGT	
	CCCACCTTAGCTCAGTTGGT	CCGACATTAGCTCAGTTGGT	
	CCGACCTTAACTCAGTTGGTA	CCGACCTTACCTCAGTTGGT	
	CCGACCTTAGCTCAGTCT	CCGACCTTAGCTCAGTGGG	
	CGACCTTAACTCAGTTGGT	CCGACCTTAGCTCAGTTGGA	
	CCGACCTTATCTCAGTTGG	CCAACCTTAGCTCAGTTGGTA	
	CCGACCTTAGCTCAGTTGAT	CCGACGTTAGCTCAGTTGGT	
	CCGACCTTAGCTCAGTTGGT	CCGACCTTAGCTCAGCTGGT	
	CCGACCTTAGCTTAGTTGG	CCGACCTTATCTCAGTTCGT	
	GACCTTAGCTCAGTTGGT	CCGACCTTAGCTCACTTGGA	
	CCGACCTTAGCTCAGTTGTT	CCGGCCTTAGCTCAGTTGGT	
	ACGACCTTAGCTCAGTTGGT	TCGACCTTAGCTCAGTTGG	
	CCGACCTTAGCTTAGTTGGT	CCGACCTTAGCTCAGTTTGT	
	CCGACCTTACCTCAGTTGG	CCGACCTTAGCTCAGATC	
	CCGACCTTAGCTCAGTGGGT	CCGACCTTAGCTCAGTTGGG	

Supplementary Table 6 … (continued): Small RNA sequences in synthetic hexaploid with sequence similarity to non-conserved or novel known miRNAs. Vmatch identified these sequences and are complete matches with 2 or fewer mismatches to known miRNAs in miRBase v. 20. For novel known miRNAs with sequence similarity to sequences in the small RNA libraries, the species it has been annotated in is given after the miRNA name.

miR7528 (L. japonicus)	CCAAAATGCTAATCTGAGGC		
	AATGATAATCTGGAGCTT		
miR7546 (L. japonicus)	GAAAAGCCCGAACGACCT		
miR7714 (B. distachyon)	CTAATATGGATCGAAGGGAGTACC	TAATATGGATCGAAGGGAGTA	AATATGGATCGGAGGGAGTAGC
	TAATATGGATCGGAGGGAGTA	ATGGATCGAAGGGAGTAC	TTTCGGATCAATATCACT
	AATATGAATCGGAGGGAGTAG	CTAATATGGATCGGAGGGAGTA	TAATATGGATCGAAGGGAGTACC
miR7717 (B. distachyon)	CATTGAGATTCGTGGATA	AAAGATGACTGAGAAATAACT	
	TCATTGAGATTCGTGGAT	AGACGACTAAGAAATAACT	
	AGATGACTGAGAAATAACG		
miR7721 (B. distachyon)	AAAGTATGGCTTAGAATT		
miR7732 (B. distachyon)	CCAAGATCTCTGAGGCTC		
	TCCAAGATCTCTGAGGCTC		
miR7757 (B. distachyon)	ATAGCTGAATGTTTTGTT	TACAAAACCTTCAGCTAT	TACAAAACCTTCAGCTAACCA
	TACAAAACCTTCAGCTATCC	TACAAAACCTTCAGCTATC	CACAAAACCTTCAGCTATCCG
	CACAAAACCTTCAGCTATC	ATAGTTGAATGTTTTGTT	CACAAAACCTTTAGCTATCCA
	CACAAAACCTTCAGCTATCCC	GATAGTTGAATGTTTTGTTTA	CACAAAACCTTCAGCTATCC
	CACAAAACCTTCAGCTATCCA	CACAAAACCTTCAGCTATT	ATAGTTGAATGTTTTGTTTA
	GATAGCTGAATGTTTTGTTTA	TACAAAACCTTCAGCTATCCA	
	GATAGCTGAATGTTTTGTT	CACAAAACCTTCAGCTATCCT	
miR7761 (B. distachyon)	TCTTGATCAACAGACAGCT		
miR7765 (B. distachyon)	TTACAAGAAGTGGGACTAAAACGC		
	CATTTTCGTCCAACGAGTT		
	ACAAGATGTAGGACTAAAA		
miR7767 (B. distachyon)	CCCCAAGATGAGTGCTCT		
	CCCCAAGATGAGTGCTCTC		
miR7768 (B. distachyon)	CGGCGTCCTCGACCTGGA		
miR7782 (B. distachyon)	AACCGCTCTGATACCATG	CCTGCTCTGATACCATGTTGTGA	
	ACCGGCTCTGATACCATGTT		
	ACCTGCTCTGATACCATGTTGTGT		
miR7990 (S. tuberosum)	TTTCCAATGATCGTATCTTT		
miR8021 (S. tuberosum)	AATCAAGACTCAAACTCGAGACCT		
miR8029 (S. tuberosum)	CCTTTTTCTTTGTTTCGGAGC		
miR8030 (S. tuberosum)	TTGCTTTGGTCTCTGGTT		
miR8044 (S. tuberosum)	AAAATATGGTTGGAGATA		
miR8155 (C. papaya)	AACCTGGCTCTAATACCA	TAACCTGGCTCTGATACCA	CAACCTGGCTCTGATACCA
	AAACCTGGCTCTGATACC	AAACTTGGCTCTGATACCA	AATCCTGGCTCTGATACCA
	AAACCTGGCTCTGATACCA	AACCTGGCTCTGATACCA	AACTTGGCTCTGATACCA

CHAPTER 4: REGULATION OF THE HULLESS TRAIT IN *AVENA SATIVA*

4.1 Introduction

4.1.1 Oat

The Poaceae family of grasses is composed of many economically important cereals including corn (*Zea mays*), rice (*Oryza sativa*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), and oat (*Avena sativa*). Although these other cereals rank higher than *Avena sativa* in world grain production, oat has a unique nutritional profile of protein, oil, soluble fiber, and antioxidants. These qualities are sought after in both animal feed markets and for human consumption.

These desirable nutritional attributes are located in the groat, or seed. Protection by an insoluble fibrous hull lowers the nutrient density due to its indigestibility by animals and humans. Despite all oats having a hull, not all oats are classified as hulled or covered. Hulless oats have a thin hull (the lemma and palea) that protects the groat. This "hull" is easily separated upon harvest and does not require the intensive dehulling process, which can result in rancidity when groats are cracked. Hulless oat, however, has deficiencies that inhibit its worldwide commercialization and replacement of the hulled oat (Valentine, 1995).

The oat panicle has branches, or spikelets, with two to three florets per spikelet in a hulled oat plant or as many as ten florets per spikelet in a hulless oat plant. Hulled oats are classified as all having a determinate spikelet whereas hulless oats have an indeterminate spikelet. Figure 1 shows the panicle, spikelet, and flower morphology of

hulled and hulless oat. The multiflorous nature of an indeterminate spikelet produces variation in seed size. Note there are exceptions, where hulless oat cultivars are classified as having determinate spikelets. VAO-44, 45, and 46 are determinate Canadian oat cultivars from complex crosses with wild oat and other hulless cultivars. In comparison with indeterminate hulless VAO-47, these determinate hulless VAO lines have few large seeds. Fewer florets allow fewer seeds to share the nutrients, allowing increased seed size. The small kernels of VAO-47 can be blown out during harvest, reducing yield by ten percent; therefore, hulless cultivars with determinate spikelets are favored (V. D. Burrows, 2011).

Many hulless oat cultivars also exhibit incomplete expression of the hulless trait. Incomplete expression of hullessness results in mosaic plants that have both hulled and hulless kernels. In a multiflorous spikelet, seeds from primary and secondary florets are hulless but seeds from higher-ranked positions on the spikelet are hulled (Figure 1 (F)). Partial hullessness has also been observed where the lemma and palea at the base of the floret is tough as in a hulled cultivar, but the floral tissue at the apex is flimsy and papery. Incomplete expression also exhibits a temperature dependence, as complete expression of hullessness in eight hulless genotypes was observed in plants growing at 25 °C compared to incomplete expression at 20 °C and 15 °C (Lawes & Boland, 1974). During harvest, many of the hulless kernels are blown out due to the seeds' reduced weight and size, reducing the yield of these mosaic plants. In addition, the contamination of hulled seeds in the hulless sample requires dehulling of the entire sample. Hulless kernels can be damaged by the violent dehulling process and lead to rancidity. Complete expression of the hulless trait is preferred.

From a breeder's perspective, a reduced yield, temperature dependence, and damage of hulless oat seed as a result of incomplete expression of the hulless trait reduces the utility of many hulless oat cultivars. The hulless trait of oat must be better understood to improve breeding hulless oat to overcome incomplete expression and indeterminacy (V. D. Burrows, 2011).

There has been a model proposed to explain the hulless trait in oat (Jenkins & Hanson, 1976). The hulless trait is incompletely dominant, controlled by a "switch" gene and a dosage effect of three "modifier" genes, concluded from crossing hulled and hulless plants and observing the segregation ratios of the F2 population. The model does not explain incomplete expression nor have the four genes of the model been identified. Genes controlling hullessness and indeterminacy have been identified in other grasses and are candidates for controlling the hulless trait in *Avena sativa*.

In the Poaceae grasses, investigations into miR172 and its interactions with AP2/ERF family transcription factors have been performed in *Hordeum vulgare, Triticum aestivum, Oryza sativa, and Zea mays.*

4.1.2 MiR172

miR172 is a conserved miRNA family functioning primarily by the translational repression mode of target regulation (X. Chen, 2004; Gandikota et al., 2007). In the most recent release of miRBase (Kozomara & Griffiths-Jones, 2014), there are 152 miR172 genes deposited. Arabidopsis has 5 miR172 genes annotated, and in the grasses, *Sorghum bicolor* has 6, *Oryza sativa* has 4, and *Zea mays* has 5 to name a few. There are 182 mature miR172 sequences across 35 plant species with annotated pairs originating from the 3' and the 5' arm of the hairpin (Supplementary Table 1). There are also

identical miR172 sequences from the same species, indicating identical DCL products from different precursor genes. Among the mature sequences, there is clear redundancy. The most abundant sequence across all annotated miR172s is 'AGAATCTTGATGATGCTGCAT', which may indicate this mature miR172 is the most important in recognizing the AP2 domains of AP2 transcription factors to regulate floral patterning.

4.1.3 AP2/ERF Transcription Factor Family

The AP2/ERF transcription factor family is one of the larger families of plant transcription factors. The DNA binding domain that characterizes this family is conserved across both lower and higher plants, as well as protists and cyanobacteria (Wessler, 2005). Transcription factors belonging to this family are classified by their structure and binding motifs and are involved in a wide range of biological processes such as hormone response, redox, metabolism, signaling, and stress (Dietz et al., 2010). The domain in common to all transcription factors in this family is the AP2 DNA binding domain; it is approximately 60 amino acids long and folds into 3 antiparallel beta-sheets and a parallel alpha-helix (M. D. Allen, Yamasaki, Ohme-Takagi, Tateno, & Suzuki, 1998). This superfamily is classified into four or five subfamilies, all containing an AP2 domain; however, they differ in the number of AP2 domains as well as amino acid identities at certain positions in the domain (Sakuma et al., 2002). The differences specify the element recognized by the transcription factor and therefor the genes they target.

Two of the better-studied subfamilies are the AP2 (APETALA2) family and the ERF (ethylene response factor) family. The founding member of the AP2 transcription

factor family is the Arabidopsis transcription factor APETALA2 and is the most-studied gene of the AP2/ERF transcription factor superfamily through its discovery as a Class A gene in the ABC model of floral development (Bowman et al., 1991; Drews et al., 1991). Members of the Group 1 AP2-like gene family encode two AP2 DNA binding domains whereas members of the ERF gene family encode one AP2 domain (S. Kim, Soltis, Wall, & Soltis, 2006; Licausi, Ohme-Takagi, & Perata, 2013).

The ERF (Ethylene Response Factor) binds to a GCC-box. This sequence is usually upstream of genes involved in ethylene response and environmental stress such as pathogen attack and wound healing (Ohme-Takagi & Shinshi, 1995). Just as AP2 is regulated post-transcriptionally by miR172, ERF transcription factors can be post-transcriptionally regulated as well. In the grasses, stress induces alternative splicing to exclude an exon with a stop codon resulting in a functional transcription factor (Egawa, Kobayashi, Ishibashi, Nakamura, & Takumi, 2006; Xue & W., 2004). The AP2 (APETALA2) subfamily of transcription factors can contain two AP2 DNA binding domains (S. Kim et al., 2006). However, a distinguishing binding element of the AP2 subfamily has not been identified despite the early discovery of Arabidopsis APETALA2's role in flower development (Licausi et al., 2013).

4.1.4 Interaction of MiR172 and AP2/ERF Transcription Factors in the Grasses

Translational inhibition and cleavage of AP2-like transcription factors by miR172 was first described in Arabidopsis (Aukerman & Sakai, 2003; X. Chen, 2004). In the AP2 family of transcription factors, miR172 primarily targets the 3' end of the coding region of the transcript, away from the two AP2 DNA binding domains (Q. H. Zhu & Helliwell, 2011). Figure 2 shows a protein alignment of AP2-like genes IDS1

(IDETERMINATE SPIKELET1) and SID1 (SISTER OF INDETERMINATE SPIKELET1) from maize, Q (wheat), and SNB (SUPERNUMERARY BRACT) from rice. A phylogenetic tree shows the relationship of these proteins (Figure 2 (B).

4.1.5 Hullessness in Barley

Barley (*Hordeum vulgare*) is one of the closest related species to oat. Similar to oat, barley occurs in nature in both hulled and hulless forms. Hulled cultivars are used for animal feed, whereas hulless barley is preferred for human consumption as there is no need for expensive dehulling. Like hulled, or covered oat, hulled barley has a tough lemma and palea, which adhere to the seed. Hulless barley also has a lemma and palea; however, these floral organs are easily separated from the seed upon threshing. Unlike a complex system controlling hullessness in oat, the single gene NUD is responsible for hulless barley. Through positional cloning of the NUD locus in hulled and hulless barley cultivars, an ethylene response factor (ERF) transcription factor was identified as the NUD protein product. In naked barley, there exists a 17-kb deletion in the NUD locus, which includes its one AP2 domain (Taketa et al., 2008).

Although it is not known if miR172 targets NUD, miR172 does regulate barley spikelet development. A barley line was transformed with a *Zea mays* Ds element. DNA sequencing revealed that transformed indeterminate plants had the Ds element insert in a miR172 gene; the miR172 gene contained a 3.6 kilobase insert in the sequence of the mature miR172. An insert into the mature miR172 sequence inhibits miR172 from properly regulating flower and spikelet development (Brown & Bregitzer, 2011).

4.1.6 Wheat and the Q

Wheat (*Triticum aestivum*) is more closely related to barley than oat, but it has similar allo-hexaploid genome complexity to oat. Fragile glumes is one of the essential characteristics of domesticated wheat in order to harvest and process wheat on a global scale. One gene has been found to be responsible for the free-threshability of wheat, the Q gene. The Q gene has two alleles: the q and Q. q is the more primitive allele whereas Q is the allele of cultivated wheat. The Q gene is an AP2-like gene, which encodes an AP2 transcription factor (Simons et al., 2006). Differences at the DNA level account for two amino acid substitutions in the two proteins, resulting in differences in transcriptional rate and homodimer formation (Abdollahi et al., 2012; Simons et al., 2006), suggesting the sequence and structural differences between Q and q confer higher functionality to Q.

Indeterminacy in wheat is characteristic of all ploidy levels; however, spikelet number is positively associated with increasing ploidy level (Shitsukawa, Kinjo, Takumi, & Murai, 2009). More indeterminacy at higher ploidy levels is associated with the Q allele. Like other AP2 transcription factors, the Q gene contains a miR172 binding site. The domestication, Q allele has a SNP in the 3' end of the miR172 binding site (Shunzong, 2013), which suggests miR172 cannot bind or not bind as efficiently to Q. Nonetheless, sequence differences between Q and q alleles of the Q gene results in structural and functional differences in the gene, the Q allele being the preferred allele due to its role in domestication of this staple food.

4.1.7 Spikelet Abnormality in Rice

Rice (*Oryza sativa*) is another staple cereal crop where a miR172/IDS1 relationship has been established. Two other genes involved in spikelet meristem fate

have been identified to affect the miR172/IDS1 interaction (D. Y. Lee & An, 2012). In 2012, a new IDS1-like gene, SNB (SUPERNUMERARY BRACT) was discovered that acts synergistically to ensure correct time and position of spikelet meristem identity. In this study, mutant lines showing abnormal spikelets was identified from a T-DNA insertion mutant population and screened for the gene insertions. Mutants in SNB, IDS1, and mutants in both genes were identified. Double snb/ids1 mutants showed even more spikelet abnormality than either ids1 mutant or snb mutant and spikelets of the snb/ids1 mutants had open hulls due to mutations in the miR172 target site.

Transgenic mutants overexpressing miR172 showed a milder phenotype to the ids1/snb double mutants, suggesting SNB and IDS1 are only 2 of many AP2 genes involved in the development of the panicle. This overexpressing transgenic had reduced SNB and IDS1 transcript levels compared to the wildtype; however, transcripts were not depleted completely, which indicated the implication of a cleavage pathway as well as translational repression. SNB and IDS1 are negative regulators of spikelet meristem fate: once spikelet meristem is initiated, miR172 expression increases, which leads to depletion of SNB and IDS1 transcripts in addition to the translational inhibition of SNB and IDS1 (D. Y. Lee & An, 2012). A third AP2/ERF family transcription factor was identified as positively regulating SNB and IDS1 in order for correct floral organ identity and spikelet meristem development; however, the gene is classified in the ERF subfamily and it has not been identified as a target of miR172 (Ren et al., 2013).

4.1.8 Floral Differentiation and Indeterminacy in Corn

Although corn is more evolutionarily distant from oat (compared to rice, wheat, and barley), *Zea mays* has the most comprehensive study of miR172's effect on IDS1.

IDS1 (INDETERMINATE SPIKELET 1) was identified as controlling spikelet determinacy (Chuck et al., 1998). Consistent with the ABCDE model of flower development, IDS1 as a Class A gene, is expressed in the lemma and palea (outside whorl). IDS1 needs to be controlled to ensure properly timed conversion of spikelet meristem to floral meristem. In the Ts6 mutant (mutation in miR172 binding site of IDS1), improper floral differentiation resulted. The ts4 mutant has Helitron transposon insertions in the miR172 promoter regions, resulting in indeterminacy and an increase in IDS1 protein expression. A non-functional miR172 cannot bind to target IDS1 to repress translation; therefore, translation of IDS1 increases. This supports the model that miR172 targets AP2 genes primarily through translational repression. Double ts4/ids1 mutants remained indeterminate and the authors suggested the involvement of another gene in this complex system governing floral identity.

More recently in 2008, the USDA Berkeley group discovered SID1 (SISTER of IDS1), a negative auto-regulator of IDS1 (Chuck et al., 2008). SID1 protein shares 96% amino acid identity with IDS1 protein and SID1 is orthologous to rice SNB. In the ts4 miR172 mutant and an ids1 mutant, RT-PCR indicated the increase in expression of SID1. Also in the ts4 mutants, SID1 cleavage was detected by RNA cleavage assays: primarily cleavage beyond the 3' end of the miR172 binding site with fewer clones showing traditional cleavage between bases 10 and 11 of the miRNA. Wild-type only showed traditional cleavage. miR172 functions post-transcriptionally to cleave AP2 targets as well as through translational repression (Chuck et al., 2008; Chuck et al., 1998).

4.1.9 IDS1, NUD, and MiR172 in *Avena sativa*

IDS1, NUD, and miR172 are candidate genes that may have roles in conditioning the hulless trait in oat. These genes were cloned from the hulled oat cultivar GS7 (PI 5057-1) and the hulless oat cultivar Provena (PI 5051-3-2) to identify polymorphisms that could influence miR172 binding or miR172 precursor expression. Small RNAs were also sequenced from a Diversity Set of 7 hulless and 4 hulled oat cultivars. Two early flower development stages were collected under two different growth temperatures for each cultivar, resulting in 44 small RNA libraries. Using JMP Genomics (SAS) we modeled the effects of phenotype, stage, and temperature by differential expression of miR172 and other small RNAs.

4.2 Materials and Methods

4.2.1 Tissue and DNA/RNA Isolation

For the sequencing of genomic miR172, IDS1, and NUD, root and shoot tissue were harvested from Provena and GS7 oat cultivars. Seeds were plated on filter paper with 1% bleach solution and stored in the dark for 3 days in a 19°C growth chamber. A CTAB method was used for DNA extraction and the DNA was quantified using the Nanodrop (Thermo Fisher Scientific, Waltham). For small RNA sequencing of miR172, Diversity set cultivars (see Table 1) were grown in a 19°C and a 26°C Celsius growth chamber, both chambers with 18 hr. photoperiod. Flower tissue from the pre-A and A stages (see Figure 1) were collected separately on ice, and stored in the -80°C freezer. Pre-A stage of floral development pertains to pre-anthesis or 0 days post anthesis (DPA). Stage A is 1-3 DPA. A modified protocol (Emir Islamovic, UC Berkeley) using the MiraZol Small RNA Isolation Kit (BiooScientific, Austin, TX) was used for small RNA

extraction. RNA size, quality, and quantity of the enriched small RNA from the Mirazol protocol was assessed and measured on the Agilent Bioanalyzer 2100 (Santa Clara, CA) using the RNA Nano 6000 chip and reagents.

4.2.2 MiR172 Genome Walking

The miR172 gene was extended using nested PCR per manufacturer's instructions (Clonetech Genome Walker Universal Kit, Mountain View). Nested primers (As172GW-1R, As172GW-2R, As172GW-3F, As172GW-4F) were designed from a 150 base pair known miR172 sequence from the oat cultivar Leggett using primer design requirements specified by the protocol. Amplified 3' and 5' fragments were run on a 1.5% agarose/EtBr gel and all visible bands cut and labeled for size of band. Bands were cleaned-up using the Qiagen Gel Extraction kit (Qiagen, Valencia). Primer sequences can be found in Table 2.

4.2.3 Genomic IDS1 Amplification

IDS1 was PCR amplified from DNA using Q5 High-Fidelity DNA Polymerase (NEB, Ipswich, MA) and primers in Table 2 (Oat_IDS1-43F and Oat_IDS1-1243R) following manufacturers recommendations for concentration of polymerase master mix (1X) and primers (.5 μM). The PCR conditions were as follows: denaturation step of 98°C for 30 seconds 30 cycles of 98°C for 10 seconds annealing, 70°C for 15 seconds extension, and 72°C for 2 minutes elongation, with a 4°C hold. PCR products were run on a 1.5% SBE/EtBr gel and visualized on a UV imager. A 3-kilobase band was cut and purified using Qiagen Gel Extraction kit (Qiagen, Valencia, CA) to prevent Q5 proofreading polymerase from degrading tailing sequences. Internal PCR was performed on the gel slices using primers Oat_IDS1-621F and Oat_IDS1-745R (Table 2) to ensure

the purified PCR product was the IDS1 gene. Primers were at a final reaction concentration of .5 **µ**M and EconoTaq (Lucigen, Middleton) was at a reaction concentration of .25 units/**µ**L. PCR was done using 94°C for 30 seconds denaturation, 30 cycles of 94°C for 30 seconds annealing, 58°C for 30 seconds extension, and 72°C for 1 minute elongation, with a 4°C hold. UV visual inspection of amplified product run on a 1% SBE/EtBr gel size 120 bases supported IDS gene amplification.

4.2.4 Genomic NUD Amplification

Similar to the PCR amplification of IDS1, NUD was PCR amplified from DNA using primers AsNUD-1F and AsNUD-891R (Table 2) at a reaction concentration of .5 **µ**M. PCR using Q5 High-Fidelity DNA Polymerase (NEB, Ipswich) was done using 98°C for 30 seconds denaturation, 30 cycles of 98°C for 10 seconds annealing, 63°C for 15 seconds extension, and 72°C for 1 minute 30 second elongation, with a 4°C hold. The PCR reaction concentration of Q5 High Fidelity DNA Polymerase Master Mix was 1X as recommended by the manufacturer. The amplified PCR product was run on a 1% SBE/EtBr gel and visualized on a UV imager. A gel slice corresponding to about 900 bases was extracted using the Qiagen Gel Extraction Kit (Qiagen, Valencia). PCR on the gel extracted DNA was done using internal primers at a final concentration of .5 **µ**M (see Table 2 for sequences of AsNUD-1F and Oat_NUD-673R) and EconoTaq (Lucigen, Middleton) at a PCR reaction concentration of .25 units/**µ**L. The internal PCR conditions were as follows: denaturation step of 94°C for 30 seconds 30 cycles of 94°C for 30 seconds annealing, 58°C for 30 seconds extension, and 72°C for 1 minute elongation, with a 4°C hold. After running gel electrophoresis of the PCR product on a 1% SBE/EtBr gel, visualization on a UV gel imager showed a PCR amplified fragment of

about 700 bases.

4.2.5 Cloning and DNA Sequencing

Prior to cloning, A's were added to the IDS1 and miR172 genome walker PCR products through incubation at 72°C for 20 minutes using EconoTaq (Lucigen, Middleton) and dNTPs. IDS1 and miR172 genome walker products were ligated into the pSC-A-amp/kan vector and transformed into Agilent StrataClone SoloPack according to manufacturer's instructions (Agilent StrataClone PCR kit). The Agilent StrataClone Blunt Cloning kit (Agilent, Santa Clara) was used to ligate genomic miR172 and NUD PCR products into pSC-B-amp/kan vector. Competent cells were plated on LB agar with kanamyacin and 40 uL X-gal (Thermo Fisher Scientific, Waltham). Both the pSC-A-amp/kan and pSC-B-amp/kan vectors (Agilent, Santa Clara) contain a lacZ gene, making blue-white screening of colonies possible. 46 clones for each gene (23 from each Provena and GS7) were screened with internal primers to identify colonies that contained the gene insert, as was done for internal PCR checks of gel slices. Positive clones were grown up overnight followed by plasmid DNA isolation (GenScript's QuickCleanII Plasmid Miniprep Kit, Piscataway Township). DNA concentration was quantified using the Epoch BioTek Plate Reader (BioTek, Winooski). DNA was diluted to 100 ng/uL and then sent to Eurofins MWG Operon for Sanger Sequencing. The T3 and T7 primers, as well as internal sequencing primers were used for sequencing the clones.

4.2.6 IDS1, NUD, and Genomic MiR172 Sequence Analysis

Geneious Pro software package version 5.6.7 (Biomatters, New Zealand) was used to manually inspect sequence quality and trim sequences based on the chromatogram. Sequenced fragments were assembled *de novo* into the longest fragment

representing a consensus sequence of the insert from each clone. The Geneious de novo assembler uses a heuristic, greedy BLAST-like algorithm to merge the highest scoring mate pairs into a contig. The final consensus sequence and unassembled fragments were BLASTed against NCBI using BLASTN, default parameters. If the hit was similar to IDS1, NUD, or miR172 in another species, the consensus sequence or unassembled fragment was retained for future analysis. With known IDS1 and NUD genomic sequence from Leggett, Geneious performed reference assemblies to known sequences. Mapping consensus sequences to the reference uses a seed-and-expand style mapper. Geneious was also used to create phylogenetic trees of the relationships between the three homologues of each cultivar for IDS1. The neighbor-joining method was used to build the tree, based on the HKY genetic distance model. The outgroup used was the gene model GRMZM5G862109 for IDS1 from MaizeGDB.

4.2.7 Genomic MiR172 Sequence Amplification

After assembling miR172 genome walker fragments to the 150 base pair known miR172 fragment from Leggett, conserved regions were identified between Provena and GS7. New primers (As172-364F and As172-2039R) were designed to PCR amplify a 1.7 kilobase fragment containing the mature miR172 sequence (Table 2). PCR amplification was done using Q5 High-Fidelity DNA Polymerase Master Mix (NEB, Ipswich) at a reaction concentration of 1X, a denaturation temperature of 98°C for 30s, 30 cycles of 98°C annealing for 10 seconds, 63°C extension for 15 seconds, 72°C elongation for 1 minute 30 seconds, followed by a 4°C hold. The PCR product was subjected to gel electrophoresis on a 1% SBE/EtBr gel. A UV imager was used to cut the band corresponding to about 1.7 kb. Internal PCR for genomic miR172 using EconoTaq

(Lucigen, Middleton) failed a gradient PCR; therefore, Q5 High-Fidelity DNA Polymerase Master Mix (NEB, Ipswich) was used for internal PCR for genomic miR172. PCR conditions and reaction concentrations were the same as listed above except for an extension temperature of 69°C. Visualization of the PCR product on a 1% SBE/EtBr gel verified the expected size of 150 bases.

4.2.8 Small RNA Sequencing and Processing

The Illumina TruSeq sRNA library preparation kit was used to ligate RNA adapters, perform reverse transcription, ligate indexed adapters, and PCR amplify cDNA per manufacturer's instructions (Illumina, San Diego). The only exception to the protocol is the use of 18 PCR cycles instead of the recommended 11. The use of the High Sensitivity DNA chip on the Agilent Bioanalyzer 2000 (Agilent, Santa Clara) is a quality control check of presence of small RNAs in each sample prior to pooling samples of different indexes. 24 samples were pooled equimolarly based on the lowest molarity sample. Pooled samples were run on a TBE PAGE gel (Invitrogen, Calsbad) and bands in the size range 140-160 will be cut out, DNA was extracted, purified, and ethanol precipitated (per Illumina Truseq sRNA Library Prep kit). Illumina sequencing by synthesis was done on the HiSeq2000 instrument at D.H. Murdock Research Institute (DHMRI) at the North Carolina Research Campus (NCRC) in Kannapolis, North Carolina or on the HiSeq2500 by Eurofins MWG Operon.

FASTQ files were filtered for quality using the fastx quality filter tool of the FASTX-Toolkit (Hannon Lab, CSHL). The filter parameter was –q 15 –p90. In house python scripts trimmed adapters and filtered for length. Duplicate reads were collapsed using the collapse_reads_md.pl from miRDeep2 (Friedlander et al., 2012) for input into

miRDeep-P (X. Yang & Li, 2011). Total read counts before and after pre-processing are shown in Table 3.

4.2.9 MiR172 Expression

To create a high confidence annotation set for JMP Genomics (SAS, Cary), all mature miR172 sequences were filtered from all downloaded mature miRNAs from miRBase release 20 (Kozomara & Griffiths-Jones, 2014). Vmatch (Kurtz, 2011) was used to create a suffix tree and perform the matching tasks of the small RNA sequences to known miRNAs in miRBase. The mkvtree program indexed the miR172 sequences and the vmatch program solved the matching tasks over the index. Parameters used were specifying a minimum length direct (-d) match returned of 18 (-l 18) constraining the number of mismatches or hamming distance of the matches returned to three or fewer (-h 3). The final annotation set had 4174 sequences.

JMP Genomics 6.1 (SAS, Cary) was used to assess distribution, perform correlation with Pearson's coefficient and a principal component analysis on log2-transformed raw count data of small RNA. The Basic miRNA-seq Workflow was used for normalization schemes methods IQR (Interquartile Range), STD (Standard), MEAN, MEDIAN, KDL (Kernel Density Loess), and KDQ (Kernel Density Quantile). Other normalization methods available for Next-generation sequencing data were also performed: TMM (Trimmed Mean of M component), KDMM (Kernel Density of M component), and Upper Quantile Scaling. See Table 4 for a description of the normalization techniques used in JMP Genomics. ANOVA parameters of log2-normalized data include FDR as the multiple testing method and an alpha of .05. FDR calls for adjusted p-values using the linear step-up method of Benjamini and Hochberg.

A mean difference filter of 2 was also used.

4.3 Results and Discussion

4.3.1 Genomic Variation in IDS1

In sequencing IDS1 from Provena and GS7 oat cultivars, three homologues of each cultivar were identified. For Provena, PIDS-4 is representative of 4 clones PIDS-3 representative of 2 clones, and PIDS-6 representative of one clone. For GS7, GIDSb4 represents consensus of 7 full-length individual clone assemblies, GIDSb20 is representative of 3, and GIDSb11 is the only sequenced clone of that homologue. We assume each of the three homologues represents genomic IDS1 from each of the three subgenomes in hexaploid oat (the A, C, and D genomes); however, a phylogenetic tree of the genomic IDS sequences does not allow us to distinguish from which subgenome the genomic sequence originated from (Figure 3). It has been proposed the D genome is a result of a duplication of the A genome so we may infer the A and D genomic IDS sequences are represented by the group GIDSb4, GIDSb11, PIDS-3, PIDS-6 whereas PIDS-4 and GIDSb20 sequences originate from the C genome. There may be more genomic copies of IDS1; however, we were unable to identify more than three from each cultivar. Figure 4 shows a reference assembly of all homologues to a reference oat cultivar Leggett IDS1 clone. Three IDS1 homologues have been cloned from hulled *Avena sativa* cultivar Leggett as well. Black tic marks within the assembly sequence represent sequence polymorphisms (SNPs and insertions) different from the reference sequence. The reference genomic IDS1 has annotation of 10 exons supported by prior RNA-seq assemblies; however, we cannot identify the transcription start site from the annotation and the possible open reading frames extend beyond the annotated start of the

exon (Figure 5). Therefore, we cannot compare the amino acid sequence of the homeolgues or differences in promoter sequences that may influence rate of transcription.

4.3.2 Coding Variation in IDS1

miR172 is known to bind to IDS1 in the last exon. Figure 6 shows the binding site of miR172. There are no SNPs or indels in the binding site of any of the homologues that could inhibit miR172 from binding. There is also complete conservation of the miR172 binding site in a coding sequence from Provena (data not shown).

RNA-seq was also performed on bulk flower tissue from Provena and GS7 under the 19 degrees Celsius growth temperature, sequenced by Operon, and assembled using Trinity. A five-step BLAST (wheat, barley, rice, corn, Brachypodium) followed to obtain annotations for the assembled contigs. Only one contig from Provena had annotation to IDS1 from *Zea mays*. The contig is shown in Figure 7, in a *de novo* assembly to the inferred exons of the IDS homologues. The contig extends beyond the stop codon at the last exon and also has sequence variation at the beginning. SNPs present in the alignment cannot identify the cDNA sequence as originating from any of the Provena IDS homeolgues' exons. In addressing the variation at the beginning of the cDNA sequence, upon alignment of the cDNA sequence not aligning to exon regions with the genomic clones (Figure 8), the sequence aligns with the intronic region immediately before the 9th exon of PIDS-3 and PIDS-4 (with the exception of a SNP at the first position). This suggests a variation in splicing pattern from the inferred annotation of these two homologues, where intronic sequence immediately before exon 9 is not spliced out as predicted. Although this is likely not the whole IDS1 cDNA sequence, the mRNA sequence would be different for one or both of these homologues,

resulting in different proteins. Without cDNA annotated as IDS1 from GS7, we cannot eliminate that these IDS1 homologues from GS7 have a comparable splicing pattern as the two homologues cloned from Provena.

4.3.3 Genomic Variation in NUD

Three homologues for the NUD gene were identified through cloning. From Provena, PNUD-2 and PNUD-10 each represent 3 clone assemblies of the gene and PNUD-8 represents 4. GNUD-8 assembly consensus sequence is identical to 5 other clone assemblies whereas GNUD-12 and GNUD-9 each represent 2 clone assemblies of NUD from oat cultivar GS7. Figure 9 shows a reference assembly of the NUD homologues from Provena and GS7 using an annotated NUD consensus sequence from *Avena sativa* cultivar Leggett. The Leggett reference sequence has been annotated with the full exon (inferred from RNA-seq). At the genomic level, we do not see large deletions in any of the Provena NUD homologue exons as was seen with NUD from hulless barley.

4.3.4 MiR172 Binding in NUD

Without evidence of the assembled transcripts of NUD, we cannot transfer the exon annotation from Leggett with confidence. However, miRNAs regulate their targets primarily by targeting the coding regions so we looked at where miR172 would bind to the Leggett NUD exon (Figure 10). In this reference assembly, miR172 would bind to the coding region of Leggett NUD with 6 mismatches and a 1 base asymmetrical bulge. These binding characteristics are the same for two homologues from each cultivar. For GNUD-9 and PNUD-8, miR172 would bind with 7 mismatches in addition to the asymmetrical bulge. Because miRNAs bind to their respective targets with high

complementarity, canonically with fewer than 4 mismatches (Rhoades et al., 2002), this discredits the regulation of NUD by miR172.

4.3.5 Genomic MiR172

Only one miR172 gene was sequenced for each of the cultivars. This gene is very similar between Provena and GS7 as is shown in the alignment of the sequences to a pre-miR172 gene from barley. The consensus sequence G1722-14 is representative of 9 clones whereas P1722-8 consensus represents 10 clone assemblies. Prior small RNA sequencing in oat has revealed more than one mature miR172 sequence despite cloning only one miR172 gene. Only discovering one miR172 gene is most likely technical due to the primer design. The design of primers for Genome Walking for miR172 was difficult due to the short 150 base pair known sequence from Leggett. Primer sets meeting the primer design criteria were very near the mature miR172 sequence. This suggests 5' and 3' nucleotides adjacent to the mature miR172 vary between miR172 genes. The alignment of the oat miR172 to the barley pre-miR172 also supports this (Figure 11(A)). A 43 base pair fragment including the 21 base pair miR172 is conserved between the oat and barley pre-miR172 genes (Figure 11(B)), suggesting this barley pre-miR172 is not homologous to the miR172 gene sequenced from Provena and GS7.

The barley pre-miR172 gene is 1,018 bases long but the cloned pre-miR172 genes from oat are about 800 bases longer; without a RNA-seq contig consensus sequence representing the miR172 transcript in either Provena or GS7, we cannot annotate the gene's correct length. With the genomic sequences we have, we can predict how the non-coding RNA would fold, and identify the hairpin for the miR172 sequence: GGAATCTTGATGATGCTGCAT (Figure 12). Identification of the fold-back

secondary structure of a miRNA precursor fulfills the biogenesis criteria for annotation (Meyers et al., 2008). Although there may be more than one miR172 gene that is processed into this exact mature sequence, we can fully annotate the first miR172 from *Avena sativa*.

4.3.6 Quality Control of Log2-transformed Raw Count Data

The number of unique sequences with count data across the 44 samples represents 24,799,611 observations. The Experiment Input Engine of JMP Genomics log2-transformed the read counts. Quality control of the log2-transformed sequence intensities was assessed with a distribution analysis and a correlation/PCA analysis. Due to the many sequences with only one read count that may be attributed to sequencing errors and low quality reads, the distribution is skewed (Figure 13). A Principal Component Analysis (PCA) shows a decomposition of variance components with the principal component of stage contributing the most variability to the dataset (Figure 14). A 3D scatterplot of the variance components is shown, with samples from the same stage in development colored the same (Figure 15). A heat map and corresponding dendrogram of the correlation matrix (Pearson) with rows and columns sorted by hierarchical clustering is depicted in Figure 16. Positive correlations are colored red and negative correlations are colored blue. To demonstrate the effect of the principal component of stage to the correlation of samples, the dendrogram and labels of samples are colored by stage of floral development. With the exception of four samples, the principal component of stage is clustered together. From the original covariance matrix in testing for outliers, Mahalanobis distance is not significant for any of the samples (Figure 17).

4.3.7 Normalized Distributions

There is no standard for miRNA-seq normalization. Methods for normalization have been adapted from microarrays and RNA-seq. No method is ideal for every dataset and each has its own bias. Figure 18 shows a side-by-side comparison of distributions of the log2-transformed data for different normalization methods implemented in the Basic miRNA/miRNA-seq Workflow and Figure 19 shows the distributions of log2-transformed intensities normalized by scaling methods. Of the normalizations presented here, Loess normalization with kernel density estimation removes the skewedness of the distribution. However, KDL normalization results in a reduction in the overall weighted average proportion of variance attributed to each principal component (Figure 20). Upon inspection of the normalized log transformed read counts, only sequences starting with an 'A' were normalized. It is unclear if it is the Loess normalization or the kernel density estimate implementation of Loess normalization that is unable to normalize such a large dataset with such a high occurrence of missing data.

Using scaling methods for normalization, the TMM shows increased variability between experiments and Upper Quantile Scaling normalized only the upper quantile resulting in increased variability if comparing samples from the two different quantile distributions shown in Figure 19. KDMM normalization appears to have little effect on the distribution, data correlation due to the scaling factors (Figure 21).

Although the assumption of normality was violated, we proceeded with an ANOVA with log2-transformed normalized data from KDMM scaling, KDQ, STD and IQR normalization schemes. Despite IQR having no apparent shift in normalized data or change in data correlation, IQR was effective at normalizing this dataset with all missing

data filtered out (only sequences present in all samples were normalized). The PCA histograms of weighted proportions of variances and heat map of clustered Pearson correlation coefficients of these normalization methods are shown in Figures 22-24.

4.3.8 ANOVA of MiR172

Our annotation set was the high confidence miR172 sequences from the 44 samples using the Vmatch parameters discussed in Materials and Methods. This allowed us to perform the ANOVA on only our sequences of interest. The Difference Chooser was used to do hypothesis testing on the variables and interactions seen in Table 5. No significant differences were found in expression of normalized miR172 sequences. The mean difference filter was changed from 2 to 1, a less stringent parameter for fold change, and several miR172 sequences passed the multiple testing significance threshold. The volcano plots of IQR and KDQ normalized miR172s are difficult to interpret due to sequences with –log10 (p) values over 200 and the horizontal significance threshold of only 2.5 (Figure 25). The horizontal red dotted line represents the significance threshold with false discovery correction specified in the ANOVA parameters.

Figure 26 and 27 depict selected multiple testing comparisons performed on STD and KDMM normalized miR172 intensities. The red and green points represent the sequences shown in the dendrogram of clustered standardized LSmeans (Figure 28). These two sequences were found to be significant in more than one multiple testing comparison across IQR, KDMM, and STD normalization methods. The miR172 sequence represented by the green point in the volcano plots ‘CTCTGTATCACAATGACGCTGCAT‘ was determined to be significantly differentially expressed by all four normalization methods. This miR172 is up-regulated

in hulled oats in the phenotype*temperature*stage effect of H26pA-N26pA and H26A-N26A in addition to two other multiple testing comparisons. The sequence represented by the red point in the volcano plots is 'AGTATGAGAATTTTGATGTGGCAA', determined significant by IQR, STD, and KDMM for the comparison of H26A-N26A. In contrast to 'CTCTGTATCACAATGACGCTGCAT' being up-regulated in the hulled phenotype in the H26A-N26A comparison, 'AGTATGAGAATTTTGATGTGGCAA' is up-regulated in the hulless phenotype as shown by its location in the upper left corner of the volcano plot (Figure 27) and a negative value for the difference of comparison. The statistics of all significant miR172 sequences including the difference of comparison, t-statistic, p-value, and –log10 (p-value) are also shown in Table 6. Although a high R^2 value with low residual sample variance indicates that the model explains the sample variance observed, the miR172 annotation set is small. In a comparison flagged significant for differential expression of two sequences, for example H26A-N26A, one sequence may exhibit high R^2 and low residual variance whereas the other sequence may exhibit the low R^2 and high residual variance (shown in Table 6). By modeling the effects of phenotype, temperature, and stage of floral development with expression data from one family of miRNAs, we cannot use the R^2 value and residual sample variance as indicator of how well the observations explain the effect. Stem-loop qRT-PCR of the miR172 family members identified as differentially expressed in these comparisons can validate if the expression of these sequences should be part of a model to explain the mixed effect of phenotype, temperature, and stage in these hulled and naked oat cultivars.

4.3.9 ANOVA of Missing Data Filtered Subset of Small RNA Intensities

There may be other miRNAs present in the dataset that are differentially expressed that may have an effect on the severity of the indeterminacy character in hulless oats. We filtered for sequences missing in no more than 28 samples. If a sequence was expressed in hulled cultivars only, the sequence intensity for those 28 samples, representing the 7 hulless cultivars, would be considered missing data. The different normalization methods show different clustering of the standardized LSmeans of sequences with intensities deemed significant in at least one comparison (Figure 29-32). Due to IQR and KDQ-normalized intensities having such a low p-value, the volcano plots and interpretation of the standardized LSmeans cluster pattern across the comparisons is difficult. On the contrary, STD and KDMM normalizations allow for better discernment of patterns from volcano plots. The clustering pattern of standardized LSmeans of KDMM-normalization was selected for further study by reason of the expression pattern of the green cluster in two comparisons of the phenotype*stage effect and one comparison of the phenotype*temperature*stage effect as seen in their respective volcano plots (Figure 33).

The KDMM normalized, filtered subset represented approximately 300,000 observations. After performing the ANOVA with a mean difference filter of 2 to identify sequences whose intensities exhibit significant, high fold-change differential expression, 3,024 sequences were determined differentially expressed in the at the –log 10(p-value) threshold of 3.98. The volcano plots show the green clusters (identified by hierarchical clustering of standardized LSmeans shown in Figure 32). Sequences in the green clusters are up-regulated in the naked phenotype as shown in the comparison of HpA-NpA and

H26pA-N26pA and preferentially at stage pA displayed by the NA-NpA comparison (Figure 33).

In the HpA-NpA comparison, there are 190 sequences up-regulated in the naked oat phenotype and 24 in the hulled phenotype. For the H26pA-N26pA comparison, there are 143 sequences significantly up-regulated in the naked phenotype but only 5 in the hulled phenotype. Using Vmatch (parameters –l 18, -d, -p, -evalue 4), only two sequences have sequence similarity to known plant miRNAs annotated in miRBase v.20. A member of the conserved miR396 family was up-regulated in the naked phenotype for both these comparisons. miR396 is known to target the GRF (GROWTH-REGULATING FACTOR) family of transcription factors to control cell growth (Rodriguez et al., 2010). The other sequence, significantly up-regulated in the naked phenotype of the Hpa-NpA comparison, is a member of a monocot-specific miRNA family, miR444. miR444 is involved in flower development through targeting MADS-box genes (L. Wu et al., 2009). These two sequences are not unique to naked cultivars; they are present in hulled oat small RNA libraries. With implications in floral development, it is not surprising that these sequences are present in both phenotypes.

4.4. Conclusions

Three homologues corresponding to the three subgenomes of each *Avena sativa* cultivar were identified for each IDS1 and NUD. At the genomic level no large insertions or deletions were observed that could explain hullessness in *Avena sativa*. The miR172 binding site was conserved across all homologues of IDS1, in both hulless Provena and hulled GS7. If it was disrupted in one or more of the Provena homologues, miR172 would not be able to translationally inhibit IDS1, leading to higher levels of the

protein and resulting in the hulless, indeterminate phenotype. The potential binding site in NUD also was completely conserved; however, with seven mismatches and a two-nucleotide asymmetrical bulge, we hypothesize miR172 does not target NUD.

The RNA-seq dataset from bulk flower tissue of GS7 and Provena needs to be investigated further. The subgenomic complexity of the oat genome may have resulted in a mis-assembly. Support for this is that the cDNA contig identified as IDS1 from Provena does not assembly to any genomic homologue in particular. Contigs may need to be re-assembled using different methods to get three homeologous transcripts of IDS1 and NUD from each cultivar. With full-length cDNA, we can compare the annotations of these genes to the reference annotations from the Leggett cultivar.

With IDS1, we can determine if differences in the cDNA between homologues affects the secondary structure of the mRNA and if miR172 binding may be affected. If miR172 cannot bind to the mRNA from the two Provena homologues, there will be no translational inhibition of IDS1. If there is a difference in the coding sequence of the two Provena homeolgues compared to GS7 homologues, the amino acid sequences may differ. This may, in turn, affect protein folding and functionality of IDS1 as a transcription factor.

Likewise with NUD and miR172, we were not able to identify large insertions or deletions that would eliminate the AP2 domain of Provena NUD or interfere with hairpin folding of Provena pre-miR172. A more comprehensive inspection of cDNA assemblies may lead to differences in splicing, amino acid sequence, and functionality of the NUD protein. cDNA of pre-miR172 will enable annotation of the true length of the miR172 gene already sequenced and hairpin sequences of other pre-miR172s required for

annotation of the miR172 family in *Avena sativa*. With genomic and transcriptome sequences, we can then begin to explore expression.

Expression of IDS1 and NUD need to be quantified in Provena and GS7. However, because miR172 targets IDS1 through translational repression as seen in other species, qRT-PCR of IDS1 mRNA may not show reduced mRNA levels as would be expected if miR172's primary mechanism was cleavage. Some reduced level of IDS1 mRNA may be observed due to overlap of the cleavage/inhibition pathways (D. Y. Lee & An, 2012). Alternatively, the feedback mechanism of IDS1 down-regulating miR172 through co-repressors may be observed. In Provena, high levels of IDS1 mRNA may correlate with low levels of miR172 as has been observed in an indeterminate kiwi (Varkonyi-Gasic, Lough, Moss, Wu, & Hellens, 2012). Hulless oat may exhibit a mis-regulation of miR172. An antibody for IDS1, initially designed for immunolocalization of IDS1 in maize (Chuck, Cigan, Saeteurn, & Hake, 2007), could be used in ChIP-seq of IDS1 to determine if a pre-miR172 is its target.

Implementing JMP Genomics in the differential expression analysis of small RNAs has identified candidate miR172 sequences that are differentially expressed in early floral development of hulled and hulless oat under different growth temperatures. Stem-loop qRT-PCR of these miR172 sequences will support or discredit the model. However, other differentially expressed small RNAs have also been identified by JMP analysis. In the KDMM-normalized, filtered dataset, the hierarchical clustering revealed a cluster (blue in the dendrogram) that has significant up-regulation in samples grown in 26 °C and another cluster (red in the dendrogram) that is differentially expressed at stage A (1-3 days post anthesis). Identifying the small RNAs from these clusters may reveal

temperature responsive and stage-dependent miRNAs and siRNAs.

With the current filtered and normalized sequence intensities from these 44 samples, we can explore many more comparisons with agronomical value. The cultivars 58-2 and VAO-2 are bald oat cultivars. 58-2 is hulled and VAO-2 is hulless but both are sparse in trichomes. Trichomes are the tiny hairs on the groat, which attract fungal spores, causing disease. Trichomes also become dislodged upon handling and can become respiratory irritants for handlers. Although the bald conditioning gene Gt-1 is responsible for this trait (V. D. Burrows et al., 2001), differential expression of small RNAs may also influence oat baldness.

The small seed size of indeterminate hulless oat is a deficiency. Comparing the small RNA profile in determinate VAO-44, 45, and 46 to indeterminate VAO-47 may identify differentially expressed small RNAs responsible for seed size in determinate oat cultivars.

Small RNA profiles vary between cultivars and although a comparison of two cultivars would reduce statistical power, the nearly isogenic lines 141-C and 141-N would eliminate noise from cultivar-specific expression of small RNAs. Modeling the effect of phenotype, temperature, and stage with these NILs is an appropriate test case in implementing different normalization schemes; however, normalization of miRNA-seq is innately more complicated due to its lack of a standard.

Normalization schemes used in miRNA-seq analysis have been adopted from microarray and RNA-seq; however, it is questionable how effective these normalizations have been on this dataset of raw read counts in the presence of missing data. This normalization issue needs to be examined and tested extensively before we can put value

to the differential expression analysis. These normalization schemes need to be evaluated with larger datasets, including replicates, and a more balanced experimental design.

Despite the preliminary nature of differential expression of small RNAs in this study, JMP Genomics has the capabilities and the statistical expertise to continue exploring this dataset with different normalization methods including mixed-effect normalization. The utility of building workflows to implement multi-step normalization and quality control, depending on the amount of missing data present may be of great applicability.

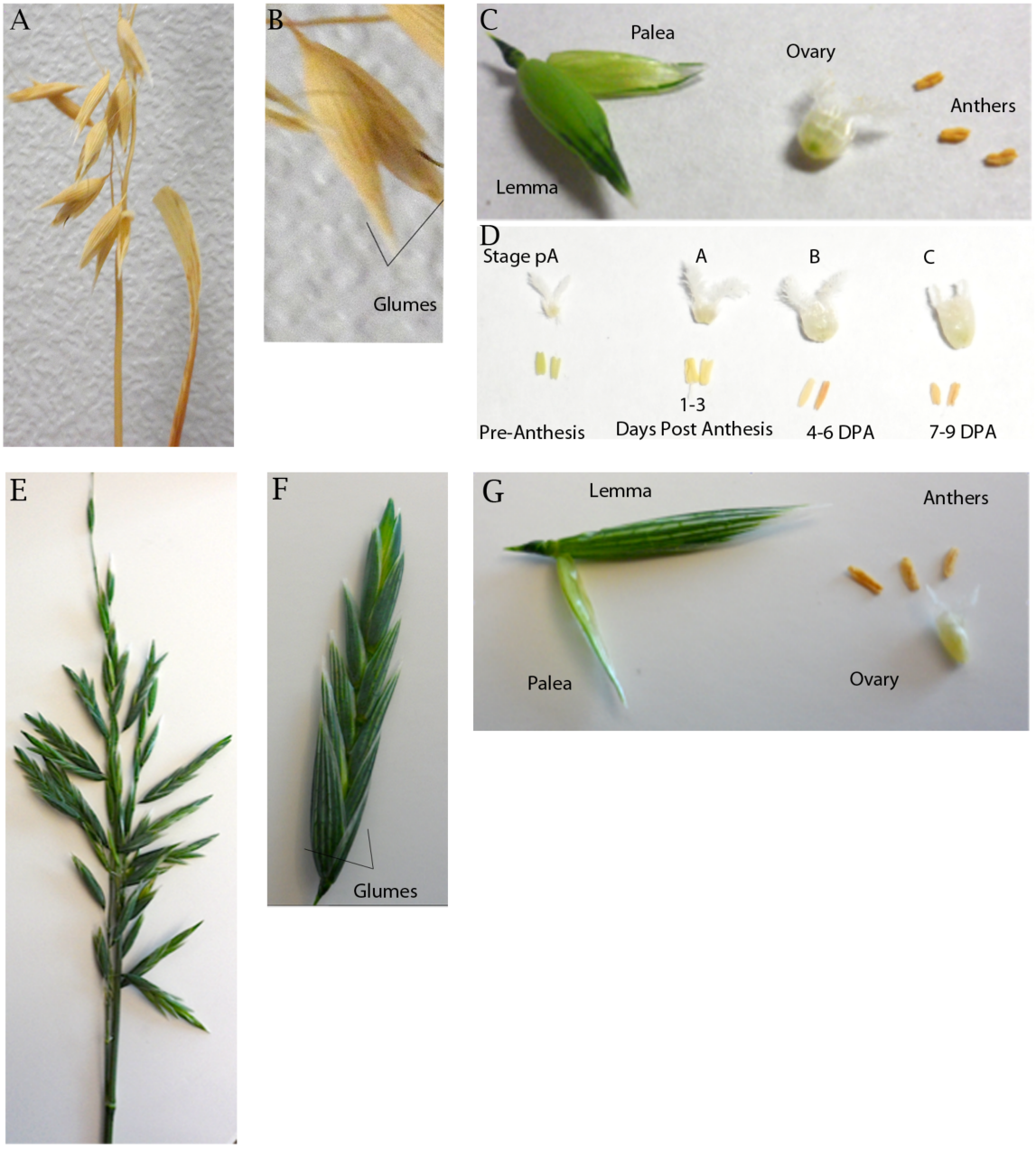


Figure 1: Morphology of hulled and hulless oat. The panicle (A, E) is the branch to which spikelets are attached. Hulled oats have a determinate spikelet (B) whereas hulless oats are classified as having indeterminate spikelets (F). Flowers are protected by sterile bracts, glumes. The flower can be hulless (G) with a papery lemma and palea or hulled (C) with a hardened lemma and palea. (D) shows the ovary and anthers during floral development.

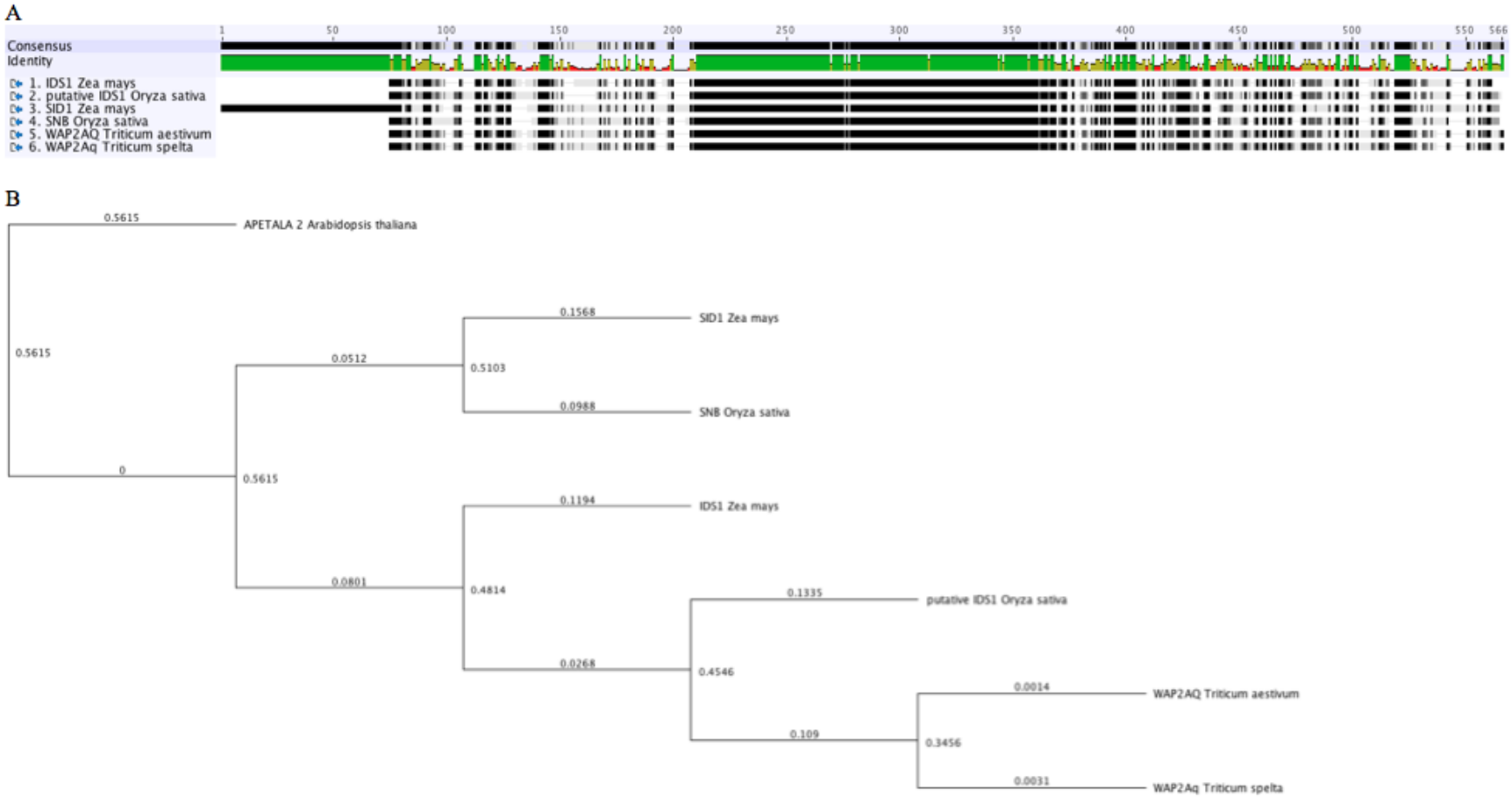


Figure 2: Relationships of AP2-like genes in the grasses. (A) shows a protein alignment from Geneious 5.6.7 (Biomatters, New Zealand) using a Blosum62 matrix and a gap penalty of 12 and extension penalty of 3. Black bars represent perfect identity between amino acids. Using APETALA2 as the outgroup, (B) shows a phylogenetic tree demonstrating the relationship of these AP2-like proteins. Geneious used a Blosum62 matrix, Jukes-Cantor evolutionary distance model, and the Neighbor-Joining method to build the tree. All sequences used were obtained from NCBI Protein database with the exception of IDS1 from rice, which was obtained from the MSU Rice Genome Annotation Project release 7. Identifiers of sequences used in the protein alignment and phylogenetic tree construction are given after the name:IDS1 Zea mays (gi|163937834), putative IDS1 Oryza sativa (LOC_Os03g60430.2), SID1 Zea mays (gi|225703094), SNB Oryza sativa (gi|87130803), WAP2Aq Triticum spelta (gi|385862166), WAP2AQ Triticum aestivum (gi|385862164), and APETALA 2 Arabidopsis thaliana (gi|334187226).

Table 1: Diversity Set of oat cultivars used for small RNA sequencing.

Hexaploid Cultivars (*Avena sativa*)	**ID**	**Classification**
Leggett		Hulled
58-2	CN 18943	Hulled
GS7	PI 5057-1	Hulled
VAO-141 C	CN 18941	Hulled
VAO-141 N	CN 18942	Hulless
VAO-47	CN 108633	Hulless
Provena	PI 5051-3-2	Hulless
VAO-2	CN 19196	Hulless
VAO-44	CN 108630	Hulless, determinate
VAO-45	CN 108631	Hulless, determinate
VAO-46	CN 108632	Hulless, determinate

Table 2: Primers used in the sequencing of NUD, miR172, and IDS1.

Primer Name	**Sequence**
AsNUD-1F	ATGGTACAGTCCAAGCAGAAG
Oat_NUD-673R	GCTGTCAGTACTGGGCGTTGG
AsNUD-891R	TCAGATGACGAGGCTGCCTTC
As172-364F	CGATGGCGATTGCTAACCAC
As172-2039R	CGGTCGCGGTAGAAATAGGG
As172GW-1R	GATTCCCACACGCACGGATGGGAGGCA
As172GW-2R	TGCAGCATCATCAAGATTCCCACACGCAC
As172GW-3F	TGATTGATTAGCAACAAGCTGCTGCACGCA
As172GW-4F	GAACAAGTGTTCTTAAGCAGGCAGGCAGGC
Oat_IDS1-43F	ACGTCGAGCTCGTCCGTCCT
Oat_IDS1-621F	CGGCGTCACGCTCCACAAGT
Oat_IDS1-745R	CCGCCCTGTCGTACGCTCTT
Oat_IDS1-1243R	AGGCGGAGGTGGTGTTTGCG

Table 3: Read counts of small RNA samples before and after pre-processing.

Hexaploid Cultivars (*Avena sativa*)	Growth Temperature (Degrees Celsius)	Stage (Days Post Anthesis)	Total Reads (FASTQ)	Pre-processed Reads	JMP Sample ID
Leggett	19	pre- or 0 DPA	2,353,860	531,462	H_19_pA_1
		1-3 DPA	2,604,118	558,827	H_19_A_1
	26	pre- or 0 DPA	4,453,319	976,055	H_26_pA_1
		1-3 DPA	4,498,661	1,091,580	H_26_A_1
58-2	19	pre- or 0 DPA	3,926,892	531,461	H_19_pA_11
		1-3 DPA	8,112,951	1,170,707	H_19_A_11
	26	pre- or 0 DPA	5,303,836	1,409,573	H_26_pA_11
		1-3 DPA	5,118,748	715,647	H_26_A_11
GS7	19	pre- or 0 DPA	1,994,362	445,957	H_19_pA_4
		1-3 DPA	9,004,705	1,192,008	H_19_A_4
	26	pre- or 0 DPA	6,850,250	1,089,307	H_26_pA_4
		1-3 DPA	5,053,106	858,860	H_26_A_4
VAO-141 C	19	pre- or 0 DPA	1,964,115	289,273	H_19_pA_5
		1-3 DPA	3,127,512	570,230	H_19_A_5
	26	pre- or 0 DPA	4,437,868	1,111,415	H_26_pA_5
		1-3 DPA	1,464,416	374,662	H_26_A_5
VAO-141 N	19	pre- or 0 DPA	7,394,073	1,383,066	N_19_pA_6
		1-3 DPA	4,161,704	681,575	N_19_A_6
	26	pre- or 0 DPA	8,094,904	2,110,334	N_26_pA_6
		1-3 DPA	10,971,895	1,906,904	N_26_A_6
VAO-47	19	pre- or 0 DPA	5,584,865	760,632	N_19_pA_10
		1-3 DPA	7,557,285	825,520	N_19_A_10
	26	pre- or 0 DPA	18,002,925	2,789,756	N_26_pA_10
		1-3 DPA	3,084,736	528,082	N_26_A_10
Provena	19	pre- or 0 DPA	10,022,633	1,050,935	N_19_pA_3
		1-3 DPA	4,092,181	710,789	N_19_A_3
	26	pre- or 0 DPA	4,029,277	1,015,212	N_26_pA_3
		1-3 DPA	3,881,540	629,834	N_26_A_3
VAO-2	19	pre- or 0 DPA	11,357,115	1,654,708	N_19_pA_12
		1-3 DPA	13,161,687	1,602,145	N_19_A_12
	26	pre- or 0 DPA	12,033,983	1,548,181	N_26_pA_12
		1-3 DPA	7,787,638	1,249,961	N_26_A_12
VAO-44	19	pre- or 0 DPA	5,256,752	993,253	N_19_pA_7
		1-3 DPA	2,425,086	164,944	N_19_A_7
	26	pre- or 0 DPA	6,328,741	935,482	N_26_pA_7
		1-3 DPA	5,325,036	754,786	N_26_A_7
VAO-45	19	pre- or 0 DPA	4,999,933	863,522	N_19_pA_8
		1-3 DPA	4,334,267	420,398	N_19_A_8
	26	pre- or 0 DPA	4,494,718	924,412	N_26_pA_8
		1-3 DPA	14,804,161	1,623,301	N_26_A_8

Table 3… (continued): Read counts of small RNA samples before and after pre-processing.

VAO-46	19	pre- or 0 DPA	5,569,209	764,773	N_19_pA_9
		1-3 DPA	2,329,288	482,902	N_19_A_9
	26	pre- or 0 DPA	15,160,814	2,827,315	N_26_pA_9
		1-3 DPA	7,685,393	792,026	N_26_A_9

Table 4: Normalization methods available in JMP Genomics 6.1 (SAS).

	Normalization	**Description**
Basic miRNA/miRNA-seq Workflow	NONE	No normalization
	STD	Column mean=0, variance=1
	MEAN	Mean=0
	MEDIAN	Median=0
	IQR	Interquartile range=1
	KDQ (Kernel Density Quantile)	Quantile normalization based on a kernel density-selected training set
	KDL (Kernel Density Loess)	Applies kernel density as a weight for Loess modeling
Expression (Next-gen only)	KDMM	Scaling factor=kernel density estimated weighted mean of M component
	RPM Scaling	Divides raw read counts by total number of mapped reads, then multiplies by 1,000,000
	TMM	Scaling factor= weighted mean of the M component after trimming means
	Upper Quartile Scaling	Scaling factor= upper quartile divided by geometric mean of upper quartiles across columns



Figure 3:Phylogenetic tree of genomic IDS sequences. The phylogenetic analysis was performed by Geneious using an IDS1 sequence from *Zea mays* as the outgroup.



Figure 4: Reference assembly of IDS1 clones from provena and GS7 aligned to a reference clone. The reference clone is from a hulled oat cultivar Leggett and has been annotated from assemblies of RNA-seq data.

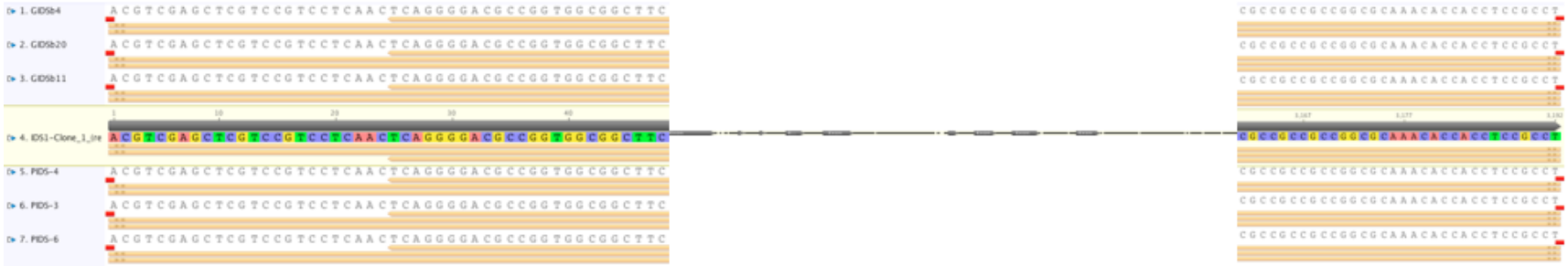


Figure 5: Depiction of possible ORFs in assembled IDS clones in Provena and GS7. The transcription start site is missing from the annotated IDS clone from Leggett as well as the stop codon.

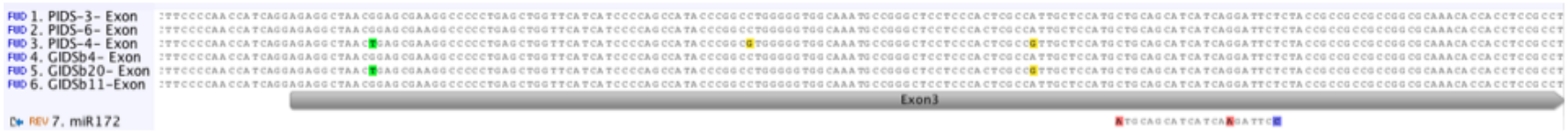


Figure 6: MiR172 binding site In IDS1. The miR172 binding site is conserved in all IDS1 clones from both Provena and GS7 oat cultivars.



Figure 7: *De novo* assembly of IDS cDNA from Provena to inferred exons of IDS1 clones. Sequence 7 is the cDNA contig identified by the 5-step BLAST as IDS1.

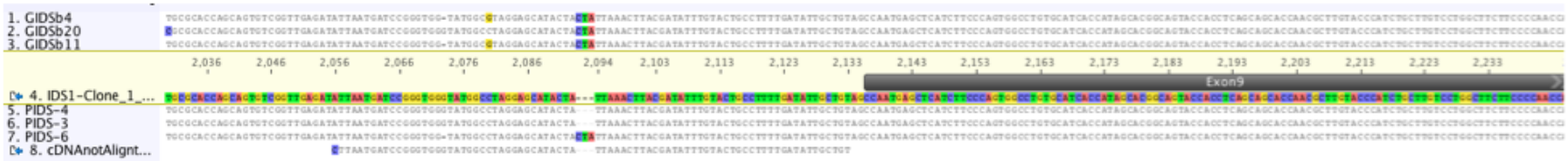


Figure 8: Assembly of genomic IDS1 with cDNA from Provena. This alignment shows that the beginning of the cDNA sequence that did not align to inferred exonic region aligns to the intron immediately before exon 9 in two Provena homologues.



Figure 9: Reference assembly of genomic NUD homologues from Provena and GS7. Consensus sequences of each clone assembly are aligned to a consensus sequence from oat cultivar Leggett.

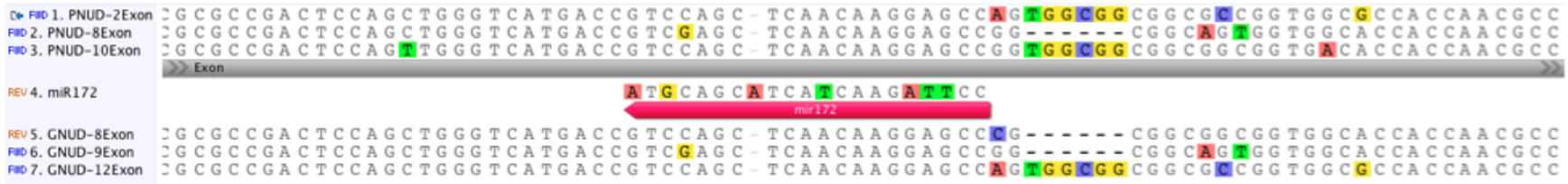


Figure 10: Alignment of NUD homologues with miR172. MiR17 2 aligns with 6 or 7 mismatches and an asymmetrical bulge to the NUD exonic region conserved in all NUD homologues.

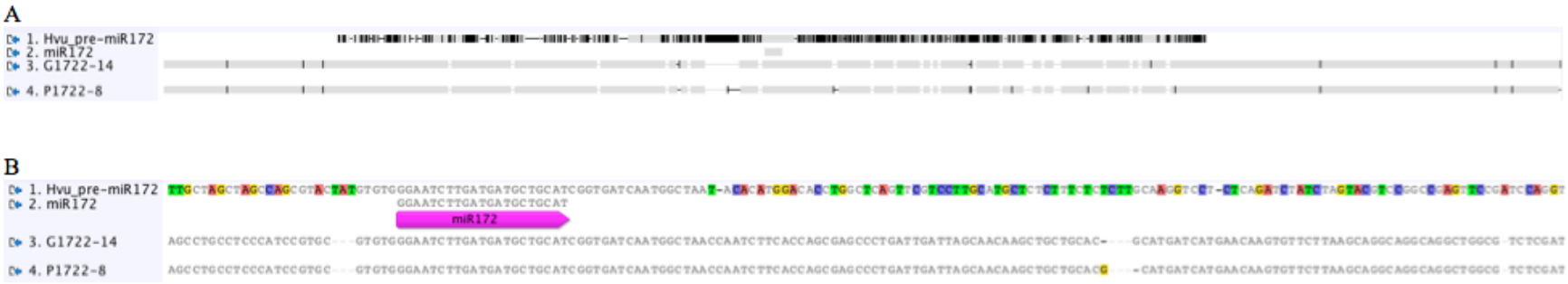


Figure 11: Reference assembly of miR172 genomic clones to the pre-miR172 gene from barley. The full-length alignment is shown in (A) and the conserved region of the pre-miR172 gene is shown in (B).



Figure 12: Predicted secondary structure of pre-miR172. Genomic miR172 cloned from Provena is on the left and from GS7 on the right. The hairpin in the middle is the conserved region of both genes that forms the hairpin of miR172. The blue line indicates the location of the mature miR172 in the hairpin.

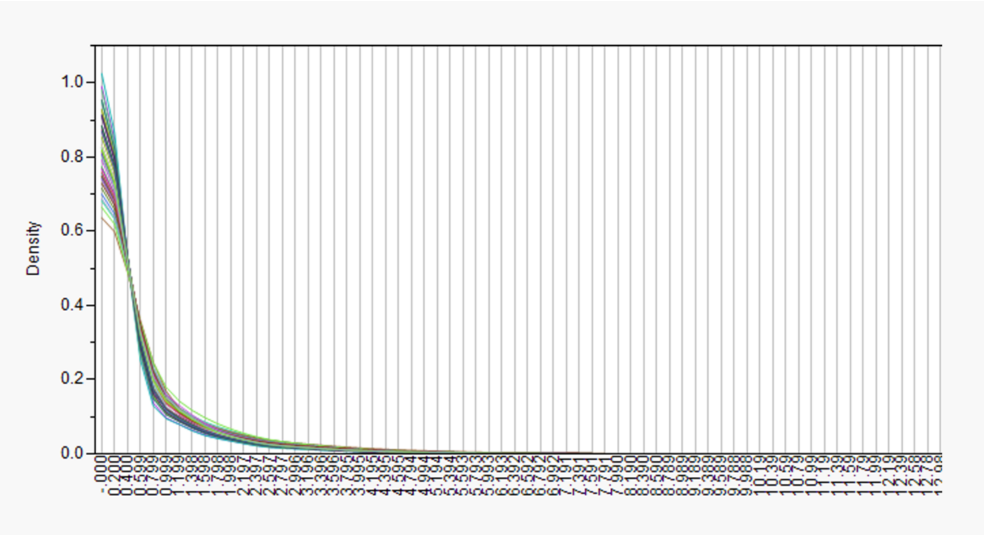


Figure 13: Quality control- log2-transformed data distribution of kernel density estimates.

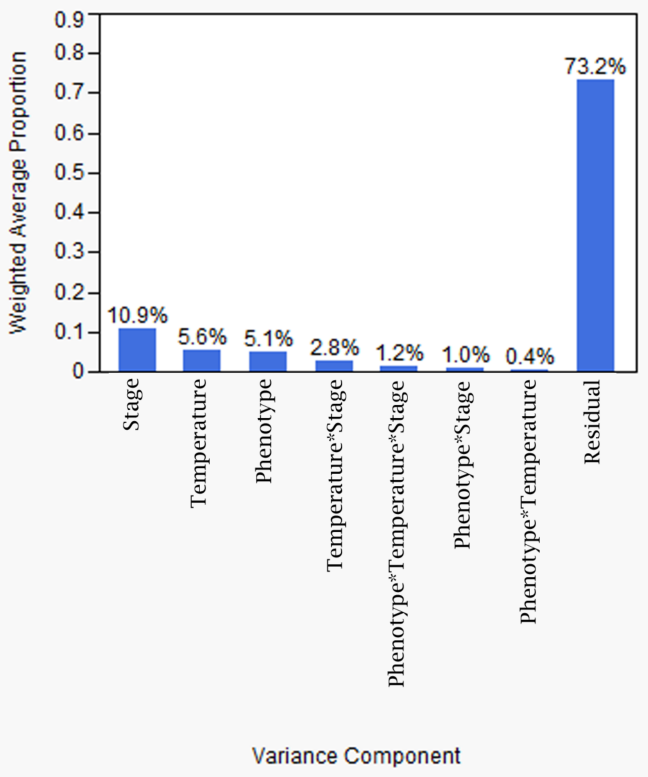


Figure 14: Quality control- log2-transformed data weighted average proportion of variance across principal components.

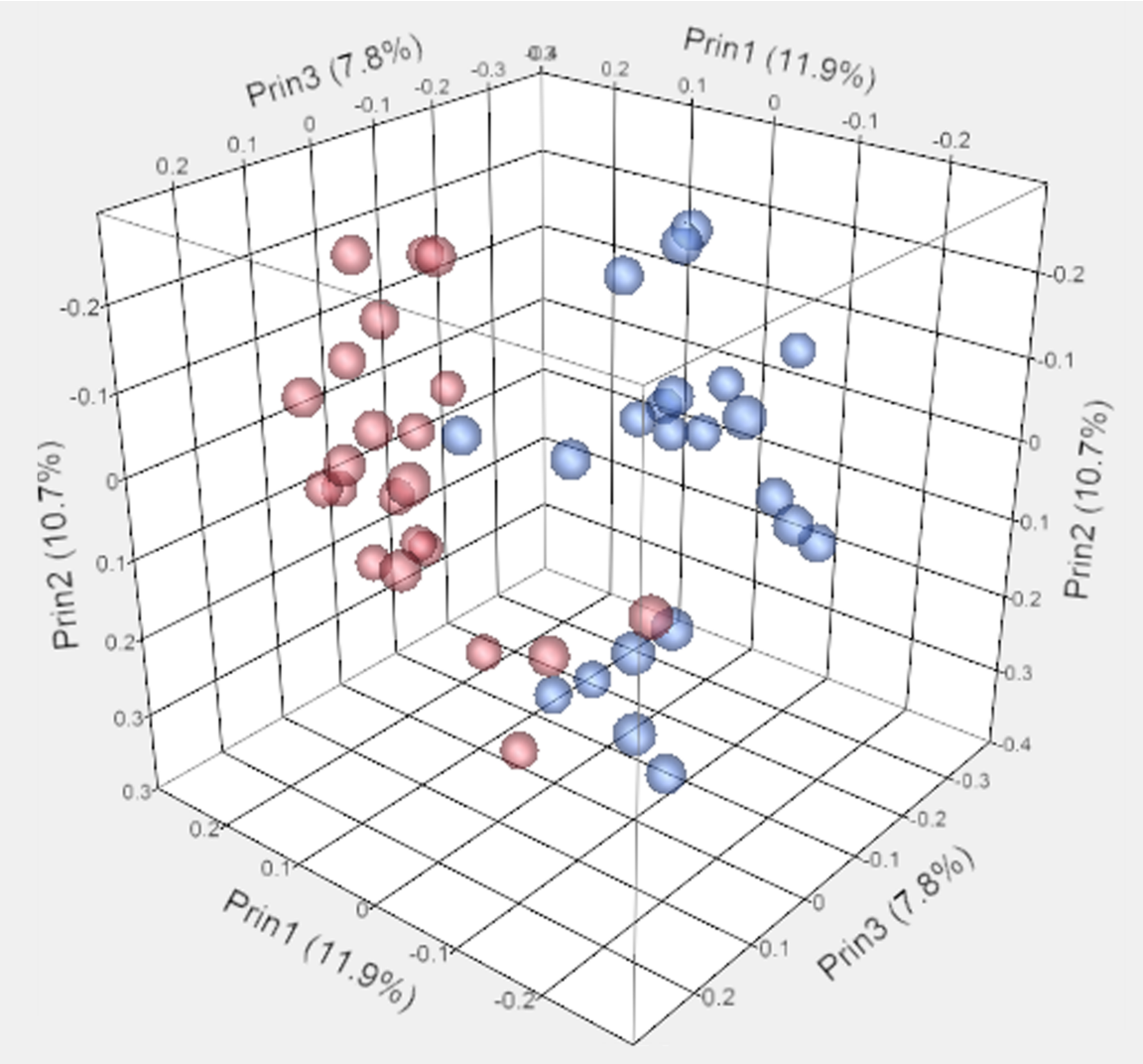


Figure 15: Quality control- log2-transformed data 3D scatterplot of principal components. Points are colored by different stage (pA or 0 DPA is blue; A or 1-3 DPA is red).

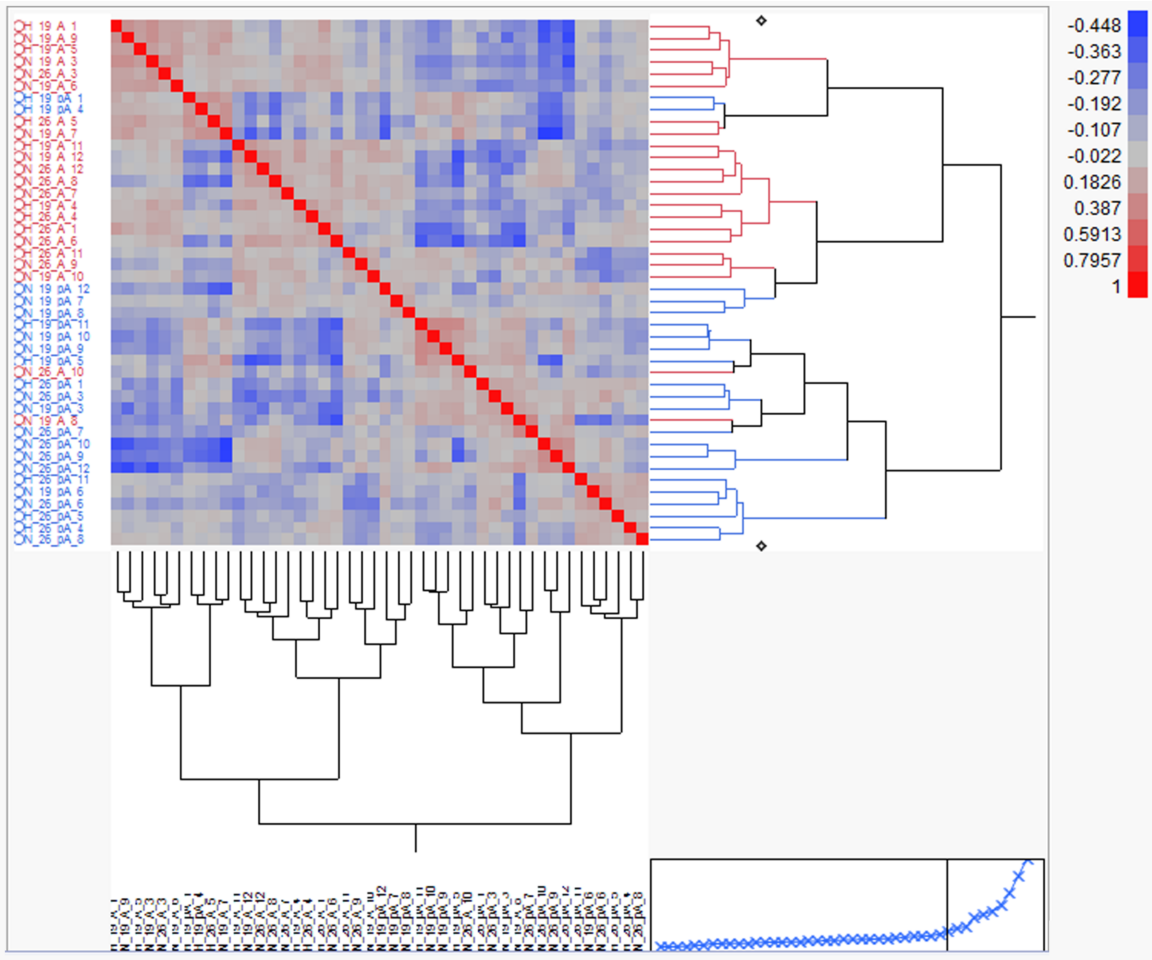


Figure 16: Quality control- log2-transformed data heat map and dendrogram of the hierarchical clustering analysis on correlations.

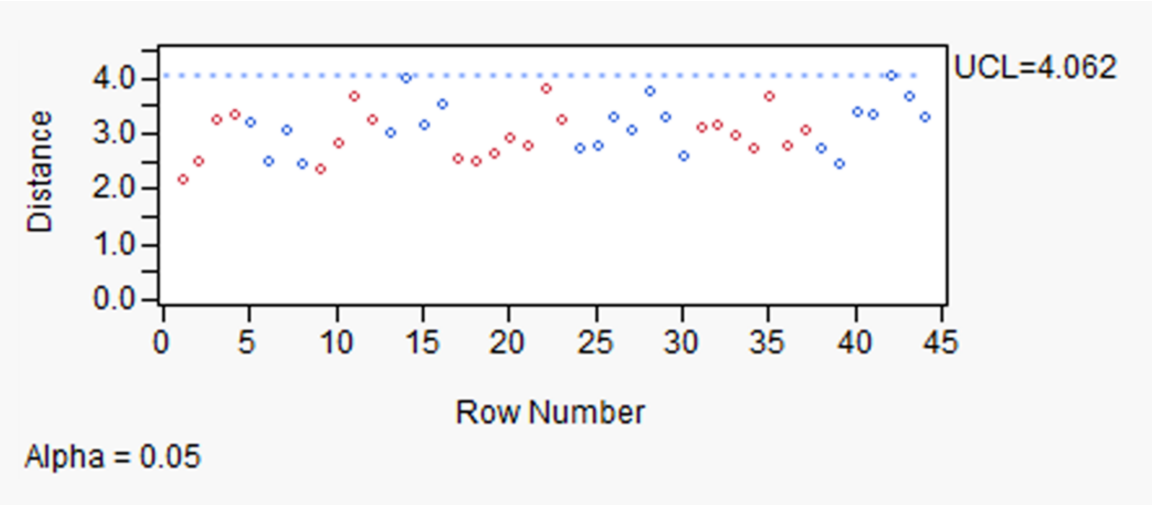


Figure 17: Quality control- log2-transformed data outlier analysis of Mahalanobis Distance from mean.

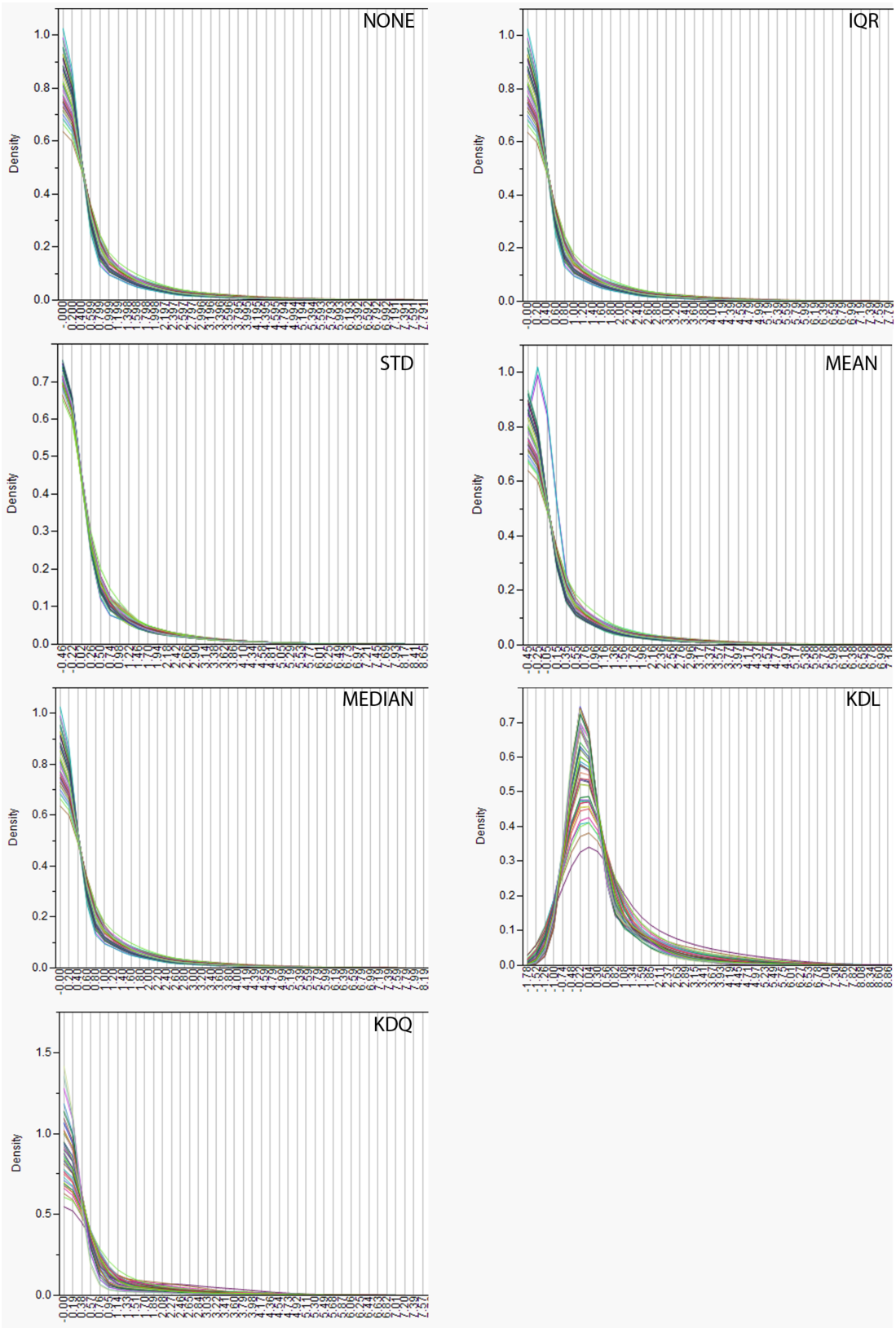


Figure 18: Normalized distributions of log2-transformed oat small RNA counts. IQR, STD, MEAN, MEDIAN, KDL, and KDQ are normalization methods from the Basic miRNA/miRNA-seq Workflow.

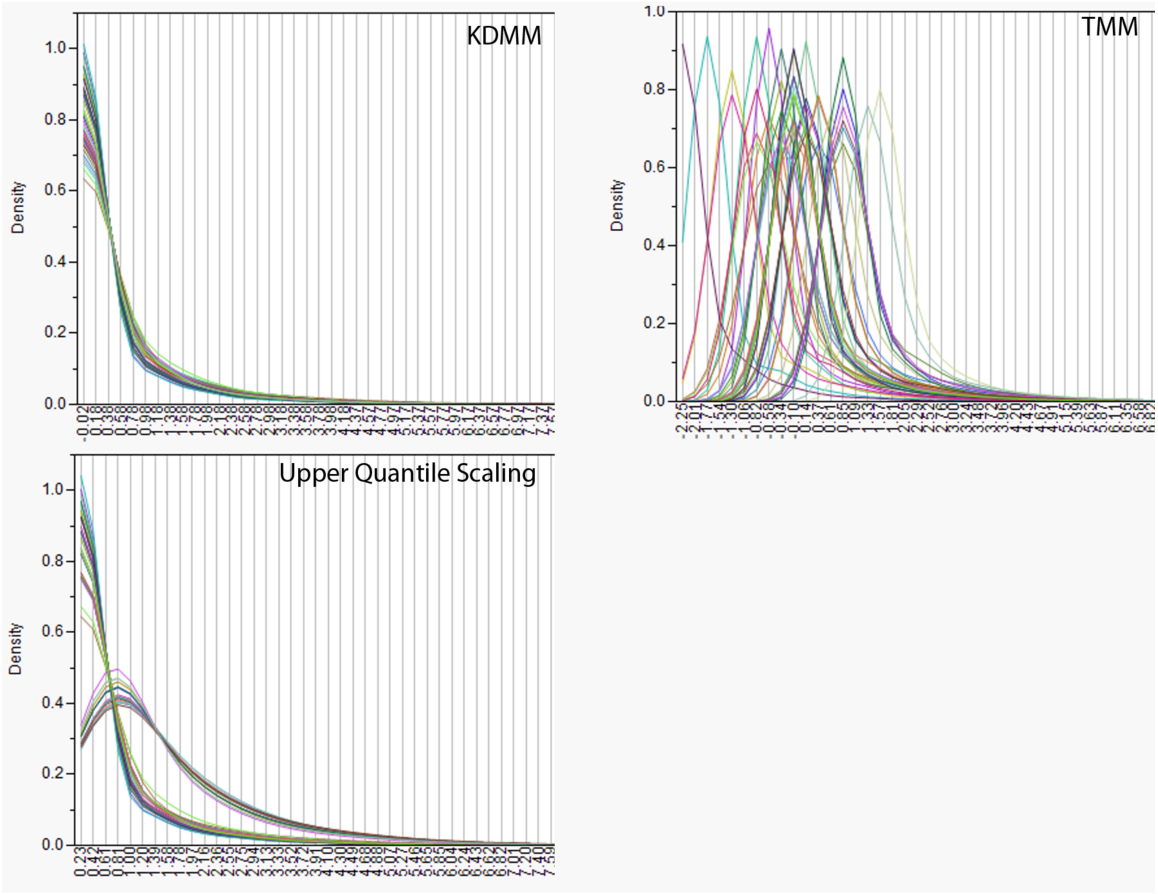


Figure 19: Normalized distributions using ccaling normalization. TMM, KDMM, and Upper Quantile Scaling are scaling methods for Next-generation sequencing data for expression.

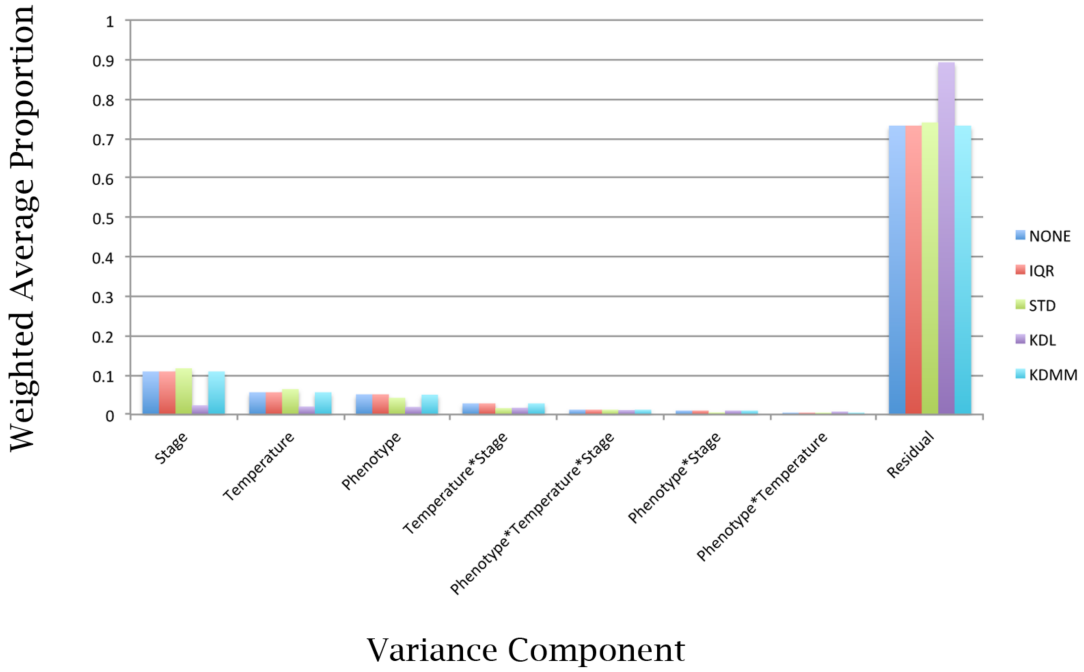


Figure 20: Normalized transformed data weighted average proportion of variance across principal components.

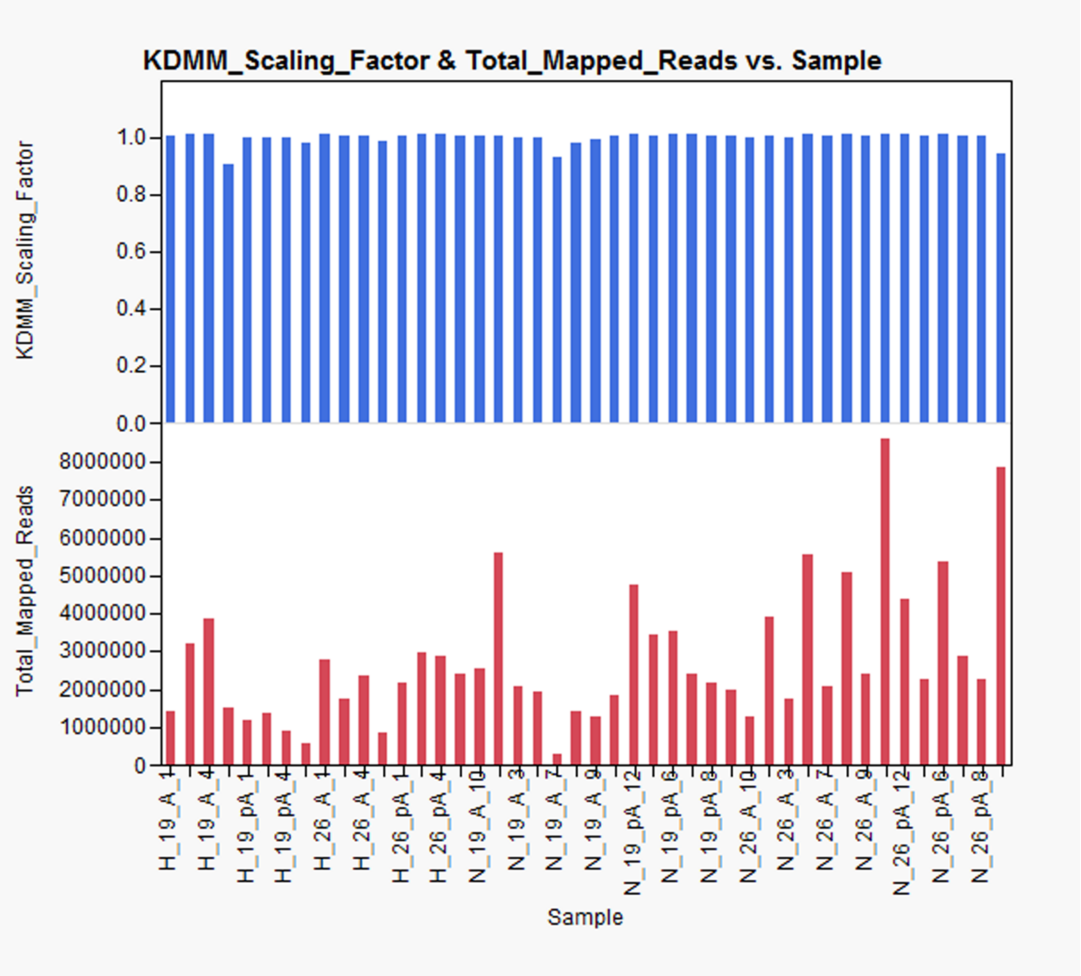


Figure 21: KDMM scaling factors.

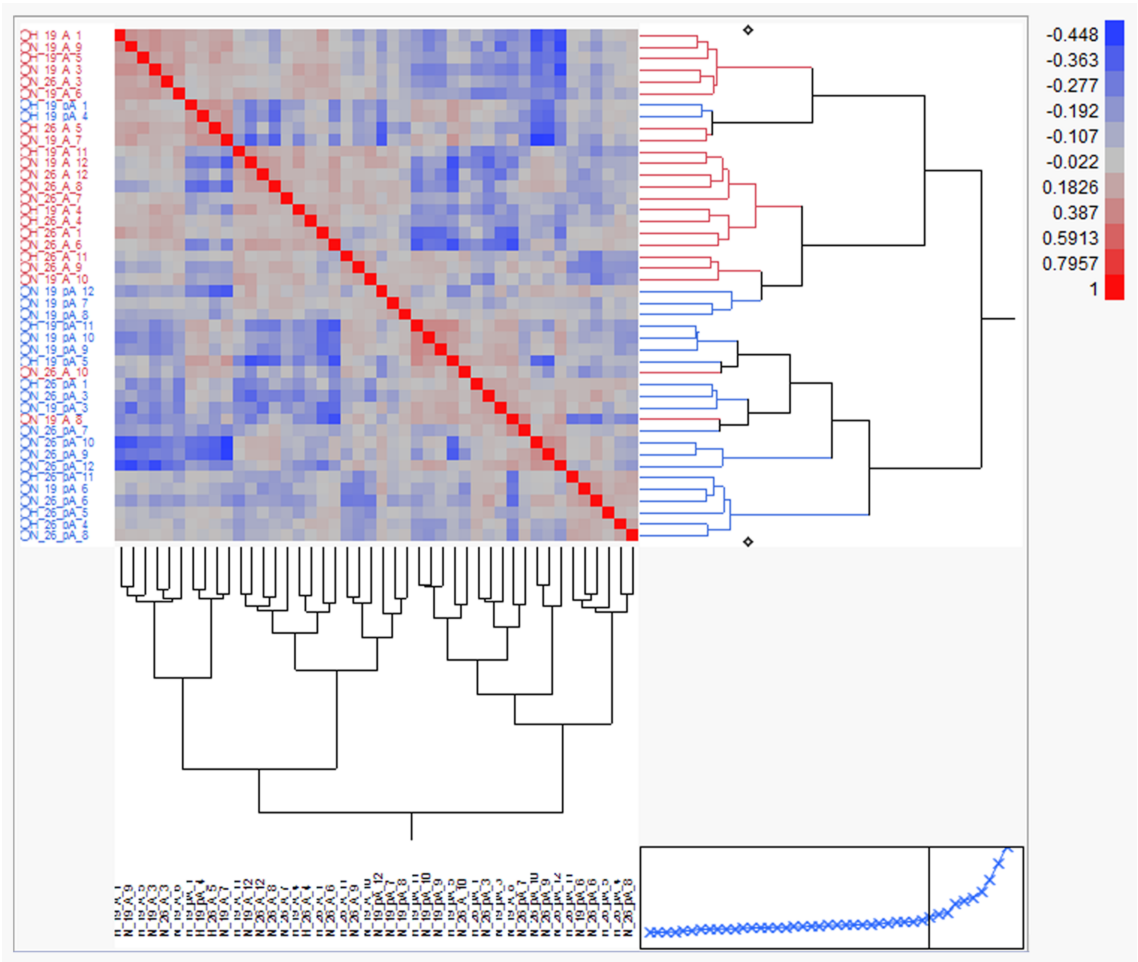


Figure 22: IQR-normalized correlation heat map and dendrogram. Samples collected from stage pA (0 DPA) are blue and A (1-3 DPA) are red.

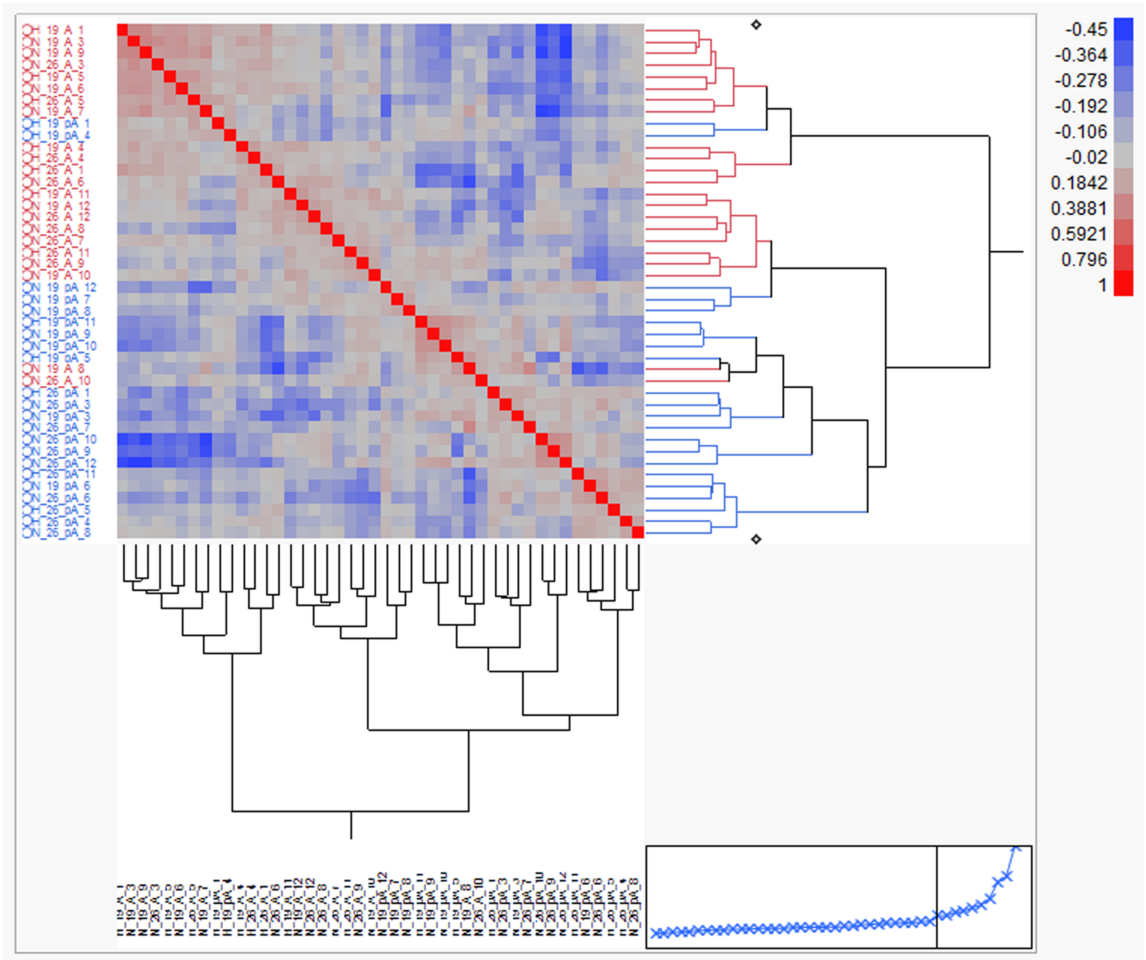


Figure 23: STD-normalized correlation heat map and dendrogram. The heat map shows the correlation of samples and the dendrogram shown the relationship of the samples from hierarchical clustering. Samples collected from stage pA (0 DPA) are blue and A (1-3 DPA) are red.

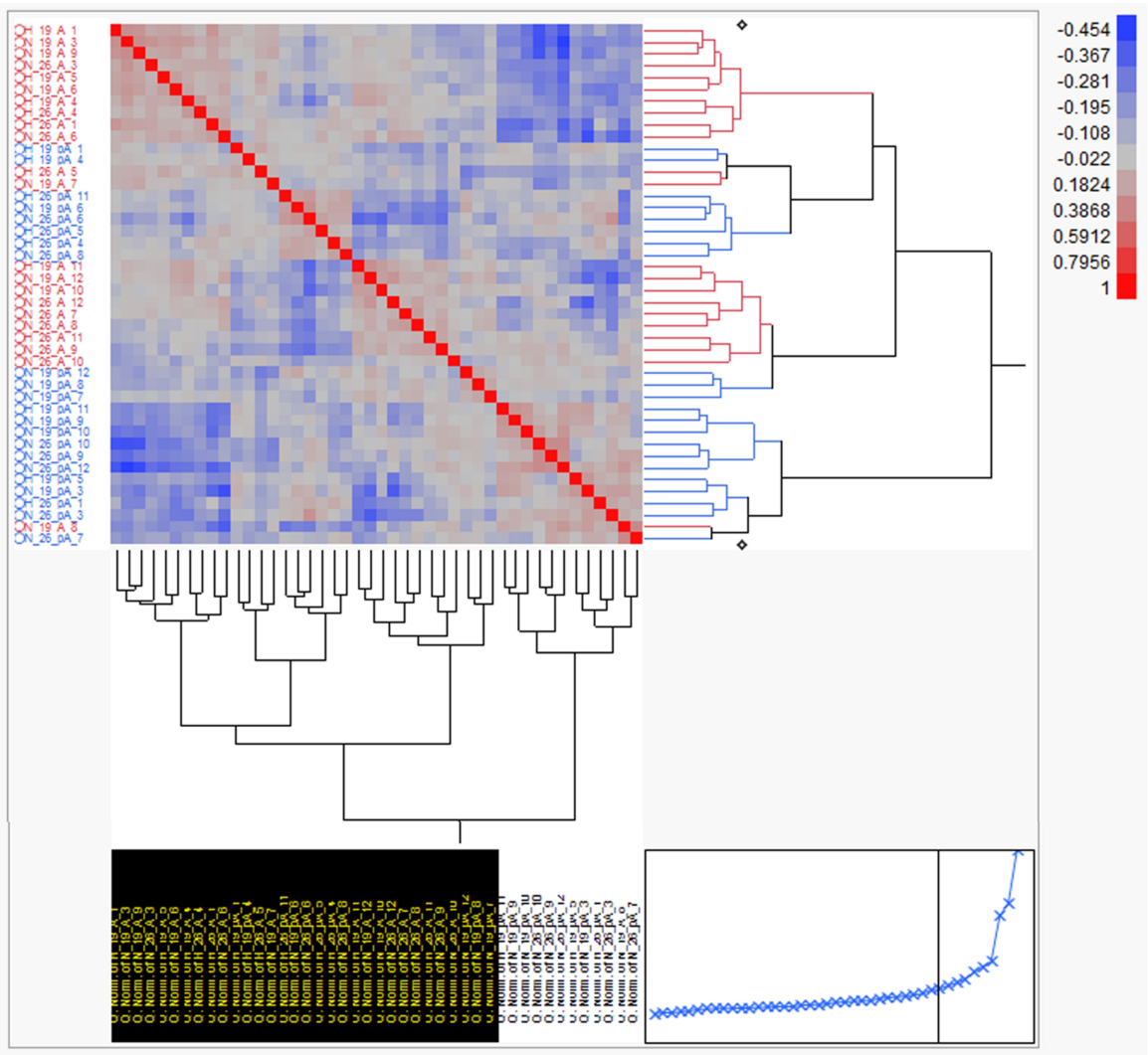


Figure 24: KDQ-normalized correlation analysis. This shows the correlation heat map and hierarchical clustering of the KDQ-normalized, log-transformed read counts. Samples collected from stage pA (0 DPA) are blue and A (1-3 DPA) are red.

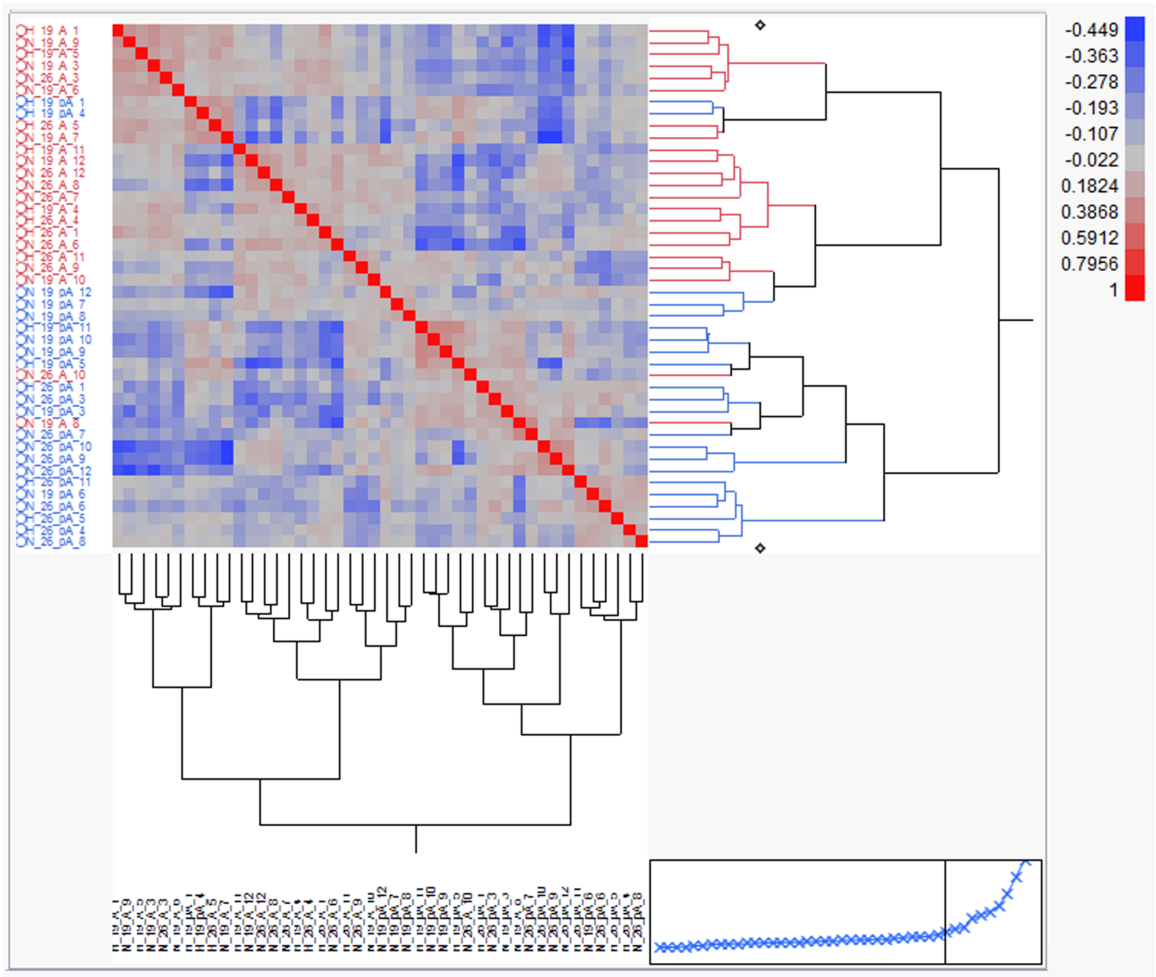


Figure 12: KDMM-normalized correlation analysis. The heat map represents the matrix of correlation coefficients between samples. Samples collected from stage pA (0 DPA) are blue and A (1-3 DPA) are red.

Table 5: Difference chooser of comparisons for JMP ANOVAs and multiple testing

Effect	Difference
Phenotype	(H) - (N)
Temp	(19) - (26)
Phenotype*Temp	(H 19) - (H 26)
Phenotype*Temp	(H 19) - (N 19)
Phenotype*Temp	(H 26) - (N 26)
Phenotype*Temp	(N 19) - (N 26)
Stage	(A) - (pA)
Phenotype*Stage	(H A) - (H pA)
Phenotype*Stage	(H A) - (N A)
Phenotype*Stage	(H pA) - (N pA)
Phenotype*Stage	(N A) - (N pA)
Temp*Stage	(19 A) - (19 pA)
Temp*Stage	(19 A) - (26 A)
Temp*Stage	(19 pA) - (26 pA)
Temp*Stage	(26 A) - (26 pA)
Phenotype*Temp*Stage	(H 19 A) - (N 19 A)
Phenotype*Temp*Stage	(H 19 pA) - (N 19 pA)
Phenotype*Temp*Stage	(H 26 A) - (N 26 A)
Phenotype*Temp*Stage	(H 26 pA) - (N 26 pA)

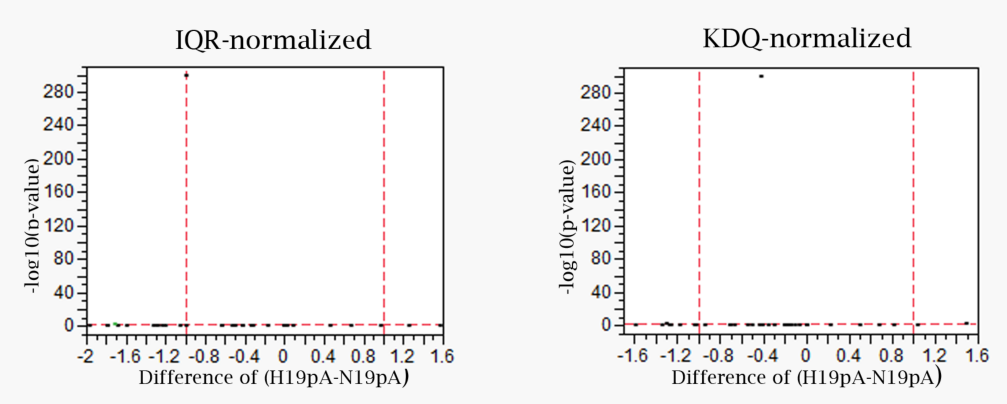


Figure 25: Volcano plot representations of miR172 differential expression. Significance of sequence significance cannot be viewed by volcano plots of IQR (A) or KDQ (B) normalized miR172 expression. The –log10 (p) significance for IQR is set at 2.696 and KDQ at 2.57.

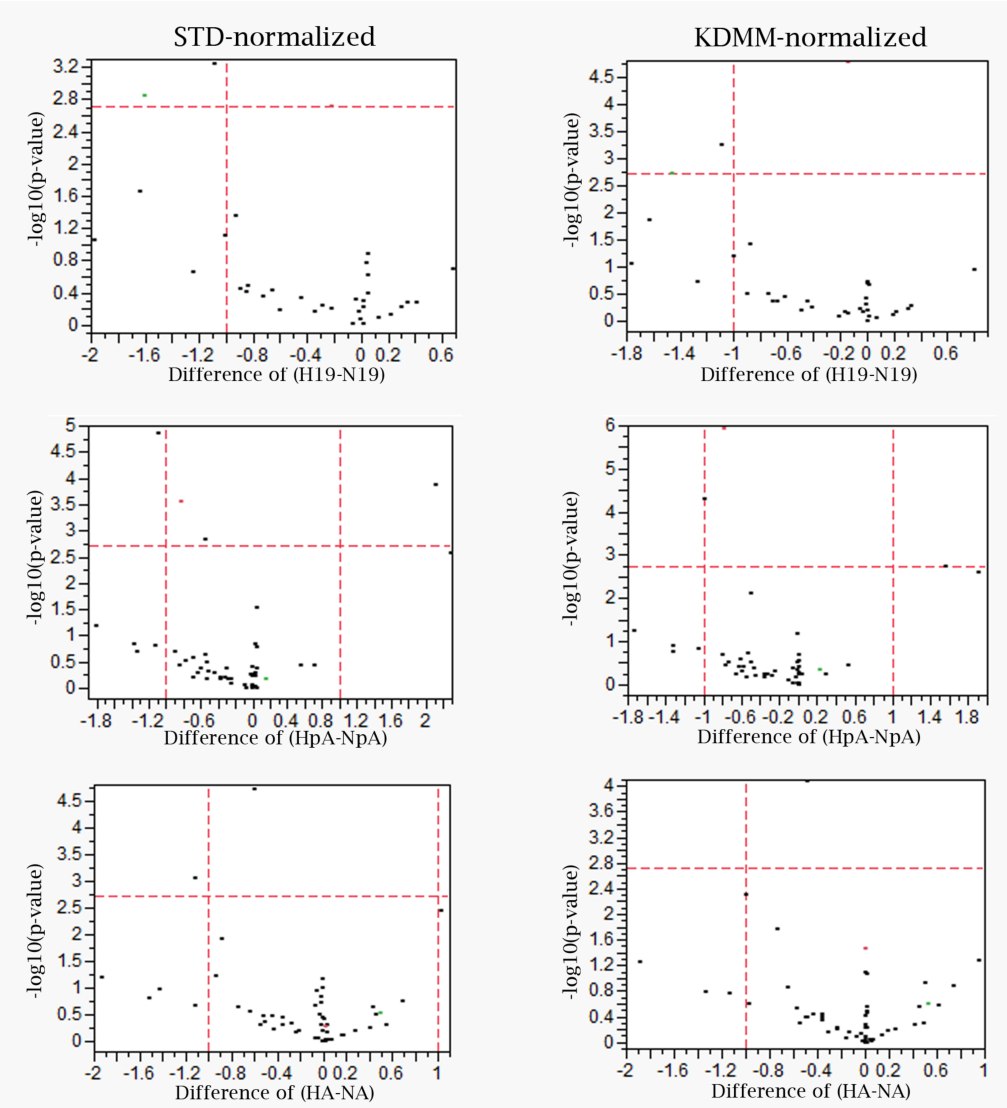


Figure 26: Volcano plot representation of miR172 differential expression of selected phenotype*temperature and phenotype *stage effects of STD- and KDMM-normalized intensities. STD with a –log10 (p) threshold of 2.72 is shown on the left panels and KDMM on the right with a –log10 (p) threshold of 2.727 and green dots were significantly differentially expressed for more than one effect. The sequences are shown in the dendrogram.

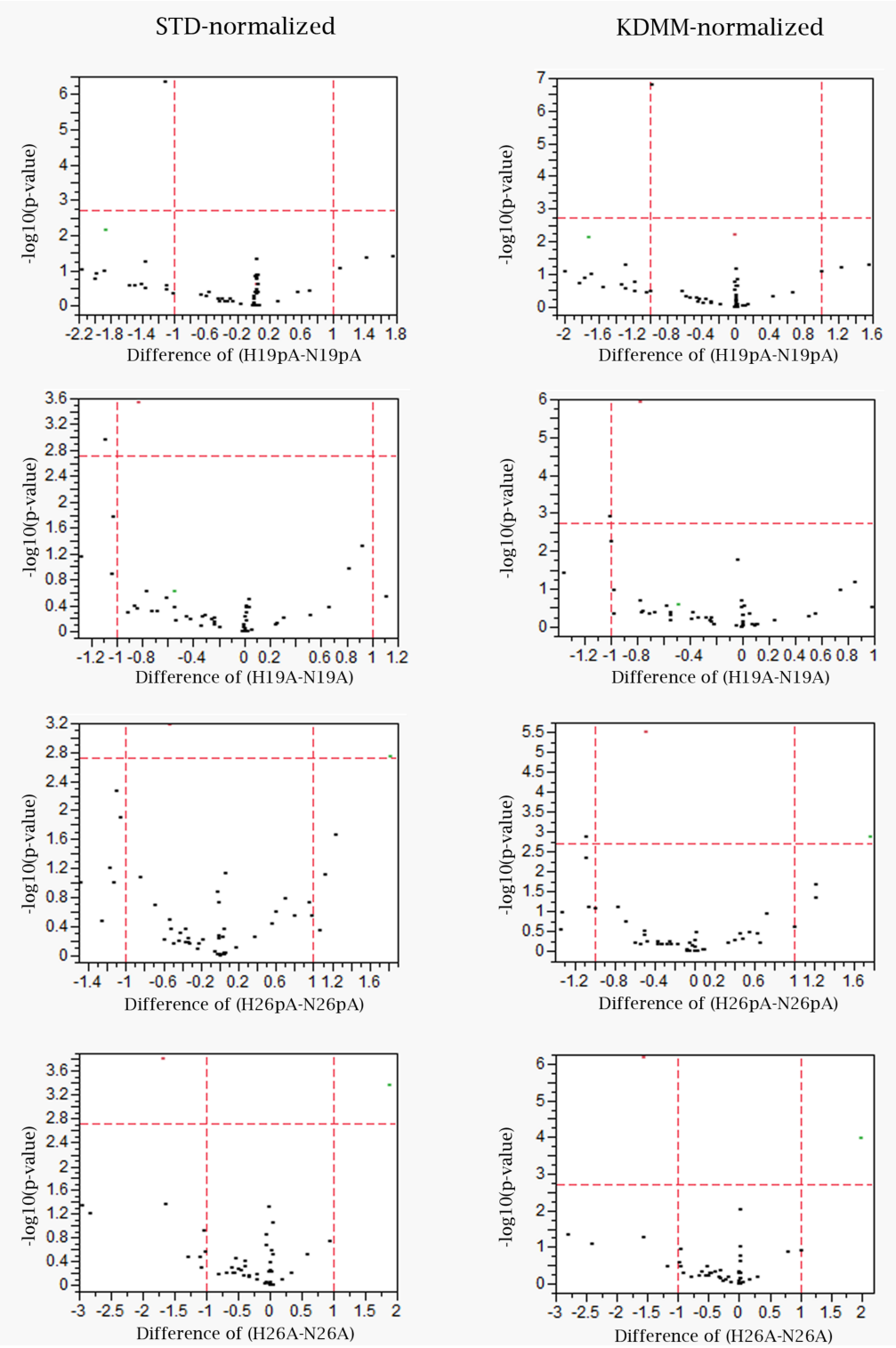


Figure 27: Volcano plot representation of miR172 differential expression of phenotype*temperature*stage effects of STD- and KDMM-normalized intensities. STD with a –log10 (p) threshold of 2.72 is shown on the left panels and KDMM on the right with a –log10 (p) threshold of 2.727 and green dots were significantly differentially expressed for more than one effect. The sequences are shown in the dendrogram.

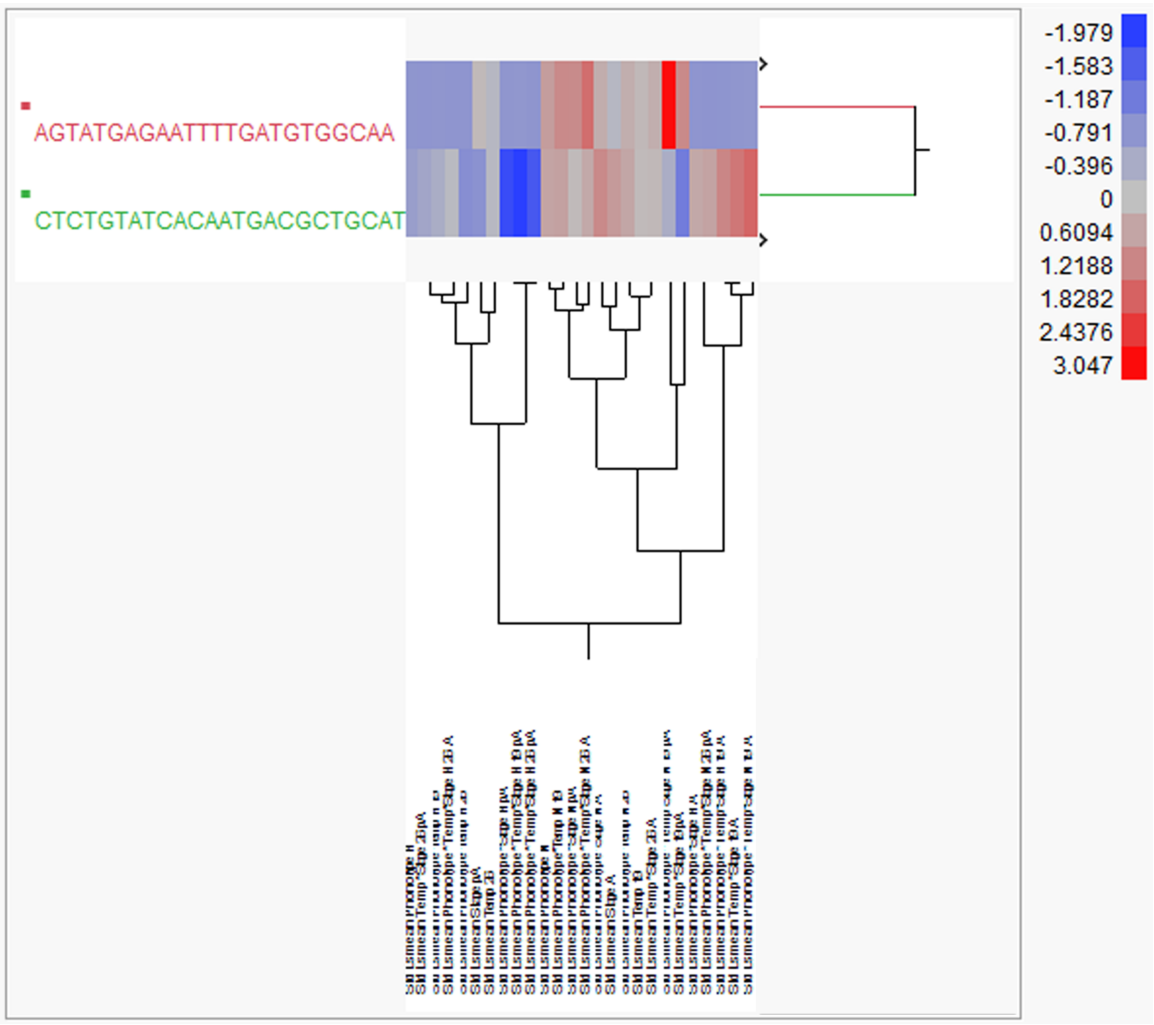


Figure 28: Heat map of clustered standardized LSmeans of KDMM-normalized miR172s. These two sequences were found to be differentially expressed in more than one test in IQR and STD-normalized miR172 expression data.

Table 6: Statistics of multiple testing of significant miR172 expression. A positive Diff of Comparison indicates the sequence is significantly up-regulated in first element of the comparison group. A negative value indicates the sequence is up-regulated in second element of the comparison group.

Effect	Comparison	MiR172 Sequence	Normalization	Residual Variance	Sample Variance	RSquare	Diff of Comparison	t-Statistic	p-Value	-log10 (p-value)
Phenotype*Temperature	(N 19)-(N 26)	TCACTTCTTCAGAATTCAGAC	IQR	0.2125	0.614379085	0.755851064	-1.025	-3.925587449	0.002014673	2.695795432
	(H 19)-(N 19)	CTCTGTATCACAATGACGCTGCAT	IQR	0.920638319	1.510681894	0.49996386	-1.467652639	-3.400117603	0.001822274	2.739386267
			STD	1.050802937	1.604030221	0.46248065	-1.609395907	-3.489935784	0.001431125	2.844322479
			KDQ	0.6959054	1.237771419	0.538687601	-1.25509711	-3.344393792	0.002114648	2.67476186
			KDMM	0.923394374	1.518268014	0.500972875	-1.46514155	-3.389230882	0.001876156	2.726731153
		TTTCCATCCCCGTCAAGATTCAAG	STD	6.32742E-07	0.298992775	0.999999295	-1.088003915	-1116.79003	0.000570044	3.244091521
			KDMM	6.16983E-07	0.299323805	0.999999313	-1.09512089	-1138.360095	0.000559243	3.252399652
Phenotype*Stage	(H A)-(N A)	CTTTAGAACATTGATGATGGGGGC	STD	1.71975E-06	0.249850913	0.999998279	-1.12720672	-701.8190092	0.000907099	3.042345298
	(H A)-(H pA)	CTCTGTATCACAATGACGCTGCAT	IQR	0.920638319	1.510681894	0.49996386	1.24872059	3.951314715	0.000401496	3.396318616
			STD	1.050802937	1.604030221	0.46248065	1.183924176	3.506582995	0.001368131	2.86387241
			KDQ	0.6959054	1.237771419	0.538687601	1.290197905	4.695719014	4.80173E-05	4.318602489
			KDMM	0.923394374	1.518268014	0.500972875	1.25627284	3.969275383	0.000381783	3.418182978
	(H pA)-(N pA)	GAGGATAAGAATATTGATGATGGG	IQR	0	0.418684355	1	1.584962501		0	300
			STD	0.000435334	0.750294579	0.999767913	2.116125043	87.83363659	0.000129597	3.88740616
			KDQ	0	0.305081089	1	1.352954741		0	300
			KDMM	0.003290555	0.409423068	0.996785178	1.551582177	23.42452633	0.001817495	2.740526884
		GAGGATCCAGATTATGCTGACGCT	IQR	0	0.8	1	2		0	300
			KDQ	0	0.601585305	1	1.734337488		0	300

Table 6… (continued): Statistics of multiple testing of significant miR172 expression.

Effect	Comparison	MiR172 Sequence	Normalization	Residual Variance	Sample Variance	RSquare	Diff of Comparison	t-Statistic	p-Value	-log10 (p-value)
Temperature*Stage	(26 A)-(26 pA)	AGTATGAGAATTTTGATGTGGCAA	STD	0.000221498	0.369326214	0.999866725	-1.023337048	-48.62039111	0.000422754	3.373912364
	(19 pA)-(26 pA)	TTTCCATCCCCGTCAAGATTCAAG	STD	6.33E-07	0.298992775	0.999999295	-1.104474172	-981.8095815	0.000648414	3.188147293
			KDMM	6.17E-07	0.299323805	0.999999313	-1.092377505	-983.3791133	0.00064738	3.188841006
Phenotype*Temperature*Stage	(H 26 A)-(N 26 A)	AGTATGAGAATTTTGATGTGGCAA	IQR	0	0.315989224	1	-1.584962501		0	300
			STD	0.000221498	0.369326214	0.999866725	-1.688533792	-80.22495963	0.000155339	3.808720218
			KDMM	8.34E-07	0.313105306	0.999999408	-1.568597732	-1214.202233	6.78E-07	6.168582496
		CTCTGTATCACAATGACGCTGCAT	IQR	0.920638319	1.510681894	0.49996386	1.968968707	4.408575059	0.000109761	3.95955388
			STD	1.050802937	1.604030221	0.46248065	1.874817757	3.929184013	0.000427158	3.369411848
			KDQ	0.6959054	1.237771419	0.538687601	1.874567417	4.827584785	3.28E-05	4.484397761
			KDMM	0.923394374	1.518268014	0.500972875	1.983223119	4.433859359	0.000102085	3.991036441
	(H 26 pA)-(N 26 pA)	CTCTGTATCACAATGACGCTGCAT	IQR	0.920638319	1.510681894	0.49996386	1.751388889	3.507413823	0.001365058	2.864848911
			STD	1.050802937	1.604030221	0.46248065	1.810748129	3.394269711	0.001851029	2.732586708
			KDQ	0.6959054	1.237771419	0.538687601	1.743619187	4.016293836	0.000334553	3.475534576
			KDMM	0.923394374	1.518268014	0.500972875	1.766591565	3.532575746	0.001275092	2.894458417
		AATCCTGGTTTTGCTGCAGAACA	KDMM	2.59E-06	0.236573938	0.999997258	-1.099584343	-482.7283263	0.001318793	2.879823282
	(H 19 A)-(N 19 A)	CTTTAGAACATTGATGATGGGGGC	STD	1.72E-06	0.249850913	0.999998279	-1.094769167	-590.3026694	0.001078462	2.96719504
		AGAATCTTGATGATGCTGCA	KDMM	1.90E-06	0.30233269	0.999998431	-1.006010637	-516.3452488	0.001232933	2.909060605
	(H 19 pA)-(N 19 pA)	AAAACATGAGAAACTTGCTGCTGC	STD	0.000166895	0.134755067	0.999449554	-1.10900067	-60.70094599	4.41E-07	6.355415961

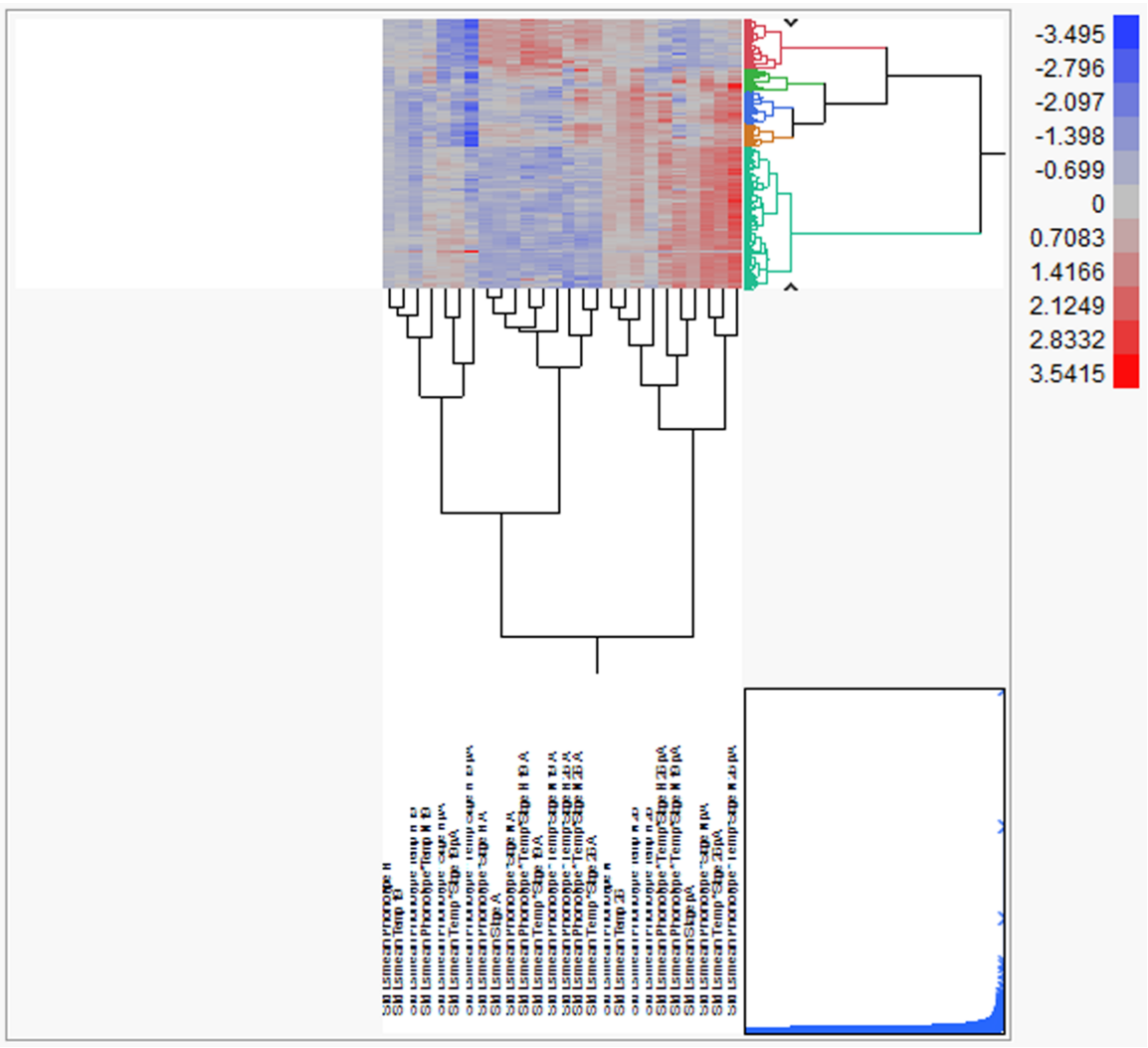


Figure 29: Heat map and dendrogram of clustered standardized LSmeans of IQR-normalized filtered dataset. Sequences missing in more than 28 of the 44 sample files were filtered out.

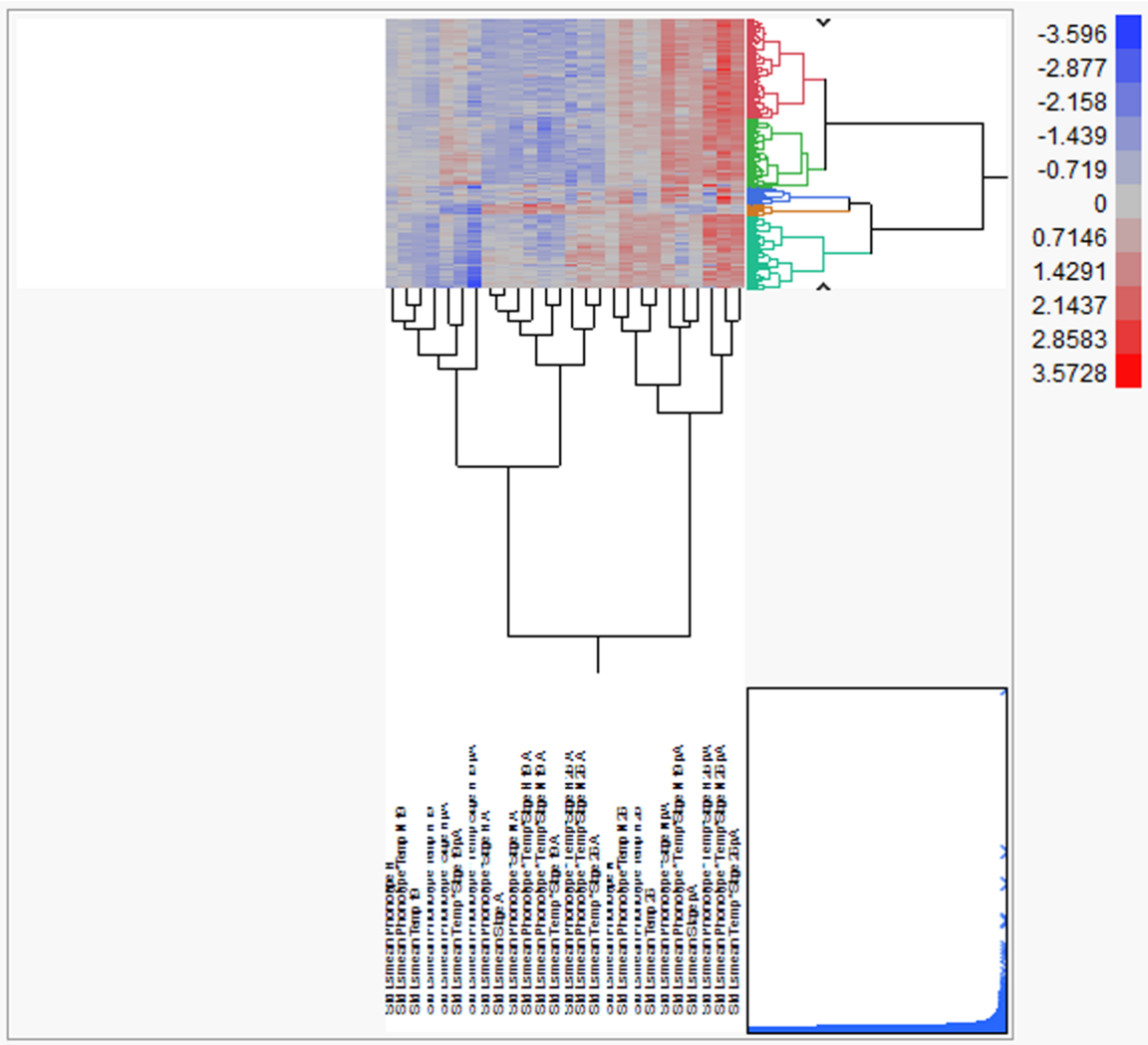


Figure 30: Heat map and dendrogram of clustered standardized LSmeans of STD-normalized filtered dataset. Sequences missing in more than 28 of the 44 sample files were filtered out.

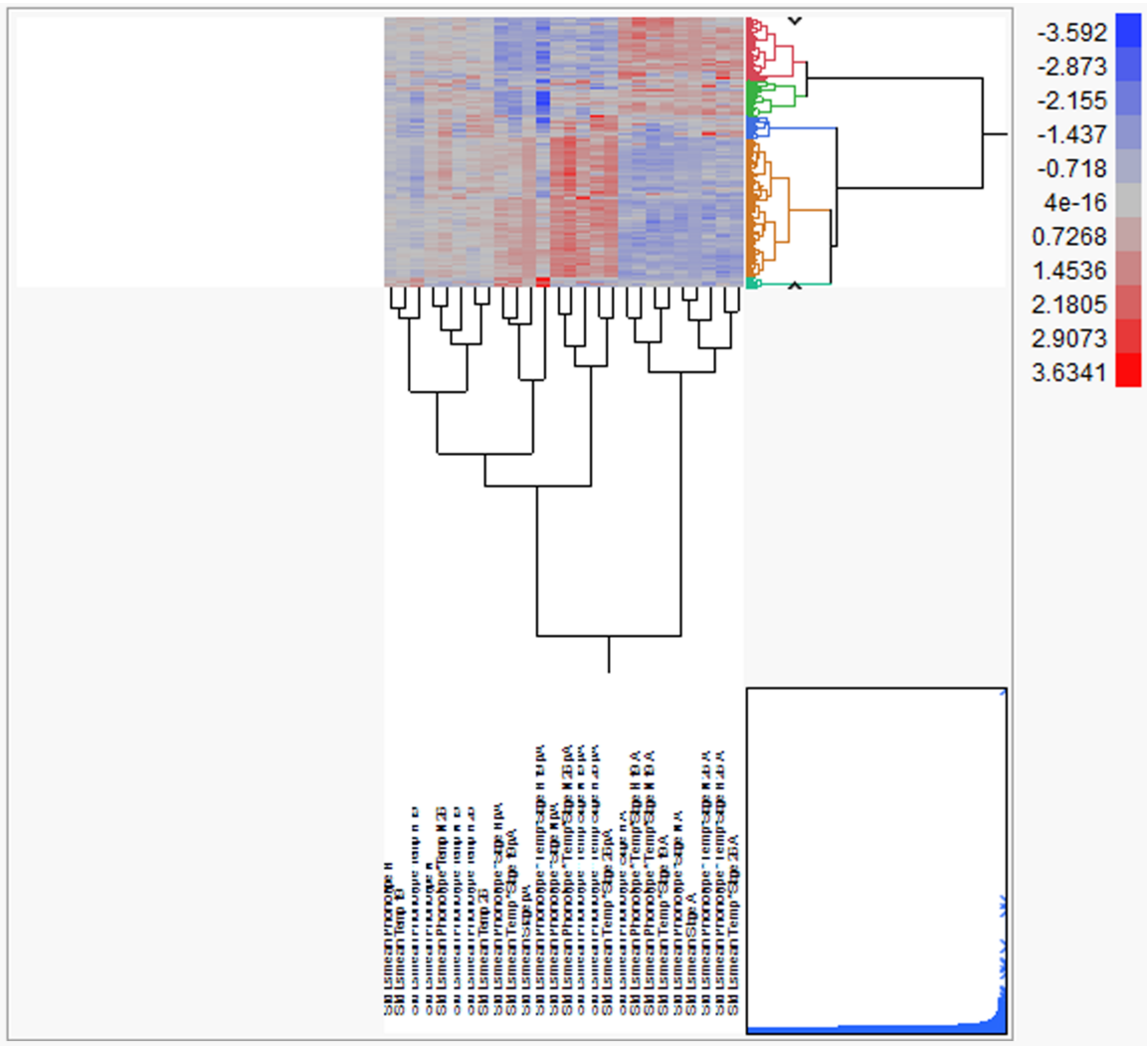


Figure 31: Heat map and dendrogram of clustered standardized LSmeans of KDQ-normalized filtered dataset. Sequences missing in more than 28 of the 44 sample files were filtered out.

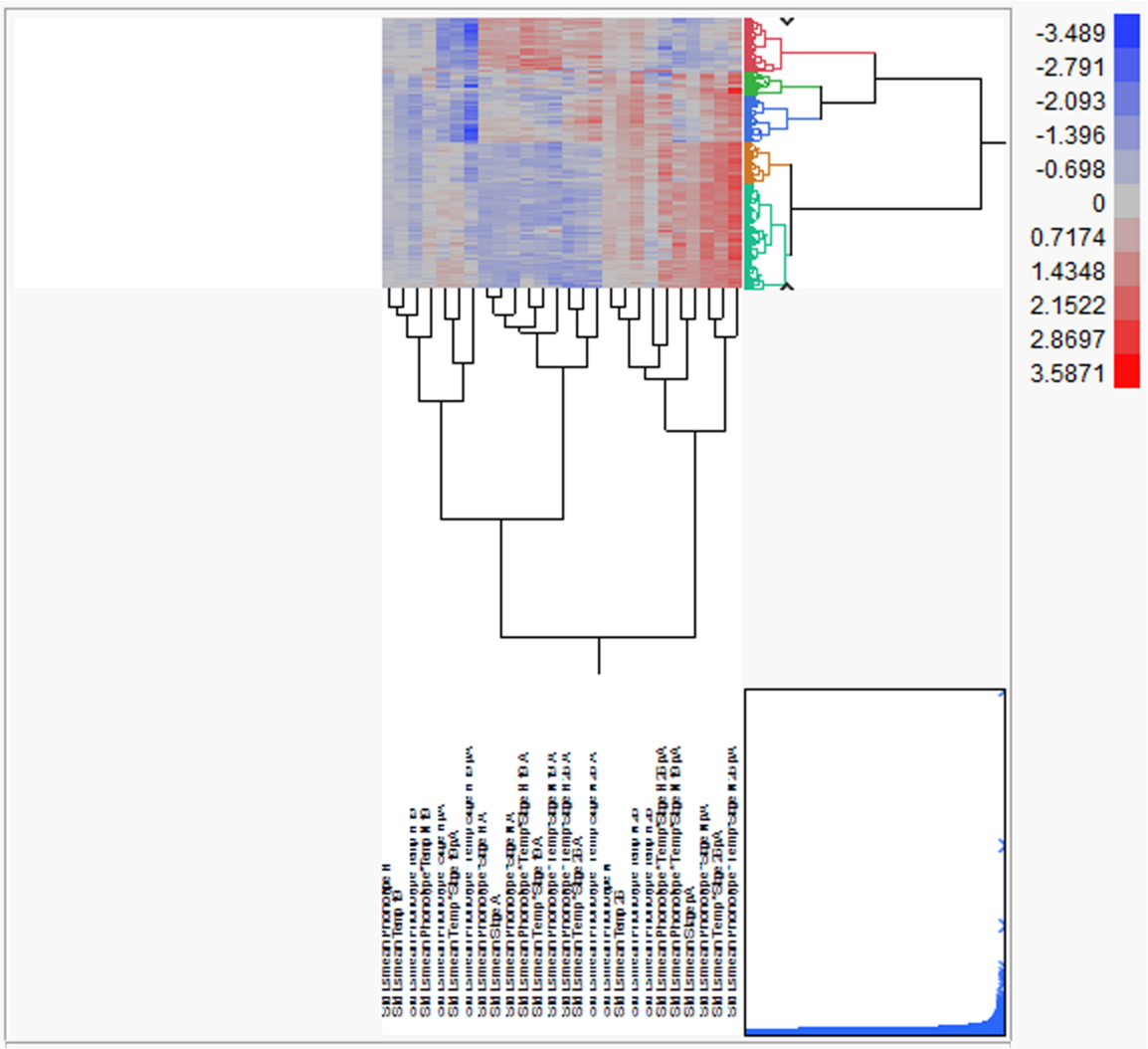


Figure 32: Heat map and dendrogram of clustered standardized LSmeans of KDMM-normalized filtered dataset. Sequences missing in more than 28 of the 44 sample files were filtered out.

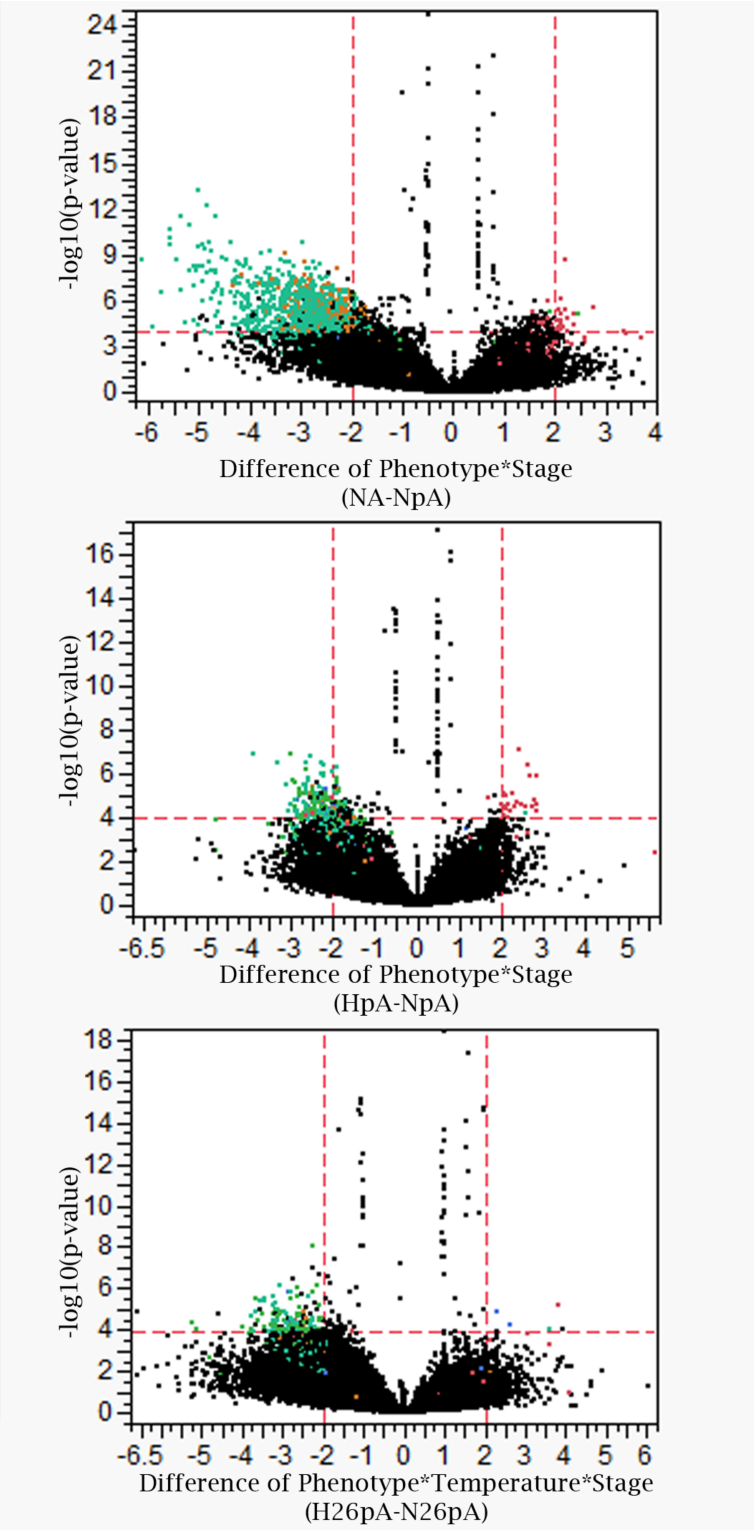


Figure 33: Select volcano plots from KDMM normalization of dataset filtered for missing data. The green clusters correspond to the green clusters of LSmeans in the KDMM Dendrogram.

Supplementary Table 1: Annotated miR172 sequences in miRBase v. 20.

Annotated miRNA Name	miRBase Identifier	Sequence
aau-miR172	MIMAT0022673	TGAGAATCTTGATGATGCTGCAT
aly-miR172a-3p	MIMAT0017522	AGAATCTTGATGATGCTGCAT
aly-miR172a-5p	MIMAT0017521	GTGGCATCATCAAGATTCACA
aly-miR172b-3p	MIMAT0017524	AGAATCTTGATGATGCTGCAT
aly-miR172b-5p	MIMAT0017523	GCAGCACCATCAAGATTCACA
aly-miR172c-3p	MIMAT0017526	AGAATCTTGATGATGCTGCAG
aly-miR172c-5p	MIMAT0017525	GGAGCATCATCAAGATTCACA
aly-miR172d-3p	MIMAT0017528	AGAATCTTGATGATGCTGCAG
aly-miR172d-5p	MIMAT0017527	GCAACATCTTCAAGATTCAGA
aly-miR172e-3p	MIMAT0017530	GAATCTTGATGATGCTGCAT
aly-miR172e-5p	MIMAT0017529	GCAGCACCATTAAGATTCAC
aly-miR172f-3p	MIMAT0031381	AGAATCTTGATGATGCTGCAG
aly-miR172f-5p	MIMAT0031380	GGAGCATCATCAAGATTCACA
aqc-miR172a	MIMAT0012571	AGAATCTTGATGATGCTGCAT
aqc-miR172b	MIMAT0012572	GGAATCTTGATGATGCTGCAT
ata-miR172	MIMAT0018221	TGAGAATCTTGATGATGCTGCAT
ath-miR172a	MIMAT0000203	AGAATCTTGATGATGCTGCAT
ath-miR172b-3p	MIMAT0000205	AGAATCTTGATGATGCTGCAT
ath-miR172b-5p	MIMAT0000204	GCAGCACCATTAAGATTCAC
ath-miR172c	MIMAT0000922	AGAATCTTGATGATGCTGCAG
ath-miR172d	MIMAT0000923	AGAATCTTGATGATGCTGCAG
ath-miR172e	MIMAT0001019	GGAATCTTGATGATGCTGCAT
bdi-miR172a-3p	MIMAT0020685	AGAATCTTGATGATGCTGCAT
bdi-miR172a-5p	MIMAT0027061	GCAGCACCACCAAGATTCACA
bdi-miR172b	MIMAT0020696	GGAATCTTGATGATGCTGCAT
bdi-miR172d	MIMAT0012182	AGAATCCTGATGATGCTGCAG
bna-miR172a	MIMAT0023633	AGAATCTTGATGATGCTGCAT
bna-miR172b	MIMAT0023631	GGAATCTTGATGATGCTGCAT
bna-miR172c	MIMAT0023632	GGAATCTTGATGATGCTGCAT
bna-miR172d	MIMAT0023630	AGAATCTTGATGATGCTGCAG
bol-miR172a	MIMAT0010171	AGAATCTTGATGATGCTGCAT
bol-miR172b	MIMAT0010172	AGAATCTTGATGATGCTGCAT
bra-miR172a	MIMAT0010166	AGAATCTTGATGATGCTGCAT
bra-miR172b-3p	MIMAT0010168	AGAATCTTGATGATGCTGCAT
bra-miR172b-5p	MIMAT0010167	GCAGCACCATTAAGATTCACA
cca-miR172	MIMAT0024521	AGAATCTTGATGATGCTGCAT
cme-miR172a	MIMAT0026181	GGAATCTTGATGATGCTGCAG
cme-miR172b	MIMAT0026124	AGAATCTTGATGATGCTGCAT

Supplementary Table 1 … (continued): Annotated miR172 sequences in miRBase v. 20

cme-miR172c	MIMAT0026125	AGAATCTTGATGATGCTGCAT
cme-miR172d	MIMAT0026127	GGAATCTTGATGATGCTGCAT
cme-miR172e	MIMAT0026126	AGAATCTTGATGATGCTGCAG
cme-miR172f	MIMAT0026099	TGAATCTTGATGATGCCGCAC
cpa-miR172a	MIMAT0031810	GGGAATCTTGATGATGCTGCA
cpa-miR172b	MIMAT0031811	GGGAATCTTGATGATGCTGCA
csi-miR172a-3p	MIMAT0014076	AGAATCTTGATGATGCTGCA
csi-miR172a-5p	MIMAT0017387	GCAGCGTCCTCAAGATTCACA
csi-miR172b	MIMAT0018457	AGAATCTTGATGATGCGGCAA
csi-miR172c	MIMAT0018458	TGGAATCTTGATGATGCTGCAG
dpr-miR172a	MIMAT0023534	AGAATCTTGATGATGCTGCAT
dpr-miR172b	MIMAT0023535	AGAATCTTGATGATGCTGCAT
egu-miR172a	MIMAT0024020	AGAATCTTGATGATGCTGCAT
egu-miR172b	MIMAT0024021	AGAATCTTGATGATGCTGCAT
egu-miR172c	MIMAT0024022	AGAATCTTGATGATGCTGCAT
egu-miR172d	MIMAT0024023	TGAATCTTGATGATGCTGCAC
egu-miR172e	MIMAT0024024	TGAATCTTGATGATGCTGCAC
egu-miR172f	MIMAT0024025	ACTGGACTTGGAGTCAGG
ghr-miR172	MIMAT0014333	AGAATCCTGATGATGCTGCAG
gma-miR172a	MIMAT0001682	AGAATCTTGATGATGCTGCAT
gma-miR172b-3p	MIMAT0001683	AGAATCTTGATGATGCTGCAT
gma-miR172b-5p	MIMAT0020921	GTAGCATCATCAAGATTCAC
gma-miR172c	MIMAT0011198	GGAATCTTGATGATGCTGCAG
gma-miR172d	MIMAT0011199	GGAATCTTGATGATGCTGCAGCAG
gma-miR172e	MIMAT0011200	GGAATCTTGATGATGCTGCAGCAG
gma-miR172f	MIMAT0018333	AGAATCTTGATGATGCTGCA
gma-miR172g	MIMAT0021656	GCAGCACCATCAAGATTCAC
gma-miR172h-3p	MIMAT0021658	AGAATCTTGATGATGCTGCAT
gma-miR172h-5p	MIMAT0021657	GCAGCAGCATCAAGATTCACA
gma-miR172i-3p	MIMAT0022996	GGAATCTTGATGATGCTGCAT
gma-miR172i-5p	MIMAT0021659	GCAGCAGCATCAAGATTCACA
gma-miR172j	MIMAT0021660	GCAGCAGCATCAAGATTCACA
gma-miR172k	MIMAT0023201	TGAATCTTGATGATGCTGCAT
gma-miR172l	MIMAT0023205	GGAATCTTGATGATGCTGCAT
lja-miR172a	MIMAT0029319	AGAATCTTGATGATGCTGCAG
lja-miR172b	MIMAT0029320	AGAATCTTGATGATGCTGCA
lja-miR172c	MIMAT0029321	AGAATCTTGATGATGCTGCA
lus-miR172a	MIMAT0027135	AGAATCTTGATGATGCTGCAT
lus-miR172b	MIMAT0027143	AGAATCTTGATGATGCTGCAT

Supplementary Table 1 … (continued): Annotated miR172 sequences in miRBase v. 20

lus-miR172c	MIMAT0027144	AGAATCTTGATGATGCTGCAT
lus-miR172d	MIMAT0027164	AGAATCTTGATGATGCTGCAT
lus-miR172e	MIMAT0027174	GGAATCTTGATGATGCTGCAG
lus-miR172f	MIMAT0027205	AGAATCTTGATGATGCTGCAT
lus-miR172g	MIMAT0027226	GGAATCTTGATGATGCTGCAT
lus-miR172h	MIMAT0027228	AGAATCTTGATGATGCTGCAT
lus-miR172i	MIMAT0027230	GGAATCTTGATGATGCTGCAG
lus-miR172j	MIMAT0027234	GCAGCATCATCAAGATTCCCA
mdm-miR172a	MIMAT0025952	AGAATCTTGATGATGCTGCA
mdm-miR172b	MIMAT0025953	AGAATCTTGATGATGCTGCA
mdm-miR172c	MIMAT0025954	AGAATCTTGATGATGCTGCA
mdm-miR172d	MIMAT0025955	AGAATCTTGATGATGCTGCAT
mdm-miR172e	MIMAT0025956	AGAATCTTGATGATGCTGCAT
mdm-miR172f	MIMAT0025957	AGAATCTTGATGATGCTGCAT
mdm-miR172g	MIMAT0025958	AGAATCTTGATGATGCTGCAT
mdm-miR172h	MIMAT0025959	AGAATCTTGATGATGCTGCAT
mdm-miR172i	MIMAT0025960	GGAATCTTGATGATGCTGCAT
mdm-miR172j	MIMAT0025961	GGAATCTTGATGATGCTGCAT
mdm-miR172k	MIMAT0025962	GGAATCTTGATGATGCTGCAT
mdm-miR172l	MIMAT0025963	GGAATCTTGATGATGCTGCAG
mdm-miR172m	MIMAT0025964	AGAATCTTGATGATGCTGCAG
mdm-miR172n	MIMAT0025965	AGAATCTTGATGATGCTGCAG
mdm-miR172o	MIMAT0025966	AGAATCTTGATGATGCTGCAG
mes-miR172a	MIMAT0024416	AGAATCTTGATGATGCTGCAT
mes-miR172b	MIMAT0029248	AGAATCTTGATGATGCTGCAT
mes-miR172c	MIMAT0029249	TGAATCTTGATGATGCTACGC
mes-miR172d	MIMAT0029250	AGAATCTTGATGATGCTGCAT
mes-miR172e	MIMAT0029251	GGAATCTTGATGATGCTGCAG
mes-miR172f	MIMAT0029252	GGAATCTTGATGATGCTGCAG
mtr-miR172a	MIMAT0011086	AGAATCCTGATGATGCTGCAG
mtr-miR172b	MIMAT0021264	AGAATCTTGATGATGCTGCAT
mtr-miR172c-3p	MIMAT0021266	AGAATCTTGATGATGCTGCAT
mtr-miR172c-5p	MIMAT0021265	GTAGCATCATCAAGATTCACA
mtr-miR172d-3p	MIMAT0029979	AGAATCTTGATGATGCTGCA
mtr-miR172d-5p	MIMAT0029978	AGTGGAGCATCATCAAGATTCACA
nta-miR172a	MIMAT0024689	AGAATCTTGATGATGCTGCAG
nta-miR172b	MIMAT0024690	AGAATCATGATGATGCTGCAT
nta-miR172c	MIMAT0024691	AGAATCTTGATGATGCTGCAT
nta-miR172d	MIMAT0024692	AGAATCTTGATGATGCTGCAT

Supplementary Table 1 … (continued): Annotated miR172 sequences in miRBase v. 20

nta-miR172e	MIMAT0024693	AGAATCTTGATGATGCTGCAT
nta-miR172f	MIMAT0024694	AGAATCTTGATGATGCTGCAT
nta-miR172g	MIMAT0024695	AGAATCTTGATGATGCTGCAT
nta-miR172h	MIMAT0024696	AGAATCTTGATGATGCTGCAT
nta-miR172i	MIMAT0024697	AGAATCTTGATGATGCTGCAT
nta-miR172j	MIMAT0024698	GGAATCTTGATGATGCTGCAT
osa-miR172a	MIMAT0001069	AGAATCTTGATGATGCTGCAT
osa-miR172b	MIMAT0001070	GGAATCTTGATGATGCTGCAT
osa-miR172c	MIMAT0001071	TGAATCTTGATGATGCTGCAC
osa-miR172d-3p	MIMAT0001084	AGAATCTTGATGATGCTGCAT
osa-miR172d-5p	MIMAT0022885	GCAGCACCATCAAGATTCAC
ppe-miR172a-3p	MIMAT0031469	AGAATCTTGATGATGCTGCAT
ppe-miR172a-5p	MIMAT0031468	GTAGCATCATCAAGATTCACG
ppe-miR172b	MIMAT0031470	AGAATCTTGATGATGCTGCAT
ppe-miR172c	MIMAT0031471	GGAATCTTGATGATGCTGCAT
ppe-miR172d	MIMAT0031472	GGAATCTTGATGATGCTGCAG
ptc-miR172a	MIMAT0001993	AGAATCTTGATGATGCTGCAT
ptc-miR172b-3p	MIMAT0001994	AGAATCTTGATGATGCTGCAT
ptc-miR172b-5p	MIMAT0022904	GGAGCATCATCAAGATTCACA
ptc-miR172c	MIMAT0001995	AGAATCTTGATGATGCTGCAT
ptc-miR172d	MIMAT0001996	GGAATCTTGATGATGCTGCAT
ptc-miR172e	MIMAT0001997	GGAATCTTGATGATGCTGCAT
ptc-miR172f	MIMAT0001998	AGAATCTTGATGATGCTGCAT
ptc-miR172g-3p	MIMAT0001999	GGAATCTTGATGATGCTGCAG
ptc-miR172g-5p	MIMAT0022905	GGAGCATCATCAAGATTCACA
ptc-miR172h-3p	MIMAT0002000	GGAATCTTGATGATGCTGCAG
ptc-miR172h-5p	MIMAT0022906	GGAGCACCATCAAGATTCACA
ptc-miR172i	MIMAT0002001	AGAATCCTGATGATGCTGCAA
rco-miR172	MIMAT0014179	GGAATCTTGATGATGCTGCAG
sbi-miR172a	MIMAT0001397	AGAATCTTGATGATGCTGCA
sbi-miR172b	MIMAT0001395	GGAATCTTGATGATGCTGCA
sbi-miR172c	MIMAT0001396	AGAATCTTGATGATGCTGCA
sbi-miR172d	MIMAT0011345	AGAATCTTGATGATGCTGCA
sbi-miR172e	MIMAT0001464	TGAATCTTGATGATGCTGCAC
sbi-miR172f	MIMAT0014045	AGAATCCTGATGATGCTGCAC
sly-miR172a	MIMAT0009143	AGAATCTTGATGATGCTGCAT
sly-miR172b	MIMAT0009144	AGAATCTTGATGATGCTGCAT
ssl-miR172	MIMAT0022514	AGAATCTTGATGATGCTGCAT
stu-miR172a-3p	MIMAT0031291	AGAATCTTGATGATGCTGCAT

Supplementary Table 1… (continued): Annotated miR172 sequences in miRBase v. 20

stu-miR172a-5p	MIMAT0031290	GTAGCATAATCAAGATTCACA
stu-miR172b-3p	MIMAT0031287	AGAATCTTGATGATGCTGCAT
stu-miR172b-5p	MIMAT0031286	GCAGCACCATCAAGATTCACA
stu-miR172c-3p	MIMAT0031289	AGAATCTTGATGATGCTGC
stu-miR172c-5p	MIMAT0031288	AGCATCTTCAAGATTCACA
stu-miR172d-3p	MIMAT0031293	GGAATCTTGATGATGCTGCAG
stu-miR172d-5p	MIMAT0031292	GGAGCATCATCAAGATTCACA
stu-miR172e-3p	MIMAT0031295	AGAATCTTGATGATGCTGCAT
stu-miR172e-5p	MIMAT0031294	GCAACATCATCAAGATTCACA
tcc-miR172a	MIMAT0020413	GGAATCTTGATGATGCTGCA
tcc-miR172b	MIMAT0020414	AGAATCTTGATGATGCTGCAT
tcc-miR172c	MIMAT0020415	GGAATCTTGATGATGCTGCAT
tcc-miR172d	MIMAT0020416	AGAATCCTGATGATGCTGCAT
tcc-miR172e	MIMAT0020417	AGAATCTTGATGATGCTGCAT
vun-miR172	MIMAT0022760	AGAATCTTGATGATGCTGCAT
vvi-miR172a	MIMAT0005699	TGAATCTTGATGATGCTACAT
vvi-miR172b	MIMAT0005700	TGAATCTTGATGATGCTACAC
vvi-miR172c	MIMAT0005701	GGAATCTTGATGATGCTGCAG
vvi-miR172d	MIMAT0005702	TGAGAATCTTGATGATGCTGCAT
zma-miR172a	MIMAT0001387	AGAATCTTGATGATGCTGCA
zma-miR172b-3p	MIMAT0001389	AGAATCTTGATGATGCTGCA
zma-miR172b-5p	MIMAT0015157	CAGCACCATCAAGATTCACA
zma-miR172c-3p	MIMAT0001390	AGAATCTTGATGATGCTGCA
zma-miR172c-5p	MIMAT0015158	CAGCACCACCAAGATTCACA
zma-miR172d-3p	MIMAT0001388	AGAATCTTGATGATGCTGCA
zma-miR172d-5p	MIMAT0015156	CAGCACCATCAAGATTCACA
zma-miR172e	MIMAT0001742	GGAATCTTGATGATGCTGCAT

REFERENCES

Abdollahi, P., Kamiya, Y., Kawaura, K., & Ogihara, Y. (2012). Overexpression of Q/q-related homoeoalleles of hexaploid wheat reveals distinct recovery of flower transformation in the apetala2 mutant of Arabidopsis. *Plant Biotechnology, 29*(3), 245-252. doi: 10.5511/plantbiotechnology.12.0224a

Aceto, S., Sica, M., De Paolo, S., D'Argenio, V., Cantiello, P., Salvatore, F., & Gaudio, L. (2014). The Analysis of the Inflorescence miRNome of the Orchid Orchis italica Reveals a DEF-Like MADS-Box Gene as a New miRNA Target. *PLoS One, 9*(5), e97839.

Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., & Axtell, M. J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Curr Biol, 18*(10), 758-762. doi: 10.1016/j.cub.2008.04.042

Allen, E., Xie, Z., Gustafson, A. M., & Carrington, J. C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell, 121*, 207-221.

Allen, E., Xie, Z., Gustafson, A. M., Sung, G.-H., Spatafora, J. W., & Carrington, J. C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat Genet, 36*, 1282-1290. doi: 10.1038/ng1478

Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., & Suzuki, M. (1998). A novel mode of DNA recognition by a [beta]-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J, 17*, 5484-5496.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Ambros, V. (2003). A uniform system for microRNA annotation. *Rna, 9*(3), 277-279. doi: 10.1261/rna.2183803

Anderson, V. L., Bauer, M. L., Canton, J. S., Berg, P., & Dhuyvetter, D. H. (1997). *An evaluation of reciprocal levels of hulless oats and barley in diets for growing and finishing steers*. North Dakota State University. http://www.ag.ndsu.nodak.edu/carringt/97beef/art5.htm.

Aryal, R., Yang, X., Yu, Q., Sunkar, R., Li, L., & Ming, R. (2012). Asymmetric purine-pyrimidine distribution in cellular small RNA population of papaya. *BMC Genomics, 13*, 682. doi: 10.1186/1471-2164-13-682

Aukerman, M. J., & Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell, 15*(11), 2730-2741. doi: 10.1105/tpc.016238

Axtell, M. J., & Bowman, J. L. (2008). Evolution of plant microRNAs and their targets. *Trends Plant Sci, 13*(7), 343-349. doi: 10.1016/j.tplants.2008.03.009

Ballen-Taborda, C., Plata, G., Ayling, S., Rodriguez-Zapata, F., Becerra Lopez-Lavalle, L. A., Duitama, J., & Tohme, J. (2013). Identification of Cassava MicroRNAs under Abiotic Stress. *Int J Genomics, 2013*, 857986. doi: 10.1155/2013/857986

Barik, S., SarkarDas, S., Singh, A., Gautam, V., Kumar, P., Majee, M., & Sarkar, A. K. (2014). Phylogenetic analysis reveals conservation and diversification of micro RNA166 genes among diverse plant species. *Genomics, 103*, 114-121.

Bartel, D. P. (2004). MicroRNAs- Genomics, Review Biogenesis, Mechanism, and Function. *Cell, 116*, 281-297.

Baum, B. R. (1977). *Oats: wild and cultivated: a monograph of the genus Avena L. (Poaceae)* (C. D. o. A. Ottawa: Biosystematics Research Institute, Research Branch Ed.): Supply and Services Canada.

Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., . . . Devos, K. M. (2012). Reference genome sequence of the model plant Setaria. *Nat Biotechnol, 30*, 555-561.

Bertollini, E., Vereist, W., Horner, D. S., Gianfranceschi, L., Piccolo, V., Inze, D., . . . Mica, E. (2013). Addressing the role of microRNAs in reprogramming leaf growth during drought stress in Brachypodium distachyon. *Molecular Plant, 6*(2), 423-443. doi: 10.1093/mp/sss160

Bologna, N. G., Mateos, J. L., Bresso, E. G., & Palatnik, J. F. (2009). A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J, 28*(23), 3646-3656. doi: 10.1038/emboj.2009.292

Borsani, O., Zhu, J., Versiues, P. e., Sunkar, R., & Zhu, J. K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell, 123*(1279--1291).

Bowman, J. L., Smyth, D. R., & Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes of Arabidopsis. *Development, 112*, 1-20.

Brodersen, P., & Voinnet, O. (2006). The diversity of RNA silencing pathways in plants. *Trends in Genetics, 22*(5), 268-280.

Brown, R. H., & Bregitzer, P. (2011). A Ds Insertional Mutant of a Barley miR172 Gene Results in Indeterminate Spikelet Development. *Crop Science, 51*(4), 1664-1672. doi: 10.2135/cropsci2010.09.0532

Broz, J., & Frigg, M. (1986). Effects of beta-glucanase on the feeding value of broiler diets based on barley or oats. *Archiv. Gefluegelkd, 50*(1), 41-47.

Burge, S. W., Daub, J., Eberhardt, J. T., Barquist, L., Nawrocki, E. P., Eddy, S. R., . . . Bateman, A. (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res, 41*(D1), D226-D232. doi: 10.1093/nar/gks1005

Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., . . . Bateman, A. (2012). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res, 41*(D1), D226-D232.

Burrows, M., & Wheeler, D. J. (1994). A Block-sorting Lossless Data Compression Algorithm. CA: System Research Center.

Burrows, V. D. (2011). Hulless Oat Development, Applications, and Opportunities. In F. H. Webster & P. J. Wood (Eds.), *Oats: Chemistry and Technology* (2 ed., pp. 31-50): AACC Press.

Burrows, V. D., McDiarmid, G., & Marder, T. (2001). Gene controlling reduced trichomes on oat groats. *Canadian Journal of Plant Science, 81*, 413-414.

Calin, G. A., & Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nature Reviews Cancer, 6*, 857-866. doi: 10.1038/nrc1997

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell, 136*(4), 642-655. doi: 10.1016/j.cell.2009.01.035

Cave, N. A., Hamilton, R. M. G., & Burrows, V. D. (1989). Naked oats (cv. Tibor) as feedstuff for laying hens. *Canadian Journal of Animal Science, 69*, 789-799.

Chen, X. (2004). A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development. *Science, 303*, 2022-2025.

Chen, Z. J., & Ni, Z. (2006). Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays, 28*(3), 240-252. doi: 10.1002/bies.20374

Chevreaux, B. (1997). MIRA: Sourceforge.

Chou, H. H., & Holmes, M. H. (2001). DNA sequence quality trimming and vector removal. *Bioinformatics, 17*(12), 1093-1104.

Chuck, G., Cigan, A. M., Saeteurn, K., & Hake, S. (2007). The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. *Nat Genet, 39*, 544-549.

Chuck, G., Meeley, R., & Hake, S. (2008). Floral meristem initiation and meristem cell fate are regulated by the maize AP2 genes ids1 and sid1. *Development, 135*(18), 3013-3019. doi: 10.1242/dev.024273

Chuck, G., Meeley, R. B., & Hake, S. (1998). The control of maize spikelet meristem fate by theAPETALA2-like gene indeterminate?spikelet1. *Genes Dev, 12*, 1145-1154.

Colaiacovo, M., Bernardo, L., Centomani, I., Crosatti, C., Guisti, L., Orru, L., . . . Faccioli, P. (2012). A Survey of MicroRNA Length Variants Contributing to miRNome Complexity in Peach (Prunus Persica L.). *Front Plant Sci, 3*, 165. doi: 10.3389/fpls.2012.00165

Colaiacovo, M., Lamontanara, A., Bernardo, L., Alberici, R., Crosatti, C., Guisti, L., . . . Faccioli, P. (2012). On the complexity of miRNA-mediated regulation

in plants: novel insights into the genomic

organization of plant miRNAs. *Biology Direct, 7*, 15.

Colombo, L., Franken, J., Koetje, E., Van Went, J., Dons, H. J., Angenent, G. C., & Van Tunen, A. J. (1995). The petunia MADS box gene FBP11 determines ovule identity. *Plant Cell, 7*(11), 1859-1868.

Cuperus, J. T., Fahlgren, N., & Carrington, J. C. (2011). Evolution and Functional Diversification of MIRNA genes. *Plant Cell, 23*(2), 431-442. doi: http://dx.doi.org/10.1105/tpc.110.082784

Czech, B., & Hannon, G. J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet, 12*(1), 19-31. doi: 10.1038/nrg2916

Dai, X., & Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res, 39*(Web Server issue), W155-159. doi: 10.1093/nar/gkr319

Dai, X., Zhuang, Z., & Zhao, P. X. (2011). Computational analysis of miRNA targets in plants: current status and challenges. *Brief Bioinform, 12*(2), 115-121. doi: 10.1093/bib/bbq065

De Koeyer, D. L., Tinker, N. A., Wight, C. P., Deyl, J., Burrows, V. D., O'Donoughue, L. S., . . . McElroy, A. R. (2004). A molecular linkage map with associated QTLs from a hulless x covered spring oat population. *Theor Appl Genet, 108*(7), 1285-1298. doi: 10.1007/s00122-003-1556-x

Debernardi, J. M., Rodriguez, R. E., Mecchia, M. A., & Palatnik, J. F. (2012). Functional Specialization of the Plant miR396 Regulatory Network through Distinct MicroRNA–Target Interactions. *PLoS Genet, 8*(1), e1002419. doi: 10.1371/journal.pgen.1002419

Devers, E. A., Branscheid, A., May, P., & Krajinski, F. (2011). Stars and Symbiosis: MicroRNA- and MicroRNA*-Mediated Transcript Cleavage Involved in Arbuscular Mycorrhizal Symbiosis. *Plant Physiol, 156*, 1990-2010.

Dietz, K. J., Vogel, M. O., & Viehhauser, A. (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma, 245*(1-4), 3-14. doi: 10.1007/s00709-010-0142-8

Drevjany, D. E., & Drevjany, L. A. (1988).

Drews, G. N., Bowman, J. L., & Meyerowitz, E. M. (1991). Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. *Cell, 65*, 991-1001.

Egawa, C., Kobayashi, F., Ishibashi, M., Nakamura, T., & Takumi, S. (2006). Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. *Genes and Genetic Systems, 81*, 77-91.

Equavena Hulless Oats. from http://www.semican.ca/english/equavena/index.htm

Fardet, A. (2010). New hypotheses for the health-protective mechanisms of whole-grain cereals: what is beyond fibre? *Nutr Res Rev, 23*(1), 65-134. doi: 10.1017/S0954422410000041

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature, 391*, 806-811.

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., . . . Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet, 39*(8), 1033-1037. doi: 10.1038/ng2079

Fric, P., Gabrovska, D., & Nevoral, J. (2011). Celiac disease, gluten-free diet, and oats. *Nutr Rev, 69*(2), 107-115. doi: 10.1111/j.1753-4887.2010.00368.x

Friedlander, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., & Rajewsky, N. (2008). Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol, 26*(4), 407-415. doi: 10.1038/nbt1394

Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades *Nucleic Acids Res, 40*(1), 37-52.

Friend, D. W., Fortin, A., Butler, G., Poste, L. M., Kramer, J. K. G., & Burrows, V. D. (1989). The feeding value of naked oats (Avena nuda) for boars and barrows: Growth, carcass, meat quality, and energy and nitrogen metabolism. *Canadian Journal of Animal Science, 69*, 765-778.

Fu, Y. B., & Williams, D. J. (2008). AFLP variation in 25 Avena species. *Theor Appl Genet, 117*(3), 333-342. doi: 10.1007/s00122-008-0778-3

Gandikota, M., Birkenbihl, R. P., Hohmann, S., Cardon, G. H., Saedler, H., & Huijser, P. (2007). The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. *Plant J, 49*(4), 683-693. doi: 10.1111/j.1365-313X.2006.02983.x

Garcia-Manzanares, A., & Lucendo, A. J. (2011). Nutritional and dietary aspects of celiac disease. *Nutr Clin Pract, 26*(2), 163-173. doi: 10.1177/0884533611399773

Gaut, B. S. (2002). Evolutionary dynamics of grass genomes. *New Phytologist, 154*, 15-28.

Gentile, A., Ferreira, T. H., Mattos, R. S., Dias, L. I., Hoshino, A. A., Carneiro, M. S., . . . Menossi, M. (2013). Effects of drought on the microtranscriptome of field-grown sugarcane plants. *Planta, 237*(3), 783-798.

German, M. A., Luo, S., Schroth, G., Meyers, B. C., & Green, P. J. (2009). Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nat Protoc, 4*(3), 356-362. doi: 10.1038/nprot.2009.8

German, M. A., Pillay, M., Jeong, D. H., Hetawal, A., Luo, S., Janardhanan, P., . . . Green, P. J. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat Biotechnol, 26*(8), 941-946. doi: 10.1038/nbt1417

Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet, 10*(2), 94-108. doi: 10.1038/nrg2504

Gong, L., Kakrana, A., Arikit, S., Meyers, B. C., & Wendel, J. F. (2013). Composition and Expression of Conserved MicroRNA Genes in Diploid Cotton (Gossypium) Species. *Genome Biol Evol, 5*(12), 2449-2459.

Gotoh, O. (1982). An improved algorithm for matching biological sequences. *Journal of Molecular Biology, 162*, 705-708.

Griffiths-Jones, S. (2004). The microRNA Registry. *NAR, 32*(Database Issue), D109-D111.

Griffiths-Jones, S., Bateman, A., Marshall, A. H., Khanna, A., & Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Res, 31*(1), 439-441.

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., & Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *NAR, 34*(Database Issue), D140-D144.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., & Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res, 33*(Database issue), D121-124. doi: 10.1093/nar/gki081

Griffiths-Jones, S., Saini, H. K., van Dongen, S., & Enright, A. J. (2008). miRBase: tools for microRNA genomics. *NAR, 36*(Database Issue), D154-D158.

Grigorova, B., Mara, C., Hollender, C., Sijacic, P., Chen, X., & Liu, Z. (2011). LEUNIG and SEUSS co-repressors regulate miR172 expression in Arabidopsis flowers. *Development, 138*(12), 2451-2456. doi: 10.1242/dev.058362

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R., & Hofacker, I. L. (2008). The Vienna RNA Website. *Nucleic Acids Res, 36 (suppl 2)*, W70-W74. doi: 10.1093/nar/gkn188

Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., . . . Tushl, T. (2008). Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods, 44*(1), 3-12.

Hannon Lab (2010). FASTX-Toolkit (Version 0.0.13). http://hannonlab.cshl.edu/fastx_toolkit/index.html. Cold Spring Harbor Laboratory.

Han, M. H., Goud, S., Song, L., & Fedoroff, N. V. (2003). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *PNAS, 101*(4), 1093-1098.

Hegarty, M. J., & Hiscock, S. J. (2008). Genomic Clues to the Evolutionary Success of Polyploid Plants. *Curr Biol, 18*(10), R435-R444.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., & Shuster, P. (1994). Fast Folding and Comparison of RNA Secondary Structures. *Manatshefte fur Chemie, 125*, 167-188.

Huang, P. J., Liu, Y. C., Lee, C. C., Lin, W. C., Gan, R. R., Lyu, P. C., & Tang, P. (2010). DSAP: deep sequencing small RNA analysis pipeline. *Nucleic Acids Res, 38*(Web Server issue), W385-391. doi: 10.1093/nar/gkq392

Huntzinger, E., & Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet, 12*(2), 99-110. doi: 10.1038/nrg2936

International Barley Genome Sequencing, C., Mayer, K. F., Waugh, R., Brown, J. W., Schulman, A., Langridge, P., . . . Stein, N. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature, 491*(7426), 711-716. doi: 10.1038/nature11543

International Brachypodium, I. (2010). Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature, 463*(7282), 763-768. doi: 10.1038/nature08747

Jenkins, G., & Hanson, P. R. (1976). The genetics of naked oats. *Euphytica, 25*, 167-174.

Jinek, M., & Doudna, J. A. (2009). A three-dimensional view of the molecular machinery of RNA interference. *Nature, 457*, 405-412.

Jones-Rhoades, M. W., & Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell, 14*(6), 787-799. doi: 10.1016/j.molcel.2004.05.027

Jones-Rhoades, M. W., Bartel, D. P., & Bartel, B. (2006). MicroRNAs and Their Regulatory Roles in Plants. *Annual Review of Plant Biology, 57*, 19-53. doi: 10.1146/

Jung, J.-H., Seo, P. J., Kang, S. K., & Park, C.-M. (2011). miR172 signals are incorporated into the miR156 signaling pathway at the SPL3/4/5 genes in Arabidopsis developmental transitions. *Plant Mol Biol, 76*(1-2), 35-45.

Kantar, M., Akpinar, B. A., Valarik, M., Lucas, S. J., Dolezel, J., Hernandez, P., & Budak, H. (2012). Subgenomic analysis of microRNAs in polyploid wheat. *Funct Integr Genomics, 12*(3), 465-479.

Katiyar-Agarwai, S., et al. (2006). A pathogen-inducible endogenous siRNA in plant immunity. *PNAS, 103*, 18002-18007.

Kidner, C. A., & Martienssen, R. A. (2004). Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature, 428*, 81-84.

Kim, S., Soltis, P. S., Wall, K., & Soltis, D. E. (2006). Phylogeny and Domain Evolution in the APETALA2-like Gene Family. *Molecular Biology and Evolution, 23*(1), 107-120. doi: 10.1093/molbev/msj014

Kim, Y. K., Yeo, J., Ha, M., Kim, B., & Kim, V. N. (2012). Retraction Notice to: Cell Adhesion-Dependent Control of MicroRNA Decay. *Mol Cell, 46*(6), 896.

Kozomara, A., & Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep sequencing data. *NAR, 39*(Database Issue), D152-D157.

Kozomara, A., & Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res, 42*(1), D68-73. doi: 10.1093/nar/gkt1181

Krichevsky, A. M., King, K. S., Donahue, C. P., Khrapko, K., & Kosik, K. S. (2003). A microRNA array reveals extensive regulation of microRNAs during brain development. *Rna, 9*, 1274-1281.

Krizek, B. A., & Fletcher, J. C. (2005). Molecular mechanisms of flower development: an armchair guide. *Nat Rev Genet, 6*(9), 688-698. doi: 10.1038/nrg1675

Kurtoglu, K. Y., Kantar, M., Lucas, S. J., & Budak, H. (2013). Unique and Conserved MicroRNAs in Wheat Chromosome 5D Revealed by Next-Generation Sequencing *PLoS One, 8*(7), e69801. doi: 10.1371/journal.pone.0069801

Kurtz, S. (2011). The Vmatch large sequence analysis software. University of Hamburg, Germany.

Kwak, P. B., & Tomari, Y. (2012). The N domian of Argonaute drive dulex unwinding during RISC assembly. *Nature Structural and Molecular Biology, 19*, 145-151. doi: 10.1038/nsmb.2232

Ladizinsky, G. (2000). A synthetic hexaploid (2n = 42) oat from the cross of Avena strigosa (2n = 14) and domesticated A. magna (2n = 28). *Euphytica, 116*, 231-235.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol, 10*, R25.

Lawes, D. A., & Boland, P. (1974). Effect of temperatre on the expression of the naked grain character in oats. *Euphytica, 23*, 101-104.

Lee, D. Y., & An, G. (2012). Two AP2 family genes, supernumerary bract (SNB) and Osindeterminate spikelet 1 (OsIDS1), synergistically control inflorescence architecture and floral meristem establishment in rice. *Plant J, 69*(3), 445-461. doi: 10.1111/j.1365-313X.2011.04804.x

Lee, R. C., Feinbaurm, R. L., & Ambros, V. (1993). The C. elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to lin-14. *Cell, 75*, 843-854.

Li, A., & Mao, L. (2007). Evolution of plant microRNA gene families. *Cell Res, 17*(3), 212-218. doi: 10.1038/sj.cr.7310113

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi: 10.1093/bioinformatics/btp324

Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., & Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics, 25*(15), 1966-1967.

Li, S., Liu, L., Zhuang, X., Yu, Y., Liu, X., Cui, X., . . . Chen, X. (2013). MicroRNAs Inhibit the Translation of Target mRNAs on the Endoplasmic Reticulum in Arabidopsis. *Cell, 153*(3), 562-574.

Li, Y., Zhang, Z., Liu, F., Vongsangnak, W., Jing, Q., & Shen, B. (2012). Performance comparison and evaluation of software tools for microRNA deep sequencing data analysis. *Nucleic Acids Res, 40*(10), 4298-4305. doi: 10.1093/nar/gks043

Licausi, F., Ohme-Takagi, M., & Perata, P. (2013). APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytologist, 199*(3), 639-649. doi: 10.1111/nph.12291

Lin, S. I., Santi, C., Jobet, E., Lacut, E., Kholti, N. R., Karlowski, W. M., . . . Echeverria, M. (2010). Complex Regulation of Two Target Genes Encoding SPX-MFS Proteins by Rice miR827 in Response to Phosphate Starvation. *Plant Cell Physiol, 51*(12), 2119-2131. doi: 10.1093/pcp/pcq170

Linares, C., Ferrer, E., & Fominaya, A. (1998). Discrimination of the closely related A and D genomes of the hexaploid oat Avena sativa L. *PNAS, 95*, 12450-12455.

Llave, C. (2002). Endogenous and Silencing-Associated Small RNAs in Plants. *The Plant Cell Online, 14*(7), 1605-1619. doi: 10.1105/tpc.003210

Lu, S., Li, Q., Wei, H., Chang, M. J., Tunlaya-Anukit, S., Kim, H., . . . Chiang, V. L. (2013). Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in Populus trichocarpa. *PNAS, 110*(26), 10848-10853.

Mallory, A. C., & Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. *Nat Genet, 38 Suppl*, S31-36. doi: 10.1038/ng1791

Manavella, P. A., & Rubio-Somoza, I. (2011). Engineering Elements for Gene Silencing: The Artificial MicroRNAs Technology. In T. Dalmay (Ed.), *MicroRNAs in Development* (Vol. 732, pp. 121-130). USA: Humana Press.

Marshall, H. G., McDaniel, M. E., & Creggar, L. M. (1992). Cultural Practices for oat: Weed and pest control. In H. G. Marshall & M. E. Sorrells (Eds.), *Oat Science and Technology* (pp. 209-210). Madison, WI: American Society of Agronomy.

Mathelier, A., & Carbone, A. (2010). MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics, 26*(18), 2226-2234.

McCaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structures. *Biopolymers, 29*, 1105-1119.

Meng, F., Liu, H., Wang, K., Liu, L., Wang, S., Zhao, Y., . . . Li, Y. (2013). Development-associated microRNAs in grains of wheat (Triticum aestivum L.). *BMC Plant Biol, 13*, 140. doi: 10.1186/1471-2229-13-140

Meng, Y., Shao, C., Wang, H., & Chen, M. (2012). Are all the miRBase-registered microRNAs true? A structure- and expression-based re-examination in plants. *RNA Biol, 9*(3), 249-253. doi: 10.4161/rna.19230

Meyers, B. C., Axtell, M. J., Bartel, B., Bartel, D. P., Baulcombe, D., Bowman, J. L., . . . Zhu, J. K. (2008). Criteria for annotation of plant MicroRNAs. *Plant Cell, 20*(12), 3186-3190. doi: 10.1105/tpc.108.064311

Naqvi, A. R., Sarwat, M., Hasan, S., & Roychodhury, N. (2012). Biogenesis, functions and fate of plant microRNAs. *J Cell Physiol, 227*(9), 3163-3168. doi: 10.1002/jcp.24052

National Institutes of Health Consensus Development Conference Statement on Celiac Disease, June 28-30, 2004. (2005). *Gastroenterology, 128*(4 Suppl1), S1-9.

Nelson, P. T., Baldwin, D. A., Scearce, L. M., Oberholtzer, J. C., Tobias, J. W., & Mourelatos, Z. (2004). Microarray-based, high-throughput gene expression profiling of microRNAs. *Nature Methods, 1*, 155-161.

Ohme-Takagi, M., & Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell, 7*(173-182).

Ozhuner, E., Eldem, V., Ipek, A., Okay, S., Sakcali, S., Zhang, B., . . . Unver, T. (2013). Boron Stress Responsive MicroRNAs and Their Targets in Barley *PLoS One, 8*(3), e59543. doi: 10.1371/journal.pone.0059543

Palatnik, J. F., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., . . . Weigel, D. (2007). Sequence and Expression Differences Underlie Functional Specialization of Arabidopsis MicroRNAs miR159 and miR319. *Dev Cell, 13*(1), 115-125.

Park, M. Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., & Poethig, R. S. (2005). Nuclear processing and export of microRNAs in Arabidopsis. *Proc Natl Acad Sci U S A, 102*, 3691-3696.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., . . . Rokhsar, D. S. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature, 457*(7229), 551-556. doi: 10.1038/nature07723

Paterson, A. H., Bowers, J. E., & Chapman, B. A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci U S A, 101*(26), 9903-9908. doi: 10.1073/pnas.0307901101

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., & Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature, 405*(6783), 200-203.

Pertea, G. (2005). SeqClean: Sourceforge.

Poulsen, C., Vaucheret, H., & Brodersen, P. (2013). Lessons on RNA silencing mechanisms in plants from eukaryotic argonaute structures. *Plant Cell, 25*(1), 22-37. doi: 10.1105/tpc.112.105643

Qu, F., Ye, X., Hou, G., Sato, S.,, Clemente, T. E., & Morris, T. J. (2005). RDR6 Has a Broad-Spectrum but Temperature-Dependent Antiviral Defense Role in Nicotiana benthamiana. *Journal of Virology, 79*(24), 15209-15217. doi: 10.1128/JVI.79.24.15209-15217.2005

Quintanar, A., Castroviejo, S., & Catalan, P. (2007). Phylogeny of the Tribe Avenae (Pooideae, Poaceae) Inferred from Plastid trnT-F and Nuclear ITS Sequences. *American Journal of Botany, 94*(9), 1554-1569.

Rajagopalan, R., Vaucheret, H., Trejo, J., & Bartel, D. P. (2006). A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev, 20*(24), 3407-3425. doi: 10.1101/gad.1476406

Ramachandran, V., & Chen, X. (2008). Degradation of microRNAs by a Family of Exoribonucleases in Arabidopsis. *Science, 321*(5895), 1490-1492. doi: 10.1126/science.1163728

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., . . . Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature, 403*, 901-906. doi: 10.1038/35002607

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., & Bartel, D. P. (2002). MicroRNAs in plants. *Genes Dev, 16*, 1616-1626.

Ren, D., Li, Y., Zhao, F., Sang, X., Shi, J., Wang, N., . . . He, G. (2013). MULTI-FLORET SPIKELET1, Which Encodes an AP2/ERF Protein, Determines Spikelet Meristem Fate and Sterile Lemma Identity in Rice. *Plant Physiol, 162*, 872-884. doi: 10.1104/pp.113.216044

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., & Bartel, D. P. (2002). Prediction of Plant MicroRNA Targets. *Cell, 110*, 513-520.

Rodionov, A. V., Tyupa, N. B., Kim, E. S., Machs, E. M., & Loskutov, I. G. (2005). Genomic Configuration of the Autotetraploid Oat Species Avena macrostachya Inferred from Comparative Analysis of ITS1 and ITS2 Sequences: on the Oat Karyotype Evolution during the Early Events of the Avena Species Divergence. *Russian Journal of Genetics, 41*(5), 518-528.

Rodriguez, R. E., Mecchia, M. A., Debernardi, J. M., Schommer, C., Weigel, D., & Palatnik, J. F. (2010). Control of cell proliferation in Arabidopsis thaliana by microRNA miR396. *Development, 137*, 103-112.

Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, A., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2002). DNA-Binding Specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochemical and Biophysical Research Communications, 290*, 998-1009.

Sanderson, M. J. (1997). A Nonparametric Approach to Estimating Divergence Times in the Absence of Rate Constancy. *Molecular Biology and Evolution, 14*(12), 1218-1231.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., . . . Wilson, R. K. (2009). The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science, 326*(5956), 1112-1115. doi: 10.1126/science.1178534

Schulz, M. H., Zerbino, D. R., Vingron, M., & Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics, 28*(8), 1086-1092.

Scotto-Lavino, E., Guangwei, D., & Frohman, M. A. (2007). Amplification of 5' end cDNA with 'new RACE'. *Nat Protoc, 1*, 3066-3061.

Shabalina, S. A., & Koonin, E. V. (2008). Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol, 23*(10), 578-587. doi: 10.1016/j.tree.2008.06.005

Shan, L., Molberg, O., Parrot, I., Hausch, F., Filiz, F., Gray, G. M., . . . Khosla, C. (2002). Structural basis for gluten intolerance in celiac sprue. *Science, 297*, 2275-2279.

Shitsukawa, N., Kinjo, H., Takumi, S., & Murai, K. (2009). Heterochronic development of the floret meristem determines grain number per spikelet in diploid, tetraploid and hexaploid wheats. *Annals of Botany, 104*(2), 243-251.

Shunzong, N. (2013). *The wheat orthologs of the barley cleistogamy gene: characterization and potential exploitation for wheat improvement.* Chiba University, Japan.

Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y. S., Gill, B. S., & Faris, J. D. (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics, 172*(1), 547-555. doi: 10.1534/genetics.105.044727

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology, 147*, 195-197.

Song, L., Axtell, M. J., & Fedoroff, N. V. (2010). RNA secondary structural determinants of miRNA precursor processing in Arabidopsis. *Curr Biol, 20*(1), 37-41. doi: 10.1016/j.cub.2009.10.076

Souret, F. F., Kastenmayer, J. P., & Green, P. J. (2004). AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Mol Cell, 15*, 173-183.

Stanton, T. R. (1955). Oat identification and classification *Tech. Bull.* Washington, DC: U.S. Department of Agriculture.

Su, C., Yang, X., Gao, S., Tang, Y., Zhao, C., & Li, L. (2014). Identification and characterization of a subset of microRNAs in wheat (Triticum aestivum L.). *Genomics, available online March 22, 2014*.

Sun, G., Stewart, N. J., Xiao, P., & Zhang, B. (2012). MicroRNA Expression Analysis in the Cellulosic Biofuel Crop Switchgrass (Panicum virgatum) under Abiotic Stress. *PLoS One, 7*(3), e32017.

Sunkar, R., Girke, T., Jain, P. K., & Zhu, J. K. (2005). Cloning and characterization of microRNAs from rice. *Plant Cell, 17*(5), 1397-1411. doi: 10.1105/tpc.105.031682

Taketa, S., Amano, S., Tsujino, Y., Sato, T., Saisho, D., Kakeda, K., . . . Takeda, K. (2008). Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. *Proc Natl Acad Sci U S A, 105*(10), 4062-4067. doi: 10.1073/pnas.0711034105

Thakur, V., Wanchana, S., Xu, M., Bruskiewich, R., Quick, W. P., Mosig, A., & Zhu, X. G. (2011). Characterization of statistical features for plant microRNA prediction. *BMC Genomics, 12*, 108. doi: 10.1186/1471-2164-12-108

Thomas, H. (1992). In H. G. Marshall & M. E. Sorrells (Eds.), *Oat Science and Technology* (pp. 473-507). Madison, WI: American Society of Agronomy and Crop Science Society of America.

Tjaden, B., Goodwin, S. S., Opdyke, J. A., Guillier, M., Fu, D. X., Gottesman, S., & Storz, G. (2006). Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res, 34*(9), 2791-2802. doi: 10.1093/nar/gkl356

Vader, L. W., Stepniak, D. T., Bunnik, E. M., Kooy, Y. M., de Haan, W., Drijfhout, J. W., . . . Koning, F. (2003). Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. *Gastroenterology, 125*(4), 1105-1113.

Valentine, J. (1995). Naked oats. In W. Welch (Ed.), *The Oat Crop: Production and Utilization* (pp. 504-532). London: Chapman and Hall.

Valentine, J., Cowan, A. A., & Marshall, A. H. (2011). Oat Breeding. In F. H. Webster & P. J. Wood (Eds.), *Oats: Chemistry and Technology* (2 ed., pp. 11-20): AACC Press.

Varkonyi-Gasic, E., Lough, R. H., Moss, S. M., Wu, R., & Hellens, R. P. (2012). Kiwifruit floral gene APETALA2 is alternatively spliced and accumulates in aberrant indeterminate flowers in the absence of miR172. *Plant Mol Biol, 78*(4-5), 417-429. doi: 10.1007/s11103-012-9877-2

Vaucheret, H. (2008). Plant ARGONAUTES. *Trends Plant Sci, 13*(7), 350-358. doi: 10.1016/j.tplants.2008.04.007

Wang, B., Sun, Y. F., Song, N., Wei, J. P., Wang, X. J., Feng, H., . . . Kang, Z. S. (2014). MicroRNAs involving in cold, wounding and salt stresses in Triticum aestivum L. *Plant Physiology and Biochemistry, 80*, 90-96.

Wang, K., Li, M., Gao, F., Li, S., Zhu, Y., & Yang, P. (2012). Identification of Conserved and Novel microRNAs from Liriodendron chinense Floral Tissues. *PLoS One, 7*(9), e44696.

Wang, W. C., Lin, F. M., Chang, W. C., Lin, K. Y., Huang, H. D., & Lin, N. S. (2009). miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics, 10*, 328.

Wang, Y., Sun, F., Cao, H., Peng, H., Ni, Z., Sun, Q., & Yao, Y. (2012). TamiR159 directed wheat TaGAMYB cleavage and its involvement in anther development and heat response. *PLoS One, 7*(11), e48445. doi: 10.1371/journal.pone.0048445

Ware, D. H., Jaiswal, P., Ni, J., Yap, I. V., Pan, X., Clark, K. Y., . . . McCouch, S. R. (2002). Gramene, a tool for grass genomics. *Plant Physiol, 130*(4), 1606-1613.

Wassarman, K. M. (2002). Small RNAs in Bacteria: Diverse Regulators of Gene Expression in Response to Environmental Changes. *Cell, 109*, 141-144.

Welch, R. W. (1995). The chemical composition of oats. In W. Welch (Ed.), *The Oat Crop: Production and Utilization* (pp. 279-320). London: Chapman and Hall.

Wessler, S. R. (2005). Homing into the origin of the AP2 DNA binding domain. *Trends Plant Sci, 10*, 54-56.

Williamson, V., Kim, A., Xie, B., McMichael, G. O., Gao, Y., & Vladimirov, V. (2013). Detecting miRNAs in deep sequencing data: a software performance comparison and evaluation. *Briefings in Bioinformatics, 14*(1), 36-45. doi: 10.1093/bib/bbs010

Wilund, K. R., Yi, M., Campagna, F., Arca, M., Zuliani, G., Fellin, R., . . . Cohen, J. C. (2002). Molecular mechanisms of autosomal recessive hypercholesterolemia. *Human Molecular Genetics, 11*(24), 3019-3030.

Wollmann, H., Mica, E., Todesco, M., Long, J. A., & Weigel, D. (2010). On reconciling the interactions between APETALA2, miR172, and AGAMOUS with the ABC model of flower development. *Development, 137*, 3633-3642.

Wu, G., Park, M. Y., Conway, S. R., Wang, J. W., Weigel, D., & Poethig, R. S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell, 138*(4), 750-759. doi: 10.1016/j.cell.2009.06.031

Wu, L., Zhang, Q., Zhou, H., Ni, F., Wu, X., & Qi, Y. (2009). Rice MicroRNA Effector Complexes and Targets. *Plant Cell, 21*(11), 3421-3435.

Xie, Z., Kasschau, K. D., & Carrington, J. C. (2003). Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. *Curr Biol, 13*, 784-789.

Xue, G. P., & W., L. C. (2004). HvDRF1 is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element. *The Plant Journal, 37*, 326-339.

Yan, Y., Wang, H., Hamera, S., Chen, X., & Fang, R. (2014). miR444a has multiplefunctions in the rice nitrate-signaling pathway. *The Plant Journal, Early View*. doi: 10.1111/tpj.12446

Yang, X., & Li, L. (2011). miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics, 27*(18), 2614-2615. doi: 10.1093/bioinformatics/btr430

Yang, X., & Li, L. (2012). Analyzing the microRNA Transcriptome in Plants Using Deep Sequencing Data. *Biology, 1*(3), 297-310. doi: 10.3390/biology1020297

Yang, X., Zhang, H., & Li, L. (2011). Global analysis of gene-level microRNA expression in Arabidopsis using deep sequencing data. *Genomics, 98*(1), 40-46. doi: 10.1016/j.ygeno.2011.03.011

Yang, Z., Ebright, Y. W., Yu, B., & Chen, X. (2006). HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res, 34*(2), 667-675. doi: 10.1093/nar/gkj474

Yant, L., Mathieu, J., Dinh, T. T., Ott, F., Lanz, C., Wollmann, H., . . . Schmid, M. (2010). Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. *Plant Cell, 22*, 2156-2170.

Yu, J., Hu, S., Wang, J., Wong, G. K.-S., Li, S., Liu, B., . . . Yang, H. (2002). A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. indica). *Science, 296*(5565), 79-92. doi: 10.1126/science.1068037

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res, 18*, 821-829.

Zhang, B., Pan, X., & Anderson, T. A. (2006). Identification of 188 conserved maize microRNAs and their targets. *FEBS Lett, 580*(15), 3753-3762. doi: 10.1016/j.febslet.2006.05.063

Zhang, B., Pan, X., Cannon, C. H., Cobb, G. P., & Anderson, T. A. (2006). Conservation and divergence of plant microRNA genes. *Plant J, 46*(2), 243-259. doi: 10.1111/j.1365-313X.2006.02697.x

Zhang, B., Pan, X., Cobb, G. P., & Anderson, T. A. (2006). Plant microRNA: a small regulatory molecule with big impact. *Dev Biol, 289*(1), 3-16. doi: 10.1016/j.ydbio.2005.10.036

Zhang, L., Hou, D., Chen, X., Li, D., Zhu, L., Zhang, Y., . . . Zhang, C. Y. (2012). Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Res, 22*(1), 107-126. doi: 10.1038/cr.2011.158

Zhang, R., Marshall, D., Bryan, G. J., & Hornyik, C. (2013). Identification and Characterization of miRNA Transcriptome in Potato by High-Throughput Sequencing. *PLoS One, 8*(2), eS7233.

Zhang, W., Gao, S., Zhou, X., Xia, J., Chellappan, P., Zhou, X., . . . Jin, H. (2010). Multiple distinct small RNAs originate from the same miRNA precursors. *Genome Biol, 11*, R81.

Zhang, X., Xia, J., Lii, Y. E., Barrera-Figueroa, B. E., Zhou, X., Gao, S., . . . Jin, H. (2012). Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. *Genome Biol, 13*(3), R20.

Zhang, X., Zhao, H., Gao, S., Wang, W. C., Katiyar-Agarwai, S., Huang, H. D., . . . Jin, H. (2011). Arabidopsis Argonaute 2 regulates innate immunity via miRNA393*-mediated silencing of a Golgi-localized SNARE gene MEMB12. *Mol Cell, 42*(3), 356-366. doi: 10.1016/j.molcel.2011.04.010

Zhou, L., Liu, Y., Liu, Z., Kong, D., Duan, M., & Luo, L. (2010). Genome-wide identification and analysis of drought-responsive microRNAs in Oryza sativa. *Journal of Experimental Botany, 61*(15), 4157-4168.

Zhu, C., Ding, Y., & Liu, H. (2011). MiR398 and plant stress responses. *Physiolia Plantarum, 143*(1), 1-9.

Zhu, Q. H., & Helliwell, C. A. (2011). Regulation of flowering time and floral patterning by miR172. *Journal of Experimental Botany, 62*(2), 487-495. doi: 10.1093/jxb/erq295 Advance

Zhu, Q. H., Spriggs, A., Matthew, L., Fan, L., Kennedy, G., Gubler, F., & Helliwell, C. (2008). A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Res, 18*(9), 1456-1465. doi: 10.1101/gr.075572.107

Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science, 244*(4900), 48-52. doi: 10.1126/science.2468181

Zuker, M., & Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamic and auxiliary information. *Nucleic Acids Res, 9*, 133-148.

Zuliani, G., Arca, M., Signore, A., Bader, G., Fazio, S., Chianelli, M., . . . Fellin, R. (1999). Characterization of a New Form of Inherited Hypercholesterolemia: Familial Recessive Hypercholesterolemia. *Arteriosclerosis, Thrombosis, and Vascular Biology, 19*, 802-809.

APPENDIX: MIRNA PREDICTIONS FROM OTHER SEQUENCING SOURCES

A.1 Introduction

Additional miRNA predictions using the modified miRDeep-P pipeline were done for hexaploid oat cultivar Leggett (Chapter 2), diploid *Avena strigosa* (Chapter 3), and tetraploid *Avena magna* (Chapter 3). 454 transcriptome sequencing was done for all three cultivars, however, tissue from Leggett was pooled with two other hexaploid oat lines. In addition to the HiSeq2000 transcriptome assemblies used in Chapter 2 for miRNA predictions in Leggett, there exists assemblies from stage A, B, and C as well.

A.2 Materials and Methods

A.2.1 Tissue and Total RNA Extraction

See Chapter 3, section 3.2.1 Total RNA Sequencing Tissue and RNA Extraction for growth conditions and RNA extraction methods used for 454 and HiSeq total RNA sequencing. See Table 1 for stages pooled for 454 sequencing and Table 2 for the stages of flower development sequenced on the HiSeq2000 for Leggett.

A.2.2 454 Total RNA Sequencing

RNA from Leggett and the two other oat lines were extracted separately and pooled immediately before cDNA synthesis. Library preparations and sequencing was done by DHMRI at the NCRC in Kannapolis, NC. Equal concentrations of extracted RNA from stages A, B, and C were pooled and reverse-transcribed using protocol from Invitrogen using Superscript II (Invitrogen, Carlsbad). The Trimmer cDNA normalization kit (Evrogen, Russia) was used as a normalization process across the libraries before 454 pyrosequencing. Samples were MID-tagged.

A.2.3 454 Transcriptome Assembly

After pyrosequencing on the 454, the Bioinformatics Service Division at DHMRI in Kannapolis, NC processed the sequencing reads. After MID separation, SeqClean (Pertea, 2005) was used in quality trimming. LUCY (Chou & Holmes, 2001) was used to remove 3' homopolymers resulting in quality trimmed reads. Assembly of the quality trimmed reads into contigs was performed using MIRA (Chevreaux, 1997). See Table 1 for the N50 values of each assembly.

A.2.4 HiSeq Transcriptome Sequencing

5 **µ**g of Leggett RNA from each flower development stage (1-3 Days Post Anthesis, 4-6 DPA, and 7-9 DPA) was input to prepare the HiSeq libraries. The library preparation and sequencing was done by DHMRI at the NCRC in Kannapolis, NC. Samples were run on separate lanes of a flow cell on the Illumina HiSeq2000.

A.2.5 HiSeq Transcriptome Assembly

The Bioinformatics Service Division also performed the assemblies of Illumina HiSeq reads. See Chapter 3, section 3.2.2 Total RNA Sequencing and Assemblies for software and parameters used in transcriptome assembly from HiSeq reads. Summary statistics of the assembly of the transcriptome assemblies from Leggett are shown in Table 2.

A.2.6 Small RNA Sequencing and Analysis

Methods for *Avena strigosa* small RNA sequencing are described in Chapter 3.2.3 Small RNA Sequencing Tissue Collection and RNA Extraction and section 3.2.4 Small RNA Sequencing and Sequence Pre-processing. Plants of tetraploid oat *Avena magna* were grown as specified in 3.2.1 Total RNA Sequencing Tissue and RNA Extraction; but

small RNA extraction of *Avena magna* was done as described in Chapter 3, section 3.2.3. *Avena sativa* cultivar Leggett plants were grown under conditions specified in Chapter 2, section 2.2.2 Tissue and RNA Isolation and small RNAs were sequenced per 2.2.4 Small RNA Sequencing and Analysis. The pre-processing of small RNA reads were done as described in 2.2.4 Small RNA Sequencing and Analysis. Table 3 shows the number of reads after each step of pre-processing.

A.2.7 Identification of Candidate miRNAs and Corresponding Precursors

A modified miRDeep-P pipeline was used to identify miRNA precursors. 454 and HiSeq data were indexed and aligned to the corresponding pre-processed small RNA libraries by Bowtie-1.0.0, allowing 1 mismatch despite the recommendation of perfect match by miRDeep-P. Due to the genome complexity and pooling of cultivars, the unassembled, cleaned 454 reads from LUCY were used for precursor prediction as well. See Chapter 2, section 2.2.6 MiRNA Discovery in the HiSeq Transcriptome for additional parameters used in the modified miRDeep-P pipeline. The naming scheme for candidate miRNAs is the same used in Chapter 2 and 3.

A.2.8 Additional Candidate MiRNAs in *Avena strigosa* (AA)

Only two candidates were identified from the 454 Mira assemblies; therefore, the cleaned and trimmed LUCY 454 reads were also input for the miRDeep-P pipeline. 6 more miRNAs (Table 4), including the only predicted conserved miRNA identified in *Avena strigosa* by miRDeep. Similar to classifying known miRNAs in Chapter 3, Vmatch was used, allowing four mismatches to detect remote sequence similarity. Of 20 conserved miRNA families, a candidate miRNA sequence with high sequence similarity to annotated miR159 family members was identified (Figure 2), however, this candidate

sequence has a mismatch to its predicted precursor. The sequence that would be produced from the hairpin does not exist in the *Avena strigosa* small RNA library. This is also the case for the other two candidate miRNAs with one mismatch to their predicted precursors. These candidates are highlighted in Table 4. Predicted hairpins for all candidate miRNA sequences are in Figures 1-7.

A.2.9 Additional Candidate MiRNAs in *Avena magna* (CCDD)

Table 5 shows candidate miRNA sequences with predicted precursors from the 454 sequencing of flower tissue cDNA. The inclusion of predictions from the 454 cleaned reads added only 3 new predictions. Figures 8-14 show the structures of candidate miRNAs predicted from the 454 assemblies and 454 cleaned unassembled reads. miRBase was searched for sequence similarity to the predicted miRNAs to determine if any of the candidate miRNAs belonged to families of known miRNAs. The predicted precursor for a member of miR166 was predicted from both the 454 cleaned reads and the 454 transcriptome assemblies, ama_cand9x67420 (Figure 8). Four hairpins have a mismatch between the predicted precursor for the candidate miRNA due to allowing for 1 mismatch during the Bowtie alignment. These are highlighted in Table 5. The sequences that would be produced by the hairpin exist for three of the four candidate miRNAs, however, they have read counts of 2 or 3. One, ama_cand15x4 has sequence similarity to annotated miRNAs from *Brachypodium* and the non-conserved miR5049 family. This candidate miRNA may be a member of the miR5049 family. The significance of low read counts call for stem-loop RT-PCR to validate classification as miRNAs. The miRNA sequence produced from the hairpin for ama_cand13x11 does not exist in the *Avena magna* small RNA library, suggesting an error in 454 assemblies or

that the hairpin is not processed by DCL1 into a miRNA. Although Rfam did not filter out pre-ama_cand13x11, it may represent a non-coding RNA not annotated in Rfam.

A.2.10 Additional Candidate MiRNAs in *Avena sativa*, Cultivar Leggett (AACCDD)

The modified miRDeep-P pipeline identified 17 candidate miRNAs and their precursors in the HiSeq cDNA assemblies, 12 in the 454 cleaned reads, and 3 candidate miRNAs in the 454 MIRA assemblies. The additional candidate miRNA sequences identified for *Avena sativa*, cultivar Leggett are given in Table 6. Two candidate miRNAs, leg_cand10x23 and leg_cand15x8 were predicted from HiSeq assemblies of pooled stages 1-3 and 4-6 DPA of Leggett in Chapter 2. The hairpin predicted in Chapter 2 for leg_cand10x23 had a mismatch between it and its precursor as well as the hairpin predicted from 454 cleaned reads (Figure 15). A third hairpin (Figure 16), different from both hairpin 1 and 2, was predicted for leg_cand10x23 in the HiSeq assemblies from 1-9 DPA. With no mismatches, hairpin 3 is most likely the true precursor for the candidate miRNA. For leg_cand15x8, all hairpins predicted, one from Chapter 2, one from 454 cleaned reads, and two from the HiSeq assembly 1-9 DPA, have a mismatch between it and the candidate miRNA sequence. The hairpins of leg_cand15x8 are shown in Figures 17-19.

Only three candidate miRNAs were predicted from the 454 assemblies (Figures 20-22) with two overlapping with predictions from the 454 cleaned reads. The lack of multiplexing each cultivar, or line of oat in these libraries may have contributed to such little overlap in predictions between the 454 assemblies and cleaned reads.

The hairpins for the candidate miRNAs predicted from the 454 clean reads that do not overlap the assembly predictions are shown in Figures 23-30. Only 4 candidate

miRNAs had hairpin predictions in both the HiSeq and 454 clean reads. These are leg_cand10x23 and leg_cand15x8 discussed above, as well as leg_cand35x5 and leg_cand37x4. For the two latter candidate miRNAs, their respective hairpin prediction from the 454 clean reads was identical in sequence to the prediction from HiSeq transcriptome assemblies; therefore, only one hairpin is shown for each candidate miRNA.

Hairpins predicted from the HiSeq assemblies are shown in Figures 31-40. Hairpins for candidate miRNAs with read counts below 4 are not shown.

A.3 Summary

As seen in Table 4, 5, and 6, highlighted sequences have a mismatch between the predicted precursor and the candidate miRNA sequence. This suggests errors in the assemblies or that the hairpins are not the origin of the candidate miRNAs. For many of these hairpins, the miRNA that would be produced by the hairpin does not exist in the small RNA library. For Leggett, this is the concern for 14 of 15 hairpins with mismatches to the candidate miRNA sequence. Aside from subgenomic complexity challenging the assemblies of HiSeq and 454 cDNA sequences of Leggett, the assembly of 454 clean reads has an additional challenge: cDNA was pooled from multiple oat cultivars and MID-tagged for stage. Mismatches between candidate miRNA sequences and their hairpins in the 454 clean reads could be a result of MID-tagging for stage and not cultivar: the hairpin predicted might be from one of the other two oat lines that were pooled with Leggett. False positive predictions are also reasonable excuses for mismatches between candidate miRNAs and their predicted hairpins in Leggett, but also *Avena strigosa,* and *Avena magna*.

For Leggett, *Avena strigosa*, and *Avena magna*, the candidate miRNAs with 0 mismatches to their respective hairpins are priority for validating with stem-loop RT-PCR to annotate as novel miRNAs. Conserved miRNA family predictions do not need to be experimentally validated prior to annotation. Only miR166 predicted from *Avena magna* can be annotated. All others candidate miRNAs remain "candidates" until they can be experimentally validated.

Table 1: Sequencing statistics of 454 transcriptome assemblies. For miRNA prediction, the three Leggett libraries were pooled. The '*' indicates Leggett tissue was pooled with two other hexaploid oat lines.

Library	Lines	Number of Cleaned LUCY Reads	Number of Assembled Sequences	N50
1-3 days post anthesis (DPA)	Leggett*	20,011	3,703	467
4-6 DPA	Leggett*	74,229	12,650	591
7-9 DPA	Leggett*	34,447	6,255	491
1-9 DPA	*A. strigosa*	71,189	11,216	656
1-9 DPA	*A. magna*	142,478	18,601	662

Table 2: Sequencing statistics of HiSeq2000 assemblies for *Avena sativa* cultivar Leggett. These assemblies were pooled for miRNA prediction.

Library	Lines	Number of Sequences	N50
1-3 days post anthesis (DPA)	Leggett	28,668	1,695
4-6 DPA	Leggett	31,142	1,665
7-9 DPA	Leggett	44,656	1,813

Table 3: Small RNA pre-processing. This table shows the raw reads of the pooled libraries and the corresponding growth conditions.

19° Celsius	26° Celsius	Lines	Raw reads	Quality Filter	Adapter Trimmed, 18-30	Non-Redundant	Number of Singlets	Filtered Sequences for miRNA Discovery
0-9 DPA	0-9 DPA	Leggett	13,909,958	11,407,867	11,007,928	3,296,119	2,638,498	657,621
0-6 DPA	1-9 DPA	*A. strigosa*	16,002,344	15,111,759	6,620,254	1,759,976	1,390,365	369,611
1-9 DPA, mature embryo, roots, shoot 20°/15° Celsius light/dark		*A. magna*	15,159,616	8,675,352	6,379,744	1,828,985	1,447,654	381,331

Table 4: Candidate miRNAs in *Avena strigosa* predicted from 454 sequencing. Only two hairpins from the MIRA assemblies passed plant filtering criteria; therefore the cleaned 454 reads were also used for more predictions. Highlighted rows indicate a mismatch between the candidate miRNA and its predicted precursor.

Predicted miRNA Candidate ID	Number of Reads	Sequence	miRBase Hit (<3 mismatches)	454 Transcriptome Cleaned Reads	454 MIRA Transcriptome Assembly
str_cand13x1274	1274	GGTAGTTCGACCGCGGAATT		X	
str_cand14x123	123	TTTGGATTGAAGGGAGCT	miR159	X	
str_cand15x12	12	GAATTCCACCACGTTCCGTGG		X	
str_cand16x11	11	AGCTAGACTGGTGGCGGCGCT			X
str_cand17x8	8	CCTAGCTGCTGAGAGTCTCCCCCA		X	
str_cand18x7	7	AGCCAGAAATCTGATCCA		X	X
str_cand19x5	5	CCTAGCTGCTGAGAGTCTCCCCC		X	



Figure 3: Pre-str_cand13x1274.



Figure 2: Pre-str_cand14x123. This small RNA sequence is a member of the miR159 family. The red "G" indicates the candidate sequence does not match the precursor identically.

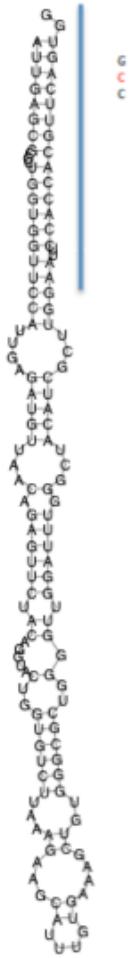


Figure 3: Pre-str_cand15x12. The red "C" indicates the candidate sequence does not match the precursor identically.

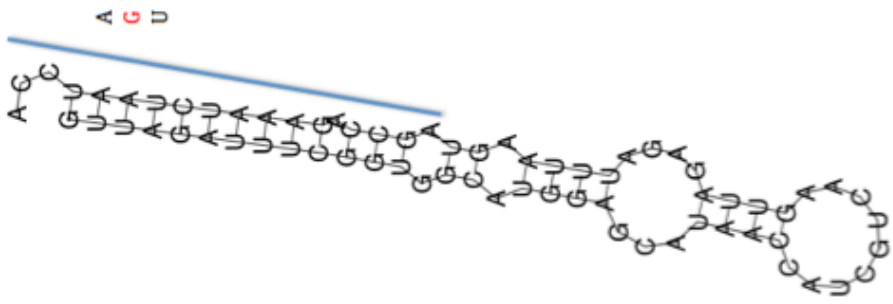


Figure 6: Pre-str_cand18x7.

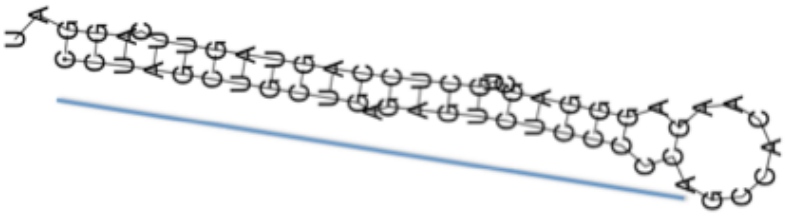


Figure 5: Pre-str_cand17x8.

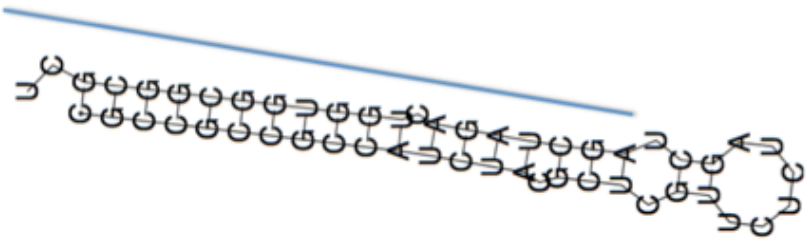


Figure 4: Pre-str_cand16x11.

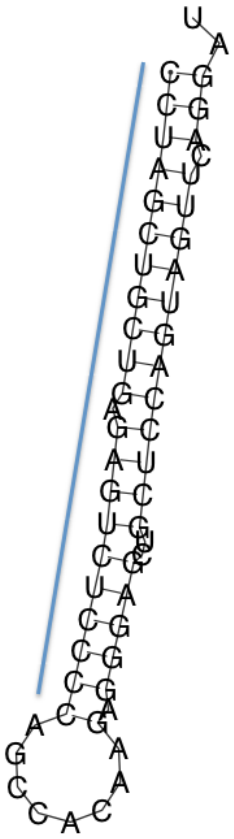


Figure 7: Pre-str_cand19x5.

Table 5: Candidate miRNAs in *Avena magna* predicted from 454 Sequencing. The cleaned 454 reads were also used for more predictions. Highlighted rows indicate a mismatch between the candidate miRNA and its predicted precursor.

Predicted miRNA Candidate ID	Number of Reads	Sequence	miRBase Hit (<3 mismatches)	miRBase Hit (3 mismatches)	miRBase Hit (4 mismatches)	454 Transcriptome Cleaned Reads	454 MIRA Transcriptome Assembly
ama_cand9x67420	67420	TCGGACCAGGCTTCATTCCCC	miR166			X	X
ama_cand10x49	49	TAAAAGAGGTCTAGATACACT					X
ama_cand11x44	44	TATAAAGAAAGTCATGAACATT					X
ama_cand12x17	17	TTGTCTAGATTCGTATGTATC				X	
ama_cand13x11	11	TCGCAGATTTGTCTAAATTTG				X	
ama_cand14x10	10	TCCAAGACGTGGACTATCACCT					X
ama_cand15x4	4	ACTTATTATGGATCAGAGGGA	bdi-miR5203	miR5049	bdi-miR5181	X	

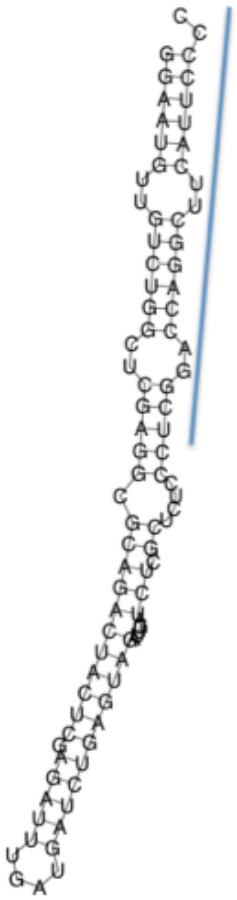


Figure 8: Pre-ama_cand9x67420. This candidate miRNA is a member of the conserved miR166 family.

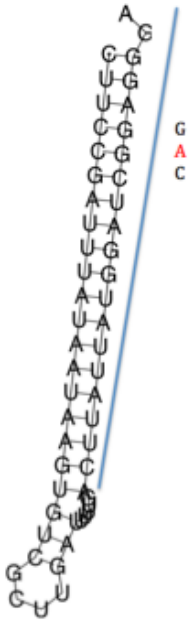


Figure 9: Pre-ama_cand15x4. This candidate miRNA has sequence similarity to miR5203 annotated in Brachypodium. The red "A" indicates a mismatch between the small RNA sequence and the predicted precursor.

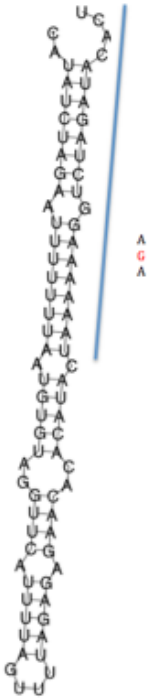


Figure 10: Pre-ama_cand10x49. The red "G" indicates a mismatch between the small RNA sequence and the predicted precursor.

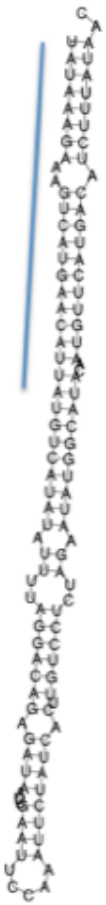


Figure 11: Pre-ama_cand11x44.

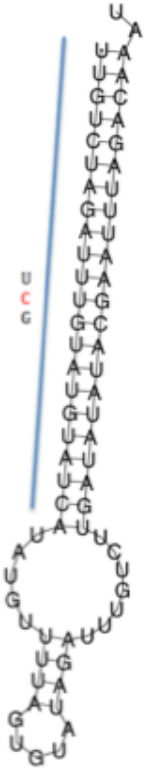


Figure 12: Pre-ama_cand12x17. The red "C" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

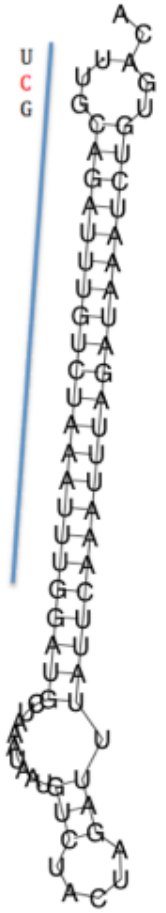


Figure 13: Pre-ama_cand13x11. The red "C" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

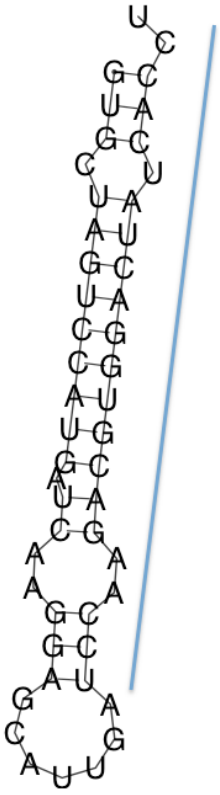


Figure 14: Pre-ama_cand14x10.

Table 6: Candidate miRNAs in *Avena sativa* cultivar Leggett predicted from HiSeq and 454 transcriptome sequencing. 454 predicted hairpins are cleaned reads or MIRA assemblies of Leggett, pooled with two other hexaploid oat lines of the same development stages (1-9 DPA). Highlighted rows indicate a mismatch between the candidate miRNA and its predicted precursor.

Predicted miRNA Candidate ID	Number of Reads	Sequence	miRBase Hit (3 mismatches)	miRBase Hit (4 mismatches)	HiSeq Transcriptome Assembly	454 Transcriptome Cleaned Reads	454 MIRA Transcriptome Assembly
leg_cand18x112	112	GTTGAGAACAATAATATGGAT			X		
leg_cand19x95	95	GAGGATGAGGATGAGGATGAC			X		
leg_cand10x23 (AB)	42	TCTGTTCCATAATATAAGCCT			X	X	
leg_cand20x33	33	GTGCTGTGGCCGGCCGCC				X	
leg_cand21x25	25	TTGTCTAGATTCGTATGTATC				X	
leg_cand15x8 (AB)	24	TCCGTTCCATAATATAAGCCT			X	X	
leg_cand22x16	16	GAAGGGGAGGGCCTGAAGCCCT			X		
leg_cand23x16	16	AAATTTGAGTCAACTAATATGGGT					X
leg_cand24x15	15	CCACCAAGATCTCTGAGGCCC		bdi-miR7732	X		
leg_cand25x14	14	TCCAAGTAGCTGTCGCACG			X		
leg_cand26x12	12	AATGGCCAAGTAACTTTGCATGAG				X	
leg_cand27x11	11	TTAATTTGGATCGGAGGGAGT		bdi-miR7714, mtr-miR5205		X	X
leg_cand28x11	11	AACGATGTGGTAGAAAGCA				X	
leg_cand29x10	10	TGAGGAGGATGAGGATGAGGAT			X		
leg_cand30x9	9	TCCATGTCCCGTCTGCCCG				X	X
leg_cand31x7	7	CTCCGTCCAGAAATAAGTGAC				X	
leg_cand32x6	6	TAACGAACGAACGATTTGA			X		
leg_cand33x6	6	GTTTTGACGAGGTTGCTGGGGCA			X		
leg_cand34x6	6	TTATATTGTGGAACAGAGGGA			X		
leg_cand35x5	5	CGCAGATTTGTCTAGATTCGC	bdi-miR5164		X	X	
leg_cand36x5	5	AATTAATGACAATCTCTATGGACT				X	
leg_cand37x4	4	TGTTCTGACCAGATTGGATGCCTGCT			X	X	

Table 6… (continued): Candidate miRNAs in *Avena sativa* cultivar Leggett predicted from HiSeq and 454 transcriptome sequencing.

leg_cand38x4	4	CAGTTAGAGCTGATGGAACC			X		
leg_cand39x3	3	AGGTGCATCGCGGACATCTGCT			X		
leg_cand40x2	2	TGCTCGAAAGATCCCAGACACG			X		
leg_cand41x2	2	GGCGGAGGGGCTGGAGGACC			X		

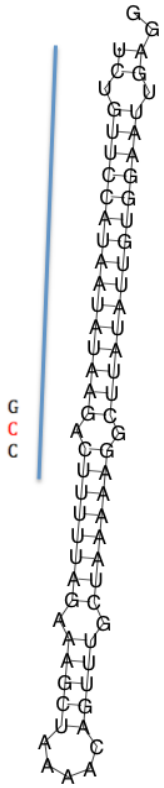


Figure 15: Pre-leg_cand10x23 hairpin 2. Hairpin 1 was predicted from Chapter 2. The red "C" indicates a mismatch between the small RNA sequence and the predicted hairpin. This hairpin was predicted from 454 sequencing.

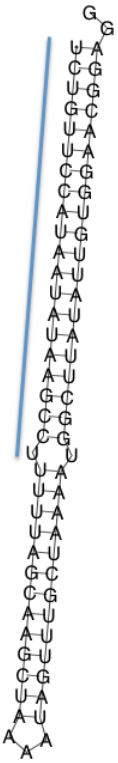


Figure 16: Pre-leg_cand10x23 hairpin 3. Hairpin 1 was predicted from Chapter 2. This hairpin was predicted from HiSeq transcriptome assemblies of 1-6 DPA.

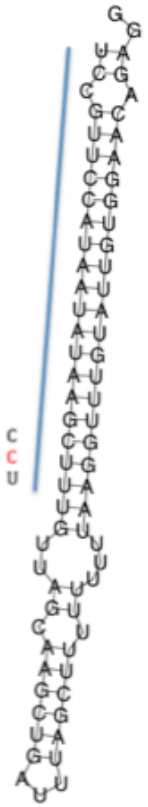


Figure 17: Pre-leg_cand15x8 hairpin 2. Hairpin 1 was predicted from Leggett cDNA assemblies from Chapter 2. The red "C" indicates a mismatch between the candidate miRNA sequence and its predicted precursor. This hairpin was predicted form 454 cDNA sequencing.



Figure 18: Pre-leg_cand15x8 hairpin 3. Hairpin 1 was predicted from Leggett cDNA assemblies from Chapter 2. The red "U" indicates a mismatch between the candidate miRNA sequence and its predicted precursor. This hairpin was predicted from HiSeq transcriptome assemblies.

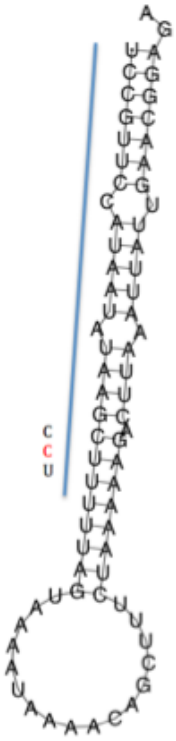


Figure 19: Pre-leg_cand15x8 hairpin 4. Hairpin 1 was predicted from Leggett cDNA assemblies from Chapter 2. The red "C" indicates a mismatch between the candidate miRNA sequence and its predicted precursor. This hairpin was predicted from HiSeq transcriptome assemblies.

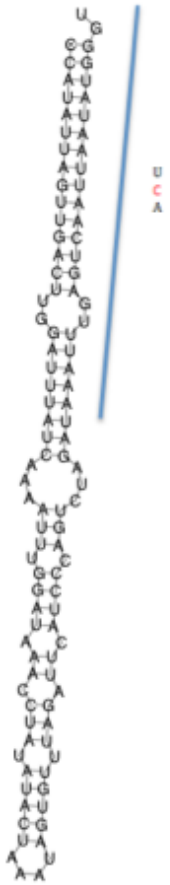


Figure 20: Pre-leg_cand23x16. The red "C" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

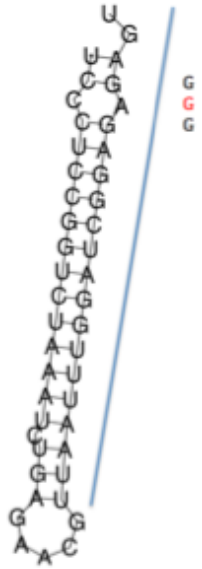


Figure 21: Pre-leg_cand27x11. The red "G" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

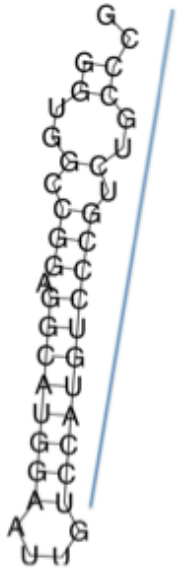


Figure 22: Pre-leg_cand30x9.

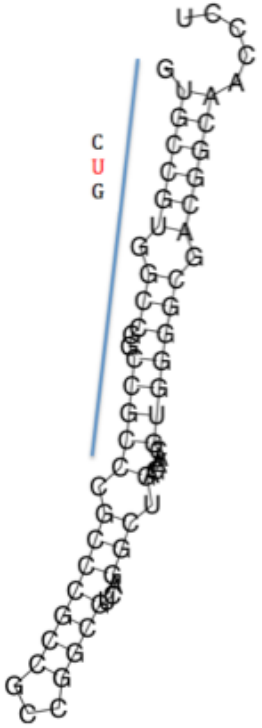


Figure 23: Pre-leg_cand20x33. The red "U" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

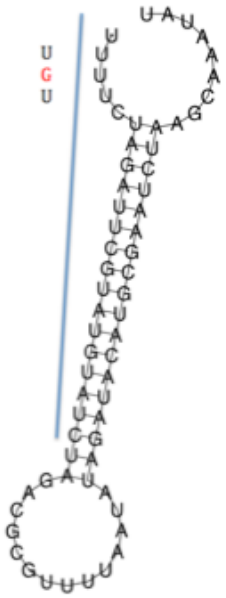


Figure 24: Pre-leg_cand21x25. The red "G" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

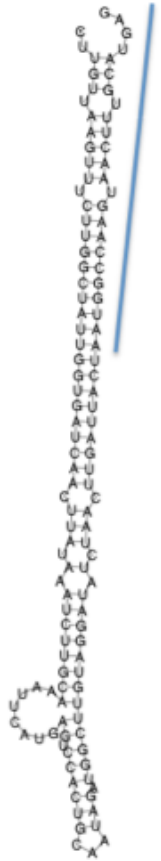


Figure 25: Pre-leg_cand26x12.



Figure 26: Pre-leg_cand28x11.

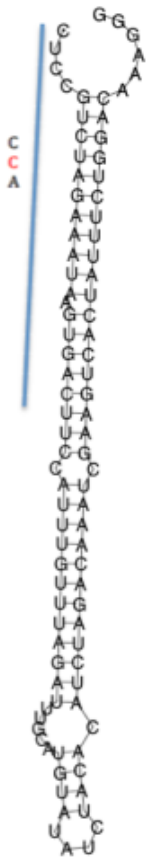


Figure 27: Pre-leg_cand31x7. T The red "C" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.



Figure 28: Pre-leg_cand35x5. The red "C" indicates a mismatch between the candidate miRNA sequence and its predicted precursor. This hairpin was identified as precursor for this candidate miRNA from HiSeq assemblies and 454 clean reads.

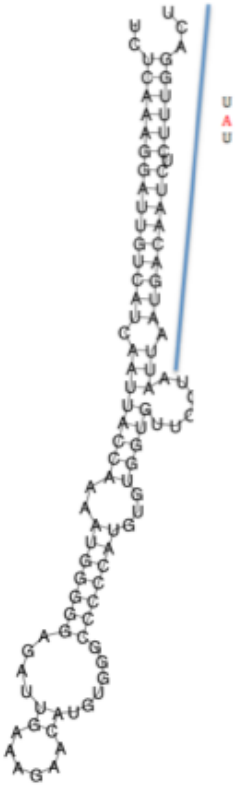


Figure 29: Pre-leg_cand36x5. The red "A" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.



Figure 30: Pre-leg_cand37x4. This hairpin was identified as a miRNA precursor from HiSeq assemblies and 454 clean reads.

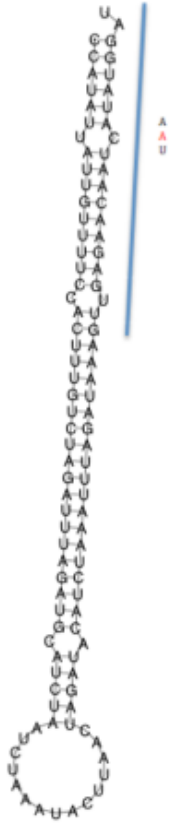


Figure 31: Pre-leg_cand18x112. The red "A" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

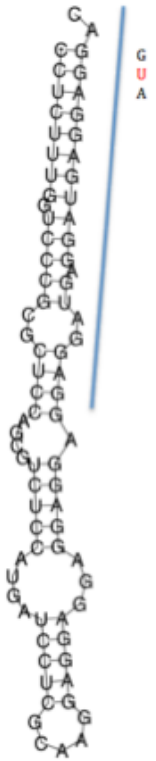


Figure 32: Pre-leg_cand19x95. The red "U" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

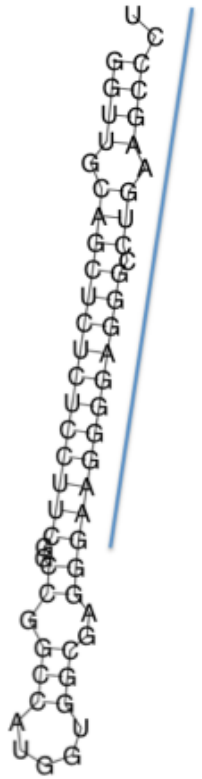


Figure 33: Pre-leg_cand22x16.

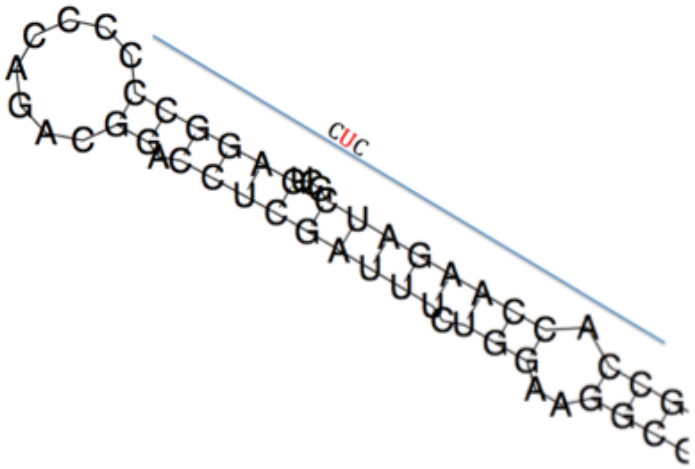


Figure 34: Pre-leg_cand24x15. The red "U" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

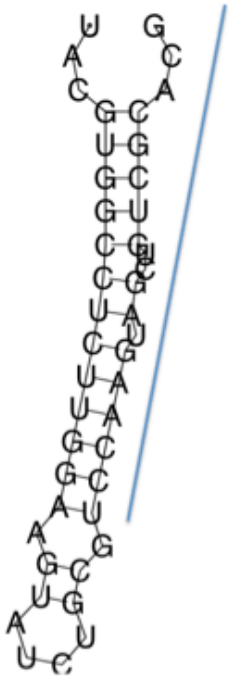


Figure 35: Pre-leg_cand25x14.

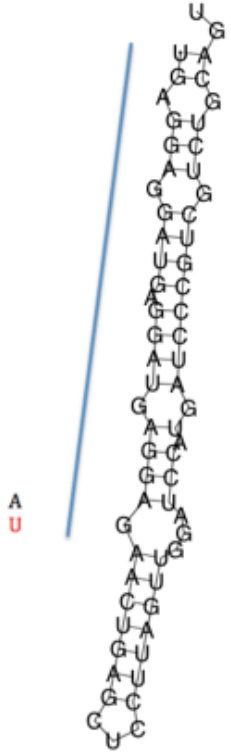


Figure 36: Pre-leg_cand29x10. The red "U" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

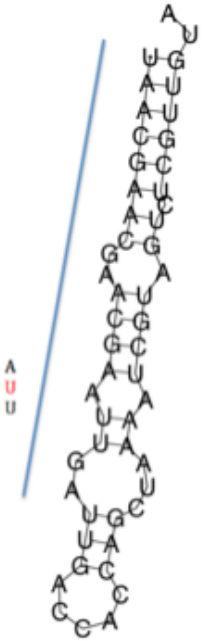


Figure 37: Pre-leg_cand32x6. The red "U" indicates a mismatch between the candidate miRNA sequence and its predicted precursor.

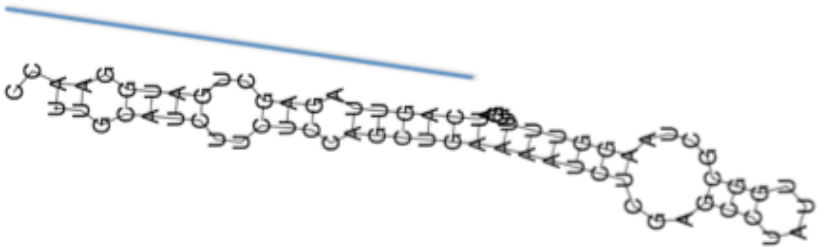


Figure 40: Pre-leg_cand38x4.

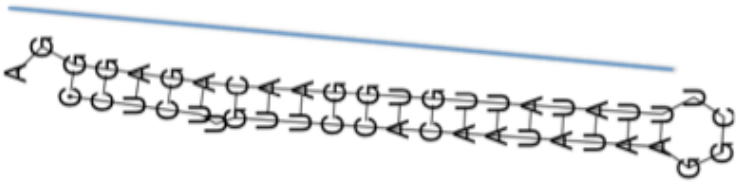


Figure 39: Pre-leg_cand34x6.

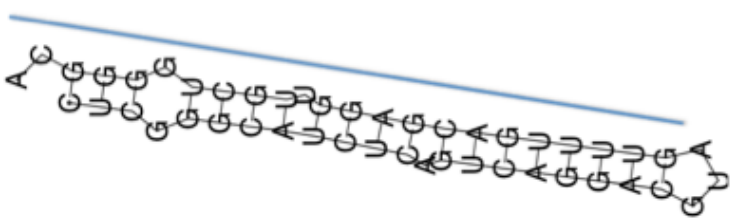


Figure 38: Pre-leg_cand33x6.