PLANT NONCODING RNAS AND COMPARATIVE ANALYSIS OF INSECT TOLERANCE LOCI

by

TUĞDEM MUSLU

Submitted to the Graduate School of Engineering and Natural Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy of Science in Molecular Biology, Genetics and Bioengineering

Sabancı University

December 2021

PLANT NONCODING RNAS AND COMPARATIVE ANALYSIS OF INSECT TOLERANCE LOCI

APPROVED BY:

Assoc. Prof. Dr. MERAL YÜCE.....
(Thesis Supervisor)

Dr. BALA ANI AKPINAR.....
(Thesis Co-Supervisor)

Prof. Dr. ALİ KOŞAR.....

Asst. Prof. Dr. BAHAR SOĞUTMAZ ÖZDEMİR.....

Asst. Prof. Dr. HASAN KURT.....

Prof. Dr. LEVENT ÖZTÜRK.....

DATE OF APPROVAL: 17/12/2021

ABSTRACT

PLANT NONCODING RNAS AND COMPARATIVE ANALYSIS OF INSECT RESISTANCE LOCI

TUĞDEM MUSLU

Molecular Biology, Genetics and Bioengineering, Ph.D. Thesis, 2021

Supervised by: Assoc. Prof. Dr. Meral Yüce

Co-supervised by: Dr. Bala Anı Akpinar

Keywords: cereals, microRNA, pangenome, insect resistance, noncoding RNA

As the growing world population tremendously increases the demand for food every year, challenging environmental conditions continue to threaten agricultural productivity. While plants try to cope with many abiotic stresses such as drought, salinity, heat, and cold, they also have to cope with many biotic stress factors such as insect pests of plants. Obtaining higher-yielding and more resilient crops thus is a necessity to achieve global food security and sustainability. Advances in genome sequencing techniques over the past 15 years revolutionized the way we perceive genomes and comparative genomics studies enabled the analysis of many conserved and diverged inter/intra-specific genomic features of a wide range of organisms.

Within the scope of this thesis, we constructed pangenome miRNomics of *Brachypodium* using *de novo* genome assemblies of 54 lineages and presented the conservation levels of microRNA families among lineages and their putative targets to uncover the molecular basis of agronomic traits in different lineages. In the second chapter, we carried out a comparative and evolutionary analysis of coding and non-coding features of 4 insect tolerance loci within wheat among other cereals, barley, rye, rice, and oat, and showed different levels of synteny in homologous insect tolerance loci across *Poaceae* species. In the third chapter, we have carried out miRNA and lncRNA identification in 3 barley cultivars, aiming to unravel the interaction network of barley mRNA-miRNA-lncRNAs. The results of this thesis will shed a light on future gene cloning and editing studies of monocots for food safety and security.

ÖZET

BİTKİLERDE KODLANMAYAN RNA VE BÖCEK TOLERANSI LOKUSLARININ KARŞILAŞTIRMALI ANALİZİ

Tuğdem Muslu

Moleküler Biyoloji, Genetik ve Biyomühendislik, Doktora Tezi, 2021

Tez Danışmanı: Doç. Dr. Meral Yüce

Eş Tez Danışmanı: Dr. Bala Anı Akpınar

Anahtar sözcükler: Tahıl, mikroRNA, pangenom, kodlanmayan RNA, böcek toleransı

Artan dünya nüfusu gıda talebini her geçen yıl artırırken, zorlu çevresel koşullar tarımsal verimliliği tehdit etmeye devam ediyor. Bitkiler kuraklık, tuzluluk, sıcak ve soğuk gibi birçok abiyotik stresle baş etmeye çalışırken, birçok biyotik stres faktörüyle de karşı karşıya gelmektedir. Daha yüksek verimli ve daha dayanıklı gıda ürünleri elde etmek, küresel gıda güvenliğini ve sürdürülebilirliğini sağlamak için bir zorunluluktur. Son 15 yılda genom dizileme tekniklerindeki ilerlemeler, genomları algılama şeklimizde bir devrim yarattı ve karşılaştırmalı genomik çalışmaları, birçok canlıda evrimsel olarak korunan veya korunmayan, türler arası veya tür içi birçok genomik özelliğin analizini mümkün kıldı.

Bu tez kapsamında, 54 soyun *de novo* genom dizisi kullanarak *Brachypodium distachyon* pangenom miRNomiklerini oluşturduk ve farklı soylardaki agronomik özelliklerin moleküler temelini ortaya çıkarmak için soylar arasındaki mikroRNA ailelerinin koruma düzeylerini ve tahmin edilen mRNA hedef proteinlerini gösterdik. İkinci bölümde, daha önce buğdayda tanımlanan 4 böcek toleransı lokusunun arpa, çavdar, pirinç ve yulaf türlerindeki homologlarında bulunan kodlanan ve kodlanmayan genomik dizilerinin karşılaştırmalı ve evrimsel analizini yaptık ve Buğdaygiller arasında homolog böcek toleransı lokuslarının farklı sinteni seviyeleri gösterdik. Üçüncü bölümde, arpa mRNA-miRNA-lncRNA'larının etkileşim ağını çözmeyi amaçlayarak 3 arpa çeşidinde miRNA ve lncRNA analizi gerçekleştirdik. Bu tezin sonuçları, gıda güvenliği ve sürdürülebilirliği için gelecekteki monokotiledon bitkilerde gen klonlama ve düzenleme çalışmalarına ışık tutacaktır.

ACKNOWLEDGEMENT

Having completed this journey, I would like to express my sincere thanks to those distinguished people for their outright contributions and help during this period.

First, I would like to thank my advisor, Assoc. Prof. Dr. Meral Yüce for giving a helping hand at the most difficult times, for her endless support, and for being the mentor a student can ever ask for. I would like to express my infinite gratitude to my co-advisor, Dr. Bala Anı Akpınar, for her guidance and endless patience. I am so lucky to work together; this could not have been possible without her. I also would like to thank Prof. Dr. Hikmet Budak for sharing his knowledge, wisdom, and support. I cannot thank him enough for motivating me every time I decided to quit.

I owe a deep sense of gratitude to my thesis committee: Prof. Dr. Levent Öztürk, Prof. Dr. Ali Koşar, Asst. Prof. Dr. Bahar Soğutmaz Özdemir and Asst. Prof. Dr. Hasan Kurt for their insightful comments and guidance.

I would like to especially thank Dr. Kadriye Kahraman. I am always thankful for our friendship and for being able to walk this path together, I wouldn't have gotten to this point without her. I would like to thank Dr. Halise Büşra Çağırıcı and Sezgi Bıyıklıoğlu Kaya for sharing their knowledge and for their companionship.

I would like to express my gratitude to the NS and ASP team of Sabancı University, for making me a part of the team, for their support and friendship. They made this journey so much better.

I cannot thank enough to my parents, my brother, my aunt, and my dear grandparents for always loving and supporting me. Everything I am today, I owe to them.

Above all, I would like to thank my dear husband, Bolkan, for the biggest love and support one could ever have, for raising me when I was closest to giving up, and for making me feel that nothing can hurt me when he is around.

To the little girl who used to experiment

with a magnifying glass and flashlight,

this is for you…

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABA……Abscisic acid hormone
AGO……..……..Argonaute
Arg…….…...……Arginine
bdi……...……..…….……Brachypodium distachyon
BLAST……...……Basic Local Alignment Search Tool
bp……...…...Basepair
C……..……..Cytosine
CDS……Coding Sequences
CNCI……...…Coding/Noncoding Index
CNV…….……..Copy number variant
CPC…….……Coding Potential Calculator
cv……Cultivar
DCL…….……...……..Dicer-like
DNA……...……..Deoxyribonucleic acid
E.…….……Embryophyta
EDF……...…….....Extremely Delayed Flowering
En/Spm…….Enhancer/Suppressor mutator
ENA……...……..…European Nucleotide Archive
EST……...……...……Expressed Sequence Tag
FAO……...……..Food and Agriculture Organization
FT……...……Flowering locus T
G……...……..Guanine
GFF……...……..……Gene-finding / General feature format
Gly……..……Glycine
GO…….……..Gene Ontology
HEN……..……..HUA Enhancer
hr……..……hour
HST……..……..HASTY
hvu……...……Hordeum vulgare
IBSC…….……International Barley Sequencing Consortium
IRGSP……...International Rice Genome Sequencing Project

IWGSC……..….International Wheat Genome Sequencing Consortium
iTRAQ……..……Isobaric Tag for Relative and Absolute Quantitation
JA……..Jasmonic acid
Leu……..……Leucine
lncRNA……...……..Long noncoding RNA
LRR……..……Leucine-rich repeat
Lys……Lysine
M…….……Magnoliophyta
MAFFT……...……Multiple Alignment using Fast Fourier Transform
Mb……...……Megabase
Met……Methionine
MFEI……...…..……Minimum folding energy index
MIPS……...……..……Munich Information Center for Protein Sequence
MITE...…….……Miniature Inverted Repeat Transposable Element
miRNA……..MicroRNA
MRG……Morf related gene
mRNA……..Messenger ribonucleic acid
MTERF…...……Mitochondrial transcription termination factor family
NB ……Nucleotide-binding
NCBI…….……National Center for Biotechnology Information
ncRNA……..……Noncoding RNA
NGS……...……Next-Generation Sequencing
NLR…….……Nucleotide binding-site leucine-rich repeat
nt……..…….……..…….…….…..……..Nucleotide
ORF……..……..…..……..…..…..…….…..……Open reading frame
OWBM…….……..Orange Wheat Blossom Midge
PAV……..……..Presence/Absence Variation
PHD ……..……Plant homeodomain
PIK……..……Phosphatidylinositol Kinase
PPR…….……Pentatricopeptide repeat-containing protein
premiRNA……...…….…….....Precursor microRNA
primiRNA……..……..……..…….…….……..Primary microRNA
psRNA……..…….....…..…….....……Plant small RNA
QTL……...……..Quantitative trait locus

RefSeq…… Reference Sequence
RGA……Resistance Gene Analog
RISC……..RNA-induced Silencing Complex
RNA……..ribonucleic acid
rRNA……ribosomal RNA
SBP…….....……Squamose promoter binding protein
seq……Sequence
Ser…….....……Serine
siRNA……...…….....……Small interfering ribonucleic acid
SNP……..Single nucleotide polymorphism
sRNA……..……..small RNA
tRNA……...Transfer RNA
UPE……..……Unpaired energy
USD……...……United States dollar
USDA……..……United States Department of Agriculture
WSS……...……..Wheat Stem Sawfly

1. GENERAL INTRODUCTION

Food security is defined as a situation where all people, in every country, at all times, have access to safe and nutritious food that is sufficient in quantity (World Food Summit, 1996). Approximately one of every three individuals in the world did not have access to enough food in 2020 and around 768 million people were chronically hungry, 118 million more people than in the previous year. The prevalence of undernourishment continues to increase in recent years despite all joint efforts committed to ending world hunger by 2030 (FAO-Food and Agricultural Organization of the United Nations, 2020; World Food Programme, 2022). The world population is expected to reach 9.7 billion by 2050, and the rising food demand will increase the pressure on agricultural production (FAO-Food and Agricultural Organization of the United Nations, 2020).

To forestall the food shortage and to feed the world, immediate action should be taken to improve agricultural production. Climate change has dramatic effects on agricultural production causing many abiotic stresses such as extreme drought conditions, floods, heatwaves, high salinity, nutrient immobilization, and land degradation (Arora, 2019). The projected increase of more than 4°C in the world's temperature in upcoming years will pose an even greater risk to the food security (Govindaraj, Pattanashetti, Patne, & Kanatti, 2018), especially a decline in the production, quality, and access of major cereal crops such as wheat, maize, and rice (Arora, 2019). These cereals are the main crops that supply more than 42% of the calorie intake of the global population, therefore sustainable production of them holds a great importance (Matres et al., 2021). Extreme weather conditions will cause the availability of water resources to decrease and every living thing to suffer from water shortage in the future years (da Silva, de Albuquerque, de Azevedo Neto, & da Silva Junior, 2013). The consequence of water shortage, drought, is one the most detrimental abiotic stresses limiting crop production causing a 9-10% decline

globally (J. Zhang et al., 2018) and this percentage would be much higher without irrigation (da Silva et al., 2013). Irrigation to overcome the effects of drought in arid or semi-arid regions may result in another stress, soil salinity, which is a global threat to crop production. Salinity affects a total area of 1 billion hectares and it is estimated that this number increases by 1.5 million hectares every year (Carillo, Annunziata, Pontecorvo, Fuggi, & Woodrow, 2011; Ivushkin et al., 2019; Munns & Tester, 2008). Besides the mentioned abiotic stress factors, the major crops also face many biotic stress factors caused by living organisms, such as insects, fungi, viruses, and bacteria. Both abiotic and biotic stresses can alter plant response against other stress conditions (P. Pandey, Irulappan, Bagavathiannan, & Senthil-Kumar, 2017).

Wheat Stem Sawfly (WSS), *Cephus Cinctus Norton* (Hymenoptera: Cephidae) is the most damaging insect pest of wheat in northern Great Plains (Cockrell et al., 2017). Female WSS oviposit in the internodes of host wheat stem, wherein approximately 7 days the larvae will hatch. Due to the cannibalistic nature of larvae, only one larva survives and feeds on vascular bundles and pith tissue. Eventually, the larva starts to move down the stem, creates a lodge by cutting a notch at the lower part of the wheat stems (Biyiklioglu et al., 2018). Feeding on stem tissue and lodging causes a significant reduction in photosynthesis yield (Macedo, Peterson, Weaver, & Morrill, 2005), kernel weight, and grain quality (Morrill, Kushnak, & Gabor, 1998; Shrestha, Briar, & Reddy, 2018). Breeding solid-stemmed wheat is the only mechanism against WSS damage as it controls infestation but because the solid-stemmed varieties are less-yielding this is not desired by the producers (Beres, Cárcamo, Yang, & Spaner, 2011; Halise B. Cagirici, Biyiklioglu, & Budak, 2017). Another major insect pest of wheat in the Northern Hemisphere is Orange Wheat Blossom Midge (OWBM), *Sitodiplosis mosellana* (Thambugala et al., 2021). Adult midges lay eggs on the wheat spikes and hatched larvae feed on kernels causing detrimental damage to kernels (Blake et al., 2014). Antixenosis, abnormal/prevented oviposition and antibiosis, larval growth suppression are two host resistance mechanisms against OWBM damage (Kassa et al., 2016). In recent years, the major quantitative trait loci (QTLs) and causal genes for WSS and OWBM tolerance were identified (Blake et al., 2011; Hao et al., 2019; Kassa et al., 2016; Nilsen et al., 2020; Thambugala et al., 2021; Walkowiak, Gao, Monat, & et al., 2020; Lijing Zhang et al., 2020) and there are many studies focused on the improvement of those insect tolerance characteristics.

Traditional breeding strategies alone are insufficient to address the need for the development of stress-tolerant crops due to limitations in terms of labor and time. Hence, it is not surprising that the interest in plant research is increasing in recent years. Rapid progress in molecular biology and computational engineering has opened a door to a new breeding era and developed a better understanding of plants' mechanisms to cope with stress conditions. Advances in sequencing technologies have tremendously increased the availability of plant genome assemblies. While the reference genomes of many plant species have been available, understanding the complex mechanisms, genetic pathways, gene functions, and conducting experiments with many crop plants are still a challenge due to their large stature, numerous growth requirements, long generation times, and large genome sizes. This is where the model organisms come into play. *Arabidopsis thaliana* was the main plant model organism for decades which provided much insight into fundamental processes, but not all the information we have learned from *A. thaliana* can be transferred to the monocot cereals (Borrill, 2020). Rice (*Oryza sativa)* and maize (*Zea mays*) are two important monocot grass model organisms with a large research community. Rice is the first cereal with a complete genome sequence, and its small genome size makes it a good choice as a model organism (J. Yu et al., 2002). However, its high growth requirements and long generation time make rice a challenging model organism. On the other hand, maize has a moderately sized genome compared to other grasses (Strable & Scanlon, 2009). While being a large plant makes maize a favorable model plant for morphological analysis, this characteristic may be a challenge for the generation of many individuals and creating multiple generations (Brkljacic et al., 2011). Therefore, there was a need for a new model organism, smaller in size, with simple growth conditions, short generation time, and small genome size. *Brachypodium distachyon*, a member of the *Pooideae* subfamily along with wheat, barley, oat, and rye, was proposed as a model organism first in 2001 (Draper et al., 2001). Its small genome size, ease of generation, and nature amenability to genetic transformation make *Brachypodium* to be defined as the model organisms of monocots for the plant biology (Scholthof, Irigoyen, Catalan, & Mandadi, 2018).

Crop development techniques mostly rely on sequencing and phenotyping for the identification of genes and QTLs. Although phenotypic variations in plants were described for centuries (Mir, Reynolds, Pinto, Khan, & Bhat, 2019), the availability of

genome and transcriptome sequences in recent years revealed the importance of genetic variation. Evolution, breeding, domestication, mutation, and selection are the main driver of genetic diversity that provides adaptation and phenotypic variation, thus the continuation of the agricultural production (Temesgen, 2021). Therefore, identification of interspecific and intraspecific variations holds a great promise on the determination of evolutionary conserved and functionally important characteristics as well as genotypic diversity among the individuals of a species. While the reference genomes are widely used and sequenced over and over resulting in less error and more completeness, only one individual of a species cannot reflect the complete genetic variability of a species. Many traits may be specific to only some individuals because of copy number variant (CNVs), single nucleotide polymorphisms (SNPs), and presence/absence variants (PAVs) (Hurgobin & Edwards, 2017). To completely reflect the intraspecific variations and genetic content of a species, the construction of pangenomes was suggested for the first time by Tettelin et al. in 2005 (H. Tettelin et al., 2005). Pangenomes are the entire gene repertoire of a species that cannot be found in a single representative genome.

More knowledge about the protein-coding genes of the genome led to the discovery of the importance of non-coding parts of the genome. Non-coding sequences, which were previously referred to as “noise” and constitute approximately 98% of genomes, were discovered to also influence the phenotypic landscape of the organisms heavily (K. C. Wang & Chang, 2011). Each cell in an organism has the same genomic data, but the expression of the genes does the differentiation. Noncoding RNAs are transcribed from noncoding DNA and not translated into proteins. They play a role in gene regulation in various ways including splicing, translational inhibition, mRNA destruction and RNA editing (Brant & Budak, 2018; Tyagi, Sharma, & Upadhyay, 2018). MicroRNAs are a group of small, non-coding RNA molecules and they negatively regulate the gene expression through translational repression or transcript cleavage (Budak & Akpinar, 2015). Due to their wide-ranging influence, and regulatory process, miRNA manipulation represents a promising opportunity for crop improvement, as they could enable efficient development of desired characteristics through gene regulation (Djami-tchatchou, Sanan-mishra, Ntushelo, & Dubery, 2017). The canonical miRNA biogenesis pathway involves the transcription of miRNA genes by DNA-dependent RNA polymerase II into primary miRNAs (pri-miRNAs), which are hairpin structures with a loop, an upper stem, miRNA-miRNA* region, and a lower stem (J. Wang, Mei, & Ren, 2019). Pri-miRNAs are

recognized by a Dicer-like (DCL) family enzyme, mostly DCL1, and cleaved into precursor miRNA (pre-miRNA) hairpin and then into miRNA-miRNA* duplex (Axtell, Westholm, & Lai, 2011; Budak & Akpinar, 2015). The length of mature miRNAs usually varies between 21-24 nucleotides depending on which DCL family protein processes. HUA Enhancer 1 (HEN1) methylates 3' end of miRNA-miRNA* duplex (B. Yu et al., 2005) and then the complex, possibly through HASTY (HST) is exported from nucleus to cytosol (Axtell et al., 2011; Budak & Akpinar, 2015; Guleria, Mahajan, Bhardwaj, & Yadav, 2011). The miRNA-miRNA* duplex separates via helicase enzyme and the guide strand (miRNA) is loaded on RNA-induced silencing complex (RISC) upon binding the Argonaute (AGO) protein (Lucas & Budak, 2012; J. Wang et al., 2019). After the assembly of RISC, miRNA mediates mRNA cleavage or translational inhibition through binding its mRNA target (Budak, Kantar, Bulut, & Akpinar, 2015).

Besides sRNAs, there is an increasing interest in another type of non-coding RNA, long non-coding RNAs (lncRNAs). LncRNAs are over 200 nucleotides long, non-coding transcripts that work together with sRNAs at various levels of the gene expression regulation (Budak, Kaya, & Cagirici, 2020; Fatica & Bozzoni, 2014; Mercer, Dinger, & Mattick, 2009; Ponting, Oliver, & Reik, 2009). LncRNAs are not well conserved at the primary sequence level but in the function (K. C. Wang & Chang, 2011). The identification of lncRNAs is somehow arbitrary but their sizes distinguish them from other noncoding RNAs (Rinn & Chang, 2012). Their ability to fold into complex secondary and tertiary structures makes lncRNAs more versatile compared to sRNAs and provides an additional potential way of use in the crop improvement studies (Xin et al., 2011). As lncRNAs can also serve as sRNA precursors (Liu, Wang, & Chua, 2015), lncRNA-miRNA-mRNA interaction holds great importance in revealing the importance of those non-coding transcripts.

In this thesis, we have investigated different monocot grass species in terms of the conservation of non-coding RNAs revealing both interspecific and intraspecific variations. We have used a homology-based *in silico* miRNA identification approach to determine the conservation and functional annotation of miRNAs among cereals. In chapter I, we have constructed the pangenome miRNomics of a model organism, *Brachypodium distachyon*, by identifying miRNA families from *de novo* genome assemblies of 54 lineages by using a homology-based *in silico* method. To our knowledge, no such comprehensive miRNA identification has been performed for a

species before, and this is the first pangenome study from a non-coding genome perspective. Sequence conservation and genic collinearity are usually high among cereal genomes, considering this information, in chapter II we have carried out the comparative analysis of four insect tolerance loci previously discovered in wheat. We have determined the homologous QTLs of these loci in other important cereal crops barley (*Hordeum vulgare*), rye (*Secale cereale*), rice (*Oryza Sativa*), and oat (*Avena sativa)* and identified the miRNA content and conservation in these homologous loci. In the last chapter, we have identified genome and transcriptome-derived miRNAs from different barley cultivars under normal and salt treatment conditions, identified lncRNA transcripts, and carried out target analysis to construct lncRNA-miRNA-mRNA interactions aiming to discover both intraspecific variations and salt-responsive noncoding RNAs. Overall, this study focuses on the identification and comparative analysis of noncoding features in cereal crops.

,

2. GENERAL MATERIALS AND METHODS

2.1. miRNA IDENTIFICATION

miRNA identification was carried out in all three chapters using a homology-based *in silico* approach (Bala Ani Akpinar, Kantar, & Budak, 2015; Lucas & Budak, 2012). The first step consists of the selection of sequence homology to a reference miRNA sequence. In chapters 1 and 2, all *Viridiplantae* mature miRNA sequences, in chapter 3 only high confidence and experimentally proved *Viridiplantae* mature miRNA sequences were obtained from miRBase (v21, June 2014) (Kozomara & Griffiths-Jones, 2010) to be used as a reference set. In brief, using an in-house SUmirFind script, candidate miRNA sequences with at most 1-3 mismatches to the reference miRNAs were selected using National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST). For chapters 1 and 2, the maximum number of mismatches was defined as 1, whereas for high confidence miRNAs in chapter 3, the maximum number of mismatches was 3. Another in-house script, SUmirFold, was used to generate potential miRNA precursor sequences by extracting and folding them using the UNAFold v3.8 algorithm (Markham & Zuker, 2008), and evaluating them for the presence of known pre-miRNA fold characteristics, including GC content and MFEI. Potential pre-miRNA sequences that passed the previous evaluation were tested for additional criteria: (1) no mismatches were allowed at Dicer cut sites, (2) no multi-branched loops were allowed in the hairpin containing the mature miRNA sequence, (3) mature miRNA sequence could not be located at the head portion of the hairpin, and (4) no more than four and six mismatches were allowed in miRNA and miRNA*, respectively, using SUmirScreen to

eliminate false-positives (Busra Cagirici, Sen, & Budak, 2021; Lucas & Budak, 2012). In chapter 2, we have used a fully automated version of this miRNA pipeline, mirMachine (H. Busra Cagirici, Sen, & Budak, 2021), which identifies the potential miRNA precursor sequences (pre-miRNAs) and fold them using RNAfold (Lorenz et al., 2011).

2.2. POTENTIAL mRNA TARGET ANALYSIS OF IDENTIFIED miRNAs

Potential mRNA targets of the putative miRNAs were identified using the online web-tool psRNATarget, with user-defined query and target options (http://plantgrn.noble.org/psRNATarget) (Dai, Zhuang, & Zhao, 2018). The psRNATarget was run with default parameters in all 3 chapters, except two in Chapter 1 and 2: (1) Maximum UPE, the binding energy between miRNA and its target where lower values indicate stronger miRNA-miRNA target interactions, was set to 25, and (2) Expectation, which, like e-value in the blast, indicates the significance of miRNA-miRNA target pairs, was limited to a maximum of 3.

2.3. IDENTIFICATION OF REPETITIVE ELEMENTS AND CONSERVATION AMONG PRECURSOR MIRNA SEQUENCES

In chapters 1 and 3, query of pre-miRNA sequences of all identified miRNA families was searched against the library of *Poaceae* repetitive elements (MIPS-REdatPoaceae version 9.3) downloaded from ftpmips.helmholtz-muenchen.de/plants/REdat using RepeatMasker tool version 4 with default parameters (http://www.repeatmasker.org/) (Smit, Hubley, & Green, 2016). Pre-miRNA sequences that contain repetitive elements by more than half of their lengths are considered ‘repetitive’.

3. PANGENOME MIRNOMICS OF BRACHYPODIUM

3.1. INTRODUCTION

Genome sequencing has come a long way since the Sanger method was developed in 1975 (Sanger & Coulson, 1975) and reached a major milestone in 2005 with the emergence of next-generation sequencing (NGS), which enabled the sequencing of huge quantities of DNA data faster and cheaper than ever before (Mardis, 2017; Slatko, Gardner, & Ausubel, 2018). The advent of high throughput NGS technology over the past 15 years has not only tremendously increased the number of de novo genome assemblies, but also revolutionized the way we perceive the genomic analysis (Shendure et al., 2017). The availability of whole-genome sequences opened the possibility of studying comparative genomics, which revealed conserved and diverged genomic features of different organisms.

Life on Earth evolves and adapts to different environmental conditions through the genetic heterogeneity of organisms driven by mutations. Genes with crucial functions among all organisms are conserved during evolution while other genes differentiate species or cause intra-species variations (Salk, Schmitt, & Loeb, 2018; Hervé Tettelin, Riley, Cattuto, & Medini, 2008). Apart from gene-level variations, single nucleotide polymorphisms (SNPs) contribute equally to intra-specific genetic variations (Hurgobin & Edwards, 2017). Even though characterizations of such variants are mostly carried out using a single reference genome-based approach, highly polymorphic regions, presence/absence variations (PAVs), and copy number variations (CNVs) that diverge from the reference genome are unavoidably lost (Gordon et al., 2017; Zhao et al., 2018).

It has been previously shown that 20% of the genes in an agronomically important crop, Brassica oleracea, are affected by PAVs; therefore, capturing non-reference sequences is vital in studying genetic diversity within the species (Golicz et al., 2016). A more detailed technique to capture intra-specific variations was pioneered by Tettelin et al. (2005) in which the term "pan-genome" was used for the first time (H. Tettelin et al., 2005).

Pan-genome can be defined as the whole genomic repertoire or union of entire genes of all individuals belonging to a single species or a phylogenetic clade. This repertoire is more extensive than any of a single strain of the species and can be categorized into two: (1) the core genome, and (2) the dispensable genome (Hervé Tettelin et al., 2008). The core genome consists of genes present in all individuals of the clade and those genes have been shown to be mainly responsible for basic biological activities and the phenotype. On the other hand, cloud genes of the dispensable genome are present in only a few strains and are not essential for main biological functions, but they beneficially contribute to the genetic diversity (H. Tettelin et al., 2005; Hervé Tettelin et al., 2008).

With an ever-increasing global food demand, attaining higher-yielding and more resilient crops, which is the ultimate aim of agricultural studies, has become greatly dependent upon a good understanding of molecular mechanisms. Plants are exposed to various biotic and abiotic stresses because of their sessile nature; hence, consequently, they develop various response mechanisms to cope with stress conditions (Akpinar, Avsar, Lucas, & Budak, 2012). The construction of plant pan-genomes explores the genetic diversity and enables the characterization of gene variants. Model organisms such as *Brachypodium distachyon* have a great place in genomic studies, revealing complex networks controlling molecular mechanisms like stress response (Rasool, Ahmad, Rehman, Arif, & Anjum, 2015). Recently the pan-genome of *Brachypodium distachyon* was constructed from de novo genome assemblies of 54 lineages. The pangenome was shown to include twice as many genes as the genome of an individual and the genes forming the dispensable genome were shown to function in selective advantages such as adaptation, defense, and development (Gordon et al., 2017).

Besides protein-coding genes, which comprise only a small fraction of the genome in most species, the phenotypic landscape of organisms is also heavily influenced by non-coding sequences. Many studies have shown that plant ncRNAs are differentially expressed under unfavorable environmental conditions for adaptation and enhancing

growth and development (Waititu, Zhang, Liu, & Wang, 2020). MicroRNAs (miRNAs) are 18–24 nucleotide long, endogenous non-coding RNAs that negatively regulate gene expression and control key developmental plasticity, disease resistance, and stress response mechanisms (Budak & Akpinar, 2015). In general, miRNAs regulate the expression of numerous genes through translational repression or transcript cleavage (Schwab et al., 2005).

Because of their regulatory role in gene expression, interest has increased in identifying miRNAs and their functions in plants over the last decade. However, despite the vast amount of plant genome miRNA identifications and characterizations, to the best of our knowledge, miRNAs have not yet been explored at the pan-genome level. In this study, we explored the miRNA contents of the de novo genome assemblies of 54 *Brachypodium* lineages (Gordon et al., 2017) to gain insight into the miRNA evolution of *Brachypodium* through the identification of conserved and rare miRNA families, the variation of microRNA abundance and function, and to uncover the molecular basis of agronomic traits that are common or specific to some of the lineages. As a monocot model organism, exploring intraspecific variations in terms of noncoding features in *Brachypodium* holds great importance in future monocot crop studies for crop improvement.

3.2. MATERIALS AND METHODS

3.2.1. Datasets

De novo genome assemblies of 54 *Brachypodium distachyon* lineages were obtained from the study by Gordon et al. (Gordon et al., 2017). The coding sequences and annotations of rice orthologs of 54 *Brachypodium distachyon* lineages were also retrieved from the public repository of the same study (Nordberg et al., 2014).

3.2.2. miRNA Identification and Potential mRNA Target Analysis of Identified miRNAs

miRNA Identification was carried out for each genome separately, following homology-based *in silico* miRNA identification as previously described. Potential mRNA targets of the miRNAs were identified using psRNATarget and for each lineage, targets were searched among annotated CDS sequences from 54 lineages of the *Brachypodium distachyon* (Gordon et al., 2017). Additionally, rice annotations of CDS of each lineage obtained from Gordon et al. 2017 were also used for target prediction of some miRNAs.

Clustering of target transcripts was done based on 90% sequence similarity using the CD-HIT-EST tool (Huang, Niu, Gao, Fu, & Li, 2010). Representative sequences from each cluster were retained and representative target transcripts within each group were compared to *Viridiplantae* proteins using the blastx tool on a local server (e-value of 1 × 106 and a maximum target of 1). Gene Ontology (GO) annotations for the biological process were then obtained using Blast2GO software (Götz et al., 2008) following mapping, annotation, and GO slim steps for plants. Other computational analyses were carried out with in-house Python 3 scripts.

3.2.3. Construction of the heatmap

The heatmap was drawn on Heatmapper Pairwise Comparison (www.heatmapper.ca/pairwise, accessed on 29 January 2021) (Babicki et al., 2016). The data for the heatmap was generated using custom Python 3 scripts. For each pairwise comparison, the extent of conservation was defined and calculated as: the extent of miRNA conservation of lineage 1 with lineage 2 = common miRNAs (lineage1, lineage 2)/all miRNAs (lineage 1). From this perspective, the calculation differs for (lineage1, lineage2) and (lineage2, lineage1) comparisons.

3.2.4. Identification of Repetitive Elements and Conservation among Precursor miRNA Sequences

The repetitive element content of precursor miRNA sequences was identified, as previously described. Alignment of precursor miRNA sequences of each miRNA family

was done using the MAFFT-auto-alignment tool with default parameters (https://mafft.cbrc.jp/) (Katoh & Standley, 2013). Aligned sequences were analyzed for the conservation of pre-miRNA sequences.

3.3. RESULTS

3.3.1. miRNA Identification in Each Lineage

To explore the conservation and diversity of miRNA families across the *Brachypodium* pangenome, *de novo* genome assemblies of 54 lineages, published in an earlier study (Gordon et al., 2017), were used for miRNA identification. Homology-based *in silico* miRNA identification identified a total of 115 miRNA families processed from 168,657 miRNA precursors (pre-miRNAs) across all 54 lineages, representing the pan-genome for miRNAs in *Brachypodium* with an average miRNA family count of 90 miRNA families per lineage (Appendix A: Supplementary Table S1). Mature miRNA sequences and pre-miRNA sequences identified in each lineage are shown in Appendix A: Supplementary Table S2. Bd1-1 lineage was found to be the lineage with the lowest miRNA family number, 83, and the highest number of miRNA families identified in a lineage was 93 (Appendix A: Supplementary Figure 1).

Of the 115 miRNA families, we identified 11 miRNA families, which have not been previously reported as *Brachypodium* miRNAs to the best of our knowledge, namely miR1130, miR165, miR2873, miR5161, miR5522, miR5566, miR5568, miR6197, miR6224, miR8155, and miR9783, but have been reported in related grasses (Bala A. Akpinar & Budak, 2016; Bala Ani Akpinar, Yuce, et al., 2015a; Cheah, Nadarajah, Divate, & Wickneswari, 2015; Franke et al., 2018; Jeong et al., 2013; Sakaguchi & Watanabe, 2012; Y. Zhou et al., 2016).

Familywise, we observed that the majority of miRNA families were identified in all or most of the lineages and only a few families were found in progressively fewer lineages

(Appendix A: Supplementary Figure 1). Specifically, 56 of the 115 miRNA families were identified in all 54 lineages and 67 miRNA families were present in more than 98% of them. Thirteen miRNA families were found in only one or two lineages, although it should be noted that the presence or absence of miRNA families within our approach depended on the completeness of the genome assemblies and the accuracy of our predictions, as well as computational approaches. Thus, miRNA families that were identified in only a few lineages can still be present in other lineages with less complete genome sequences.

To further explore clues into what roles highly or rarely conserved miRNAs play, we grouped the miRNA families into four groups based on their level of conservation in 54 lineages (Appendix A: Supplementary Figure 1). The first group was the "rare" miRNAs group, which consisted of 20 miRNA families identified in fewer than 10 lineages. Thirteen of these miRNA families were present in only one or two lineages and may be considered as lineage-specific miRNAs, keeping in mind the limitations of sequencing and our predictions. Ten miRNA families identified in more than 10 but fewer than 45 lineages were grouped under "moderately conserved" miRNAs. The miRNA families conserved in more than 80% of lineages were classified under the "highly conserved" miRNA families group. This group consisted of 18 miRNA families which were identified in more than 45 but fewer than 53 lineages. The fourth group, which had the highest number of miRNA families, was the "common" miRNAs group and included miRNA families identified as present in 53 or all 54 lineages.

A closer look at the 168,657 pre-miRNAs sequences, giving rise to the miRNAs, indicated substantial variations in the number of precursor sequences per miRNA family. The lowest pre-miRNA sequences counts were identified in the rare miRNA families group (Appendix A: Supplementary Figure 2). As the conservation of miRNA families increased from moderately conserved miRNA families to highly conserved miRNA families group, the pre-miRNA counts of those families were also identified to be increasing (Appendix A: Supplementary Figure 3 and Figure 4). For the common miRNA families group, some high pre-miRNA family numbers were identified (Appendix A: Supplementary Figure 5). A few miRNA families had extremely high numbers of precursors across all lineages, such as miR5174 (44,403), miR5181 (43,651), miR5049 (22,340), and miR5175 (20,694). Within each lineage, the pre-miRNA distributions were also similar (Appendix A: Supplementary Table S1). Comparison of the precursors

against known repeats from *Poaceae* species revealed that 48 miRNA families were characterized by pre-miRNA sequences containing repetitive elements by more than 50% of their lengths (repetitive pre-miRNAs hereafter). In most cases, repetitive pre-miRNAs represented all or none of the precursors within a miRNA family (Appendix A: Supplementary Table S1). Among the non-repetitive pre-miRNAs, we also explored sequence conservation within miRNA families by aligning all precursors from all lineages. We observed that while some miRNA families had precursors aligning perfectly or near-perfectly across different lineages, such as miR394, some other miRNA families, such as miR156, had precursors aligning around the mature miRNA and miRNA* regions, but otherwise contained large gaps within the alignment. In general, pre-miRNAs belonging to the miRNA families that were identified in specific phyla, such as miR7745 and miR7763 (Table 3.1), seem to retain considerable sequence homology across their entire lengths in different lineages. In contrast, miRNA families widely found across the plant kingdom, such as miR156, miR160, miR166, and miR395, appeared to have accumulated considerable sequence variation. Additionally, these miRNA families had relatively high numbers of precursor sequences, which, at least for non-repetitive pre-miRNAs, may indicate the presence of multiple copies within the *Brachypodium* genomes. Representative pre-miRNA alignments from rare, moderately conserved, highly conserved, and common miRNA families are given in Appendix A: Supplementary File S1.

Table 3.1 All miRNA families identified from *de novo* genome assemblies of 54 *Brachypodium* lineages are classified based on their presence in multiple lineages. Phylogenetic inference based on miRBase phyla data for each miRNA family is given. M column represents Magnoliophyta, C column represents Coniferophyta, and E column represents Embryophyta phyla.

		M	***C***	***E***
Rare	miR1139, miR2873, miR5063, miR5161, miR5167, miR5169, miR5179, miR531, miR5566, miR6224, miR7708, miR7709, miR7717, miR7726, miR7729, miR7745, miR7748, miR7763, miR7765, miR9494	√ (m)		
Mod. Cons.	miR398	√	√	
	miR5068, miR5184, miR7725, miR7727, miR7744, miR7766, miR9480, miR9490	√ (m)		
	miR8155	√ (e)		
Highly Cons.	miR394	√		
	miR1133, miR1436, miR5062, miR5163, miR5201, miR7715, miR7722, miR7732, miR7736, miR7740, miR7754, miR7756, miR7771, miR7775, miR7781	√ (m)		
	miR5281, miR845	√ *		
Common	miR1122, miR1127, miR1128, miR1130, miR1135, miR1432, miR1435, miR1439, miR5049, miR5054, miR5067, miR5070, miR5164, miR5165, miR5171, miR5174, miR5175, miR5176, miR5180, miR5181, miR5182, miR5183, miR5185, miR5198, miR5199, miR5200, miR5202, miR528, miR5522, miR5568, miR6197, miR7716, miR7723, miR7728, miR7731, miR7733, miR7738, miR7755, miR7770, miR7772, miR7773, miR7777, miR9481, miR9485, miR9486, miR9489, miR9493, miR9495, miR9783	√ (m)		
	miR156, miR159, miR160, miR166, miR171, miR395, miR396	√	√	√
	miR164, miR169, miR397	√	√	
	miR165	√ (e)		
	miR167	√		√
	miR172, miR393	√		
	miR2118, miR2275, miR399	√ *		

*: All monocotyledons and eudicotyledons, except *Amborella trichopoda*;
(m): Only in monocotyledons; (e): Only in eudicotyledons.

3.3.2. Pairwise Comparison of Common microRNA Families between Lineages

To explore the extent of conservation of miRNA families between lineages, we identified the number of common miRNA families between each pair of lineages. To account for differences in the total numbers of miRNA families predicted in each lineage, we described the extent of miRNA conservation in a lineage as the ratio of miRNA families shared with a second lineage over the total number of the families identified for that lineage (Figure 3.1). For instance, 96.7% of miRNA families in Arn1 lineage were shared with Mon3 lineage, making the extent of miRNA family conservation 0.967 in Arn1 with Mon3. Notably, this approach introduces a directionality; miRNA family conservation in Mon3 with Arn1 is 0.989. This analysis showed that miRNA families are shared by more than 80% in each pairwise comparison. Additionally, some lineages appeared to have more miRNA families in common than others. For example, of the miRNA families identified in Mon3, a higher number of families are shared with Arn1 than with ABR8 (Figure 3.1).

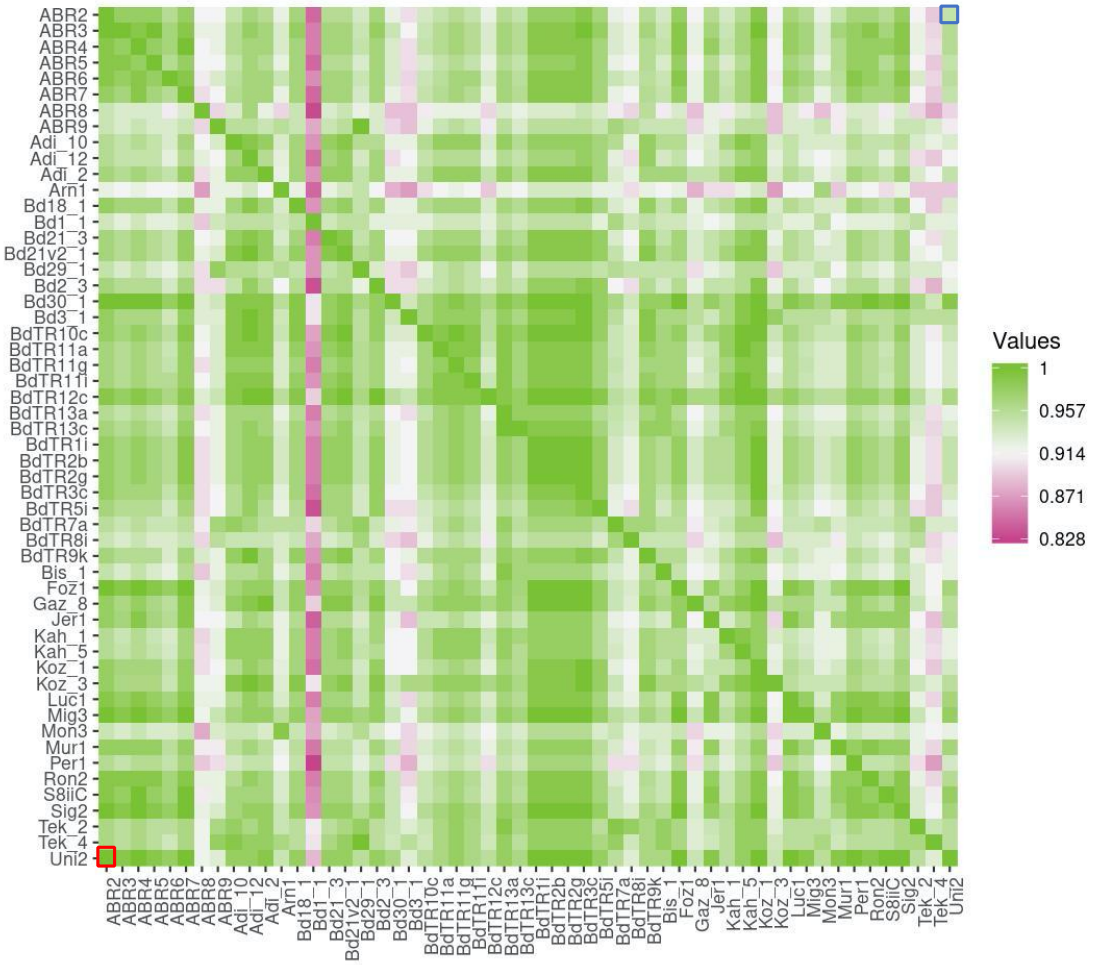


Figure 3.1 Heatmap displaying the miRNA family commonality between lineages. Darker green cells represent higher numbers of common miRNAs shared between two lineages while darker purple represents fewer numbers of common miRNA families. The heatmap is not symmetrical; it should be read from y > x. For example, the extent of conservation between ABR2 and Uni2, defined as the ratio of common miRNA families between ABR2 and Uni2 over the number of total miRNA families in ABR2 by our approach, is given on the top right corner (indicated by a blue square), while the conservation between Uni2 and ABR2, the ratio over the total number of families in Uni2, is given on the bottom left (red square).

3.3.3. mRNA Target Analysis and Determination of Potential Biological Processes Targeted by Each miRNA Groups

Potential mRNA targets of the identified mature miRNAs were predicted using the psRNATarget tool (Dai et al., 2018) among the coding sequences of each lineage (Gordon et al., 2017). Maximum UPE and expectation parameters were adjusted to allow high confidence targets to be retained. In total, mature miRNA sequences from 111 families retrieved 109,438 predicted targets in all 54 lineages (Appendix A: Table S3). No mRNA

targets for the four miRNA families, miR1139, miR5054, miR7709, and miR7748, were obtained. Two of those miRNAs, miR7709 and miR7748, were found to be lineage-specific miRNAs that have not been previously reported. While the miRNA families and their respective targets within each lineage give clues into the regulatory networks contributing to the overall characteristics of those lineages, from the pan-genome perspective, we sought to explore the regulatory roles of highly-to-rarely conserved miRNA families across all lineages. To provide a global view, we combined all predicted targets of all miRNA families from all lineages within each of the four groups previously described, to generate global groups of targets of rare, moderately conserved, highly conserved, and common miRNAs. We hypothesized that the predicted targets of common miRNA families would be associated with essential pathways, while the predicted targets of the rare miRNAs could identify dispensable but critical pathways related to species diversity. To eliminate redundancy coming from homologous transcripts of different lineages, we clustered the target transcript sequences based on sequence identity and retained only the longest cluster representative. This resulted in 123 clusters for rare miRNA targets, 115 clusters for moderately conserved miRNA targets, 647 clusters for highly conserved miRNA targets, and 6131 clusters for common miRNA targets from the starting total of 109,438 target transcripts. Gene Ontology (GO) annotations were obtained by comparing representative sequences against all *Viridiplantae* proteins using Blast2Go (Götz et al., 2008) (Figure 3.2). Biosynthetic process, cellular protein modification, and transport and nucleobase-containing com-pound metabolic processes were found to be the major biological processes putatively targeted by all four groups. It appeared that the regulation of core biological processes, such as biosynthesis and protein modification, potentially carried out by common miRNAs found in most lineages, was also contributed to heavily by moderately con-served and even rare miRNAs. Interestingly, the potential targets of only moderately conserved miRNAs were involved in other important processes such as signal transduction, response to chemicals, reproduction, and endogenous stimulus. Response to stimulus and cell communication processes were targeted only by rare miRNA families.

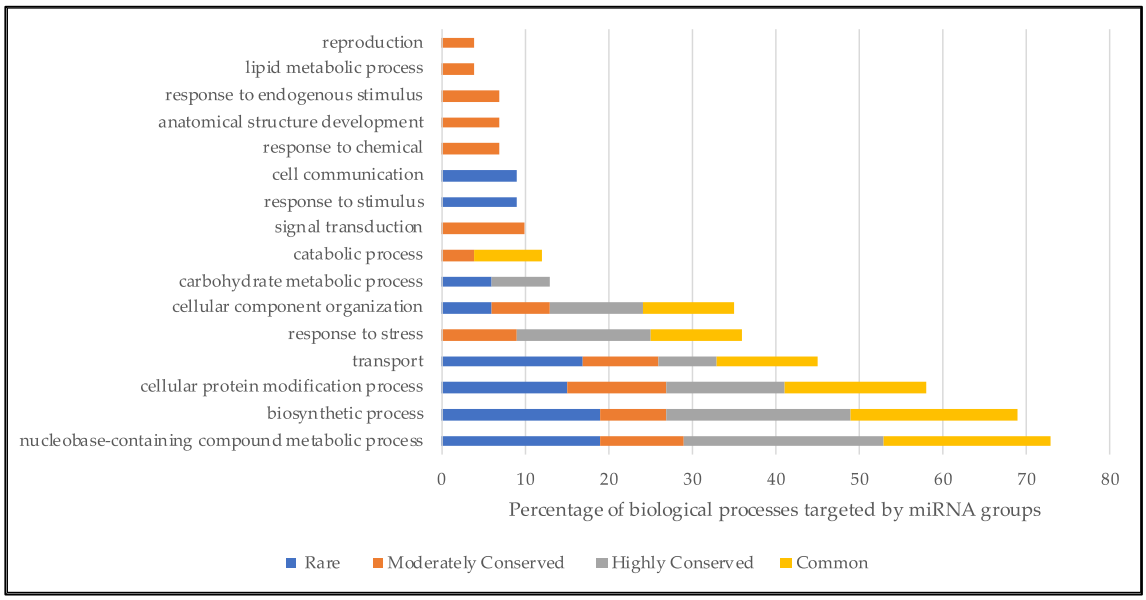


Figure 3.2 Gene Ontology analysis for biological processes for rare, moderately conserved, highly conserved, and common miRNA family groups targets, combined and clustered. Predicted targets from five biological processes appear to be collectively targeted by all four miRNA groups.

3.3.4. miRNA- mRNA Transcript Multiplicity Analysis

Within each lineage, there were instances of miRNA families targeting multiple coding sequences and, conversely, coding sequences targeted by multiple miRNA families, indicating regulatory networks. For each lineage, we extracted the number of miRNAs having one or multiple predicted targets. Similarly, we also extracted the number of predicted targets targeted by one or multiple miRNA families for each lineage. For miRNAs with multiple predicted targets, we observed that, on average, 21 miRNA families were targeting 2–5 coding sequences across lineages (Figure 3.3a). The mean number of miRNA families targeting only one coding sequence is five and in rare instances, some miRNA families putatively targeted even hundreds of targets, although we cannot rule out that some of these targets could be false positives. For target sequences, on the other hand, most predicted target sequences appeared to be targeted by only one miRNA family (Figure 3.3b).

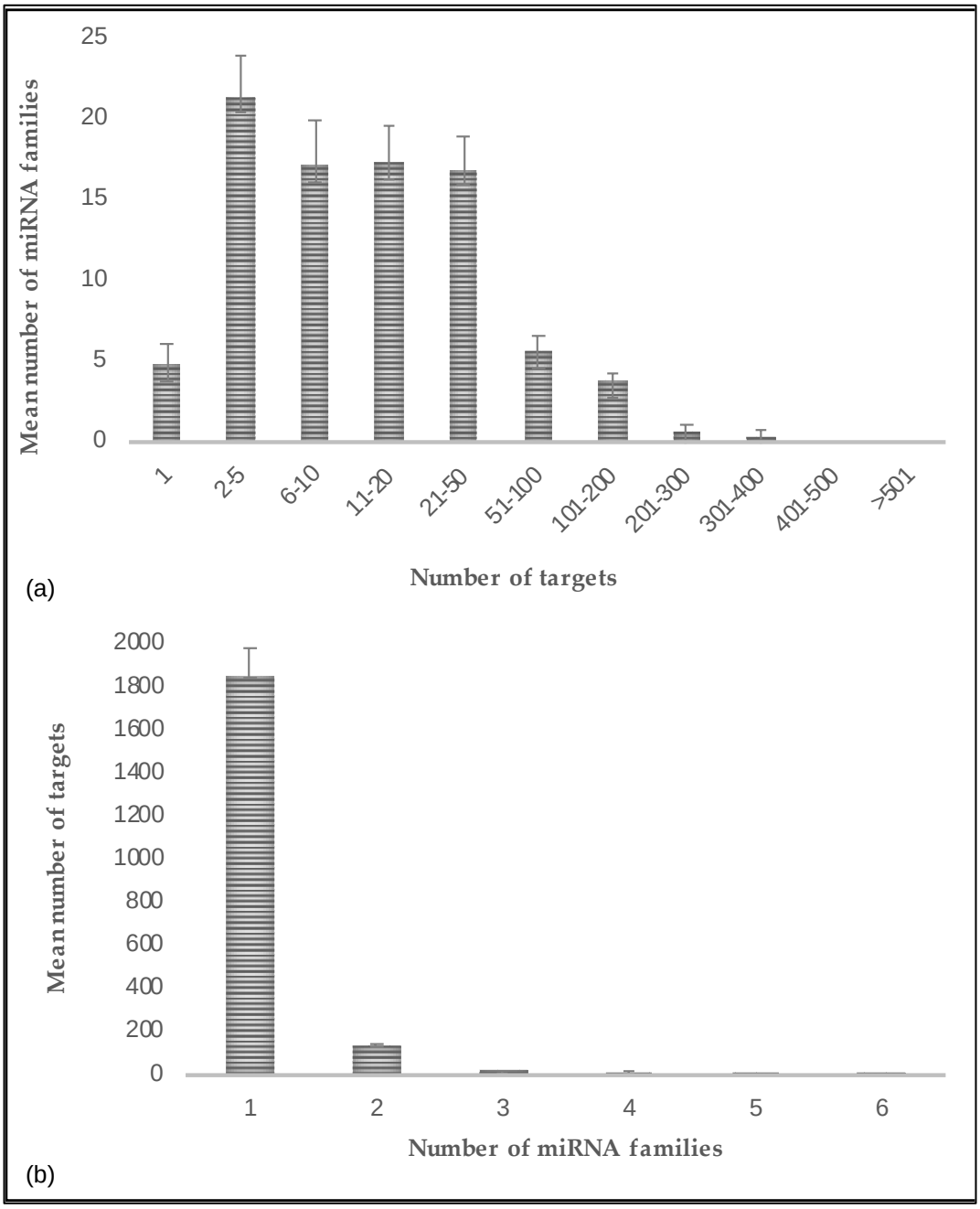


Figure 3.3 a) The number of miRNA families having multiple predicted targets were averaged across lineages. For simplicity, targets were considered within 11 bins of certain sizes as given. Most miRNA families seemed to target a minimum of 2 and a maximum of 50 coding sequences, whereas only a few miRNA families targeted only one target or had more than 50 predicted targets. (b) The number of predicted targets of only one or multiple miRNA families was averaged across lineages. In contrast to miRNAs targeting multiple targets, each target appeared to be predominantly targeted by one miRNA family.

We hypothesized that miRNAs acting together in all, or most lineages may point to the most functionally conserved networks. We listed all miRNA teams, >1 miRNA putatively acting on a single transcript in a given lineage and extracted teams that were conserved in all 54 or 53 lineages. Five miRNA teams, miR5049 and miR5174, miR156 and miR529, miR5183 and miR7777, miR1439 and miR5174, and miR397 and miR164, were predicted to be targeting at least one common transcript together in all 54 lineages. Additionally, 4 miRNA teams were predicted to be targeting the same transcript together in 53 lineages. These were miR5185 and miR2275 (except in Tek-4), miR5185 and miR2118 (except in Bd3-1), miR166 and miR165 (except in BdTR8i), and miR5175 and miR5049 (except in Tek-2). While previous findings suggested functional roles for a few of these miRNA teams, others require further studies to unravel potential pathways linked to these miRNA families seemingly functioning together in several different lineages. It should be noted that near-exact miRNA teams (such as miRxxx and miRyyy in some lineages and miRxxx, miRyyy, and miRzzz in others) were missed in our stringent approach where only exact matches were extracted. These miRNA teams may represent highly conserved regulatory networks that are finely controlled by the action of a pair of miRNAs.

3.4. DISCUSSION & CONCLUSION

3.4.1. Discussion

Advances in next-generation sequencing have enabled extensive use of high-throughput technologies in many studies and the accessibility of RNA sequences has paved the way for studying other types of RNAs besides mRNA. Such studies have revealed the importance of noncoding RNAs in plant gene expression and stress response mechanisms (Waititu et al., 2020). MicroRNAs are genome-level regulators of gene expression and identifying miRNAs and their targets provides an understanding of the complex regulatory mechanisms (Manavella, Yang, & Palatnik, 2019). While the miRNA identification of the *Brachypodium* genome has been the subject of many previous studies, to the best of our knowledge, no comprehensive miRNA identification and target analysis of multiple lineages have been conducted previously.

In this chapter, we identified miRNA families from 54 *Brachypodium* lineages, carried out the target analysis of all mature miRNAs using coding sequences of corresponding genomes, and revealed the potential targets of different miRNA groups. Eighty-five of 115 miRNAs identified in our study were shown to be conserved in more than 80% of 54 lineages (Appendix A: Supplementary Figure 1). Our predictions included many already-known and experimentally validated *Brachypodium* miRNAs (bdi-miRNAs), such as miR156, miR159, miR160, miR164, miR393, miR528, miR5176, miR5200, and miR5202, some of which are also highly conserved among other monocots (Baev et al., 2011; Kozomara & Griffiths-Jones, 2010; Unver & Budak, 2009). We identified many well-known plant miRNAs, such as miR156, miR159, miR160, miR393, and miR397, which are conserved between monocots and dicots, and widely found in most flowering plants (Jones-Rhoades & Bartel, 2004). Consistent with our classification (Table 1.1), miRNA families common to all or most *Brachypodium* lineages contained families found typically in land plants (Embryophyta), but those that were found in progressively fewer lineages contained families that are so far specific to flowering plants (Magnoliophyta). Additionally, 11 miRNA families, which were previously reported in close relatives but not in *Brachypodium*, were identified (Franke et al., 2018; Jeong et al., 2013; Unver & Budak, 2009). For instance, researchers previously identified miR5566, miR5568, and miR6224 in sorghum (Li Zhang et al., 2011); miR1130, miR9783, and miR6197 in wheat (Bala A. Akpinar & Budak, 2016; Bala Ani Akpinar, Yuce, et al., 2015a; Kurtoglu, Kantar, & Budak, 2014) and miR2873 in rice (Cheah et al., 2015), but to the best of our knowledge, none of these miRNAs have been identified in *Brachypodium*, including in recent studies (Franke et al., 2018; Jeong et al., 2013). These suggest that our approach allows robust identification of bona fide miRNA sequences using next-generation sequencing data. Notably, among the 11 miRNA families not previously reported in *Brachypodium,* three families, miR2873, miR5161, and miR5566 were specific to one or two lineages, a fact which would likely be missed in conventional approaches or lineages frequently used in studies, including the reference species. Another six families, miR165, miR1130, miR5522, miR5568, miR6197, and miR9783, were highly conserved and common to 53 lineages or more, which may point to the importance of high-throughput data in identifying RNA species, such as miRNAs, which may have developmental-stage-specific or spatio-temporal expression patterns that can be missed in small-scale studies. Overall, 91 miRNA families were identified in *Brachypodium* reference genome Bd21,

and the other 24 miRNA families identified in this study would likely be missed without the comprehensive pan-genome approach.

An inspection of the pre-miRNA sequences revealed that repetitive sequences made up the majority, if not all, of the precursors of some miRNA families those with extremely high pre-miRNA counts (Appendix A: Supplementary Figure 5). It is tempting to speculate that these families, namely miR5049, miR5174, miR5175, and miR5181, which appear to be specific to monocot species (Table 3.1) might have been generated and proliferated subsequently to a Transposable-Element (TE)-capture after the divergence of monocot species. Alternatively, these families may in fact be mis-annotated siRNAs (Axtell & Meyers, 2018). Additionally, we observed that pre-miRNA sequences for some families were identical or extremely similar in different lineages, such as miR7745 and miR7763. These mostly corresponded to relatively less conserved miRNA families (Appendix A: Supplementary Figure 2). On the other hand, other families, mostly from the common miRNA group, exhibited considerable variation among precursors from different lineages. Pre-miRNA secondary structures are critical to the correct processing of miRNAs, which mandates sequence constraints on primary sequences which are not yet fully realized in the plants (Narjala, Nair, Tirumalai, Vivek Hari Sundar, & Shivaprasad, 2020). However, it has also been observed that relatively young non-conserved miRNAs typically have few copies in the genomes and retain extensive homology to their targets beyond the mature miRNA subsequence (Fahlgren et al., 2007; Narjala et al., 2020). Such miRNAs might then be expected to retain sequence conservation toward the entire length of the respective pre-miRNA sequences. Consistently, we observed such conservation mostly in rarely or moderately conserved miRNA families, which appeared to have limited numbers of precursors in each genome. Conversely, evolution might have more time to act on the precursor sequences of miRNA families found in virtually all plants, usually with multiple copies in the genome (Fahlgren et al., 2007).

Our predictions identified five miRNA teams targeting at least one common transcript in all 54 lineages. One of these teams, miR156 and miR529, has been previously shown to be evolutionarily related (Morea et al., 2016; S. D. Zhang, Ling, Zhang, Xu, & Cheng, 2015) and, parallel to these studies, our findings showed that they target SQUAMOSA promoter binding protein (SBP)-box gene family-related transcripts in all 54 lineages. These transcripts encode plant-specific transcription factors and are involved in the plant

growth and development (Y. Li et al., 2020). Another miRNA team, miR165, and miR166 were determined to target at least one common transcript in 53 lineages, and this finding is also consistent with the literature. The miR165/166 family was shown to be present in many plants suggesting its regulatory circuit has been conserved since the last common ancestor of vascular plants (Axtell & Bartel, 2005). Wojcik et al., (2017) previously showed that miR165/166 regulates the developmental plasticity of somatic cells in vivo, affecting auxin biosynthesis, and is involved in key stress response mechanisms in Arabidopsis. We have found 12-oxophytodienoate reductase as a common target of miR165 and miR166, and this protein was previously described in Arabidopsis as a stress-regulated protein that is involved in the jasmonic acid (JA) biosynthesis (Mussig et al., 2000).

Our observations suggest that while it is fairly common for a miRNA family to target a few different sequences, it is not very common for two different families to act on the same target sequence. The miRNA families putatively acting on different target sequences may signify crosstalk between different pathways. On the other hand, target transcripts predicted to be targeted by different miRNA families may indicate strategies to fine-tune target expression in response to specific conditions or needs (Figure 3.3).

Brachypodium has a high level of genetic variation in its subspecies as a consequence of its allopolyploidization and self-fertilization (Wilson et al., 2019). Significant differences in population phenotypes, even in a small geographic range, have been observed in previous studies (Filiz et al., 2009; Gordon et al., 2017). Bd1_1 cultivar was found to have the fewest miRNA families, 83, of all the cultivars in our study. Bd1_1 cultivar from Turkey is a divergent cultivar, based on SSR markers and a late flowering phenotype, which makes this cultivar phenotypically distant from most of the lineages (Gordon et al., 2014). Of all 54 lineages, Bd1_1 is the only cultivar in which miR7772 was identified.

Based on single nucleotide polymorphisms (SNPs), Gordon et al. (2017) deduced phylogenetic relationships among the 54 *Brachypodium* lineages. In terms of miRNA family conservation among given lineages, our observations are in line with these relationships (Figure 3.1). For instance, BdTR3c is most related to the Koz1 lineage in terms of SNPs, and we identified the same 93 miRNA families in both lineages. Moreover, Arn1 and Ron2 lineages, which have 92 and 90 miRNA families, respectively, have 84 miRNA families in common based on our prediction; supporting this were

findings in Gordon et al. (2017) that they are distant in terms of SNPs. However, despite the ABR9 and Bd1_1 lineages having been shown to be closely related in terms of SNPs, we identified 78 common miRNA families, which is lower than the average common miRNA family number of 85.5. This difference may be interpreted as the result of a low miRNA family number identified from the Bd1_1 cultivar. We identified miR5200, a conserved *Brachypodium* miRNA that is known to play a role in the regulation of the FLOWERING LOCUS T (FT) (Wu et al., 2013). Target prediction analyses showed FT was targeted only by miR5200 in all lineages. Mon3 and Arn1 lineages were predicted to have an extra locus than other lineages targeted by miR5200. BdTR7a, BdTR8i, Bd1-1, Bd29-1, Tek2, and Tek-4 were identified to have fewer miR5200 pre-miRNA sequences than other lineages. All six lineages were classified into the extremely delayed flowering phenotype (EDF+) clade in Gordon et al., 2017. Two other lineages of the same clade, Arn1, and Mon3 lineages, both from Spain, were shown to be closely related based on high confidence SNPs and both have earlier flowering than other lineages of the (EDF+) clade (Gordon et al., 2017). We identified 89 common miRNAs between these two lineages and a lineage-specific miRNA, miR5161, identified only in Arn1 and Mon3 lineages, which may be further studied in the future to identify its role in the attributed phenotype.

We classified the miRNA families identified in our study into four groups based on their conservation among lineages: 20 miRNA families as rare miRNA families, 10 moderately conserved miRNA families, 18 highly conserved miRNA families, and 67 common miRNA families (Appendix A: Supplementary Figure 1). Based on this classification, we observed that most miRNA families identified in our prediction were conserved among lineages. A psRNATarget analysis to identify potential mRNA targets of the miRNA families resulted in a high number of target sequences for the 54 lineages (Appendix A: Supplementary Table S3). No target sequences were identified for miR1139, miR5054, miR7709, and miR7748 families, which may be a result of the criteria we applied during target identification analysis to eliminate low confidence target sequences.

Target sequences obtained from the psRNATarget analysis were clustered based on at least 90% similarity and the total number of targets with representative sequences was decreased to 7016 for four groups. Gene Ontology enrichment revealed the biological functions in which the four groups, rare, moderately conserved, highly conserved, and common miRNAs, were mainly involved (Figure 3.2). Moderately conserved miRNAs

were shown to regulate biological functions, such as response to chemicals and stimulus and signal transduction, which are all key processes in plants' response mechanisms. Rare miRNA families were found to be involved in cell communication and response to stimulus processes. Plants sense their environments and respond to various biotic and abiotic stresses by altering molecular processes, such as signal transduction (Koroban et al., 2016). Both rare and moderately conserved miRNA families were found to be involved in the regulation of various mechanisms which can contribute to species diversity. However, potential targets of rare miRNA families should be analyzed with caution as low genome completeness, sequencing errors and miRNA identification criteria may result in the presence or absence of some miRNA families in lineages. Hence, moderately conserved miRNA families may provide more accurate insight into lineage differentiation in terms of miRNA evolution.

3.4.2. Conclusion

The use of reference genomes provides many advantages such as being more complete and error-free along with the benefits of working as a global community, but they cannot reflect the total genetic variations of a species adequately. As the sequencing costs are decreasing day by day, more comprehensive approaches, such as the construction of pangenomes, are being frequently used. In this chapter, we have identified miRNAs from 54 different *Brachypodium* lineages constructing the pangenome miRNomics. Such a comprehensive approach provided new insights about miRNA conservation and novel *Brachypodium* miRNAs. We have identified a total of 115 miRNA families as the pangenome miRNomics of *Brachypodium* which 24 of those would have been missed by using only the *Brachypodium* reference genome, Bd21. We have proposed a candidate lineage-specific miRNA, miR5161; which may be a regulator of flowering time phenotype in *Brachypodium* which the effect of its expression on this phenotype should be observed in future studies. Another previously described flowering time-related miRNA, miR5200, has been shown to affect flowering phenotype, and the copy number of this miRNA is suggested to be effective in this phenotype.

We determined 9 miRNA teams that target at least one common transcript in more than 53 lineages. Although it is widely common for miRNAs to target multiple -even hundreds of- transcripts, it is rare that a transcript to be targeted by multiple miRNAs. Two of those miRNA teams have been previously shown to be evolutionary related and conserved in

plants whereas 7 of those miRNA teams were not previously well characterized. miRNAs that work together in the regulation of the same transcripts will help understand of the miRNA-target network. The function of proposed miRNA teams on gene expression needs to be characterized in further studies.

In this study, we have shown that there are many lineage-specific or less conserved miRNA families which have an important role in many key characteristics such as disease resistance and cell communication. These miRNAs would be likely missed in conventional reference genome approaches. The verification of the proposed lineage-specific and novel miRNAs in future studies will contribute to unraveling many gene regulation mechanisms and their consequences on plant phenotype in model organism *Brachypodium* and relative cereal crops.

4. COMPARATIVE ANALYSIS OF CODING AND NON-CODING FEATURES WITHIN INSECT TOLERANCE LOCI IN WHEAT WITH THEIR HOMOLOGS IN CEREAL GENOMES

4.1. INTRODUCTION

Ensuring global food security and agricultural sustainability becomes increasingly challenging every year, with the pressure of a continuously growing population causing a rise in food demand, dietary changes, and shortage of land, and water resources (Budak, Kantar, et al., 2015; Campi, Dueñas, & Fagiolo, 2021). Global food demand is estimated to increase by 70–85% by 2050 (FAO, 2017); therefore, obtaining higher-yielding crops is vital to meet the food demand by producing more products with fewer resources. Wheat is the most consumed cultivated crop globally, contributing 20% of the calories consumed by humans (International Wheat Genome Sequencing Consortium (IWGSC), 2018). The United Nations' Food and Agriculture Organization (FAO) estimated the global wheat production to be around 776 million tons in the 2021/2022 marketing season and holds a USD 50 billion trade market. The United States of America annually exports about 28 million metric tons of wheat, majorly produced in Montana, Kansas, and North Dakota. It is predicted that the wheat demand will increase by 60% by 2050 as the world population rapidly increases (FAO-Food and Agricultural Organization of the United Nations, 2020; "WHEAT in the World," 2020).

While the food demand is continuously increasing, there are also a variety of stress factors threatening agricultural productivity. Abiotic stress conditions such as drought, high salinity, high or low temperatures, and heavy metals can negatively affect crop productivity and are often interrelated (Dresselhaus & Hückelhoven, 2018). Additionally,

plants are vulnerable to diseases caused by biotic stress factors such as viruses, fungi, bacteria, weeds, nematodes, arachnids, and insects (Harris et al., 2003). Two such pests, wheat stem sawfly (*Cephus cinctus* Norton, *WSS*) and orange wheat blossom midge (*Sitodiplosis mosellana, OWBM*) can cause significant economic losses through yield deprivation and damaged kernels that cannot be harvested in cereals.

WSS is an endemic stem-mining insect of Northern Great plains and a significant pest of winter and spring hexaploid wheat and tetraploid durum wheat, minimizing its yield and causing economic losses (Halise B. Cagirici et al., 2017; Varella et al., 2016). Severe sawfly infestations occur especially in Montana, North Dakota, northern South Dakota, and western Minnesota, and there are currently no chemical control measures to suppress *WSS* damage (Shrestha et al., 2018).

The *WSS* produces one generation per year when the female sawflies oviposit an average of 30–50 eggs into the internode of the host plant stem, and only one larva survives (Varella et al., 2016). Larvae feed on parenchyma and vascular bundles and move down the stem as they mature. This larval feeding decreases the photosynthetic ability of the host plant and lowers the mass by up to 30% (Nilsen et al., 2017). Eventually, larvae cut the stem at the base creating a lodge to allow them to accumulate during overwintering diapause (Biyiklioglu et al., 2018). Infested stems are easily windblown and not easily picked during harvest, resulting in further yield loss (Nilsen et al., 2017). In addition, uncut infested plants suffer yield loss due to a decrease in head weight (Halise B. Cagirici et al., 2017). Plant characteristics influence the host preference of female sawflies, which is important in progeny survival as the *WSS* larva cannot switch hosts (Varella et al., 2016). Although sawfly may oviposit in other cereals, including oat, rye, and barley, larvae development rarely occurs in barley and rye, whereas larvae do not survive in oat (Cockrell et al., 2017). Host–plant resistance provided by solid-stemmed and semi-solid-stemmed cultivars is the only effective management strategy against *WSS* infestation as pith development in the culm lumen reduces larva survival and interferes with insect oviposition (Szczepaniec, Glover, & Berzonsky, 2015). A major QTL, *SSt1,* identified on the long arm of the 3B chromosome has been associated with stem solidness in tetraploid durum wheat, common wheat, and wild emmer wheat (Nilsen et al., 2017). This QTL and its orthologue in hexaploid wheat, *Qss.msub-3BL,* have been the most favorable tools of management against *WSS* in the wheat (Cook, Wichman, Martin, Bruckner, & Talbert, 2004). Recently, the *TdDof* gene (TRITD3Bv1G280530) has been identified as the causal

gene of the stem solidness phenotype conferred by the *SSt1* QTL in tetraploid wheat. Copy number variation, where the presence of additional copies coupled with increased expression of this gene, has been shown to positively regulate the stem solidness (Nilsen et al., 2020).

Orange wheat blossom midge (*OWBM*) is another damaging pest of spring wheat effective in a wide range from North America, to several European countries and Asia (Thambugala et al., 2021). Around May, the mature *OWBM* larvae drop from the wheat ears to the ground, resulting in diapause of the larvae within a cocoon for a long time. In springtime, the larvae leave the cocoon and pupate at the soil surface. Two weeks after they pupate, adult midges move to the wheat canopy, where female midges lay up to 80 eggs. The hatched larvae will eat the kernels damaging the crop (Gong et al., 2013; Kassa et al., 2016). There are two host resistance mechanisms against *OWBM* damage: (1) abnormal or prevented oviposition (antixenosis) and (2) suppression of larval growth (antibiosis) (Kassa et al., 2016). *Sm1* gene mapped on the short arm of chromosome 2B from Augusta cultivar of American wheat has been the first antibiosis gene identified to provide resistance against *OWBM*. *Sm1* is believed to inhibit larval growth through ferulic acid and/or p-coumaric acid production in seed coat (McKenzie et al., 2002; Thomas et al., 2005). The QTL containing the *Sm1* gene has been under close scrutiny and had been saturated by several molecular markers by Kassa et al. (2016) (Kassa et al., 2016). Finally, very recently, extensive genome sequencing in wheat led to the identification of a candidate gene for this locus, an NB-ARC-LRR-kinase-MSP gene, that is likely responsible for the *OWBM* resistance phenotype (Walkowiak et al., 2020).

Even though the antibiosis-related *Sm1* gene has been the only resource utilized in *OWBM* management strategies in the field so far, antixenosis-related mechanisms have been observed in common and durum wheat that can also be utilized to grow resilient crops (Lamb et al., 2002; Lijing Zhang et al., 2020). Accordingly, a major QTL on chromosome 1A, *QSm.mst-1A,* was first identified in 2011 in the spring wheat variety Reeder (Blake et al., 2011). This QTL was associated with oviposition deterrence again in a recent study and was saturated with molecular markers to facilitate cloning and marker-assisted selection efforts. Finally, another QTL on chromosome 4A has attracted attention for potential use against *OWBM* through the oviposition deterrence (Lijing Zhang et al., 2020). So far, the causal genes underlying the antixenosis responses conferred by these QTL are unknown.

In this chapter, we investigated the conservation of coding and noncoding features of 4 insect tolerance loci previously identified in wheat among other important cereal crops barley, rye, rice, and oat. As cereals retain an overall high synteny across large chromosomal segments, we hypothesized that the conservation of coding features among those loci would be high, despite ploidy differences and for noncoding features, the miRNAs with key regulatory functions would be conserved among all cereals analyzed. In addition, we identified putative microRNA (miRNA) encoding sequences within these QTLs that may contribute to the resistance response. Rather than focusing on causal or candidate genes, we sought to explore how these chromosomal regions carrying important information on insect tolerance were shaped in closely related cereal genomes. We also explored sequences within the newly released oat genome assembly that should belong to the genomic loci homologous to these insect tolerance loci.

4.2. MATERIALS & METHODS

4.2.1. Molecular Markers of OWBM and WSS Resistance Loci

Molecular markers delineating the major QTL conferring stem solidness against WSS oviposition (Qss.msub-3BL in hexaploid bread and SSt1 in tetraploid wheat; 3BL-QTL hereafter) were taken from Nilsen et al. (Nilsen et al., 2017). Markers linked to the Sm1 locus, associated with resistance against OWBM, were taken from Kassa et al. (Kassa et al., 2016). Markers for two additional QTLs on 1A and 4A chromosomes in wheat contributing to OWBM resistance were taken from Hao et al. (for QSm.hebau-4A) and Thambugala et al. (for QSm.mrc-1A), respectively (Hao et al., 2019; Thambugala et al., 2021). Sequences of the molecular markers were retrieved from the wheat 90K array (S. Wang, Wong, Forrest, & et al., 2014), CerealDB database (www.cerealsdb.uk.net, last accessed 23.09.2021), and Wheat URGI website (http://wheat-urgi.versailles.inra.fr/, last accessed 23.09.2021). For single nucleotide polymorphism (SNP) probes, both variants were kept.

4.2.2. Datasets used in the study

Cereal species with published genome sequences were used in this study. *T. aestivum* cv. Chinese Spring (hexaploid, AABBDD) IWGSC RefSeq v2.1 genome assembly and annotations were downloaded from Wheat URGI website (http://wheat-urgi.versailles.inra.fr/Seq-Repository, last accessed 21.07.2021; also available at NCBI project no. PRJNA669381). *T. turgidum* ssp. *durum* cv. Svevo (tetraploid, AABB) chromosome sequences, coding sequences and GFF files were downloaded from NCBI (BioProject: PRJEB22687, Assembly: GCA_900231444.1; https://www.ncbi.nlm.nih.gov/assembly/GCA_900231445.1 (last accessed 21.07.2021); LT934111.1[chr1A], LT934114.1[chr2B], LT934116.1[chr3B], LT934117.1[chr4A]). *T. turgidum* ssp. *dicoccoides* genotype Zavitan (tetraploid, AABB) chromosome sequences, coding sequences and GFF files were downloaded from NCBI (BioProject: PRJNA310175, Assembly: GCF_002162165.1; https://www.ncbi.nlm.nih.gov/assembly/GCF_002162155.1 (last accessed 21.07.2021); NC_041380.1[chr1A], NC_041383.1[chr2B], NC_041385.1[chr3B], NC_041386.1[chr4A]). *Hordeum vulgare* cv. Morex IBSC v2 chromosome, coding sequences, and GFF files were downloaded from EnsemblPlants (http://ftp.ebi.ac.uk/ensemblgenomes/pub/release-51/plants/fasta/hordeum_vulgare/, last accessed 14.11.2021). Similarly, *Oryza sativa* ssp. *japonica* cv. Nipponbare IRGSP 1.0 genome, coding sequences, and GFF files were downloaded from EnsemblPlants (http://ftp.ebi.ac.uk/ensemblgenomes/pub/release-51/plants/fasta/oryza_sativa/, last accessed 14.11.2021) (Howe, Contreras-Moreira, De Silva, & et al., 2019). *Secale cereale* line Lo7 pseudomolecules were downloaded from e!DAL IPK (https://doi.org/10.5447/ipk/2020/33 and https://doi.org/10.5447/ipk/2020/29, last accessed 14.11.2021) (Rabanus-Wallace, Hackauf, Mascher, & et al., 2021). Lastly, *Avena sativa* v1.0 genome assembly was obtained from The Oat Genome Project website (Avena Sativa v1.0).

4.2.3. Homology Searches Using BLAST+

BLAST 2.11.0+ standalone package was used for all blast searches (Camacho et al., 2009). Blast databases were generated from a chromosome or whole-genome sequences using makeblastdb. Marker sequences were blasted against these databases using blastn (-outfmt “6 std qlen slen”) and matches with at least 90% coverage of the marker

sequences and at least 95% (for wheat genotypes), 90% (for barley and rye), 85% (for oat), or 80% (for rice) sequence identity to the genomic sequences were retained, based on evolutionary relationships. Markers that map to different locations with the same coverage and sequence identity were not used.

In cases where the number of molecular markers matching the genomic sequences above the filtering criteria given above did not allow identification of homologous regions, homologous transcripts were used instead. Specifically, for the wheat 3BL-QTL, transcript isoforms extracted from the homologous regions (detailed below) of *T. aestivum* Chinese Spring 3B (309 isoforms), *T. durum* Svevo 3B (316 isoforms), *T. dicoccoides* Zavitan 3B (258 isoforms), rye 6R (252 isoforms), and barley 3H (248 isoforms) chromosomes were combined (1383 isoforms in total). Rice IRGSP 1.0 CDSs were blasted against a database constructed from these combined transcripts (blastn; -evalue 1E-10, -outfmt "6 std qlen slen"). Matches were filtered for 80% sequence identity and 50% rice CDS coverage. Of the significant matches, only those that form a continuous interval on a chromosome were detained (this also compensates for the relatively low coverage threshold). Similarly, for wheat 2B-QTL, transcript isoforms from Chinese Spring 2B (452 isoforms), Svevo 2B (380 isoforms), Zavitan 2B (331 isoforms), and rye 7R (460 isoforms) were combined (1623 isoforms in total), and compared to all rice IRGSP 1.0 CDS and barley IBSC v2 CDS separately to define homologous regions in rice and barley, where molecular markers could not. For rice, filtering of the matches was conducted as above (80% identity, 50% coverage). For barley, the matches were filtered for 90% sequence identity and 80% barley CDS coverage, due to widespread conservation between wheat and barley genomes. Finally, for wheat 1A-QTL, Chinese Spring 1A (582 isoforms), Svevo 1A (545 isoforms), Zavitan 1A (559 isoforms), rye 1R (602 isoforms), and barley 6H (633 isoforms) chromosomes were combined (2921 isoforms in total). Rice IRGSP 1.0 CDSs were blasted against a database constructed from these isoforms and filtered as above.

For the wheat 4A-QTL, due to the scarcity of molecular markers saturating the putative QTL, 58 genes, from the earlier IWGSC v1.0 assembly, identified to reside within the resistance loci were taken in a recent publication (Lijing Zhang et al., 2020). The longest transcript isoforms from these genes were blasted against all coding sequences from Chinese Spring, Svevo, Zavitan, barley, rye, and rice (blastn; -evalue 1E-10, -outfmt "6 std qlen slen"). Matches with at least 80% coverage of the query transcript sequences and

at least 95% (for wheat genotypes) and 90% (for barley and rye) or 80% (for rice, coverage threshold was also lowered to 50% due to increased evolutionary distance) sequence identity to the genomic sequences were retained.

For all resistance loci (3BL-QTL, 2B-QTL, 1A-QTL, and 4A-QTL), all transcripts identified in homologous regions in wheat, barley, rye, and rice were combined and compared against oat genome assembly (blastn; -evalue 1E-10, -outfmt "6 std qlen slen"). Due to the presence of introns in the genomic contigs, matches were considered significant if the percent sequence identity was above 85% over at least 300 aligned nucleotides.

To gain insight into probable functions of the genes within identified genomic loci, protein sequences encoded by these genes were compared against (1) all annotated and reviewed *Viridiplantae* proteins from the UniProt database (https://www.uniprot.org/; 40,927 sequences, last accessed 18.10.2021), (2) the annotated proteome of the model grass *Brachypodium distachyon* (v3.0; retrieved from EnsemblPlants https://plants.ensembl.org/Brachypodium_distachyon/Info/Index, last accessed 18.10.2021 (Howe et al., 2019)) and (3) iTRAQ-based proteomic data (unpublished) obtained from semi-solid stemmed wheat variety Scholar (PI607557) and solid stemmed wheat variety Conan (PI6907549) which both carry alleles associated with stem-solidness at *QSS.msub-3BL*. The matches were filtered for at least 50% sequence similarity and 50% coverage of the wheat, barley, rye, or rice protein for Uniprot blasts and at least 75% sequence similarity and 50% coverage for the *Brachypodium* proteome for closer evolutionary distance. Values for blast parameters, including similarity and coverage, are given in Appendix B: Table S2, S3, and S8, where, in specific cases, the significance of a match can be evaluated if needed. For the miRNA targets, since the above approach did not provide useful matches to either Uniprot or *Brachypodium* proteins in general, a further step was taken to compare these against all *Viridiplantae* proteins in the NCBI database web interface. The top 5 matches were provided to gain insight into the functions of these targets (Supplementary Table S7).

4.2.4. Extracting Transcripts from Homologous Regions

The borders for regions homologous to the WSS and OWBM resistance QTLs were determined using (1) molecular markers or (2) homologous transcripts, as detailed above.

The transcripts within each region were then extracted using custom python3 scripts, using GFF files for the respective chromosomes, fasta files containing the coding sequences, and homologous region borders as input. Using the borders, all 'gene' features within the GFF file are listed, and the CDS IDs for each gene are outputted. In cases where a gene encodes several isoforms, only the longest CDS is retained. Isoforms were manually eliminated in cases where a homologous region was defined using homologous transcripts from the other species, as detailed above.

4.2.5. Comparative Analysis of Transcripts in Homologous Regions

For a comparative analysis of the organization of transcripts within the homologous regions, transcript sequences extracted from these regions as above were individually blasted against transcripts from the homologous region in Svevo for WSS-resistance loci (on 3B) and against transcripts from the homologous region in Chinese for OWBM-resistance loci (on 2B and 4A). Homologous transcripts were identified using blastn (-evalue 1E-10, -outfmt "6 std qlen slen") and significant matches were filtered for at least 80% coverage of the marker sequences and at least 95% (for wheat genotypes), 90% (for barley and rye) or 80% (for rice) sequence identity. Visualization was performed with custom python3 scripts using the matplotlib library.

4.2.6. Identification of Putative microRNAs and Target Sequences

Chromosomal sequences of the identified homologous regions were extracted from each respective chromosome with custom python3 scripts. MicroRNA identification from each homologous region was done using the mirMachine pipeline (H. Busra Cagirici et al., 2021).

Potential target transcripts of identified miRNAs from the homologous regions from each species were identified using mature miRNA sequences of each cereal crop in the psRNAtarget tool (Dai & Zhao, 2011) against the high confidence coding sequences extracted from those loci. Protein sequences of the transcript targets mapped on the homologous regions are then protein-blasted against non-redundant protein sequences of

Viridiplantae taxa using blast tool on (https://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed 24.08.2021).

4.3. RESULTS

4.3.1. The Content and Organization of the Coding Features around the SSt1 Locus

The causal gene, *TdDof*, residing within the major QTL on the long arm of chromosome 3B (Qss.msub-3BL in hexaploid bread and *SSt1* in tetraploid wheat; 3BL-QTL hereafter, for simplicity) that confers stem solidness in wheat has been recently identified (Nilsen et al., 2020). To understand how this locus has been broadly shaped in cereal genomes, molecular markers defining the 3BL-QTL (Nilsen et al., 2017) were first mapped to the 3B chromosomes in *Triticum aestivum* cv. Chinese spring (hexaploid, AABBDD, Chinese Spring hereafter) and *Triticum turgidum* ssp. *durum* cv. Svevo and *Triticum turgidum* ssp. *dicoccoides* genotype Zavitan (tetraploid, AABB, Svevo, and Zavitan, respectively, hereafter) and to the homologous 3H chromosome in *Hordeum vulgare* cv. Morex (barley). Additionally, due to known rearrangements in the rye genome, the molecular markers were mapped to the entire genome of Secale cereale line Lo7 (rye) (Devos et al., 1993). Similarly, due to the evolutionary distance and the small genome size, molecular markers were mapped to the entire *Oryza sativa* ssp. *japonica* cv. Nipponbare genome (rice) (International Rice Genome Sequencing Project, 2005). Moreover, 3BL-QTL molecular markers were compared against the recently released genome assembly of *Avena sativa* (oat) (Avena Sativa v1.0, n.d.). Using these markers, homologous regions on wheat 3B chromosomes, barley 3H chromosome, and rye 6R chromosome were identified, spanning 12–27 Megabases (Mb) (Appendix B: Table S1). Due to increased evolutionary distance in rice, a potential locus homologous to the 3B-QTL could not be identified through the molecular markers.

Chromosome subsequences for the given intervals were extracted for a comparative analysis of coding and non-coding features. For coding features, each gene's longest

isoform within the subsequences was retrieved from respective GFF files. With this approach, 309 transcripts for Chinese Spring, 316 transcripts for Svevo, 258 transcripts for Zavitan, 248 transcripts for barley, and 252 transcripts for rye were extracted. These transcripts were first used to identify the homologous region in the rice genome, which could not have been determined using the 3BL-QTL molecular markers. For this, all the transcript isoforms were combined and blasted against all rice coding sequences. Of the significant hits, transcripts between Os01t0958700-01 and Os01t0977200-00 on rice chromosome 1, spanning ~1Mb, were determined to be homologous to the 3BL-QTL in wheat. This interval contains 46 rice genes. Potential functions of the transcripts from all cereal loci were deduced based on homology to annotated proteins from the Uniprot database and/or fully annotated proteome of *Brachypodium distachyon*, the model plant for monocots (Appendix B: Table S2). Furthermore, the protein sequences identified in the homologous regions were compared to unpublished proteomics data of semi-solid/solid-stemmed wheat cultivars upon infestation (Appendix B: Table S3). Since the causal gene for stem solidness was first described in the Svevo genome, we compared the organization of these coding features within each homologous chromosomal interval in the cereal genomes with respect to Svevo, which indicated a high level of overall conservation across all regions. The distal regions of each homologous interval seemed to be more prone to the rearrangements, where conservation also tended to be lost as the evolutionary distance increased (Figure 4.1).

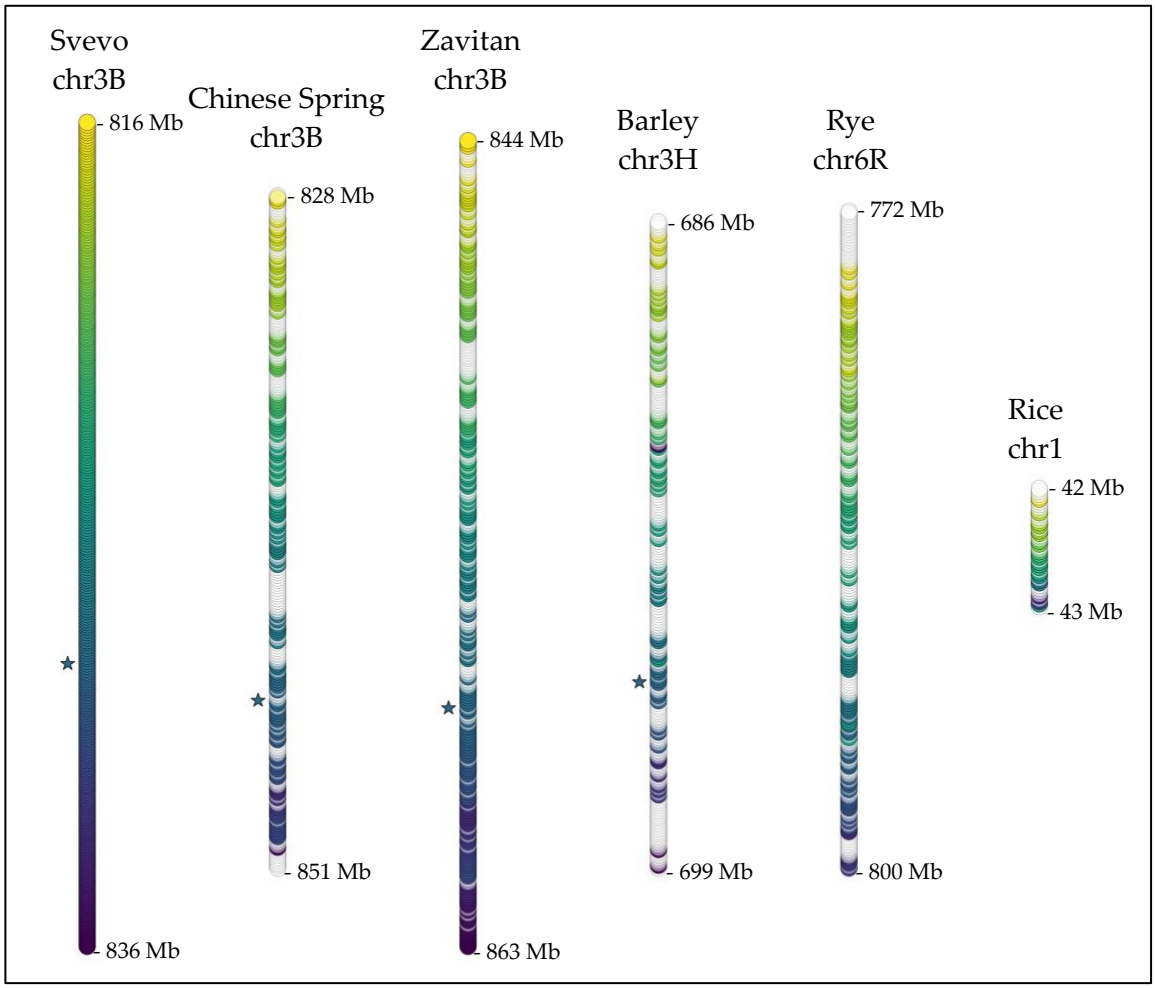


Figure 4.1 The coding organization along the 3BL-QTL in Svevo chromosome 3B and homologous regions on Zavitan 3B, Chinese Spring 3B, barley 3H, rye 6R, and rice 1 chromosomes. Stacked circles indicate genes ordered along each chromosomal segment, and colors only indicate homologous relationships to Svevo genes; genes without Svevo homologs are not colored. Color patterning is for visual purposes. The lengths of the chromosome segments are not proportionate to their physical lengths but rather to the total number of genes within each segment. Stars indicate the causal gene (TRITD_3Bv1G280530) in the Svevo 3BL-QTL and its homologs in the other chromosomal segments if present.

Currently, only a fragmented genome assembly is available in oat, which restricts direct comparisons of the genomic organization of homologous loci or features. However, molecular markers defining the 3BL-QTL matched 12 contigs that may represent a homologous region in the oat genome. Additionally, the transcript sequences identified from homologous regions on wheat, barley, rye, and rice chromosomes matched 172 oat

contigs. In total, 176 oat contigs may be associated with the genomic loci that are homologous to wheat SSt1 (Appendix B: Table S4).

4.3.2. miRNAs Identified from the SSt1 and Homologous Loci in Cereals

Homology-based miRNA identification from the given intervals of the chromosome subsequences indicated a total of 485 precursor miRNA sequences for 38 miRNA families. miR1122, miR1137, miR1120, miR1127, and miR5049 families were the miRNA families with the most abundant precursor miRNA sequences (Appendix B: Table S5). The number of miRNA families identified in Chinese Spring, Svevo, Zavitan, and rye was found to be close to each other, 23, 26, 23, 22, respectively, of which 16 were common to all those species. Relatively fewer, 16 miRNA families were identified from barley chromosome subsequence, and in rice, there were three identified miRNA families. miRNA families identified from homologous 3BL QTL were mostly conserved between closely related wheat, barley, and rye. While the relatively smaller region that could be identified in rice did not permit meaningful comparisons with other cereals, of the three families that could be identified in rice, miR1130 was common to all. We have identified eight miRNA families present only in wheat genotypes (Figure 4.2). All miRNA families identified in 3BL QTL regions have been shown in Appendix B: Table S6. mRNA target analysis of the putative miRNAs has shown that miR1127 and miR1439 of Chinese Spring, miR1118 of Svevo, and miR5049 of Zavitan have potential targets within the same 3BL-QTL homologous regions of each respective genotype. The transcript targets were then compared to all *Viridiplantae* proteins to reveal any annotated homologs. miR1118 from Svevo is predicted to target a Pik2-like protein which may be involved in insect tolerance while the other miRNA targets are mostly uncharacterized proteins (Appendix B: Table S7).

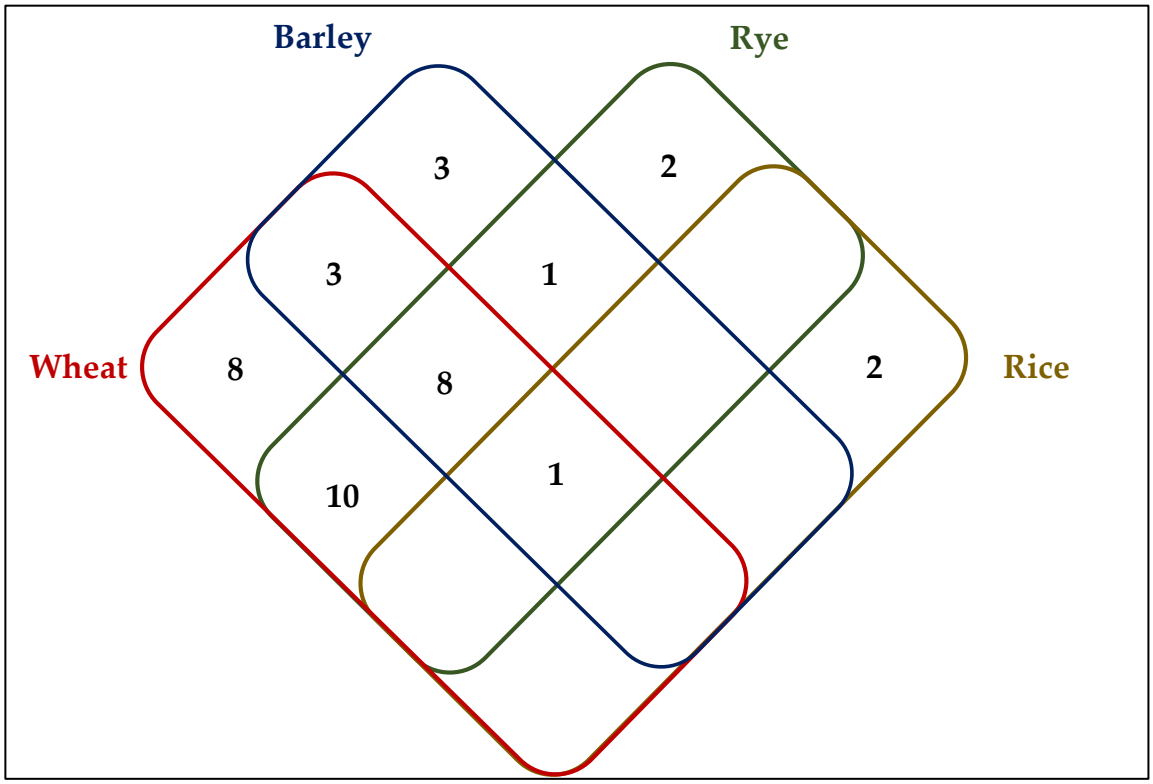


Figure 4.2 Thirty-eight miRNA families identified from homolog QTL of *WSS* 3B in Chinese Spring, barley, rye, rice, Svevo, and Zavitan. Chinese Spring, Svevo, and Zavitan miRNA families are grouped and shown as wheat miRNA families. The numbers indicate the number of miRNA families identified in each cereal species that are common (or not) with the others.

4.3.3. Content and Organization of the Coding Features across Genomic Regions Homologous to *Sm1* Locus

Extensive genome sequencing in diverse wheat genotypes has led to identifying the candidate gene for the Sm1 locus on chromosome 2B, conferring resistance to another devastating pest, OWBM (Walkowiak et al., 2020). To explore the genomic organization of this locus on cereal genomes, molecular markers associated with Sm1 were mapped on Chinese Spring 2B, Svevo 2B, Zavitan 2B, barley 2H chromosomes, as well as rye and rice genomes. While homologous regions of 25–33 Mb in length were identified in wheat 2B and rye 7R chromosomes, too few molecular markers were mapped to the barley 2H chromosome and the rice genome (Appendix B: Table S1). For barley and rice, transcript isoforms from wheat (452, 380, and 331 isoforms from Chinese Spring, Svevo, and Zavitan 2B chromosomes, respectively) and rye (460 isoforms from 7R chromosome) homologous regions were used to identify homologous regions of 24.9 Mb on barley 2H

chromosome and 7.2 Mb on rice chromosome 4, spanning 204 and 52 genes, respectively (Appendix B: Table S1). Potential functions of these transcripts were inferred based on homology to annotated proteins from the Uniprot database, fully annotated proteome of *B. distachyon*, and an unpublished proteomics study of semi-solid and solid stemmed wheat cultivars (Appendix B: Tables S3 and S8). The organization of the genes in the homologous regions was compared with respect to Chinese Spring since the candidate gene was first defined in the hexaploid background (Walkowiak et al., 2020). With respect to Chinese Spring, extensive rearrangements, including an apparent inversion was evident in the rye 7R chromosome. The conservation of the coding features in the proximal regions of Sm1 also appeared to be lost in rye 7R (Figure 4.3).

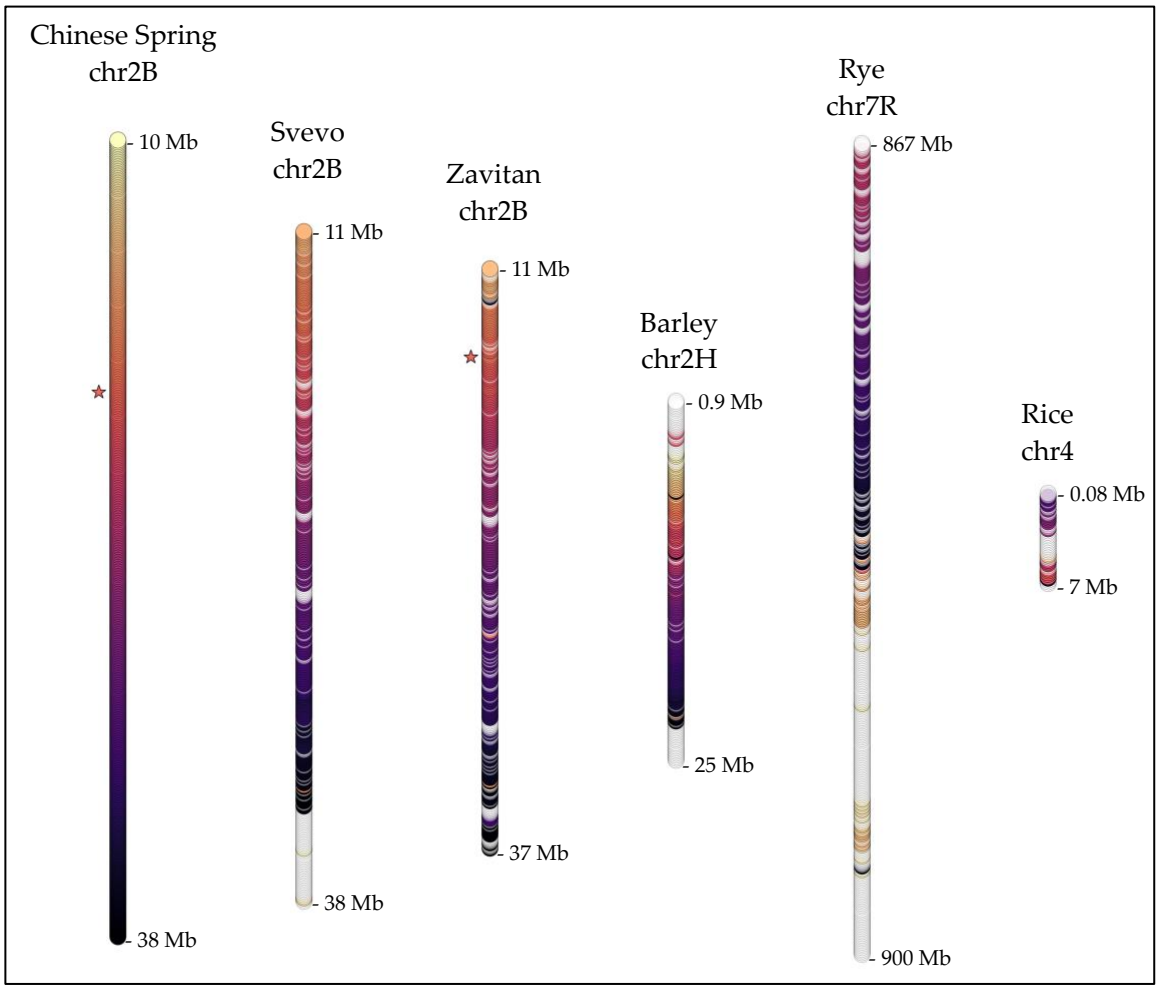


Figure 4.3 The organization of the coding features along the Sm1-homologous regions in Chinese Spring chromosome 2B, Svevo chromosome 2B, Zavitan chromosome 2B, barley chromosome 2H, rye chromosome 7R, and rice chromosome 4. Colors and patterning are as in Figure 4.1, except in reference to Chinese Spring. Stars indicate the best Chinese Spring match (TraesCS2B03G0071700.2) to the Sm1 candidate gene described in CDC Landmark in Walkowiak et al. (Walkowiak et al., 2020) and its homolog in Zavitan 2B.

The comparison of the oat genome assembly to the molecular markers associated with the Sm1 locus identified three contigs. These contigs were among the 276 oat contigs that also significantly matched transcript sequences from wheat, barley, rye, and rice chromosomal regions belonging to or homologous to the Sm1 loci. These 276 may represent a genomic locus in the oat genome that is homologous to the Sm1 locus, even though it is not yet possible to deduce the genomic organization of this region compared to its relatives due to the fragmented nature of the current oat genome assembly (Appendix B: Table S4).

4.3.4. miRNAs identified from Genomic Loci Homologous to the Sm1 Locus in Cereals

miRNA identification analysis from Chinese Spring, Svevo, Zavitan, barley, rye, and rice chromosome subsequences yielded 24, 25, 25, 12, 22, and 51 miRNA families, respectively, resulting in a total number of 76 miRNA families with 659 precursor sequences (Appendix B: Table S5). miR1122, miR1120, and miR5049 were found to have the highest number of precursor miRNA sequences, and miR1128 and miR1130 were identified in all chromosome subsequences. Chinese Spring, Svevo, and Zavitan genotypes have miR1121, miR1125, and miR1136 families not identified in barley, rye, and rice, whereas miR1439 was only identified in barley, rye, and rice (Appendix B: Table S6). Figure 4.4 shows the conservation of miRNA families among the species analyzed. Three miRNA families were commonly predicted from all cereal loci analyzed. There is a vast amount of miRNA families, 44 miRNA families, predicted only in rice. We identified eight miRNA families only present in wheat genotypes. psRNAtarget analysis of mature miRNA sequences of putative miRNA families against coding sequences of homologous regions revealed five miRNA families—miR1120, miR1127, miR1135, miR2118, and miR9782—with at least one transcript target mapped on these homologous regions. Putative targets of miR2118 in Zavitan were predicted to be a disease resistance protein RPM-like gene, whereas a Morf Related Gene 1 (MRG-1)-like protein was predicted to be targeted by miR1127 of Chinese Spring and Zavitan. Functional aspects of the putative targets of other 3 miRNA families remain mostly unidentified as the target transcripts have not shown homology to well-characterized plant proteins (Appendix B: Table S7).

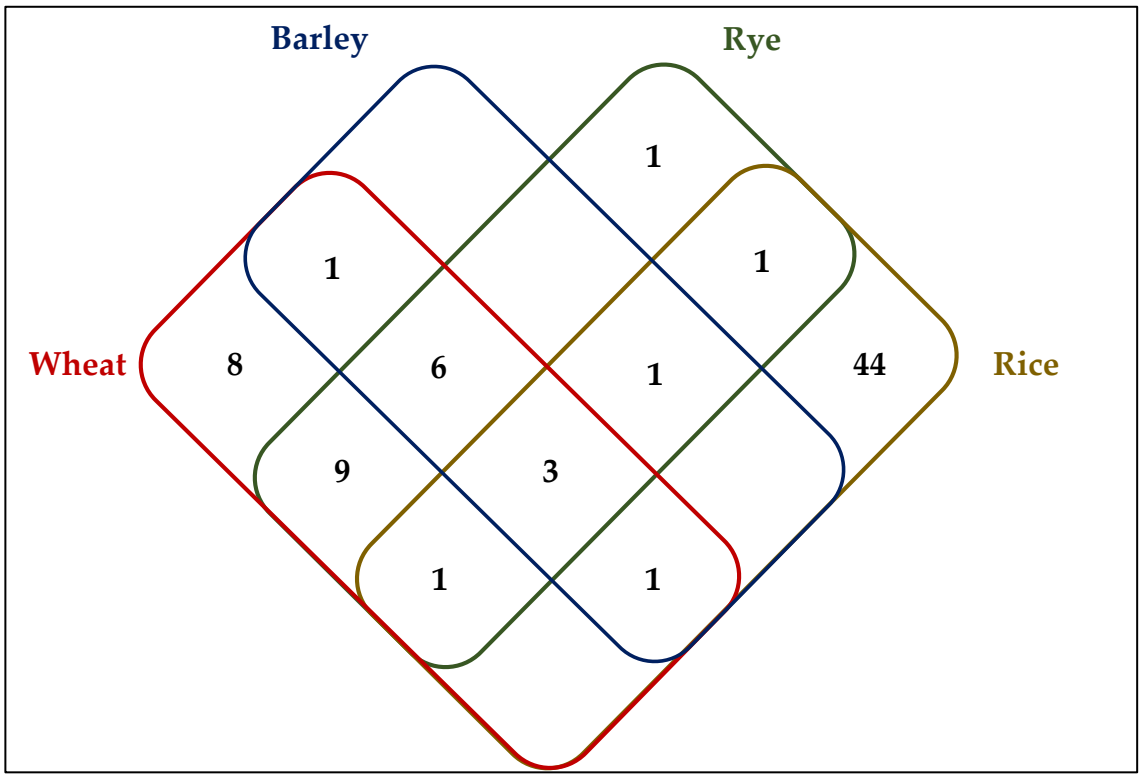


Figure 4.4. Seventy-six miRNA families identified from Chinese Spring, Svevo, Zavitan (all shown as wheat), barley, rye, and rice homologs of the 2B QTL region. Apart from 44 miRNA families identified only in rice, the remaining 22 miRNA families are primarily conserved in wheat, barley, and rye. The numbers indicate the number of miRNA families identified in each species that are common (or not) with the others.

4.3.5. Comparison of the Coding Features across Genomic Regions Homologous to Additional Loci Associated with OWBM Resistance

In addition to the antibiotic resistance provided by the Sm1 locus, additional loci have also been associated with OWBM resistance through different mechanisms. One known major QTL, associated with oviposition deterrence, resides on chromosome 1A in wheat (Blake et al., 2011; Thambugala et al., 2021). Using molecular markers mapped on this 1A-QTL (Thambugala et al., 2021), regions of 28–62 Mb homologous to this QTL were identified in wheat chromosomes 1A, barley chromosome 1H, and rye chromosome 1R (Appendix B: Table S1). In rice, homologous transcripts from wheat (582, 545, and 556 isoforms on Chinese Spring, Svevo, and Zavitan 1A chromosomes, respectively), barley (633 isoforms on 1H), and rye (602 isoforms on 1R) identified a 9 Mb region on chromosome 5, that contained 193 genes. The organization of these genes along the homologous regions, with respect to Chinese Spring 1A, suggested overall conservation

of the gene sequences and orders, except for an apparent inversion in Svevo 1A (Figure 4.5).

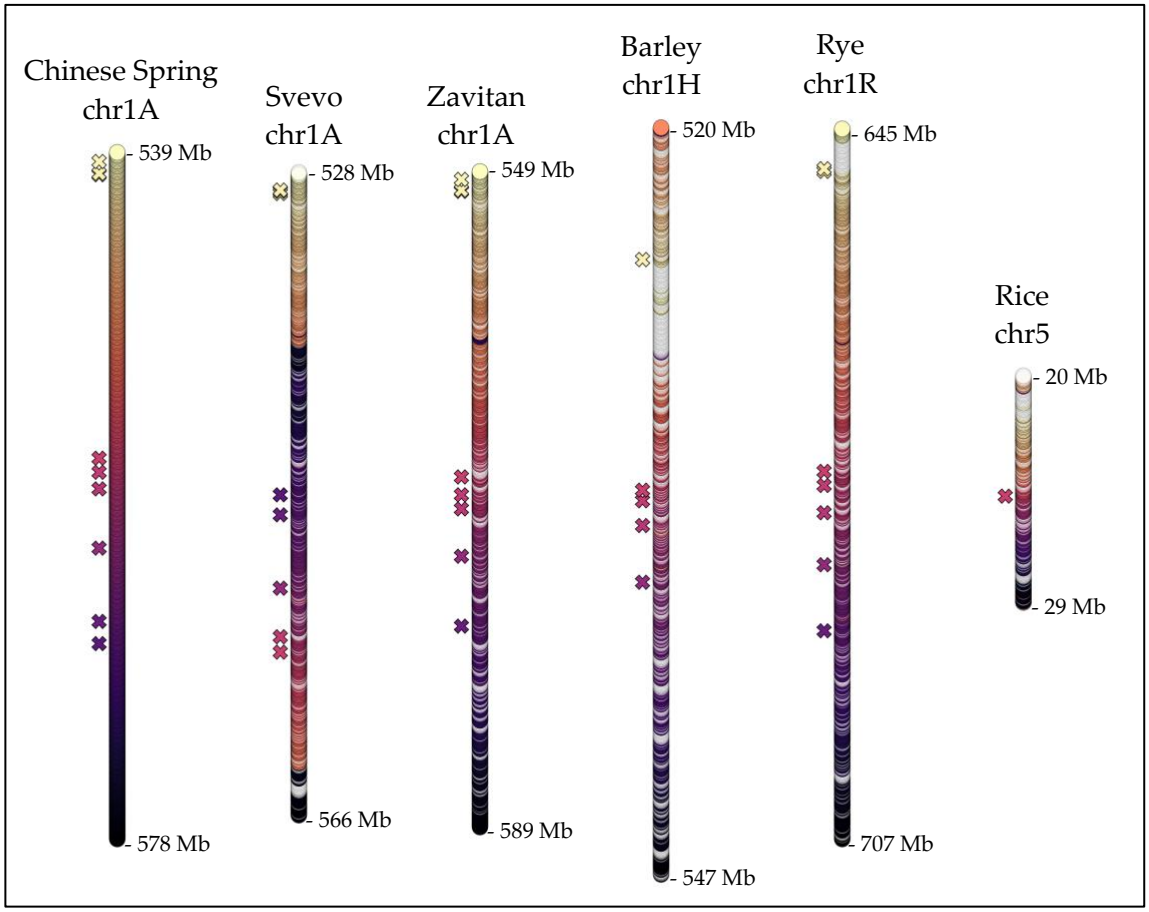


Figure 4.5 The organization of the coding features along an additional QTL for oviposition deterrence against OWBM, on wheat 1A chromosomes and homologous regions on barley chromosome 1H, rye chromosome 1R, and rice chromosome 5. Colors and patterning are as in Figure 4.3, in reference to Chinese Spring. 'X' symbols indicate genes homologous to potential candidates for the 1A-QTL proposed in a recent study (Thambugala et al., 2021).

Another major QTL for OWBM resistance was mapped to the long arm of chromosome 4A in wheat, where the candidate gene has not been identified yet (Hao et al., 2019; Lijing Zhang et al., 2020). While relatively few markers have been mapped to this resistance loci, a 4.9 Mb genomic region has been defined on the previous assembly of the Chinese Spring genome (IWGSC v1.0 assembly) that included 58 genes (Lijing Zhang et al., 2020). Therefore, the longest isoforms encoded by these genes were used to identify the transcripts that may be associated with the 4AL-QTL in the current version of the Chinese Spring genome (IWGSC v2.1 assembly) as well as homologous transcripts in Svevo, Zavitan, barley, rye, and rice. These transcripts identified 5.49–9.5 Mb regions on

homologous wheat 4A and rye 7R chromosomes and a relatively small region of only 3.75 Mb on the barley 7H chromosome. In rice, only a few homologous transcripts were identified on chromosome 6 (Appendix B: Table S1). These regions contained 54 genes in Chinese Spring 4A, 75 genes in Svevo 4A, 36 genes in Zavitan 4A, 18 genes in barley 7H, 34 genes in rye 4R, and 7 genes in rice chromosome 6. Comparative analysis indicated shared patterns of rearrangements in durum wheat Svevo and Zavitan and rye and barley (Figure 4.6). Putative functions for the transcripts identified in 1A-QTL and 4AL-QTL homologous loci are given in Appendix B: Tables S3 and S8.

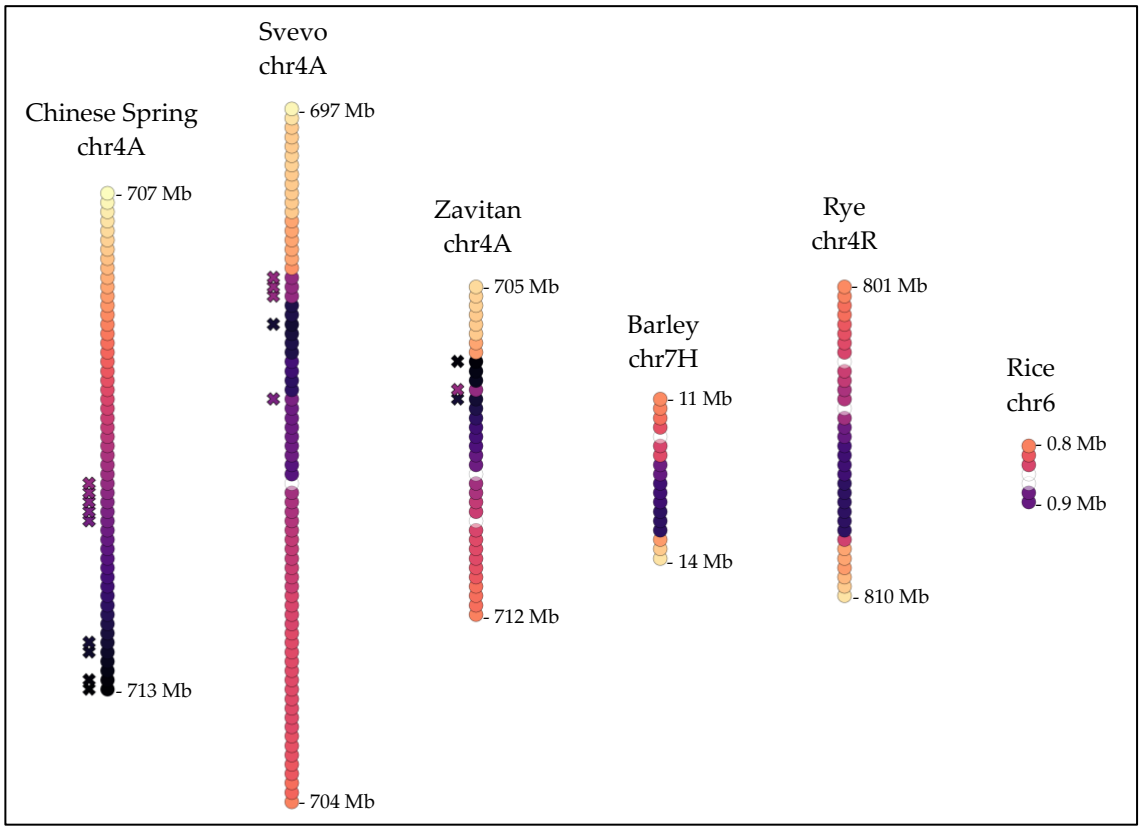


Figure 4.6 The organization of the coding features along a second QTL for oviposition deterrence against OWBM, on wheat 4A chromosomes and homologous regions on barley chromosome 7H, rye chromosome 4R, and rice chromosome 6. Colors and patterning are as in Figure 4.3 in reference to Chinese Spring. 'X' symbols indicate genes homologous to potential candidates for the 4A-QTL proposed in two recent studies (Hao et al., 2019; Lijing Zhang et al., 2020).

The comparison of molecular markers associated with the 1A-QTL to the oat genome assembly identified 18 contigs that may belong to a homologous region on the oat genome. Curiously, homologous transcripts identified in wheat, barley, rye, and rice matched a baffling total of 4659 contigs from the oat genome assembly. For the 4A-QTL,

where relatively small genomic regions were comparatively analyzed, transcripts combined from all homologous loci identified 12 contigs that may represent a homologous region in the oat genome (Appendix B: Table S4).

4.3.6. The Comparison of miRNAs across Genomic Regions Homologous to Additional Loci Associated with OWBM Resistance in Cereal Genomes

Sixty-one miRNA families with 746 precursor miRNA sequences were identified from chromosome subsequences of 1A-QTL homologous regions, consisting of 35 miRNA families identified from Chinese Spring, 35 and 32 miRNA families from Svevo and Zavitan, respectively. A relatively smaller number of miRNA families, 27 for rice, 25 for rye, and 16 miRNA families for barley chromosome subsequences, were identified. miR1120, miR1122, miR5049, miR1127, and miR1137 are the miRNA families with the highest number of pre-miRNA sequences (Appendix B: Table S5). Similar to the results of homologous 2B-QTL miRNA identification, many rice miRNA families were identified. Ten miRNA families were identified to be present only in wheat genotypes. Rye has been found to have more common miRNA families with wheat genotypes than barley (Figure 4.7a). A comparison of all mature miRNA sequences against all transcripts identified from the respective homologous loci suggested that, overall, nine miRNA families may target transcripts from the same loci. Even though the majority of these transcripts did not have functionally well-characterized homologs among *Viridiplantae*, putative targets of miR1118, miR1439, and miR2275 had matches to genes that may be involved in disease resistance. Specifically, putative miR2275 targets were highly similar to resistance gene analog (RGA) type of genes, whereas the miR1118 and miR1439 families putatively targeted PIK-like isoforms. It should be noted that putative miR1439 targets in barley solely matched uncharacterized proteins, which indicate that these may include resistance-associated genes that are yet to be characterized (Appendix B: Table S7).

Due to the low number of 4AL-QTL homologous transcripts identified in rice, miRNA identification was performed by excluding rice chromosome subsequences. A total of 224 precursor miRNA sequences resulted in total of 24 miRNA families, which composed of

17 miRNA families identified from Chinese spring, 18 from Svevo, 17 from Zavitan, 6 miRNA families from barley and 15 from rye chromosome subsequences, and 5 of those miRNA families, miR1127, miR1122, miR5049, miR1120 and miR1130 are found to be common to all species (Appendix B: Table S5 and S6). Figure 4.7b shows how the identified miRNA families are shared among wheat genotypes, barley, and rye. Eight miRNA families were identified only in wheat genotypes, and the remaining miRNAs are mostly conserved among all analyzed species. None of the predicted miRNA families targeted transcripts on the 4A-QTL homologous regions.

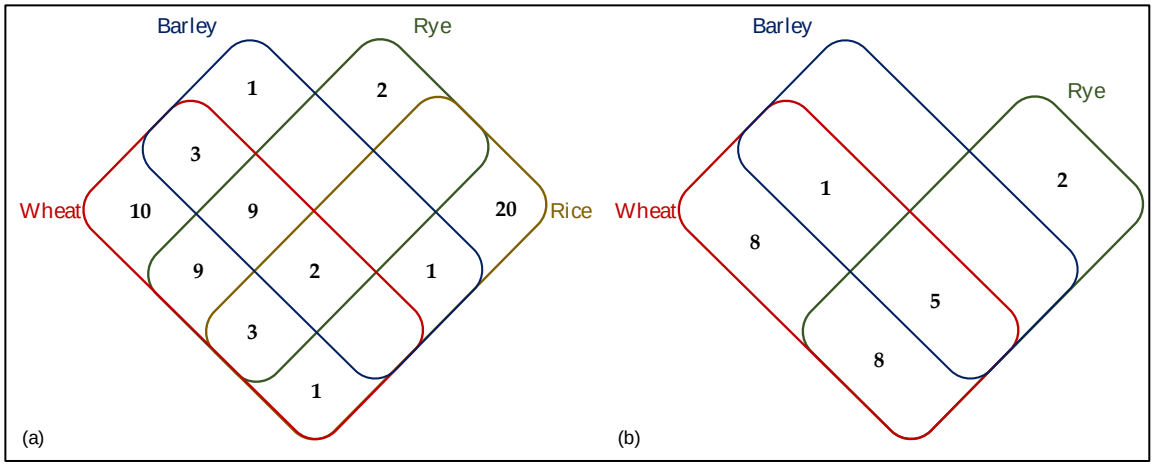


Figure 4.7 a) Distribution of 61 miRNA families identified in 1A-QTL homologs among wheat genotypes, barley, rye, and rice. Each number indicates the number of miRNA families identified in each cereal species that are shared (or not) with the others. Three miRNA families were shown to be shared among all analyzed species, whereas eight miRNAs were conserved in wheat, barley, and rye. (**b**) Distribution of 24 miRNA families identified in 4A-QTL homologs among wheat genotypes, barley, and rye. The number of miRNA families identified in each cereal and the number of miRNA families common to each other are indicated by each number. Due to smaller chromosome subsequences, 4A-QTL yielded relatively fewer miRNA families compared to other QTLs. Among 24 miRNA families, wheat and rye share 13 miRNA families, whereas barley has 6 common miRNA families with wheat.

4.4. DISCUSSION & CONCLUSION

4.4.1. Discussion

Food security of the upcoming generations depends on our ability to develop resilient crops not only to the more frequent and severe weather extremes caused by global warming but also to the devastating biotic threats by pests and pathogens that are constantly under an arms race with their hosts. In this study, we focused on two crucial wheat pathogens, Wheat Stem Sawfly (WSS; Cephus cinctus Norton) and Orange Wheat Blossom Midge (OWBM, Sitodiplosis mosellana), and explored the content and organization of coding and non-coding features on multiple resistance loci across cereal genomes. The focus of this study is not the causal or candidate genes, which have recently been firmly established for two of the loci already (Nilsen et al., 2020; Walkowiak et al., 2020), but rather to provide a comparative overview of these loci across essential cereal crops and search cues into how evolution has acted on them.

Modern grass genomes, including cereals, are believed to have evolved from an ancestor with 7 predicted proto-chromosomes, through whole-genome duplications and chromosomal rearrangements (Murat, Armero, Pont, Klopp, & Salse, 2017; Pont & Salse, 2017). The extant species still retain considerable synteny across large chromosomal segments (Murat et al., 2017; The International Brachypodium Initiative, 2010), to an extent where the syntenic relationships help build reference assemblies and/or inferred gene orders for closely related complex genomes (Mayer, Martis, Hedley, & et al., 2011). Functionally important loci can be expected to remain in such well-conserved syntenic loci. Stem solidness trait has been utilized as the primary defense strategy against WSS in wheat. The causal gene, *TdDof,* of the major QTL on chromosome 3BL controlling stem solidness was recently identified and characterized (Nilsen et al., 2020). When this QTL and homologous regions in related *Poaceae* species were compared, a high level of

synteny was observed (Figure 4.1). While the sequence conservation of the genes seemed to be lost towards the distal end, the overall order of the genes seemed to be well preserved. The *TdDof* gene is identified in the tetraploid durum wheat and is present in the hexaploid bread wheat genome. A putative homolog was also identified in the barley genome; however, this does not inform upon stem solidness in barley as the copy number, not the mere presence, of *TdDof* regulates pith filling of the stems (Nilsen et al., 2020). In rye, the region homologous to the wheat 3BL-QTL was clearly involved in the ancestral translocation between chromosomes 3 and 6 (Devos et al., 1993) yet, this ancestral translocation did not seem to disrupt the homology to a great extent (Figure 4.1). The long arm of chromosome 6R also appears to be enriched in nucleotide-binding site leucine-rich repeat (NLR) genes involved in the pathogen resistance (Rabanus-Wallace et al., 2021). It is, thus, tempting to wonder that, beyond stem solidness, whether this locus has been under purifying selection during the *Poaceae* evolution for its importance in biotic stress responses.

Several genomic loci have also been associated with resistance against another devastating pest, OWBM. Most notably, the Sm1 locus on the short arm of chromosome 2B in wheat has long been known to confer resistance through antibiosis. A candidate gene for this locus has also been recently suggested (Walkowiak et al., 2020). Compared to the 3BL-QTL, the Sm1 locus appeared to be less conserved across *Poaceae* (Figure 4.3). Despite regions of similar physical length, Svevo and Zavitan might have lost homologous sequences to the proximal and distal parts of the Sm1 locus in Chinese Spring. Additionally, a major re-arrangement might have occurred in the rye, after rye and barley had diverged from their last common ancestor. In rye, complex ancestral rearrangements that translocated the most telomeric part of the short arm of ancestral chromosome 2 to the telomeric part of the long arm of ancestral chromosome 7 might have disrupted the grass homology within this locus. A proposed model that involves multiple steps of translocation to the formation of the modern 7R chromosome may explain the extensive loss of homology for this region (Devos et al., 1993). In rice, conservation of only a few genes at the microscale was observed. Interestingly, a homologous sequence to the *Sm1* candidate gene was found only in the wild genotype Zavitan but not in the domesticated Svevo cultivar (Figure 4.3).

At the sequence level, extensive conservation across cereals was evident for another known QTL on wheat chromosome 1A that provides resistance by interfering with the

oviposition of OWBM eggs. Nevertheless, an apparent inversion on the Svevo 1A chromosome was observed (Figure 4.5). A previous study has proposed 11 candidate genes for the 1A-QTL based on the functional annotation (Thambugala et al., 2021). Several of these had homologous sequences within each chromosomal segment analyzed, further supporting that these loci may be structurally and functionally very well preserved. The identification of the candidate gene that is responsible for oviposition deterrence will await further experimental studies. Despite limited genomic data on another QTL on wheat chromosome 4A that contributes to the OWBM resistance (Hao et al., 2019; Lijing Zhang et al., 2020), a comparison of homologous genomic regions in cereals revealed an interesting observation. Genome-specific rearrangements might have occurred twice within *Triticeae*: Once after the hybridization of the D-genome to the tetraploid wheat ancestor, giving rise to the hexaploid bread wheat and once after the divergence of barley and rye from their last common ancestor with wheat (Figure 4.6). A recent inversion might have shaped the Chinese Spring QTL compared to the tetraploid Svevo and Zavitan, and a more complex rearrangement might have shaped the common ancestor or rye and barley.

The proteomics approach provides a unique chance to investigate plant stress responses in more depth; thus, studies on complex abiotic and biotic stress responses on cereals hold great importance (Budak, Akpinar, Unver, & Turktas, 2013). Putative functional annotation of homologous insect tolerance loci has identified the conserved proteins encoded by the genes in these loci (Appendix B: Table S2, S3, and S8). While the causal genes for 3BL-QTL and 2B-QTL have been identified (Nilsen et al., 2020; Walkowiak et al., 2020) and the 1A-QTL has been under scrutiny for some time (Blake et al., 2011), the 4A-QTL has very recently been proposed and relatively little is known on this potential QTL that confers oviposition deterrence against OWBM. The inferred functions of the proteins encoded by the genes within this QTL may provide interesting insight into potential candidates (Appendix B: Table S8). Among these, ENHANCED DISEASE RESISTANCE 2-like was found across all cereal loci, along with growth-related GROWTH-REGULATING FACTOR 5 and SENESCENCE-RELATED GENE 1. In addition, CYCLOPS, usually together with calmodulin-like proteins, was found across all loci. Interestingly, CYCLOPS proteins and calcium and calmodulin-dependent kinases are believed to constitute an ancient signaling complex that is required for infection of symbiotic bacteria (Yano et al., 2008). Homologous sequences to the stress-related

chaperone ClpB1 (Lee et al., 2007) were found in wheat and barley. Similarly, sequences homologous to 7-DEOXYLOGANETIN GLUCOSYLTRANSFERASE involved in the biosynthesis of iridoids, important defense compounds against herbivores (Nagatoshi, Terasaka, Nagatsu, & Mizukami, 2011), were found in wheat barley and rye. Of the candidate genes that were proposed in the recent studies (Hao et al., 2019; Lijing Zhang et al., 2020), three (TraesCS4A03G1085400.1, TraesCS4A03G1086100.1, TraesCS4A03G1092700.1; CS annotation v2.1) were similar to BISDEMETHOXYCURCUMIN SYNTHASE, which is involved in flavonoid biosynthesis, had homologs only in wheat genotypes. Another candidate gene, TraesCS4A03G1086200.1, matched 2-HYDROXYISOFLAVANONE DEHYDRATASE. Svevo genes TRITD_4Av1G250060 and TRITD_4Av1G250090 also matched this protein, which is involved in isoflavone biosynthesis and is targeted by bacterial virulence factors to facilitate infection in soybean (H. Zhou et al., 2011). The 4A-QTL and homologous loci in cereals may encode genes that are involved in the synthesis of secondary metabolites and defense compounds (Appendix B: Table S8).

Non-coding genomes have gained interest after discovering their regulatory functions on the coding parts of the genomes (Budak et al., 2020). Many important plant microRNAs have previously been shown to be well conserved among *Viridiplantae* for their vital role as gene regulators. Manipulation of miRNAs by molecular techniques thus holds great promise for the future of economically important crops (Djami-tchatchou et al., 2017). Homology-based miRNA identification from the chromosome subsequences of OWBM and WSS QTLs of Chinese Spring, Svevo, Zavitan, barley, rye, and rice have shown the conservation of miRNA families as well as the emergence of new miRNA families. 1A-QTL and 2B-QTL homologous regions resulted in many rice-specific miRNA families which are not present in other cereals. The small homologous region of 3BL-QTL on rice chromosome 1 resulted in a lower number of miRNA families compared to other loci (Figure 4.2). Wheat, barley, and rye belong to the *Triticeae* tribe and have a more recent ancestor after divergence with rice (Hands & Drea, 2012). The presence of miRNA families identified only in rice may be explained by their evolutionary distance. miR1136 was only identified in Svevo 2B-QTL, as the causal gene for this locus was first identified in Svevo, this miRNA may have a function in insect tolerance regulation. In 2B-QTL, we have identified 44 rice-specific miRNA families (Figure 4.4). As the homologous 2B-QTL region defined for rice corresponds to nearly a whole chromosome, the high miRNA

family number was expected. Rice belongs to the Oryzoideae subfamily it is expected for wheat, barley, and rye to have more common miRNA families among them than with rice. In 2A-QTL, 3BL-QTL, and 4A-QTL regions, eight miRNA families, and in 1A-QTL region, ten miRNA families were found to be present in only wheat genotypes (Figure 4.2, Figure 4.4, and Figure 4.7). Some of those miRNAs are wheat miRNAs which may be emerged after the divergence of wheat from barley and rye.

Only a small fraction of identified miRNAs were found to have a transcript target mapped on the homologous regions of resistance QTLs. Transcript targets of four miRNA families, miR1118, miR1439, miR2118, and miR2275, in some genotypes, have been predicted to be disease-resistance gene-like proteins, even though most of these transcripts only matched with uncharacterized proteins (Appendix B: Table S7). Plant miRNAs involved in plants' intracellular resistance mechanisms and miRNA-mediated expression of the insect-related genes are crucial for plant defense against a wide range of pathogens. Identifying such miRNAs will lead to the development of miRNA-based plant resistance mechanisms using molecular techniques and insect-resistant breeding (Yang, Zhang, Yang, Schmid, & Wang, 2021). A more comprehensive analysis of the transcript targets, including those out of QTL regions, may help unravel the function of identified microRNAs in the biotic stress conditions (Brant & Budak, 2018).

4.4.2. Conclusion

In this chapter, we examined the organization of coding and non-coding features of loci associated with WSS and OWBM resistance in an evolutionary comparative way among cereals. As it is the only known tolerance mechanism for WSS, the causal stem solidness gene, *Tdof*, holds great importance. We have compared the homologous regions of the QTL associated with *Tdof* and show a high level of conservation among all species, including the overall order of the genes in the genomic loci. However, evolutionary changes were shown to be more frequent on the distal ends of the homologous regions. Comparative analysis of *Sm1* locus revealed major rearrangements in rye chromosome 7R which resulted in the loss of homology in this region. When compared with respect to Chinese Spring, a homologous sequence to the antibiosis gene, Sm1, is only identified in

Zavitan genotype, and not in the Svevo cultivar, both have the BBAA genome. We have also described a potential genomic locus in the oat genome which may be homologous to Sm1. Homologous region analysis for 1A-QTL, the locus associated with oviposition deterrence of OWBM eggs, presented a potential inversion on the Svevo 1A chromosome even though overall extensive conservation among cereals was shown. Interestingly, the conservation of 4A-QTL revealed possible complex rearrangements for *Triticeae*. While we presented some candidate genes at homologous 4A-QTL, further experimental analysis is needed.

We observed an overall high synteny in homologous insect tolerance loci among cereals in terms of coding features. In addition to the coding features, the non-coding regions of the homologous regions were also found to support the evolutionary relationship that the comparative analysis presented. We have identified evolutionary conserved bona fide plant miRNAs such as miR1122, miR1127, miR1130, and miR5049 in wheat, barley, and rye species showing the functional importance of those miRNAs. We have identified species-specific miRNAs, indicating the emergence of the miRNAs after the divergence of this species from their last common ancestor. Moreover, we have identified miRNAs targeting transcripts in these identified homologous loci. While some of those miRNAs, miR2118, miR118, and miR1439 are shown to target important disease resistance genes, the functional annotation of the other miRNA targets is yet to be characterized. Identification of proteins encoded by the genes within the insect tolerance QTLs may provide potential candidates for this phenotype. Putative functional annotation of insect tolerance loci with proteomics approach using three protein annotation data has identified and annotated many proteins falling in these loci. For 4A-QTL, we have identified the conserved proteins encoded by the genes in these loci and identified flavonoid biosynthesis function. As oviposition deterrence may exhibit oviposition deterrence activity against female midges, these results will provide potential candidates for this resistance mechanism. As future work, novel miRNAs identified to be species-specific need to be validated. Expression analysis of miRNAs found to target transcripts from these loci will reveal their effect on gene regulation and insect-tolerance phenotype. Moreover, future characterization of the candidate proteins identified in these loci will provide insight into insect tolerance. The results of this study have already shed light on our future cloning, editing, and engineering of some of the genes, proteins, and metabolites in wheat, barley, and oat.

5. BARLEY NONCODING RNAs: GENOME AND TRANSCRIPTOME DERIVED ANNOTATION

5.1. INTRODUCTION

Drought is the most devastating abiotic stress faced by agricultural lands across the globe. It has been shown to cause major losses in crop yields every year, and with the changing climate, the incidence of drought conditions is only expected to increase (Seleiman et al., 2021). One method that is commonly implemented to mitigate the effects of drought stress is irrigation. However, this has also been shown to induce negative impacts by enhancing soil salinity, limiting crop productivity through ion toxicity and cellular dehydration (Wichelns & Qadir, 2014). To quantify this, every year approximately 1.5 million hectares of irrigated land becomes unsuitable for agriculture due to high soil salinization, which amounts to one-fifth of cultivated land globally suffering from high levels of salinity (Carillo et al., 2011; Munns & Tester, 2008). This places significant strain on agricultural yield, as each year the area available for crop cultivation decreases (Shrivastava & Kumar, 2015).

Barley (*Hordeum vulgare* L.) is one of the world's earliest domesticated cereal crop species (Pourkheirandish & Komatsuda, 2007). Its high fiber, vitamin, and mineral content make it an ideal staple food source, leading to its ranking as the fourth most abundant cereal crop in the world (FAOSTAT 2018, http://www.fao.org/faostat). Barley has a higher salt tolerance than most small grain crops, allowing it to be grown in a wider range of areas, including those suffering from soil salinization (Jamshidi & Javanmard, 2018; Kosová et al., 2015). Furthermore, with booming populational trends being observed across the globe, food demand is continually increasing, and agricultural

research has therefore become focused on increasing crop productivity (Tester & Langridge, 2010). In particular, a focus has been placed on producing crop varieties capable of withstanding common biotic stresses, including drought and high soil salinity (Cattivelli et al., 2008; H. Zhang, Li, & Zhu, 2018).

However, classical crop breeding is notoriously limited by the inherent genetic variation of the target species, which in crops has largely been lost due to domestication. Breeding trials are also highly labor and time-intensive, with many trials taking upwards of 6 years to produce one variety, with no guarantee that the desired trait will be obtained (Gepts, 2002; Ulukan, 2009; Wanga, Shimelis, Mashilo, & Laing, 2021). Following advances in sequencing technologies and *in silico* tools, the process of crop breeding has, however, been greatly accelerated. This is due to high-throughput sequencing providing a genomic base on which to work upon, enabling a deeper insight into the molecular background of a plant (Moose & Mumm, 2008; Ray & Satya, 2014). High-throughput sequencing has streamlined the process of crop improvement by facilitating elucidation of the molecular mechanisms responsible for enhanced stress tolerance, such as soil salinity in barley, which increases understanding of specific stress-responsive genes in plants for targeted breeding and genetic modifications (Budak, Hussain, Khan, Ozturk, & Ullah, 2015; Perez-de-Castro et al., 2012).

Following this, advances in sequencing technology have also paved the way for the identification of non-coding RNAs (ncRNAs). Previously referred to as “noise” in genomic samples due to limited information, ncRNAs, transcribed from noncoding DNA and not translated into proteins, have recently been discovered to contribute to the regulation of several important processes, including chromosome dynamics, splicing, RNA editing, translational inhibition, and messenger RNA (mRNA) destruction (Brant & Budak, 2018; Tyagi et al., 2018). One important class of ncRNAs, known as microRNAs (miRNAs), consists of 18-24 nucleotide (nt) long, endogenous, small RNAs (sRNAs) that are key regulators of gene expression at the post-transcriptional level. They modulate a plethora of essential processes in plants, including growth, development, and stress responses, via transcriptional inhibition and/or silencing of mRNA (Budak & Akpinar, 2015; Budak, Kantar, et al., 2015).

Since their discovery, the number of reported miRNAs has continued to increase gradually, and the specific roles of many plant miRNAs have now been unraveled. However, most research has been conducted in model organisms, such as *Arabidopsis thaliana* L., although many miRNAs have also been characterized in a variety of other species (Bala Ani Akpinar et al., 2015; Budak, Kantar, et al., 2015; Guleria et al., 2011; Muslu, Biyiklioglu-Kaya, Akpinar, Yuce, & Budak, 2021; Song, Li, Cao, & Qi, 2019). Due to their wide-ranging influence, and regulatory process, miRNA manipulation represents a promising opportunity for crop improvement, as they could enable efficient development of desired characteristics through the gene regulation (Djami-tchatchou et al., 2017).

To add to this, most RNA studies have been solely focused on sRNAs. However, there is an increasing interest in longer transcript regions also present within the genome, known as long non-coding RNAs (lncRNAs) (Budak et al., 2020; Mercer et al., 2009; Ponting et al., 2009). This group consists of ncRNA transcripts over 200nt long, that have been observed to work alongside sRNAs at various levels to aid in the gene expression regulation (Fatica & Bozzoni, 2014; Liu et al., 2015). Although more novel in their discovery, the ability of lncRNAs to fold into complex secondary, and tertiary structures label them as more versatile than sRNAs, offering another potential route for increasing the capability of crop improvement (Xin et al., 2011). However, for either miRNAs or lncRNAs to be utilized for crop improvement, knowledge needs to be gathered on their presence and function within the crop of interest.

In this chapter, publicly available genome and transcriptome assemblies of *H. vulgare* were used for genome and transcriptome-guided identification of miRNA families and long non-coding regions in three *H. vulgare* varieties: Barke, Bowman, and Morex. In parallel, miRNAs and long non-coding RNAs of *Hordeum spontaneum* under salt stress conditions were identified using publicly available transcriptome data, to reveal the potential network between lncRNA, miRNA, and mRNAs. We have shown the miRNA conservation among three cultivars as well as identifying lineage-specific miRNAs. The expression of identified miRNA families in small RNA and pre-miRNA levels and their repetitive element contents were also shown. Moreover, by identifying miRNAs using *Hordeum spontaneum* transcriptome with salt treatment samples at 0 hr, 2 hr, 12 hr, and 24 hr, we have identified candidate salt responsive miRNAs of barley. Many lncRNA transcripts are identified some of which were shown to be interacting with miRNAs. As

the interaction between lncRNA and miRNA is yet to be fully discovered, this study also investigates the lncRNA-miRNA-mRNA interaction triangle, to reveal the function of lncRNAs as endogenous Target Mimics (Borah et al., 2018), contributing to new insights into plant ncRNA studies.

5.2. MATERIALS & METHODS

5.2.1. Whole Genome and Transcriptome Datasets

De novo genome assemblies of Morex, Barke, and Bowman cultivars, were obtained from Munich Information Center for Protein Sequence (MIPS) (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/sequences/) (Mayer et al., 2012). Physical map-minimum tiling path-based reference sequence (RefSeq v.1.0) of Morex cultivar is used in miRNA identification and the mRNA target analysis of identified miRNAs. Barley (H*ordeum vulgare*) whole-genome shotgun sequences are obtained from Plant Genome and System Biology website (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/) for 3 different cultivars. All analyses were conducted separately for each cultivar.

Transcriptome data of the Morex cultivar were retrieved from European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena), under the accession number SRR492921 and SRR492923. Samples including inoculation applied barley were excluded, and only two read datasets of normal conditions carried out were used for transcriptome-guided miRNA identification.

Another transcriptome dataset of *Hordeum spontaneum*, the ancestor of cultivated barley, was obtained from European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena, accession number SRP032854). The transcriptome data of this study, performed on young leaves treated with salt at four different time intervals, including 0 hr, 2 hr, 12 hr, and 24 hr, was used for *in silico* miRNA and lncRNA identification.

As a reference miRNA dataset, a list of all published *Viridiplantae* mature miRNA sequences of high confidence and experimentally verified sequences from 72 species, was collected from the miRBase database (www.mirbase.org, release v21) (Kozomara & Griffiths-Jones, 2010) as described previously by Alptekin and Budak (Alptekin & Budak, 2016) to be used in homology-based miRNA identification.

5.2.2. Homology-based *in silico* miRNA identification

A two-step procedure, previously described by Lucas and Budak (Lucas & Budak, 2012), was followed to predict the putative miRNAs from the genomic and transcriptomic sequences. For transcriptome-guided miRNA identification, Trinity 2.2.0 assembly program was used for quality trimming and adaptor removal of reads of the *Hordeum vulgare* cv. Morex and *Hordeum spontaneum* transcriptome dataset, and further *in-silico* miRNA identification analyses were carried out as in whole genome-guided miRNA identification.

5.2.3. Putative miRNA Target Prediction and Functional Annotation

Putative mRNA targets of miRNAs identified from Morex, Barke, and Bowman cultivars were determined using psRNATarget tool (Dai & Zhao, 2011) with default parameters against barley high confidence genes obtained from IBSC assembly. Target proteins were identified using BLASTx search (e-value 10^{-6}) against all non-redundant *Viridiplantae* proteins obtained from NCBI. Functional annotation of putative target proteins was performed using Blast2GO software (www.blast2go.com) (Götz et al., 2008).

5.2.4. Pre-miRNA Expression Analysis

Expression analysis of putative pre-miRNAs of all three cultivars was performed using in BLASTn search against two databases: (1) diverse transcriptome of Morex cultivar by Illumina Genome Analyzer II paired-end sequencing obtained from European Nucleotide Archive (ENA) under study supported by the Agriculture and Food Research Initiative of USDA's National Institute of Food and Agriculture with accession number SRP012567 and (2) *Hordeum vulgare* Expressed Sequence Tags (EST) data obtained from NCBI. Cutoff parameter of 95% identity and query coverage was applied to combined results.

5.2.5. Small RNA Expression Analysis

Small RNA sequences of *H. vulgare* through Illumina platform under study accession number PRJNA16859358 were downloaded from European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena). RNA-seq reads were trimmed using the Cutadapt tool (version 1.9.1) with a threshold of 15 to eliminate adaptor sequences (Martin, 2011). A blast database was constructed from combined trimmed small RNAseq reads and mature miRNA queries of all three cultivars were formed separately. BLASTn (version 2.2.31) was used to align mature miRNA query to small RNAseq database for verification of small RNA expression of putative miRNAs as *in silico* expressed miRNAs. A summary flow of work chart of miRNA identification, target and expression analysis of the putative miRNAs is shown in Figure 5.1.

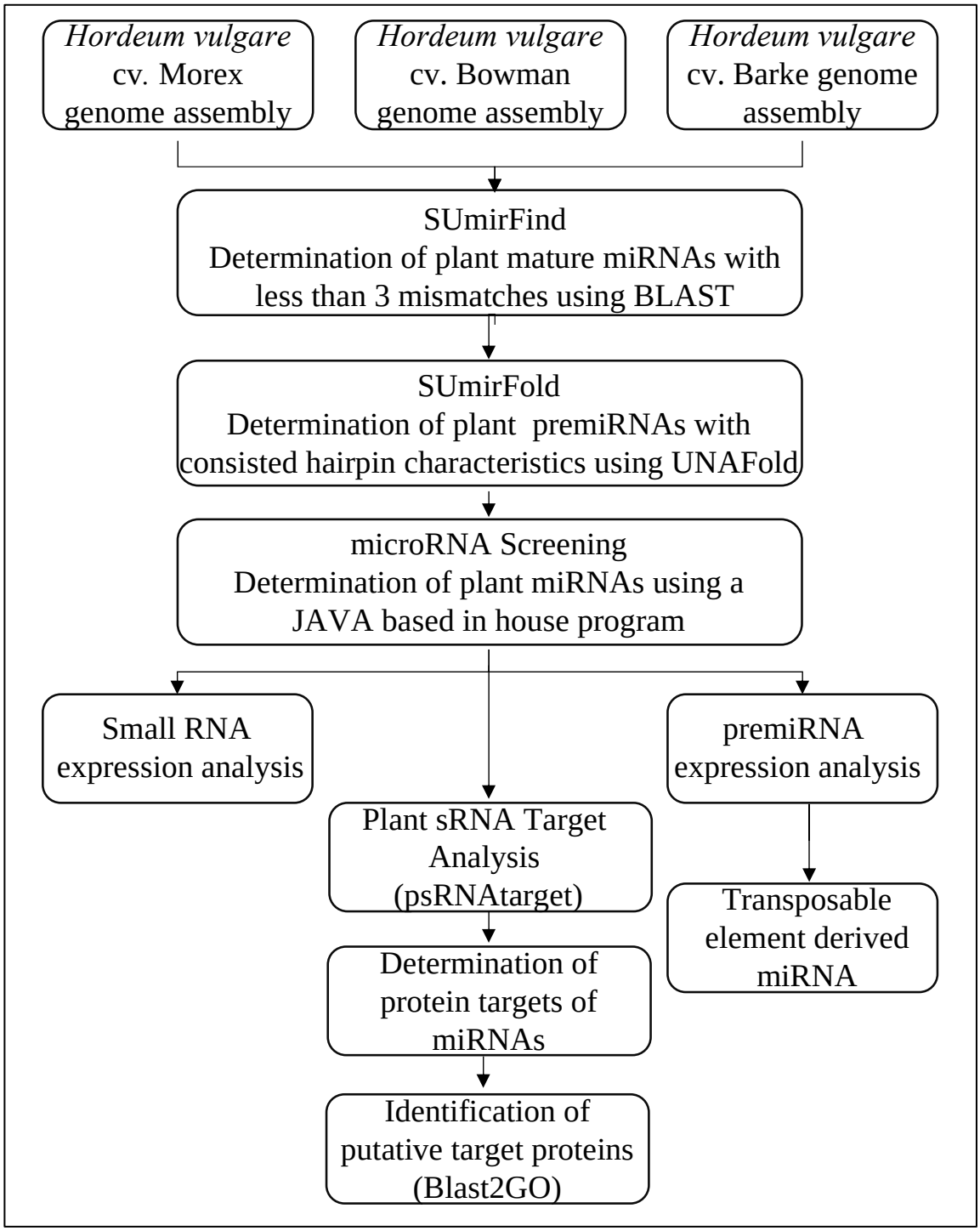


Figure 5.1 Workflow of *in silico* miRNA identification of Barley genome assembly of Morex, Barke and Bowman cultivars. Putative miRNA identification, expression, and target analysis of identified miRNAs.

5.2.6. Putative tRNA Gene Analysis

Transfer RNAs (tRNAs) are a family of non-coding RNAs in all three domains of life and identification of plant tRNAs gained interest in recent years (Zahra, Singh, Poddar, & Kumar, 2021). tRNAscan-SE program version 1.3.1 (Lowe & Chan, 2016) with default parameters for eukaryotes was used for the determination of tRNA genes from unmasked genome assemblies of *Hordeum vulgare* Morex, Barke and Bowman cultivars.

5.2.7. Identification of Transcriptome Guided Long Non-coding RNAs

Trinity assembly applied two sets of transcriptome assemblies named SRR492921 and SRR492923 obtained from European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena) which was previously mentioned in transcriptome guided miRNA analysis section. Transdecoder script (version 3.0.1) is used for detection of coding regions based on open reading frame (ORF) sizes (Bryant et al., 2017). All *Hordeum vulgare* proteins sequences identified up to date were downloaded from UniProt (www.uniprot.org) database and the outputs of transdecoder analysis were screened against these protein sequences using Blast algorithm. Any match with homology to known H*ordeum vulgare* proteins was eliminated. Additionally, all known protein sequences of close relatives of *H. vulgare* including *T. aestivum*, *T. durum*, *T. monoccocum*, *T. urartu*, *Aegilops speltoides*, *Aegilops sharonensis* and *Aegilops tauschii* were downloaded from URGI database (https://wheat-urgi.versailles.inra.fr) and output of trinity assembly of Morex cultivar transcriptome was screened against these sequences using Blastx algorithm (version 2.2.31). Again, any transcript giving a hit to protein sequences was eliminated. To distinguish coding and non-coding transcripts of Morex transcriptome, Coding-Non-Coding Index (CNCI) tool (version 2014, https://github.com/www-bioinfo-org/CNCI) which profiles adjoining nucleotide triplets was used (Sun et al., 2013). Furthermore, Coding Potential Calculator (CPC) tool (http://cpc.cbi.pku.edu.cn/) was used to evaluate the protein-coding potential of the Morex transcript (Kong et al., 2007). For the determination of homologous proteins which are evolutionarily conserved and annotation of uncharacterized sequences, a hidden

Markov model-based statistical model of Pfam (version 30.0) database was used (https://www.ebi.ac.uk/services/teams/pfam) (Figure 5.2).

Another transcriptome assembly, accession number SRP032854, was used to identify long non-coding RNAs related to salt stress. Same analyses, carried out for Morex transcriptome, were conducted separately for all datasets.

LncTar, a tool for predicting lncRNA-mRNA interaction, was used to define the interaction between lncRNAs and mRNAs (J. Li et al., 2014). It is crucial to explore RNA targets to understand the functions and action mechanisms of lncRNAs.

5.3. RESULTS

5.3.1. Genome and Transcriptome-derived Identification of miRNAs

Genome-derived miRNA identification of three barley varieties resulted in 1177, 1263 and 1183 potential mature miRNA-premiRNA pairs for Barke, Bowman and Morex, respectively. Thirty-eight miRNA families were observed to be common in all three cultivars. miR397, miR528 and miR9662 families were identified only in both Morex and Bowman cultivars, whereas miR5384 and miR2120 families were observed only in Morex and Barke cultivars, respectively. Of the total 43 miRNA families identified from all three cultivars, 10 miRNA families were present in miRbase_v21 as *Hordeum vulgare* miRNAs (www.mirbase.org). miRNA identification results from Barke, Bowman and Morex genomes showing premiRNA and mature miRNA sequences are given in Appendix C: Supplementary TableS1, S2, and S3, respectively. Besides, eleven miRNA families were observed in Morex transcriptome, and a total of 12 miRNA families were found in *Hordeum spontaneum* transcriptome 0hr of salt treatment Appendix C: Table

S4-S14). 8 miRNA families identified in Morex transcriptome were also identified in genome-guided miRNA identification (Table 5.1).

Table 5.1 Putative miRNA families identified from genome-derived and transcriptome-derived *in-silico* miRNA identification analysis. For *Hordeum Spontaneum*, miRNA families identified from 0hr salt treatment is given.

GENOME DERIVED						**TRANSCRIPTOME DERIVED**	
BARKE		**BOWMAN**		**MOREX**		**MOREX**	***Hordeum Spontaneum 0hr***
miR156	miR1121	miR156	miR1118	miR156	miR1118	miR1122	miR166
miR159	miR1122	miR159	miR1120	miR159	miR1120	miR1128	miR169
miR160	miR1125	miR160	miR1121	miR160	miR1121	miR1133	miR444
miR164	miR1127	miR164	miR1122	miR164	miR1122	miR1135	miR1120
miR166	miR1128	miR166	miR1125	miR166	miR1125	miR1137	miR1121
miR167	miR1130	miR167	miR1127	miR167	miR1127	miR1436	miR1122
miR169	miR1131	miR169	miR1128	miR169	miR1128	miR1439	miR1127
miR171	miR1133	miR171	miR1130	miR171	miR1130	miR319	miR1128
miR172	miR1135	miR172	miR1131	miR172	miR1131	miR437	miR1135
miR393	miR1136	miR393	miR1133	miR393	miR1133	miR444	miR1137
miR394	miR1137	miR394	miR1135	miR394	miR1135	miR5048	miR1436
miR395	miR1139	miR395	miR1136	miR395	miR1136		miR1439
miR396	miR1171	miR396	miR1137	miR396	miR1137		miR9662
miR398	miR1436	miR397	miR1139	miR397	miR1139		
miR399	miR1439	miR398	miR1171	miR398	miR1171		
miR414	miR2118	miR399	miR1436	miR399	miR1436		
miR437	miR2120	miR414	miR1439	miR414	miR1439		
miR818	miR2275	miR437	miR2118	miR437	miR2118		
miR1117		miR528	miR2275	miR528	miR2275		
miR1118		miR818	miR9662	miR818	miR5384		
miR1120		miR1117		miR1117	miR9662		

We have identified a total of 25 miRNA families from *Hordeum spontaneum* transcriptome guided miRNA identification as 13 miRNA families for 0 hr of salt treatment, 17 miRNA families for 2 hr of salt treatment, 22 miRNA families for 12 hr salt treatment and 20 miRNA families for 24 hr of salt treatment transcriptome data. 10 of those miRNA families were identified in all hours of salt treatment whereas some miRNA families such as miR818 and miR1133 were identified in 12 hr and 24 hr samples of salt treatment. All identified miRNA families for each dataset is shown in Table 5.2.

Table 5.2 miRNA families identified from 0hr, 2hr, 12hr and 24 hr salt treatment applied Hordeum spontaneum transcriptome data. 10 miRNA families identified to be present in all datasets.

0 hr	2 hr	12 hr	24 hr
-	miR1117	-	-
-	-	miR1118	-
miR1120	miR1120	miR1120	miR1120
miR1121	miR1121	miR1121	miR1121
miR1122	-	miR1122	miR1122
miR1127	-	miR1127	miR1127
miR1128	miR1128	miR1128	miR1128
-	miR1130	miR1130	miR1130
-	-	miR1133	miR1133
miR1135	miR1135	miR1135	miR1135
miR1137	miR1137	miR1137	miR1137
miR1436	miR1436	miR1436	miR1436
miR1439	miR1439	miR1439	miR1439
-	miR156	miR156	miR156
-	miR157	miR157	miR157
-	-		miR160
miR166	-	miR166	miR166
	miR167	miR167	-
miR169	miR169	miR169	miR169
-	miR393	miR393	-
-	miR397	-	miR397
miR444	miR444	miR444	miR444
-	-	miR818	miR818
-	-	miR854	-
miR9662	miR9662	miR9662	miR9662

The criteria to distinguish miRNAs from other small RNAs such as length of mature miRNA sequences were GC content (24-71%) and Minimum Folding Energy Index (MFEI) of hairpin structures. The average length of genome-guided mature miRNAs was found to be 20nt, as in 1474 of all 2687 mature miRNAs. The identified miRNAs for barley had a GC content of 45% which was in agreement with previous results (Dubcovsky et al., 2001; Kurtoglu et al., 2014). Other types of RNAs such as tRNAs (0.64), rRNAs (0.59), or mRNAs (0.62-0.66) presented lower MFEI values than the miRNAs. Average MFEI of the predicted barley miRNAs were calculated as 1.08 ± 0.25.

Extensive conservation of those miRNAs among close relatives of *Hordeum vulgare* (*Triticum aestivum*, *Brachypodium distachyon*, *Oryza sativa*, *Zea mays and Sorghum bicolor*) indicated the importance of the miRNAs that were constrained during evolution. Homology-based analysis of the identified barley miRNAs showed these miRNAs were conserved in at least one close relative of barley.

5.3.2. Target Analysis of Predicted miRNAs and Functional Annotation

psRNATarget analysis identified many putative mRNA targets for Barke, Bowman and Morex genomes (Appendix C: Supplementary Table S15, S16, S17). It is known that miRNAs regulate crucial cellular mechanisms mostly by complementarily binding to their targets for cleavage or translational inhibition (Brant & Budak, 2018). In order to elucidate the function of a miRNA within a cell, tissue or plant, it is important to identify and annotate the functions of its target(s). Nucleobase-containing compound metabolic, biosynthetic, cellular protein modification processes were found to be the major biological processes targeted by the miRNAs identified in all three cultivars (Figure 5.2).

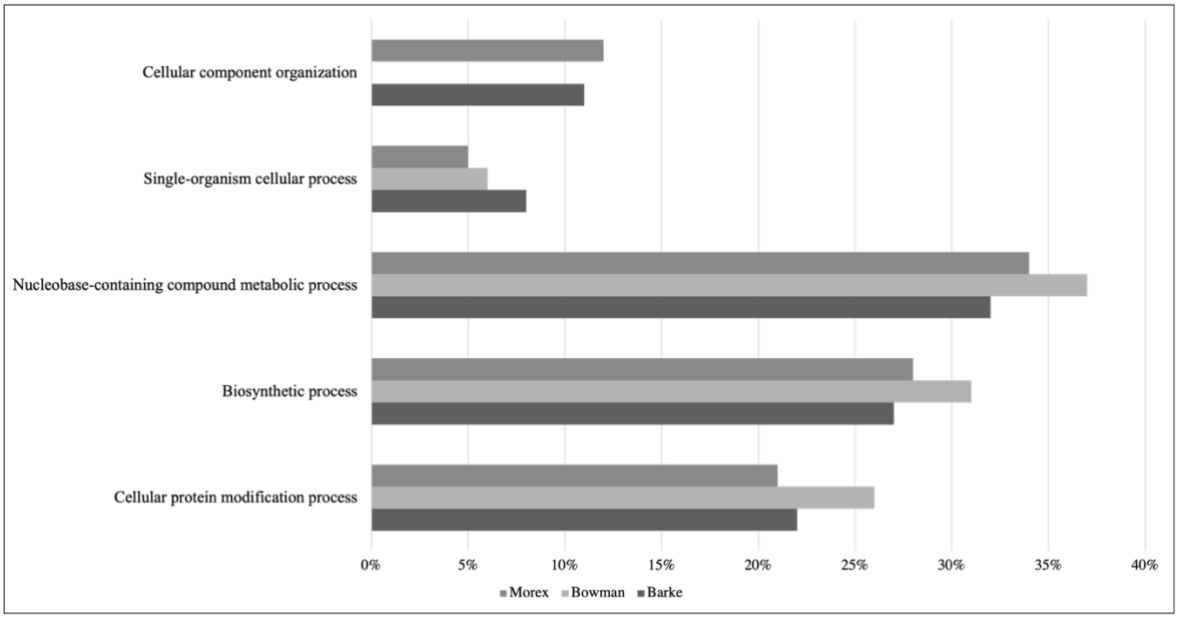


Figure 5.2 GO annotation results of biological processes of putative miRNA targets of Morex, Bowman and Barke cultivars.

Cell membrane was found as the main cellular component of miRNA targetst, in addition to nucleus and cytoplasm, as the major compartments of the cell (Figure 5.4), whereas

nucleotide binding was found to be the major molecular function assigned to the putative miRNA targets in all cultivars, followed by hydrolase activity, DNA-binding activity and kinase activity (Figure 5.5).

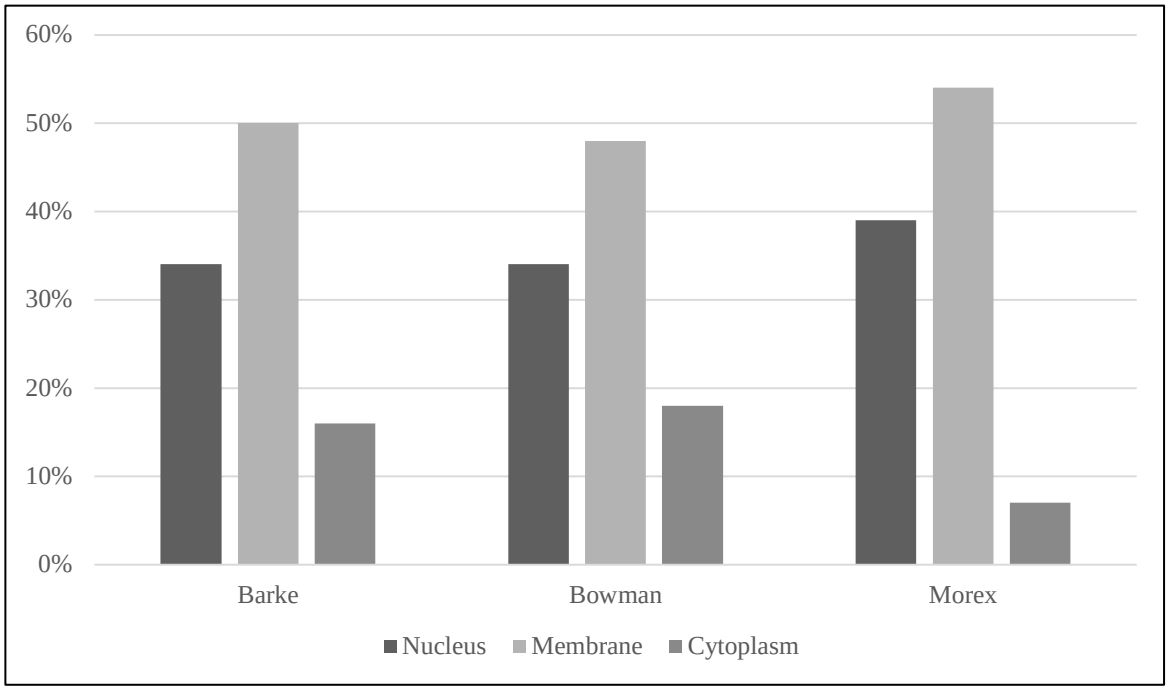


Figure 5.3 GO annotations of putative miRNA targets of Barke, Bowman and Morex for cellular components.

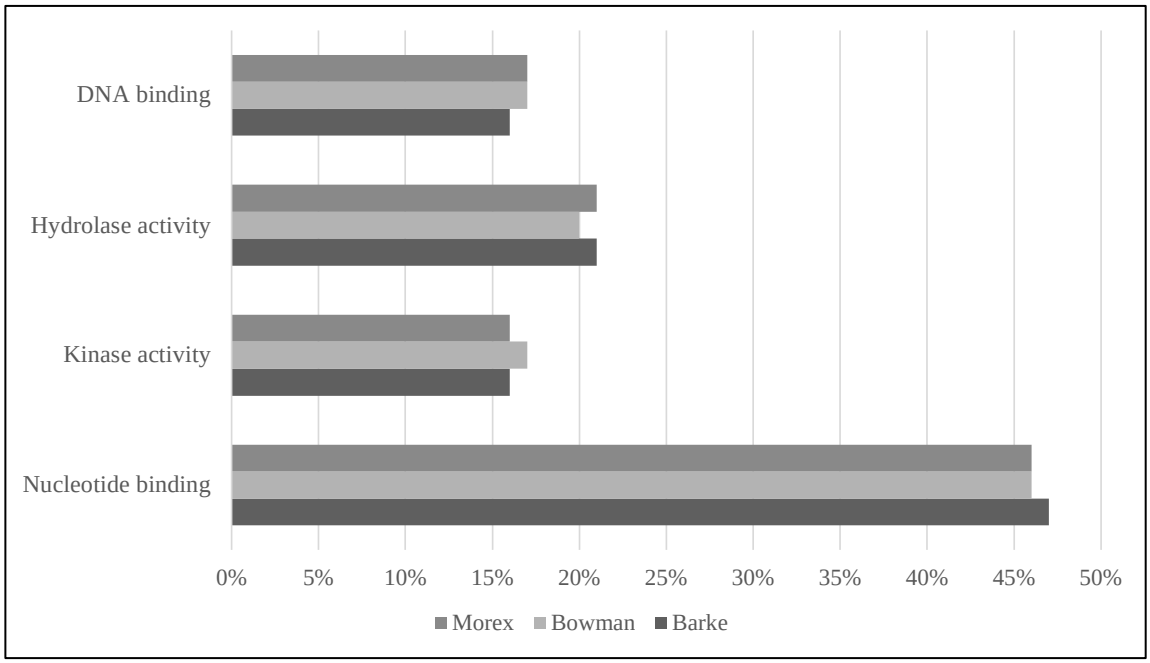


Figure 5.4 GO annotations of putative miRNA targets of Morex, Bowman and Barke cultivars for molecular functions.

The functional annotation analysis for the target proteins of the largest miRNA family (miR1436 which has the most putative mature miRNA sequences) suggested a relation

with plant's phosphatidylserine decarboxylase activity involving in phospholipid biosynthetic processes. Another target protein of the same miRNA family plays a role in cell wall biogenesis which is important as a source of dietary fiber and malting quality of barley. miR5384 was found to be a Morex cultivar-specific miRNA targeting four different proteins. One of its targets involved in fatty acid biosynthesis and another one involved in zinc-ion binding (http://www.uniprot.org/).

5.3.3. Determination of Repetitive Elements

It was found that 86%, 87% and 88% of Morex, Bowman and Barke cultivar miRNAs, respectively, were transposable element (TE)-related, as demonstrated by Mayer et al previously. The major TE superfamilies for Bowman miRNAs were MITEs and En-Spm, whereas En-Spm was the most predominant TE superfamily for Morex and Barke, (Figure 5.6).

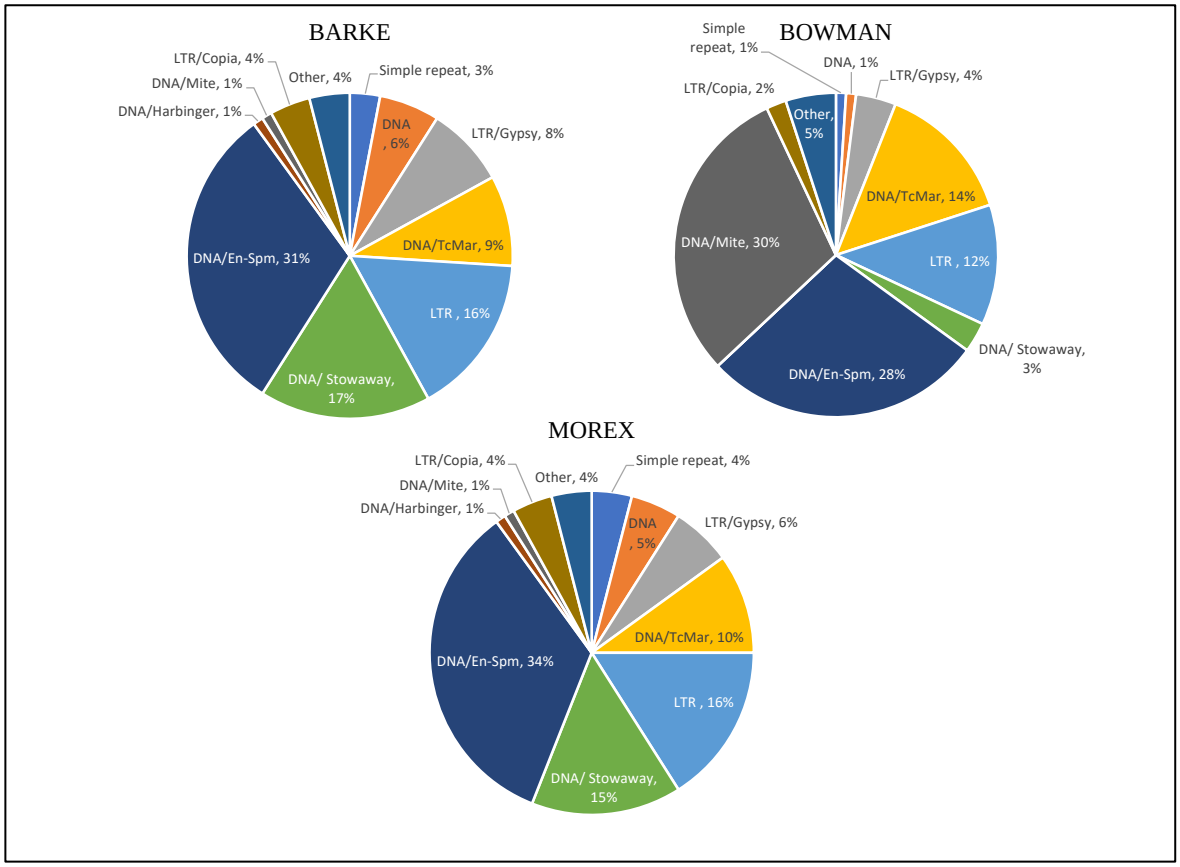


Figure 5.5 Repeat class distribution of transposable elements of Barke, Bowman and Morex cultivars.

5.3.4. Expression Analysis

Twenty-one identified miRNA families were expressed in all three cultivars at small RNA level under normal conditions, according to the small RNA expression analysis. miR1133 was found to be expressed only in Barke and Bowman cultivars whereas miR397, miR398, miR528 and miR9662 were not expressed in Barke cultivar at small RNA level (Appendix C: Supplementary Table S18, S19, S20). For pre-miRNA expression, eleven miRNA families were expressed in all three cultivars. miR2118 was expressed only in Barke whereas miR1122 was specifically expressed in Bowman. Two miRNA families, miR397 and miR1137, were expressed in Bowman and Morex cultivars at precursor miRNA level (Appendix C: Supplementary Table S21, S22, S23). For salt treatment, 18, 23, 21 and 22 miRNA families were identified to be expressed in premiRNA level in 0 hr, 2hr, 12hr and 24 hr samples, respectively. Table 5.3 shows the premiRNA expression of miRNA families identified. 15 miRNA families were identified to be expressed at all hours of salt treatment at premiRNA level.

Table 5.3 Salt treatment applied transcriptome miRNA families which were identified to be expressed at premiRNA level. 15 miRNA families were identified to be expressed at premiRNA level at all times of salt treatment.

0 hr	2 hr	12 hr	24 hr
-	miR1117	miR1117	miR1117
miR1118	miR1118	miR1118	miR1118
miR1120	miR1120	miR1120	miR1120
miR1121	-	miR1121	miR1121
miR1122	miR1122	miR1122	miR1122
miR1128	miR1128	miR1128	miR1128
miR1130	miR1130	miR1130	miR1130
miR1133	miR1133	miR1133	miR1133
miR1135	miR1135	miR1135	miR1135
miR1136	miR1136	miR1136	miR1136
miR1137	miR1137	miR1137	miR1137
-	miR1139	-	-
miR1436	miR1436	miR1436	miR1436
miR1439	miR1439	miR1439	miR1439
-	miR156	miR156	miR156
-			miR160
miR166	miR166	miR166	miR166
-	miR167	miR167	
miR169	miR169	miR169	miR169
-	miR2275		
-	miR393	miR393	
-	miR397		miR397
miR818	miR818	miR818	miR818
miR9662	miR9662	miR9662	miR9662
miR1117	-	-	-
-	miR1121	-	-
-	-	-	miR1139
-	-	-	miR1125
miR160	-	-	-
-	-	miR5384	-

5.3.5. tRNA Genes of Barley Genome

Genome assemblies of three barley cultivars were used to determine the distribution of tRNA genes. For Barke and Morex cultivars, tRNAMet was found to be extremely abundant followed by tRNAArg while for Bowman cultivar tRNASer followed by tRNAGly was the most abundant amino acid. Long terminal repeat (LTR) sequences identified through TE analysis showed that 73%, 92% and 65% of all tRNA genes were located on LTRs. tRNA genes determined from Barke, Bowman and Morex cultivars are shown (Appendix C: Supplementary Table S24, S25, S26).

5.3.6. High Confidence mRNA and lncRNAs of Barley Genome and LncRNA-mRNA Interaction

Potential coding sequences of the transcriptome data were eliminated by several analyses. Seventy-nine transcripts out of 27,979 were identified as high confidence lncRNAs in Morex transcriptome (Appendix C: Supplementary Table S27). According to the lncTar analysis, 79 lncRNA transcripts targeted 18580 mRNA transcripts in Morex transcriptome. miRNA-based target analysis showed eight miRNA families targeting only mRNA transcripts, whereas only miR1128 family was identified to be targeting both lncRNA and mRNA transcripts (Appendix C: Supplementary Table S28). The lncRNA-miRNA-mRNA interaction network for Morex miR1128 is shown in Appendix C: Supplementary Figure 1. Numerous lncRNA transcripts were identified from 0 hr, 2 hr, 12 hr and 24 hr salt treatment samples of *H. Spontaneum* transcriptome (Appendix C: Supplementary Table S29-S32). From the analysis of *H. spontaneum* transcriptome, 37 miRNAs were found to be interacting at least once with both lncRNA and mRNA transcripts whereas 18 of those miRNAs were found to be non-interacting. LncRNA-miRNA-mRNA regulatory network created for all identified miRNAs in salt treated samples for *H. spontaneum* are shown in Appendix C: Supplementary Figure 2, 3, 4, 5 and 6. Predicted functions of putative targets of miRNA families targeting both lncRNAs and mRNAs in salt treated transcriptomes, except for the 0hr sample, were presented in Appendix C: Supplementary Table S33.

5.4. DISCUSSION AND CONCLUSION

5.4.1. Discussion

Plants are among the most affected organisms by the outcomes of global warming and changing climate patterns. Because they hold the bottom of the food chain, any alteration in plant biomass will directly or indirectly affect the living organisms. Abiotic stress is the most crucial outcome of climate change that limits crop productivity worldwide. Plants try to adapt themselves in response to changing environmental conditions by altering the expression profiles of some genes, for example, elevating secretion levels of abscisic acid hormone (ABA) to cope with environmental stress. miR167 and miR393, two miRNA families we identified in all cultivars, targeting ABA receptors of mRNA were shown in previous studies as the important regulators of such plant adaptation processes (H. Chen, Li, & Xiong, 2012; Ferdous, Hussain, & Shi, 2015; Shinozaki & Yamaguchi-Shinozaki, 2000). Therefore, modification of plant biological pathways could be a feasible way to overcome biotic/abiotic stresses, such as drought stress, salinity, pathogens, nutrient deficiency, and eventually increase crop yield, environmental adaptation, and plant endurance in agriculture. Identification of non-coding genome of plant species including miRNAs and lncRNAs, which act as an on-and-off switch in many biological processes and developmental pathways, holds a great promise on securing agricultural crop production for future by enabling altered gene expression.

In this study, *in silico* miRNA identification and characterization of *H. vulgare*, using barley cultivars Morex, Barke and Morex data, and target analysis via psRNATarget tool were carried out. According to the findings, the largest group of miRNA families based on the number of potential mature miRNA sequences were identified as miR1117, miR1120, miR1122, miR1436 and miR1439 in all analyzed cultivars, Barke, Bowman and Morex. Congruently with these findings, miR1120 was previously reported to be one of the largest miRNA families in *T. aestivum*, a close relative of *H. vulgare* and precursor

structure of miR1120 shows great similarity to its wheat orthologue (Budak & Akpinar, 2015). Previously, it was presented that miR1120 functioned as an important regulator of seed and leaf development (Alptekin & Budak, 2016; Han et al., 2014) and its mature miRNA level varied during development, however miRNA/pri-miRNA levels stayed constant. These findings indicate the importance of this miRNA in developmental stages of barley (Kruszka et al., 2013) and explains the reason behind its conservation among a wide range of plant species. Among 43 identified putative miRNA families, 38 families were found to be common in all three cultivars whereas the other 5 families were found to be missing in one or two of the cultivars (Table 5.1). Among those, only miR397 were previously identified as barley miRNAs whereas miR528, miR9662, miR2120 and miR9662 were identified in other cereals, but to our knowledge, not in barley.

The origin of miRNAs and their mRNA targets are not fully established yet, however transposable elements (TEs) are one of the possible scenarios of the origin of miRNA genes, hypothesized by Smalheiser and Torvik in 2005 (Borchert et al., 2011; Smalheiser & Torvik, 2005). DNA transposable elements, the first identified class II transposons, are the major contributors of plant genome content. Particularly, miniature inverted-repeat transposable elements (MITEs) are strongly associated with miRNA evolution as they have the capability of forming hairpin structures (Alptekin & Budak, 2016; Alptekin, Langridge, & Budak, 2017; Budak & Akpinar, 2015; Piriyapongsa & Jordan, 2008). Around 80% of grass genomes are composed of TEs (Diao & Lisch, 2006) and the repetitive contents of putative barley pre-miRNAs identified in this study were in agreement with previous findings. Retrotransposons that contain long terminal repeats (LTRs) are the most present class of plant retrotransposons. Activity of transposable elements may vary under different biotic/abiotic stress conditions due to methylation processes which often results in altered gene activity (Bennetzen, 2000; Peng & Zhang, 2009). The role of transposable elements in stress adaptiveness was defined as increasing fitness and evolutionary adaptation of their hosts (Chadha & Sharma, 2014). Parallel to this information, we have identified 86% of Morex, 87% of Bowman and 88% of Barke miRNAs are transposable element related. Barke and Morex cultivar miRNAs show great similarity in terms of their TE contents, where in Bowman Miniature Inverted-repeat Transposable Elements (MITEs) seem to be enriched and in contrast, Stowaway, which is a type of MITEs, seems to be much scarcer in comparison to Barke and Morex cultivars (Figure 5.6).

Little is known about eukaryotic tRNA which is one of the most ancestral RNA. tRNAs are formed by a series of duplications, mutations, and re-organization events during evolution (Michaud, Cognat, Duchêne, & Maréchal-Drouard, 2011). It was hypothesized earlier that high copy number of tRNA genes may be a result of its presence in repetitive regions (Tanaka et al., 2014). We showed that more than 65% of all tRNA genes were present on LTRs in all three cultivars with the highest percentage of 92% in Bowman cultivar. A close relative of barley, *T. aestivum,* was previously defined as rich in tRNALys, however, a study on Haruna Nijo cultivar of barley showed tRNALeu as the most abundant tRNA gene (Sato et al., 2016; Tanaka et al., 2014). in this study tRNAMet and tRNASer were found to be large in number in barley.

Biotic/abiotic stress conditions alter the expression of related genes controlled by miRNAs. Salt stress is one of the most concerned abiotic stresses affecting hectares of agricultural lands every year. Understanding the expression pattern of salt-responsive miRNAs is vital to overcome the negative effects of salinity. Expression of miR1133 which involves in lncRNA-miRNA-mRNA network was previously shown to have an upregulated pattern under salt stress treatment (Lu et al., 2011). The target of miR1133, ubiquitin carboxyl-terminal hydrolase, was also found to be predominately expressed at especially stem tissues in various studies (L. Chen, Ren, Zhong, Jiang, & Li, 2010; S. Li et al., 2016). Another miR1133 target is pentatricopeptide repeat-containing protein. Arabidopsis plants with mutated pentatricopeptide repeat protein (PPR) displayed deficient resistance to abiotic stress including salinity as well as impaired ABA signaling that caused stress sensitivity in plants (Laluk, Abuqamar, & Mengiste, 2011). Jiang et al. (Jiang et al., 2015) showed that nucleo-cytoplasmic localized PRR acts as a positive regulator of high salinity (Sharma & Pandey, 2016). Barke specific miR2120 was found to be expressed in premiRNA levels but not in smallRNA levels under normal conditions. The homolog miR2120 in *T. aestivum* was shown to target cytochrome P450 which is a part of defense mechanism evolution of plants. Plant cytochrome P450 is also involves in biosynthesis of various plant hormones and secondary metabolites and xenobiotic pathways such as assimilation of carbon sources (Du et al., 2016). miR9662, identified in Morex and Bowman genomes, was found to be not expressed in premiRNA and sRNA levels under normal conditions. miR9662, previously only identified in wheat (Chu et al., 2016; T. Li et al., 2015) were identified in Bowman and Morex cultivars was found to target Mitochondrial transcription termination factor family (MTERF). MTERFs are

regulatory genes which mutations in those genes are shown to result in altered gene expression of mitochondrial and chloroplast genes. MTERF mutants were previously shown to have a diverse range of phenotypes such as resistant phenotype or lethality. Chan et al.(Chan, Crisp, Estavillo, & Pogson, 2010) proposed that under stress conditions chloroplast senses and transmits the message to the nucleus in order to cope with the stress. Thus, the expression of organelle genes holds great importance in the plant defense mechanism (Quesada, 2016), therefore the presence of miR9662 in Bowman and Morex cultivars needs to be verified in future studies. Morex cultivar-specific miR5384 had PHD finger proteins and 3-ketoacyl-CoA synthase genes as its mRNA targets. PHD finger proteins are also regulatory proteins that control plant development such as germination, flowering time as well as meiosis (Mouriz, López-González, Jarillo, & Piñeiro, 2015). Another target, 3-ketoacyl-CoA synthase, was shown to be the precursor in cuticular wax production of plants in a previous study (Weidenbach et al., 2014). Moreover, miR5384 was shown to be expressed in premiRNA level at 12 hr of salt treatment, but not in 0 hr, 2 hr or 24 hr, indicating this miRNA may be expressed under salt treatment.

The function of lncRNAs is yet to be discovered. It has been proposed that lncRNAs, like miRNAs, target mRNA to act as regulatory elements and to enhance stability of mRNAs, can be precursors of miRNAs and can mimic miRNA targets to control the miRNA function. Thus, comprehension of lncRNAs-mRNA interaction is vital to unfold many key regulatory mechanisms. In this study, lncRNA-mRNA interaction was investigated via lncTar computational tool. Target analysis of the identified miRNAs provided a basis to relate lncRNAs to miRNAs to see whether the lncRNAs were precursor of miRNAs. Many functional lncRNAs were previously shown to involve in gene regulation by the act of lncRNA transcription (Kornienko, Guenzl, Barlow, & Pauler, 2013). It was found that many identified miRNAs were complementary to lncRNAs and mRNAs, which may be speculated as those lncRNAs may suppress the regulatory function of miRNAs by binding to miRNAs and preventing the binding of miRNAs to the target mRNAs. In Morex transcriptome, miR1128 was found to have both a lncRNA target and mRNA targets. The expression pattern of lncRNAs is also condition-dependent like miRNAs and the possible indirect involvement of lncRNAs indicates the complexity of gene expression regulation process.

5.4.2. Conclusion

In this chapter we have identified 38 common miRNA families from Barke, Morex and Bowman genomes, 11 miRNA families from Morex transcriptome and 13 miRNA families from *H. spontaneum* transcriptome 0 hr of salt stress application. We have also identified miRNAs from transcriptome data of samples from different times of salt application. We have shown different expression profile of identified miRNAs under normal conditions and salt stress application which will provide an insight of salt stress related miRNAs of barley. We have identified some cultivar specific miRNAs which need to be validated in future studies. LncRNAs holds a potential, yet to be fully characterized, as regulatory transcripts of miRNA functions. To shed light on this, we determined lncRNAs from 2 different transcriptome data, one under normal conditions and one salt stress applied. A total of 83,089 lncRNA transcripts were identified which will be beneficial in future studies. Different hours of salt treatment greatly altered the number of lncRNAs identified. We have matched the identified miRNAs to their target lncRNAs and mRNA transcripts and for the miRNAs targeting both lncRNAs and mRNAs, we have created lncRNA-miRNA-mRNA interaction profiles. We have used only high confidence and experimentally verified plant miRNA sequences as the reference miRNA sequences in our miRNA identification. Therefore, our miRNA prediction can detect bona fide and novel miRNAs, therefore the validation and expression analysis of these miRNA families by qRT-PCR and verification of identified lncRNA via RNASeq studies in future studies hold great importance.

6. GENERAL CONCLUSION

Cereals preserve high synteny across their chromosomes despite high ploidy and genome size differences (Murat et al., 2017). This synteny provides the construction of reference assemblies and/or characterization of inferred gene orders in closely related complex genomes (Mayer et al., 2011). The use of the comparative genomics approach in cereals, therefore, provides a powerful tool for gene discovery and identification of evolutionary changes (Strable & Scanlon, 2009). Here in this thesis, we analyzed grain crops, wheat, barley, rye, rice, and oat, and model organism *Brachypodium* aiming to unravel the conservation and evolution of noncoding features in monocots. We have identified many putative novel miRNAs, lineage-specific miRNAs as well as showing high conservation of many miRNAs among the cereals. We have carried out a comprehensive miRNA identification and characterization analysis in an important monocot model organism, *Brachypodium distachyon*, by aiming for a better understanding of the effect of intraspecific variations in noncoding genomes on specific traits and for developing a broad miRNA conservation perspective. In this context, we have identified three lineage specific miRNAs which were not previously identified as *Brachypodium* miRNAs. One of those, miR5161 was shown to be present only in two lineages, Arn1 and Mon3 which were shown to have earlier flowering phenotype than other lineages. Earlier flowering phenotype is an important characteristic especially for seeds under stress environments. As temperature is the main controller of plant development, fluctuating temperatures due to climate change becomes a major challenge for critical stages of development. Selection of varieties with appropriate flowering times therefore is a good strategy for adaptation. Thus, validation and differential expression of miR5161 in future studies may be promising.

We have shown many known, highly conserved, and experimentally validated plant miRNAs to be common in all *Brachypodium* lineages. Some of these miRNAs, miR156, miR159, miR160, miR166, miR171, miR395 and miR396 are well conserved plant miRNAs, present in Magnoliophyta, Coniferophyta and Embryophyta and we identified all of those miRNAs in Barke, Bowman and Morex genomes as well.

As hypothesized, well-conserved miRNAs have been identified to target essential proteins. Nucleobase-containing compound metabolic processes are found to be the major target of miRNAs, followed by biosynthetic processes and cellular protein processes. Pangenome approach provided determination of response to stimulus and cell communication as biological processes as targets of rare miRNA families. Comparative analysis of three cultivars in chapter III didn't allow us to comprehensively analyze rare miRNA families indicating the importance of pangenome approach as the more individuals of a species are compared, the more genetic variation of that species can be determined. Moreover, we have shown 24 miRNA families identified in our analysis which were not present in *Brachypodium* reference genome. This shows how many miRNAs we may not have detected in other studies that were not inclusive, rather than the pangenome.

We have identified high repetitive sequence percentage in many premiRNA sequences, especially the ones with extremely high premiRNA counts. Evolution acts on premiRNA abundance of miRNAs as creating multiple copies of miRNA genes in the genome. Young, newly emerged miRNAs usually retain extensive target homology with low premiRNA counts whereas highly conserved plant miRNAs usually have high premiRNA counts (Fahlgren et al., 2007; Narjala et al., 2020). We have shown many highly conserved monocot miRNAs with high premiRNA counts, and it may be speculated that these miRNAs have been generated and proliferated subsequently to a Transposable-Element capture after the divergence of monocot species. miR1122 and miR1128 are among the most abundant miRNAs in terms of premiRNA counts and identified in every lineage of *Brachypodium*, in genome and transcriptome derived analysis of Barke, Bowman and Morex cultivars, and from homologous insect tolerance loci 1A, 2B, 3B and 4A of Chinese Spring, Svevo, Zavitan, barley and rye. One exception, miR1122 was not identified in any rice insect tolerance loci, this may be because this miRNA may be emerged after the divergence of Pooideae and Oryzoideae from their last common ancestor. In previous studies, miR1122 is identified as stress responsive miRNA in barley

and wheat (Kumar et al., 2014; B. Pandey, Gupta, Pandey, Sharma, & Sharma, 2013), regulating a disease resistance gene PR-1 and a stress tolerance gene BAK1 in wheat (Y. Wang et al., 2018) .

Both miR1122 and miR1128 have previously been shown to have multiple coding regions on different chromosomes (Alptekin, Akpinar, & Budak, 2017). miR1128 is the only miRNA identified from Morex transcriptome to be targeting both mRNA and lncRNA.

lncRNAs hold a great potential waiting to be explored, as miRNA precursors and miRNA repressors. Thus, understanding miRNA-lncRNA interactions are crucial for a better understanding of gene regulation. Here in this thesis, we have identified lncRNAs from barley transcriptome data under normal conditions and under salt treatment. We have revealed miRNAs that target lncRNAs and mRNAs and predicted function of their target proteins. miRNAs determined to target both lncRNAs and mRNAs are all well conserved monocot miRNAs, which may indicate the effect of evolution on construction of those networks.

Overall, we have carried out a complementary miRNA analysis among important monocot plants. We have identified many well-conserved and evolutionary important miRNAs with key developmental functions as well as novel or lineage specific miRNAs which contributes to species diversity. The miRNAs we have suggested to be agronomically important need to be validated and characterized to fully explore their potential in future crop improvement studies.

BIBLIOGRAPHY

Akpinar, B Ani, Avsar, B., Lucas, S. J., & Budak, H. (2012). Plant abiotic stress signaling. *Plant Signaling and Behavior*, (November), 1450–1455.

Akpinar, Bala A., & Budak, H. (2016). Dissecting miRNAs in Wheat D Genome Progenitor, Aegilops tauschii. *Frontiers in Plant Science*, *7*(606), 1–17. https://doi.org/10.3389/fpls.2016.00606

Akpinar, Bala Ani, Kantar, M., & Budak, H. (2015). Root precursors of microRNAs in wild emmer and modern wheats show major differences in response to drought stress. *Functional and Integrative Genomics*, *15*(5), 587–598. https://doi.org/10.1007/s10142-015-0453-0

Akpinar, Bala Ani, Yuce, M., Lucas, S., Vrána, J., Burešová, V., Doležel, J., & Budak, H. (2015a). Molecular organization and comparative analysis of chromosome 5B of the wild wheat ancestor Triticum dicoccoides. *Scientific Reports*, *5*(10763), 1–13. https://doi.org/10.1038/srep10763

Alptekin, B., Akpinar, B. A., & Budak, H. (2017). A Comprehensive Prescription for Plant miRNA Identification. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.02058

Alptekin, B., & Budak, H. (2016). Wheat miRNA ancestors: evident by transcriptome analysis of A, B, and D genome donors. *Functional and Integrative Genomics*, (March), 1–17. https://doi.org/10.1007/s10142-016-0487-y

Alptekin, B., Langridge, P., & Budak, H. (2017). Abiotic stress miRNomes in the Triticeae. *Functional and Integrative Genomics*, *17*, 145–170. https://doi.org/10.1007/s10142-016-0525-9

Arora, N. K. (2019). Impact of climate change on agriculture production and its sustainable solutions. *Environmental Sustainability*, *2*(2), 95–96. https://doi.org/10.1007/s42398-019-00078-w

Avena Sativa v1.0. (n.d.). The Oat Genome project. Retrieved March 13, 2021, from

https://avenagenome.org/

Axtell, M. J., & Bartel, D. P. (2005). Antiquity of microRNAs and their targets in land plants. *Plant Cell*, *17*(6), 1658–1673. https://doi.org/10.1105/tpc.105.032185

Axtell, M. J., & Meyers, B. C. (2018). Revisiting criteria for plant microRNA annotation in the Era of big data. *The Plant Cell*, *30*(2), 272–284. https://doi.org/10.1105/tpc.17.00851

Axtell, M. J., Westholm, J. O., & Lai, E. C. (2011). Vive la différence: Biogenesis and evolution of microRNAs in plants and animals. *Genome Biology*, *12*(221), 1–13. https://doi.org/10.1186/gb-2011-12-4-221

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., & Wishart, D. S. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Research*, *44*, W147–W153. https://doi.org/10.1093/nar/gkw419

Baev, V., Milev, I., Naydenov, M., Apostolova, E., Minkov, G., Minkov, I., & Yahubyan, G. (2011). Implementation of a de novo genome-wide computational approach for updating Brachypodium miRNAs. *Genomics*, *97*(5), 282–293. https://doi.org/10.1016/j.ygeno.2011.02.008

Bennetzen, J. L. (2000). Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology*, *42*(1), 251–269. https://doi.org/10.1023/A:1006344508454

Beres, B. L., Cárcamo, H. A., Yang, R. C., & Spaner, D. M. (2011). Integrating spring wheat sowing density with variety selection to manage wheat stem sawfly. *Agronomy Journal*, *103*(6), 1755–1764. https://doi.org/10.2134/agronj2011.0187

Biyiklioglu, S., Alptekin, B., Akpinar, B. A., Varella, A. C., Hofland, M. L., Weaver, D. K., … Budak, H. (2018). A large-scale multiomics analysis of wheat stem solidness and the wheat stem sawfly feeding response, and syntenic associations in barley, Brachypodium, and rice. *Functional and Integrative Genomics*, *18*, 241–259. https://doi.org/10.1007/s10142-017-0585-5

Blake, N. K., Stougaard, R. N., Bohannon, B., Weaver, D. K., Heo, H.-Y., Lamb, P. F.,

… Talbert, L. E. (2014). Registration of 'Egan' Wheat with Resistance to Orange Wheat Blossom Midge. *Journal of Plant Registrations*, *8*, 298–302. https://doi.org/10.3198/jpr2014.04.0022crc

Blake, N. K., Stougaard, R. N., Weaver, D. K., Sherman, J. D., Lanning, S. P., Naruoka, Y., … Talbert, L. E. (2011). Identification of a quantitative trait locus for resistance to Sitodiplosis mosellana (Géhin), the orange wheat blossom midge, in spring wheat. *Plant Breeding*, *130*, 25–30. https://doi.org/10.1111/j.1439-0523.2010.01809.x

Borah, P., Das, A., Milner, M. J., Ali, A., Bentley, A. R., & Pandey, R. (2018). Long non-coding rnas as endogenous target mimics and exploration of their role in low nutrient stress tolerance in plants. *Genes*, *9*(549), 1. – 17. https://doi.org/10.3390/genes9090459

Borchert, G. M., Holton, N. W., Williams, J. D., Hernan, W. L., Bishop, I. P., Dembosky, J. A., … Larson, E. D. (2011). Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins. *Mobile Genetic Elements*, *1*(1), 8–17. https://doi.org/10.4161/mge.1.1.15766

Borrill, P. (2020). Blurring the boundaries between cereal crops and model plants. *New Phytologist*, *228*(6), 1721–1727. https://doi.org/10.1111/nph.16229

Brant, E. J., & Budak, H. (2018). *Plant Small Non-coding RNAs and Their Roles in Biotic Stresses*. *9*(1038), 1–9. https://doi.org/10.3389/fpls.2018.01038

Brkljacic, J., Grotewold, E., Scholl, R., Mockler, T., Garvin, D. F., Vain, P., … Vogel, J. P. (2011). Brachypodium as a model for the grasses: Today and the future. *Plant Physiology*, *157*(1), 3–13. https://doi.org/10.1104/pp.111.179531

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., … Whited, J. L. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Reports*, *18*(3), 762–776. https://doi.org/10.1016/j.celrep.2016.12.063

Budak, H., & Akpinar, B. A. (2015). Plant miRNAs: biogenesis, organization and origins. *Functional & Integrative Genomics*, *15*(5), 523–531.

https://doi.org/10.1007/s10142-015-0451-2

Budak, H., Akpinar, B. A., Unver, T., & Turktas, M. (2013). Proteome changes in wild and modern wheat leaves upon drought stress by two-dimensional electrophoresis and nanoLC-ESI-MS/MS. *Plant Molecular Biology*, *83*(1–2), 89–103. https://doi.org/10.1007/s11103-013-0024-5

Budak, H., Hussain, B., Khan, Z., Ozturk, N. Z., & Ullah, N. (2015). From genetics to functional genomics: Improvement in drought signaling and tolerance in wheat. *Frontiers in Plant Science*, *6*(1012). https://doi.org/10.3389/fpls.2015.01012

Budak, H., Kantar, M., Bulut, R., & Akpinar, B. A. (2015). Stress responsive miRNAs and isomiRs in cereals. *Plant Science*, *235*, 1–13. https://doi.org/10.1016/j.plantsci.2015.02.008

Budak, H., Kaya, S. B., & Cagirici, H. B. (2020). Long Non-coding RNA in Plants in the Era of Reference Sequences. *Frontiers in Plant Science*, *11*(March), 1–10. https://doi.org/10.3389/fpls.2020.00276

Cagirici, H. Busra, Sen, T. Z., & Budak, H. (2021). Mirmachine: A one-stop shop for plant mirna annotation. *Journal of Visualized Experiments*, *2021*(171), 1–10. https://doi.org/10.3791/62430

Cagirici, Halise B., Biyiklioglu, S., & Budak, H. (2017). Assembly and annotation of transcriptome provided evidence of miRNA mobility between wheat and wheat stem sawfly. *Frontiers in Plant Science*, *8*(September), 1–14. https://doi.org/10.3389/fpls.2017.01653

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(421), 1–9. https://doi.org/10.1186/1471-2105-10-421

Campi, M., Dueñas, M., & Fagiolo, G. (2021). Specialization in food production affects global food security and food systems sustainability. *World Development*, *141*(105411). https://doi.org/10.1016/j.worlddev.2021.105411

Carillo, P., Annunziata, M. G., Pontecorvo, G., Fuggi, A., & Woodrow, P. (2011).

Salinity Stress and Salt Tolerance. *Abiotic Stress in Plants - Mechanisms and Adaptations*, *1*, 21–38. https://doi.org/10.5772/22331

Cattivelli, L., Rizza, F., Badeck, F. W., Mazzucotelli, E., Mastrangelo, A. M., Francia, E., … Stanca, A. M. (2008). Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crops Research*, *105*, 1–14. https://doi.org/10.1016/j.fcr.2007.07.004

Chadha, S., & Sharma, M. (2014). Transposable elements as stress adaptive capacitors induce genomic instability in fungal pathogen Magnaporthe oryzae. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0094415

Chan, K. X., Crisp, P. A., Estavillo, G. M., & Pogson, B. J. (2010). Chloroplast-to-nucleus communication: Current knowledge, experimental strategies and relationship to drought stress signaling. *Plant Signaling and Behavior*, *5*(12), 1575–1582. https://doi.org/10.4161/psb.5.12.13758

Cheah, B. H., Nadarajah, K., Divate, M. D., & Wickneswari, R. (2015). Identification of four functionally important microRNA families with contrasting differential expression profiles between drought-tolerant and susceptible rice leaf at vegetative stage. *BMC Genomics*, *16*(692), 1–18. https://doi.org/10.1186/s12864-015-1851-3

Chen, H., Li, Z., & Xiong, L. (2012). A plant microRNA regulates the adaptation of roots to drought stress. *FEBS Letters*, *586*(12), 1742–1747. https://doi.org/10.1016/j.febslet.2012.05.013

Chen, L., Ren, F., Zhong, H., Jiang, W., & Li, X. (2010). Identification and expression analysis of genes in response to high-salinity and drought stresses in Brassica napus. *Acta Biochimica et Biophysica Sinica*, *42*(2), 154–164. https://doi.org/10.1093/abbs/gmp113

Chu, Z., Chen, J., Xu, H., Dong, Z., Chen, F., & Cui, D. (2016). Identification and comparative analysis of microRNA in wheat (triticum aestivum L.) callus derived from mature and immature embryos during In vitro culture. *Frontiers in Plant Science*, *7*(1302). https://doi.org/10.3389/fpls.2016.01302

Cockrell, D. M., Griffin-Nolan, R. J., Rand, T. A., Altilmisani, N., Ode, P. J., & Peairs,

F. (2017). Host plants of the wheat stem sawfly (hymenoptera: Cephidae). *Environmental Entomology*, *00*(0), 847–854. https://doi.org/10.1093/ee/nvx104

Cook, J. P., Wichman, D. M., Martin, J. M., Bruckner, P. L., & Talbert, L. E. (2004). Identification of microsatellite markers associated with a stem solidness locus in wheat. *Crop Science*, *44*(4), 1397–1402. https://doi.org/10.2135/cropsci2004.1397

da Silva, E. C., de Albuquerque, M. B., de Azevedo Neto, A. D., & da Silva Junior, C. D. (2013). Drought and Its Consequences to Plants – From Individual to Ecosystem. In *Responses of organisms to water stress*.

Dai, X., & Zhao, P. X. (2011). PsRNATarget: A plant small RNA target analysis server. *Nucleic Acids Research*, *39*(SUPPL. 2), 155–159. https://doi.org/10.1093/nar/gkr319

Dai, X., Zhuang, Z., & Zhao, P. X. (2018). PsRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Research*, *46*, W49–W54. https://doi.org/10.1093/nar/gky316

Devos, K. M., Atkinson, M. D., Chinoy, C. N., Francis, H. A., Harcourt, R. L., Koebner, R. M. D., … Gale, M. D. (1993). Chromosomal rearrangements in the rye genome relative to that of wheat. *Theoretical and Applied Genetics*, *85*(6–7), 673–680. https://doi.org/10.1007/BF00225004

Diao, X. M., & Lisch, D. (2006). Mutator transposon in maize and MULEs in the plant genome. *Acta Genetica Sinica*, *33*(6), 477–487. https://doi.org/10.1016/S0379-4172(06)60075-9

Djami-tchatchou, A. T., Sanan-mishra, N., Ntushelo, K., & Dubery, I. A. (2017). Functional Roles of microRNAs in Agronomically Important Plants — Potential as Targets for Crop Improvement and Protection. *Frontiers in Plant Science*, *8*(378). https://doi.org/10.3389/fpls.2017.00378

Draper, J., Mur, L. A. J., Jenkins, G., Ghosh-Biswas, G. C., Bablak, P., Hasterok, R., & Routledge, A. P. M. (2001). Brachypodium distachyon. A new model system for functional genomics in grasses. *Plant Physiology*, *127*(4), 1539–1555. https://doi.org/10.1104/pp.010196

Dresselhaus, T., & Hückelhoven, R. (2018). Biotic and abiotic stress responses in crop plants. *Agronomy*, *8*(267), 8–13. https://doi.org/10.3390/agronomy8110267

Du, H., Ran, F., Dong, H. L., Wen, J., Li, J. N., & Liang, Z. (2016). Genome-wide analysis, classification, evolution, and expression analysis of the cytochrome P450 93 family in land plants. *PLoS ONE*, *11*(10), 1–24. https://doi.org/10.1371/journal.pone.0165020

Dubcovsky, J., Ramakrishna, W., SanMiguel, P. J., Busso, C. S., Yan, L., Shiloff, B. A., & Bennetzen, J. L. (2001). Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiology*, *125*(3), 1342–1353. https://doi.org/10.1104/pp.125.3.1342

Fahlgren, N., Howell, M. D., Kasschau, K. D., Chapman, E. J., Sullivan, C. M., Cumbie, J. S., … Carrington, J. C. (2007). High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS ONE*, *2*(2). https://doi.org/10.1371/journal.pone.0000219

FAO-Food and Agricultural Organization of the United Nations. (2020). FAOSTAT statistical database. Retrieved October 23, 2021, from http://www.fao.org/faostat/

FAO. (2017). The future of food and agriculture: trends and challenges. In *The future of food and agriculture: trends and challenges*. Retrieved from https://www.fao.org/3/i6583e/i6583e.pdf

Fatica, A., & Bozzoni, I. (2014). Long non-coding RNAs: New players in cell differentiation and development. *Nature Reviews Genetics*, *15*(1), 7–21. https://doi.org/10.1038/nrg3606

Ferdous, J., Hussain, S. S., & Shi, B. J. (2015). Role of microRNAs in plant drought tolerance. *Plant Biotechnology Journal*, *13*(3), 293–305. https://doi.org/10.1111/pbi.12318

Filiz, E., Ozdemir, B. S., Budak, F., Vogel, J. P., Tuna, M., & Budak, H. (2009). Molecular, morphological, and cytological analysis of diverse Brachypodium distachyon inbred lines. *Genome*, *52*(10), 876–890. https://doi.org/10.1139/G09-062

Franke, K. R., Schmidt, S. A., Park, S., Jeong, D. H., Accerbi, M., & Green, P. J. (2018). Analysis of Brachypodium miRNA targets: Evidence for diverse control during stress and conservation in bioenergy crops. *BMC Genomics*, *19*(547), 1–18. https://doi.org/10.1186/s12864-018-4911-7

Gepts, P. (2002). A Comparison between Crop Domestication, Classical Plant Breeding, and Genetic Engineering. *Crop Science*, *42*(6), 1780–1790.

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H. R., Martinez, P. A., … Edwards, D. (2016). The pangenome of an agronomically important crop plant Brassica oleracea. *Nature Communications*, *7*, 1–8. https://doi.org/10.1038/ncomms13390

Gong, Z. J., Wu, Y. Q., Miao, J., Duan, Y., Jiang, Y. L., & Li, T. (2013). Global Transcriptome Analysis of Orange Wheat Blossom Midge, Sitodiplosis mosellana (Gehin) (Diptera: Cecidomyiidae) to Identify Candidate Transcripts Regulating Diapause. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0071564

Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., … Vogel, J. P. (2017). Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. *Nature Communications*, Vol. 8. https://doi.org/10.1038/s41467-017-02292-8

Gordon, S. P., Priest, H., Des Marais, D. L., Schackwitz, W., Figueroa, M., Martin, J., … Vogel, J. P. (2014). Genome diversity in Brachypodium distachyon: Deep sequencing of highly diverse inbred lines. *Plant Journal*, *79*(3), 361–374. https://doi.org/10.1111/tpj.12569

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., … Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435. https://doi.org/10.1093/nar/gkn176

Govindaraj, M., Pattanashetti, S. K., Patne, N., & Kanatti, A. A. (2018). Breeding Cultivars for Heat Stress Tolerance in Staple Food Crops. *Next Generation Plant Breeding*.

Guleria, P., Mahajan, M., Bhardwaj, J., & Yadav, S. K. (2011). Plant Small RNAs: Biogenesis, Mode of Action and Their Roles in Abiotic Stresses. *Genomics, Proteomics and Bioinformatics*, *9*(6), 183–199. https://doi.org/10.1016/S1672-0229(11)60022-3

Han, R., Jian, C., Lv, J., Yan, Y., Chi, Q., Li, Z., … Zhang, J. (2014). Identification and characterization of microRNAs in the flag leaf and developing seed of wheat (Triticum aestivum L .). *BMC Genomics*, *15*(289), 1–14.

Hands, P., & Drea, S. (2012). A comparative view of grain development in Brachypodium distachyon. *Journal of Cereal Science*, *56*(1), 2–8. https://doi.org/10.1016/j.jcs.2011.12.010

Hao, Z., Geng, M., Hao, Y., Zhang, Y., Zhang, L., Wen, S., … Liu, G. (2019). Screening for differential expression of genes for resistance to Sitodiplosis mosellana in bread wheat via BSR-seq analysis. *Theoretical and Applied Genetics*, *132*(11), 3201–3221. https://doi.org/10.1007/s00122-019-03419-9

Harris, M. O., Stuart, J. J., Mohan, M., Nair, S., Lamb, R. J., & Rohfritsch, O. (2003). Grasses and Gall Midges: Plant Defense and Insect Adaptation. *Annual Review of Entomology*, *48*, 549–577. https://doi.org/10.1146/annurev.ento.48.091801.112559

Howe, K. L., Contreras-Moreira, B., De Silva, N., & et al. (2019). Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Research*, *48*(D1), D689–D695. https://doi.org/10.1093/nar/gkz890

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics*, *26*(5), 680–682. https://doi.org/10.1093/bioinformatics/btq003

Hurgobin, B., & Edwards, D. (2017). SNP Discovery Using a Pangenome: Has the Single Reference Approach Become Obsolete? *Biology*, Vol. 6, p. 21. https://doi.org/10.3390/biology6010021

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature*, *436*(7052), 793–800. https://doi.org/10.1038/nature03895

International Wheat Genome Sequencing Consortium (IWGSC). (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, *361*(eaar7191). https://doi.org/10.1126/science.aar7191

Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., Kempen, B., & de Sousa, L. (2019). Global mapping of soil salinity change. *Remote Sensing of Environment*, *231*(11260). https://doi.org/10.1016/j.rse.2019.111260

Jamshidi, A., & Javanmard, H. R. (2018). Evaluation of barley (Hordeum vulgare L.) genotypes for salinity tolerance under field conditions using the stress indices. *Ain Shams Engineering Journal*, *9*(4), 2093–2099. https://doi.org/10.1016/j.asej.2017.02.006

Jeong, D. H., Schmidt, S. A., Rymarquis, L. A., Park, S., Ganssmann, M., German, M. A., … Green, P. J. (2013). Parallel analysis of rna ends enhances global investigation of microrna s and target rnas of brachypodium distachyon. *Genome Biology*, *14*(145), 1–22. https://doi.org/10.1186/gb-2013-14-12-r145

Jiang, S. C., Mei, C., Liang, S., Yu, Y. T., Lu, K., Wu, Z., … Zhang, D. P. (2015). Crucial roles of the pentatricopeptide repeat protein SOAR1 in Arabidopsis response to drought, salt and cold stresses. *Plant Molecular Biology*, *88*, 369–385. https://doi.org/10.1007/s11103-015-0327-9

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

Kassa, M. T., Haas, S., Schliephake, E., Lewis, C., You, F. M., Pozniak, C. J., … McCartney, C. A. (2016). A saturated SNP linkage map for the orange wheat blossom midge resistance gene Sm1. *Theoretical and Applied Genetics*, *129*(8), 1507–1517. https://doi.org/10.1007/s00122-016-2720-4

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., & Gao, G. (2007).

CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, *35*, 345–349. https://doi.org/10.1093/nar/gkm391

Kornienko, A. E., Guenzl, P. M., Barlow, D. P., & Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biology*, *11*(59). https://doi.org/10.1186/1741-7007-11-59

Koroban, N. V., Kudryavtseva, A. V., Krasnov, G. S., Sadritdinova, A. F., Fedorova, M. S., Snezhkina, A. V., … Melnikova, N. V. (2016). The role of microRNA in abiotic stress response in plants. *Molecular Biology*, *50*(3), 337–343. https://doi.org/10.1134/S0026893316020102

Kosová, K., Vítámvás, P., Hlaváčková, I., Urban, M. O., Vlasáková, E., & Prášil, I. T. (2015). Responses of two barley cultivars differing in their salt tolerance to moderate and high salinities and subsequent recovery. *Biologia Plantarum*, *59*, 106–114. https://doi.org/10.1007/s10535-014-0465-y

Kozomara, A., & Griffiths-Jones, S. (2010). MiRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, *39*, 152–157. https://doi.org/10.1093/nar/gkq1027

Kruszka, K., Pacak, A., Swida-Barteczka, A., Stefaniak, A. K., Kaja, E., Sierocka, I., … Szweykowska-Kulinska, Z. (2013). Developmentally regulated expression and complex processing of barley pri-microRNAs. *BMC Genomics*, *14*(1), 34. https://doi.org/10.1186/1471-2164-14-34

Kumar, D., Singh, D., Kanodia, P., Prabhu, K. V., Kumar, M., & Mukhopadhyay, K. (2014). Discovery of novel leaf rust responsive microRNAs in wheat and prediction of their target genes. *Journal of Nucleic Acids*. https://doi.org/10.1155/2014/570176

Kurtoglu, K. Y., Kantar, M., & Budak, H. (2014). New wheat microRNA using whole-genome sequence. *Functional and Integrative Genomics*, *14*(2), 363–379. https://doi.org/10.1007/s10142-013-0357-9

Laluk, K., Abuqamar, S., & Mengiste, T. (2011). The arabidopsis mitochondria-

localized pentatricopeptide repeat protein PGN functions in defense against necrotrophic fungi and abiotic stress tolerance. *Plant Physiology*, *156*(4), 2053–2068. https://doi.org/10.1104/pp.111.177501

Lamb, R. J., Wise, I. L., Smith, M. A. H., McKenzie, R. I. H., Thomas, J., & Olfert, O. O. (2002). Oviposition deterrence against Sitodiplosis mosellana (Diptera: Cecidomyiidae) in spring wheat (Gramineae). *Canadian Entomologist*, *134*(1), 85–96. https://doi.org/10.4039/Ent13485-1

Lee, U., Rioflorido, I., Hong, S. W., Larkindale, J., Waters, E. R., & Vierling, E. (2007). The Arabidopsis ClpB/Hsp100 family of proteins: Chaperones for stress and chloroplast development. *Plant Journal*, *49*(1), 115–127. https://doi.org/10.1111/j.1365-313X.2006.02940.x

Li, J., Ma, W., Zeng, P., Wang, J., Geng, B., Yang, J., & Cui, Q. (2014). LncTar: A tool for predicting the RNA targets of long noncoding RNAs. *Briefings in Bioinformatics*, *16*(5), 806–812. https://doi.org/10.1093/bib/bbu048

Li, S., Fan, C., Li, Y., Zhang, J., Sun, J., Chen, Y., … Hu, Z. (2016). Effects of drought and salt-stresses on gene expression in Caragana korshinskii seedlings revealed by RNA-seq. *BMC Genomics*, *17*(1), 1–19. https://doi.org/10.1186/s12864-016-2562-0

Li, T., Ma, L., Geng, Y., Hao, C., Chen, X., & Zhang, X. (2015). Small RNA and degradome sequencing reveal complex roles of miRNAs and their targets in developing wheat grains. *PLoS ONE*, *10*(10), 1–19. https://doi.org/10.1371/journal.pone.0139658

Li, Y., Song, Q., Zhang, Y., Li, Z., Guo, J., Chen, X., & Zhang, G. (2020). Genome-wide identification, characterization, and expression patterns analysis of the SBP-box gene family in wheat (Triticum aestivum L.). *Scientific Reports*, *10*(17250), 1–14. https://doi.org/10.1038/s41598-020-74417-x

Liu, J., Wang, H., & Chua, N. H. (2015). Long noncoding RNA transcriptome of plants. *Plant Biotechnology Journal*, *13*(3), 319–328. https://doi.org/10.1111/pbi.12336

Lorenz, R., Bernhart, S. H., Siederdissen, C. H. zu, Tafer, H., Flamm, C., Stadler, P. F.,

& Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*(26), 1–14. https://doi.org/10.1093/nar/gkz164

Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, *44*, 54–57. https://doi.org/10.1093/nar/gkw413

Lu, W., Li, J., Liu, F., Gu, J., Guo, C., Xu, L., … Xiao, K. (2011). Expression pattern of wheat miRNAs under salinity stress and prediction of salt-inducible miRNAs targets. *Frontiers of Agriculture in China*, *5*(4), 413–422. https://doi.org/10.1007/s11703-011-1133-z

Lucas, S. J., & Budak, H. (2012). Sorting the wheat from the Chaff: Identifying miRNAs in genomic survey sequences of Triticum aestivum chromosome 1AL. *PLoS ONE*, *7*(7). https://doi.org/10.1371/journal.pone.0040859

Macedo, T. B., Peterson, R. K. D., Weaver, D. K., & Morrill, W. L. (2005). Wheat stem sawfly, Cephus cinctus Norton, impact on wheat primary metabolism: An ecophysiological approach. *Environmental Entomology*, *34*(3), 719–726. https://doi.org/10.1603/0046-225X-34.3.719

Manavella, P. A., Yang, S. W., & Palatnik, J. (2019). Keep calm and carry on: miRNA biogenesis under stress. *Plant Journal*, *99*(5), 832–843. https://doi.org/10.1111/tpj.14369

Mardis, E. R. (2017). DNA sequencing technologies: 2006-2016. *Nature Protocols*, *12*(2), 213–218. https://doi.org/10.1038/nprot.2016.182

Markham, N., & Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. *Methods in Molecular Biology (Clifton, N.J.)*, *453*, 3–31. https://doi.org/10.1007/978-1-60327-429-6_1

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. Retrieved from http://www-huber.embl.de/users/an-

Matres, J. M., Hilscher, J., Datta, A., Armario-Nájera, V., Baysal, C., He, W., …

Slamet-Loedin, I. H. (2021). Genome editing in cereal crops: an overview. In *Transgenic Research* (Vol. 30). https://doi.org/10.1007/s11248-021-00259-6

Mayer, K. F. X., Martis, M., Hedley, P. E., & et al. (2011). Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell*, *23*(4), 1249–1263. https://doi.org/10.1105/tpc.110.082537

Mayer, K. F. X., Waugh, R., Langridge, P., Close, T. J., Wise, R. P., Graner, A., … Stein, N. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature*, Vol. 491, pp. 711–716. https://doi.org/10.1038/nature11543

McKenzie, R. I. H., Lamb, R. J., Aung, T., Wise, I. L., Barker, P., & Olfert, O. O. (2002). Inheritance of resistance to wheat midge, Sitodiplosis mosellana, in spring wheat. *Plant Breeding*, *121*(5), 383–388. https://doi.org/10.1046/j.1439-0523.2002.745267.x

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature Reviews. Genetics*, *10*, 155–159. https://doi.org/10.1038/nrg2521

Michaud, M., Cognat, V., Duchêne, A. M., & Maréchal-Drouard, L. (2011). A global picture of tRNA genes in plant genomes. *Plant Journal*, *66*(1), 80–93. https://doi.org/10.1111/j.1365-313X.2011.04490.x

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., & Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Science*, *282*(May 2020), 60–72. https://doi.org/10.1016/j.plantsci.2019.01.007

Moose, S. P., & Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology*, *147*(3), 969–977. https://doi.org/10.1104/pp.108.118232

Morea, E. G. O., da Silva, E. M., e Silva, G. F. F., Valente, G. T., Barrera Rojas, C. H., Vincentz, M., & Nogueira, F. T. S. (2016). Functional and evolutionary analyses of the miR156 and miR529 families in land plants. *BMC Plant Biology*, *16*(40), 1–13. https://doi.org/10.1186/s12870-016-0716-5

Morrill, W. L., Kushnak, G. D., & Gabor, J. W. (1998). Parasitism of the wheat stem sawfly (Hymenoptera: Cephidae) in Montana. *Biological Control*, *12*, 159–163. https://doi.org/10.1006/bcon.1998.0629

Mouriz, A., López-González, L., Jarillo, J. A., & Piñeiro, M. (2015). PHDs govern plant development. *Plant Signaling and Behavior*, *10*(7), 1–7. https://doi.org/10.4161/15592324.2014.993253

Munns, R., & Tester, M. (2008). Mechanisms of salinity tolerance. *Annual Review of Plant Biology*, *59*, 651–681. https://doi.org/10.1146/annurev.arplant.59.032607.092911

Murat, F., Armero, A., Pont, C., Klopp, C., & Salse, J. (2017). Reconstructing the genome of the most recent common ancestor of flowering plants. *Nature Genetics*, *49*(4), 490–496. https://doi.org/10.1038/ng.3813

Muslu, T., Biyiklioglu-Kaya, S., Akpinar, B. A., Yuce, M., & Budak, H. (2021). Pan-genome mirnomics in Brachypodium. *Plants*, *10*(991). https://doi.org/10.3390/plants10050991

Mussig, C., Biesgen, C., Lisso, J., Uwer, U., Weiler, E. W., & Altmann, T. (2000). A novel stress-inducible 12-oxophytodienoate reductase from Arabidopsis thaliana provides a potential link between brassinosteroid-action and jasmonic-acid synthesis. *Journal of Plant Physiology*, *157*(2), 143–152. https://doi.org/10.1016/S0176-1617(00)80184-4

Nagatoshi, M., Terasaka, K., Nagatsu, A., & Mizukami, H. (2011). Iridoid-specific glucosyltransferase from Gardenia jasminoides. *Journal of Biological Chemistry*, *286*(37), 32866–32874. https://doi.org/10.1074/jbc.M111.242586

Narjala, A., Nair, A., Tirumalai, V., Vivek Hari Sundar, G., & Shivaprasad, P. V. (2020). A conserved sequence signature is essential for robust plant miRNA biogenesis. *Nucleic Acids Research*, *48*(6), 3103–3118. https://doi.org/10.1093/nar/gkaa077

Nilsen, K. T., N'Diaye, A., MacLachlan, P. R., Clarke, J. M., Ruan, Y., Cuthbert, R. D., … Pozniak, C. J. (2017). High density mapping and haplotype analysis of the

major stem-solidness locus SSt1 in durum and common wheat. *PLoS ONE*, *12*(4), 1–19. https://doi.org/10.1371/journal.pone.0175285

Nilsen, K. T., Walkowiak, S., Xiang, D., Gao, P., Quilichini, T. D., Willick, I. R., … Pozniak, C. J. (2020). Copy number variation of TdDof controls solid-stemmed architecture in wheat. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(46), 28708–28718. https://doi.org/10.1073/pnas.2009418117

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., … Dubchak, I. (2014). The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Research*, *42*(D1), 26–31. https://doi.org/10.1093/nar/gkt1069

Pandey, B., Gupta, O. P., Pandey, D. M., Sharma, I., & Sharma, P. (2013). Identification of new stress-induced microRNA and their targets in wheat using computational approach. *Plant Signaling and Behavior*, *8*(5), 1–9. https://doi.org/10.4161/psb.23932

Pandey, P., Irulappan, V., Bagavathiannan, M. V., & Senthil-Kumar, M. (2017). Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Frontiers in Plant Science*, *8*, 537. https://doi.org/10.3389/fpls.2017.00537

Peng, H., & Zhang, J. (2009). Plant genomic DNA methylation in response to stresses: Potential applications and challenges in plant breeding. *Progress in Natural Science*, *19*(9), 1037–1045. https://doi.org/10.1016/j.pnsc.2008.10.014

Perez-de-Castro, A. ., Vilanova, S., Canizares, J., Pascual, L., Blanca, J. ., Diez, M. ., … Pico, B. (2012). Application of Genomic Tools in Plant Breeding. *Current Genomics*, *13*, 179–195. https://doi.org/10.1109/cdc.1996.577319

Piriyapongsa, J., & Jordan, I. K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. *Rna*, *14*(5), 814–821. https://doi.org/10.1261/rna.916708

Pont, C., & Salse, J. (2017). Wheat paleohistory created asymmetrical genomic evolution. *Current Opinion in Plant Biology*, *36*, 29–37.

https://doi.org/10.1016/j.pbi.2017.01.001

Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and Functions of Long Noncoding RNAs. *Cell*, *136*(4), 629–641. https://doi.org/10.1016/j.cell.2009.02.006

Pourkheirandish, M., & Komatsuda, T. (2007). The importance of barley genetics and domestication in a global perspective. *Annals of Botany*, *100*, 999–1008. https://doi.org/10.1093/aob/mcm139

Quesada, V. (2016). The roles of mitochondrial transcription termination factors (MTERFs) in plants. *Physiologia Plantarum*, *157*(3), 389–399. https://doi.org/10.1111/ppl.12416

Rabanus-Wallace, M. T., Hackauf, B., Mascher, M., & et al. (2021). Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nature Genetics*, *53*(4), 564–573. https://doi.org/10.1038/s41588-021-00807-0

Rasool, S., Ahmad, P., Rehman, M. U., Arif, A., & Anjum, N. A. (2015). Achieving Crop Stress Tolerance and Improvement—an Overview of Genomic Techniques. *Applied Biochemistry and Biotechnology*, *177*(7), 1395–1408. https://doi.org/10.1007/s12010-015-1830-9

Ray, S., & Satya, P. (2014). Next generation sequencing technologies for next generation plant breeding. *Frontiers in Plant Science*, *5*(July), 1–4. https://doi.org/10.3389/fpls.2014.00367

Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry*, *81*, 145–166. https://doi.org/10.1146/annurev-biochem-051410-092902

Sakaguchi, J., & Watanabe, Y. (2012). miR165/166 and the development of land plants. *Development Growth and Differentiation*, *54*(1), 93–99. https://doi.org/10.1111/j.1440-169X.2011.01318.x

Salk, J. J., Schmitt, M. W., & Loeb, L. A. (2018). Enhancing the accuracy of next-

generation sequencing for detecting rare and subclonal mutations. *Nature Reviews Genetics*, Vol. 19, pp. 269–285. https://doi.org/10.1038/nrg.2017.117

Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, *94*(3), 441–448. https://doi.org/10.1016/0022-2836(75)90213-2

Sato, K., Tanaka, T., Shigenobu, S., Motoi, Y., Wu, J., & Itoh, T. (2016). Improvement of barley genome annotations by deciphering the Haruna Nijo genome. *DNA Research*, *23*(1), 21–28. https://doi.org/10.1093/dnares/dsv033

Scholthof, K. B. G., Irigoyen, S., Catalan, P., & Mandadi, K. K. (2018). Brachypodium: A monocot grass model genus for plant biology. *The Plant Cell*, *30*(8), 1673–1694. https://doi.org/10.1105/tpc.18.00083

Schwab, R., Palatnik, J. F., Riester, M., Schommer, C., Schmid, M., & Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. *Developmental Cell*, *8*(4), 517–527. https://doi.org/10.1016/j.devcel.2005.01.018

Seleiman, M. F., Al-Suhaibani, N., Ali, N., Akmal, M., Alotaibi, M., Refay, Y., … Battaglia, M. L. (2021). Drought stress impacts on plants and different approaches to alleviate its adverse effects. *Plants*, *10*(2), 1–25. https://doi.org/10.3390/plants10020259

Sharma, M., & Pandey, G. K. (2016). Expansion and function of repeat domain proteins during stress and development in plants. *Frontiers in Plant Science*, *6*, 1–15. https://doi.org/10.3389/fpls.2015.01218

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: Past, present and future. *Nature*, *550*, 345–353. https://doi.org/10.1038/nature24286

Shinozaki, K., & Yamaguchi-Shinozaki, K. (2000). Molecular responses to dehydration and low temperature: Differences and cross-talk between two stress signaling pathways. *Current Opinion in Plant Biology*, *3*(3), 217–223. https://doi.org/10.1016/S1369-5266(00)00067-4

Shrestha, G., Briar, S. S., & Reddy, G. V. P. (2018). Plant defense elicitors: Plant fitness versus wheat stem sawfly. *PeerJ*. https://doi.org/10.7717/peerj.5892

Shrivastava, P., & Kumar, R. (2015). Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. *Saudi Journal of Biological Sciences*, *22*, 123–131. https://doi.org/10.1016/j.sjbs.2014.12.001

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next Generation Sequencing technologies (and bioinformatics) in cancer. *Curr Protoc Mol Biol.*, *122*(1), 1–15. https://doi.org/10.1002/cpmb.59.Overview

Smalheiser, N. R., & Torvik, V. I. (2005). Mammalian microRNAs derived from genomic repeats. *Trends in Genetics*, *21*(6), 322–326. https://doi.org/10.1016/j.tig.2005.04.008

Song, X., Li, Y., Cao, X., & Qi, Y. (2019). MicroRNAs and Their Regulatory Roles in Plant-Environment Interactions. *Annual Review of Plant Biology*, *70*, 27.1-37.1. Retrieved from https://doi.org/10.1146/annurev-arplant-050718-

Strable, J., & Scanlon, M. J. (2009). Maize (Zea mays): A model organism for basic and applied research in plant biology. *Cold Spring Harbor Protocols*, *4*(10), 1–10. https://doi.org/10.1101/pdb.emo132

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., … Zhao, Y. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research*, *41*(17). https://doi.org/10.1093/nar/gkt646

Szczepaniec, A., Glover, K. D., & Berzonsky, W. (2015). Impact of solid and hollow varieties of winter and spring wheat on severity of wheat stem sawfly (Hymenoptera: Cephidae) Infestations and Yield and Quality of Grain. *Journal of Economic Entomology*, *108*(5), 2316–2323. https://doi.org/10.1093/jee/tov207

Tanaka, T., Kobayashi, F., Joshi, G. P., Onuki, R., Sakai, H., Kanamori, H., … Handa, H. (2014). Next-generation survey sequencing and the molecular organization of wheat chromosome 6B. *DNA Research*, *21*, 103–114. https://doi.org/10.1093/dnares/dst041

Temesgen, B. (2021). Role and economic importance of crop genetic diversity in food security. *International Journal of Agricultural Science and Food Technology*, *7*, 164–169. https://doi.org/10.17352/2455-815x.000104

Tester, M., & Langridge, P. (2010). Breeding Technologies to Increase. *Science*, *327*, 818–822. Retrieved from http://www.sciencemag.org/cgi/content/abstract/327/5967/818

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., … Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." *Proceedings of the National Academy of Sciences*, *102*(39), 13950–13955. https://doi.org/10.1073/pnas.0506758102

Tettelin, Hervé, Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, Vol. 12, pp. 472–477. https://doi.org/10.1016/j.mib.2008.09.006

Thambugala, D., Pozniak, C. J., Kumar, S., Burt, A. J., Wise, I. L., Smith, M. A. H., … McCartney, C. A. (2021). Genetic analysis of oviposition deterrence to orange wheat blossom midge in spring wheat. *Theoretical and Applied Genetics*, *134*(2), 647–660. https://doi.org/10.1007/s00122-020-03720-y

The International Brachypodium Initiative. (2010). Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature*, *463*(7282), 763–768. https://doi.org/10.1038/nature08747

Thomas, J., Fineberg, N., Penner, G., McCartney, C., Aung, T., Wise, I., & McCallum, B. (2005). Chromosome location and markers of Sm1: A gene of wheat that conditions antibiotic resistance to orange wheat blossom midge. *Molecular Breeding*, *15*(2), 183–192. https://doi.org/10.1007/s11032-004-5041-2

Tyagi, S., Sharma, A., & Upadhyay, S. K. (2018). Role of Next-Generation RNA-Seq Data in Discovery and Characterization of Long Non-Coding RNA in Plants. *Next Generation Plant Breeding*.

Ulukan, H. (2009). The evolution of cultivated plant species: Classical plant breeding

versus genetic engineering. *Plant Systematics and Evolution*, *280*, 133–142. https://doi.org/10.1007/s00606-008-0118-8

Unver, T., & Budak, H. (2009). Conserved microrna s and their targets in model grass species brachypodium distachyon. *Planta*, *230*(4), 659–669. https://doi.org/10.1007/s00425-009-0974-7

Varella, A. C., Talbert, L. E., Hofland, M. L., Buteler, M., Sherman, J. D., Blake, N. K., … Weaver, D. K. (2016). Alleles at a quantitative trait locus for stem solidness in wheat affect temporal patterns of pith expression and level of resistance to the wheat stem sawfly. *Plant Breeding*, *135*, 546–551. https://doi.org/10.1111/pbr.12398

Waititu, J. K., Zhang, C., Liu, J., & Wang, H. (2020). Plant non-coding rnas: Origin, biogenesis, mode of action and their roles in abiotic stress. *International Journal of Molecular Sciences*, *21*(21), 1–22. https://doi.org/10.3390/ijms21218401

Walkowiak, S., Gao, L., Monat, C., & et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature*, *588*(7837), 277–283. https://doi.org/10.1038/s41586-020-2961-x

Wang, J., Mei, J., & Ren, G. (2019). Plant microRNAs: Biogenesis, homeostasis, and degradation. *Frontiers in Plant Science*, *10*, 1–12. https://doi.org/10.3389/fpls.2019.00360

Wang, K. C., & Chang, H. Y. (2011). Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell*, *43*(6), 904–914. https://doi.org/10.1016/j.molcel.2011.08.018

Wang, S., Wong, D., Forrest, K., & et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal*, *12*(6), 787–796. https://doi.org/10.1111/pbi.12183

Wang, Y., Shi, C., Yang, T., Zhao, L., Chen, J., Zhang, N., … Chen, F. (2018). High-throughput sequencing revealed that microRNAs were involved in the development of superior and inferior grains in bread wheat. *Scientific Reports*,

8(1), 1–18. https://doi.org/10.1038/s41598-018-31870-z

Wanga, M. A., Shimelis, H., Mashilo, J., & Laing, M. D. (2021). Opportunities and challenges of speed breeding: A review. *Plant Breeding*, *140*(2), 185–194. https://doi.org/10.1111/pbr.12909

Weidenbach, D., Jansen, M., Franke, R. B., Hensel, G., Weissgerber, W., Ulferts, S., … Schaffrath, U. (2014). Evolutionary Conserved Function of Barley and Arabidopsis 3-KETOACYL-CoA SYNTHASES in Providing Wax Signals for Germination of Powdery Mildew Fungi. *Plant Physiology*, *166*(3), 1621–1633. https://doi.org/10.1104/pp.114.246348

WHEAT in the World. (2020). Retrieved October 23, 2021, from https://wheat.org/wheat-in-the-world/

Wichelns, D., & Qadir, M. (2014). Achieving sustainable irrigation requires effective management of salts, soil salinity, and shallow groundwater. *Agricultural Water Management*, *157*, 31–39. https://doi.org/10.1016/j.agwat.2014.08.016

Wilson, P. B., Streich, J. C., Murray, K. D., Eichten, S. R., Cheng, R., Aitken, N. C., … Mur, L. (2019). Global diversity of the Brachypodium species complex as a resource for genome-wide association studies demonstrated for agronomic traits in response to climate. *Genetics*, *211*(1), 317–331. https://doi.org/10.1534/genetics.118.301589

World Food Programme. (2022). HungerMap LIVE : Global insights and key trends. Retrieved January 4, 2022, from https://hungermap.wfp.org/

World Food Summit. (1996). World Food Summit Plan of Action. *In: Rome Declaration on World Food Security and World Food Summit Plan of Action*, (13-17 November).

Wu, L., Liu, D., Wu, J., Zhang, R., Qin, Z., Liu, D., … Mao, L. (2013). Regulation of FLOWERING LOCUS t by a MicroRNA in Brachypodium distachyon. *Plant Cell*, *25*(11), 4363–4377. https://doi.org/10.1105/tpc.113.118620

Xin, M., Wang, Y., Yao, Y., Song, N., Hu, Z., Qin, D., … Sun, Q. (2011). Identification

and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biology*, *11*(61). https://doi.org/10.1186/1471-2229-11-61

Yang, X., Zhang, L., Yang, Y., Schmid, M., & Wang, Y. (2021). miRNA Mediated Regulation and Interaction between Plants and Pathogens. *International Journal of Molecular Sciences*, *22*(2913), 1–13. https://doi.org/10.3390/ijms22062913

Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., … Parniske, M. (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. *PNAS*, *105*(51), 20540–20545. https://doi.org/10.1073/pnas.0806858105

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., … Chen, X. (2005). Methylation as a Crucial Step in Plant microRNA Biogenesis. *Science*, *307*(5711), 932–935. https://doi.org/10.1126/science.1107130.Methylation

Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., … Yang, H. (2002). A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. indica). *Science*, *296*, 79–92.

Zahra, S., Singh, A., Poddar, N., & Kumar, S. (2021). Transfer RNA-derived non-coding RNAs (tncRNAs): Hidden regulation of plants' transcriptional regulatory circuits. *Computational and Structural Biotechnology Journal*, *19*, 5278–5291. https://doi.org/10.1016/j.csbj.2021.09.021

Zhang, H., Li, Y., & Zhu, J. K. (2018). Developing naturally stress-resistant crops for a sustainable agriculture. *Nature Plants*, *4*, 989–996. https://doi.org/10.1038/s41477-018-0309-4

Zhang, J., Zhang, S., Cheng, M., Jiang, H., Zhang, X., Peng, C., … Jin, J. (2018). Effect of drought on agronomic traits of rice and wheat: A meta-analysis. *International Journal of Environmental Research and Public Health*, *15*(839). https://doi.org/10.3390/ijerph15050839

Zhang, Li, Zheng, Y., Jagadeeswaran, G., Li, Y., Gowdu, K., & Sunkar, R. (2011). Identification and temporal expression analysis of conserved and novel microRNAs in Sorghum. *Genomics*, *98*, 460–468. https://doi.org/10.1016/j.ygeno.2011.08.005

Zhang, Lijing, Geng, M., Zhang, Z., Zhang, Y., Yan, G., Wen, S., … Wang, R. (2020). Molecular mapping of major QTL conferring resistance to orange wheat blossom midge (Sitodiplosis mosellana) in Chinese wheat varieties with selective populations. *Theoretical and Applied Genetics*, *133*, 491–502. https://doi.org/10.1007/s00122-019-03480-4

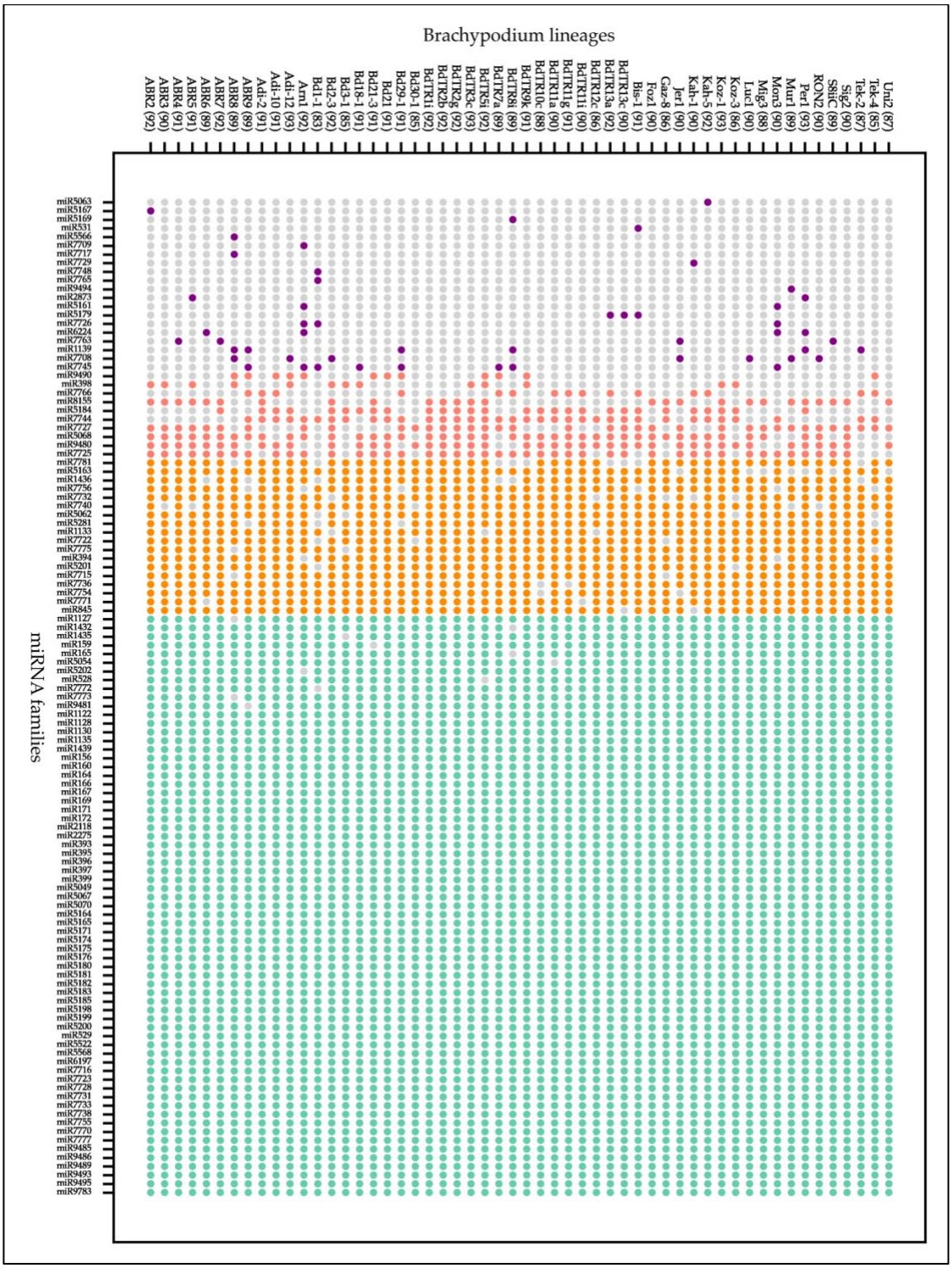
Zhang, S. D., Ling, L. Z., Zhang, Q. F., Xu, J. Di, & Cheng, L. (2015). Evolutionary comparison of two combinatorial regulators of SBP-Box genes, MiR156 and MiR529, in plants. *PLoS ONE*, *10*, 1–12. https://doi.org/10.1371/journal.pone.0124621

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., … Huang, X. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics*, *50*(2), 278–284. https://doi.org/10.1038/s41588-018-0041-z

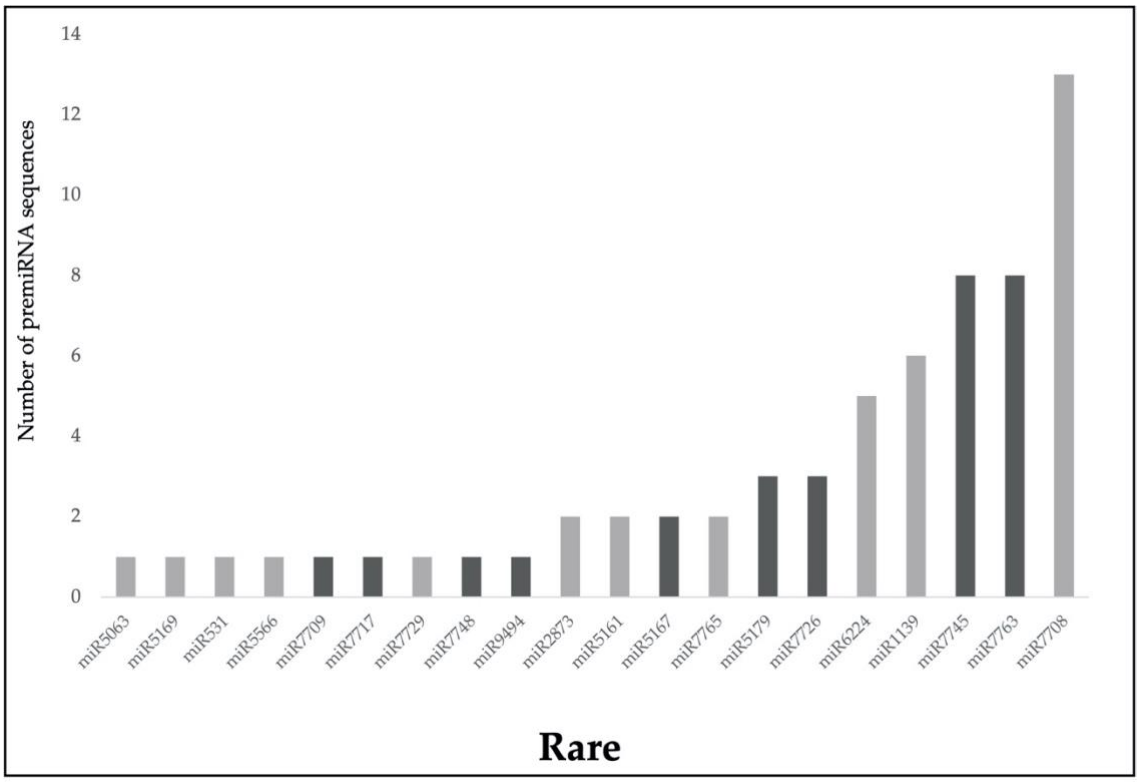
Zhou, H., Lin, J., Johnson, A., Morgan, R. L., Zhong, W., & Ma, W. (2011). Pseudomonas syringae type III effector HopZ1 targets a host enzyme to suppress isoflavone biosynthesis and promote infection in soybean. *Cell Host and Microbe*, *9*, 177–186. https://doi.org/10.1016/j.chom.2011.02.007

Zhou, Y., Xu, Z., Duan, C., Chen, Y., Meng, Q., Wu, J., … Li, X. (2016). Dual transcriptome analysis reveals insights into the response to Rice black-streaked dwarf virus in maize. *Journal of Experimental Botany*, *67*(15), 4593–4609. https://doi.org/10.1093/jxb/erw244

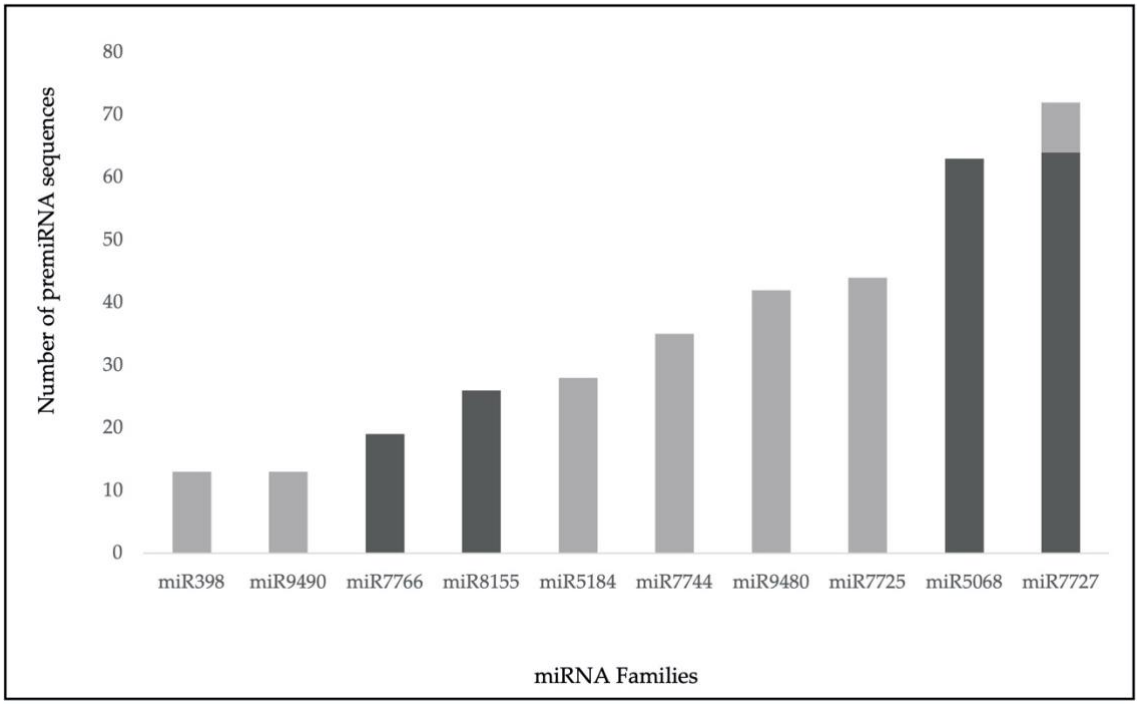
APPENDIX A



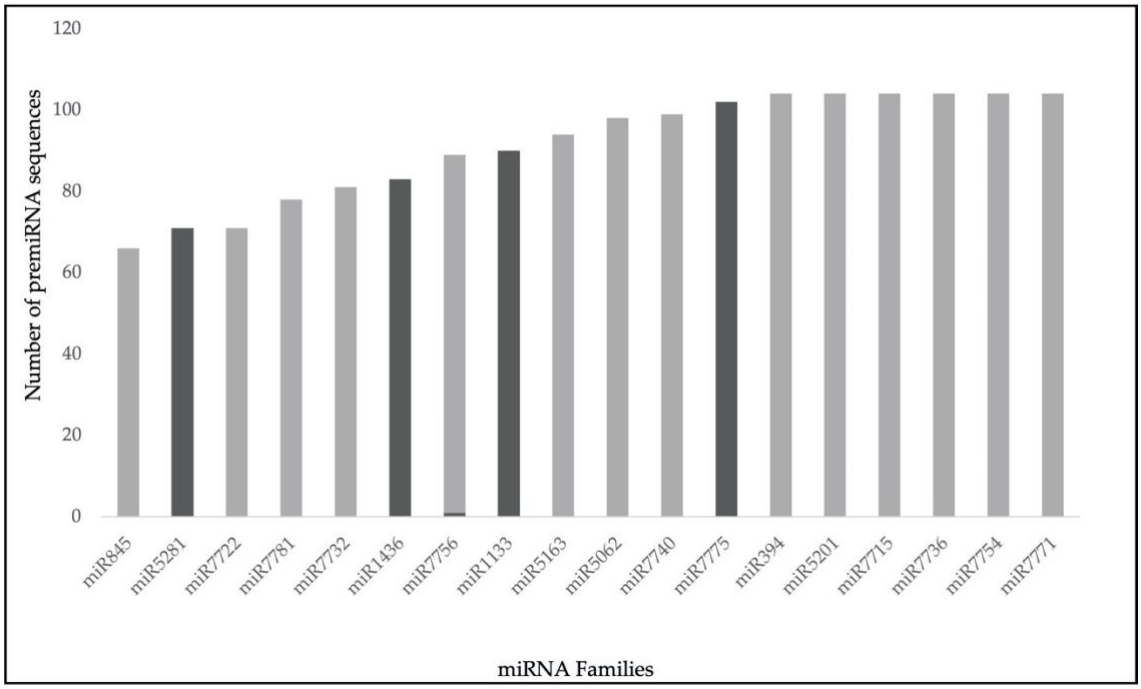
Supplementary Figure 1 The distribution of miRNA families across lineages. The x-axis shows 115 miRNA families identified in 54 *Brachypodium* lineages, shown in y-axis. Conservation of miRNA families among lineages increased along the x-axis and miRNA families were classified into four groups based on this conservation: Rare miRNA families (purple), moderately conserved miRNA families (pink), highly conserved miRNA families (orange), and common miRNA families (green). The total number of miRNA families identified in each lineage is also shown in parenthesis next to lineage names.



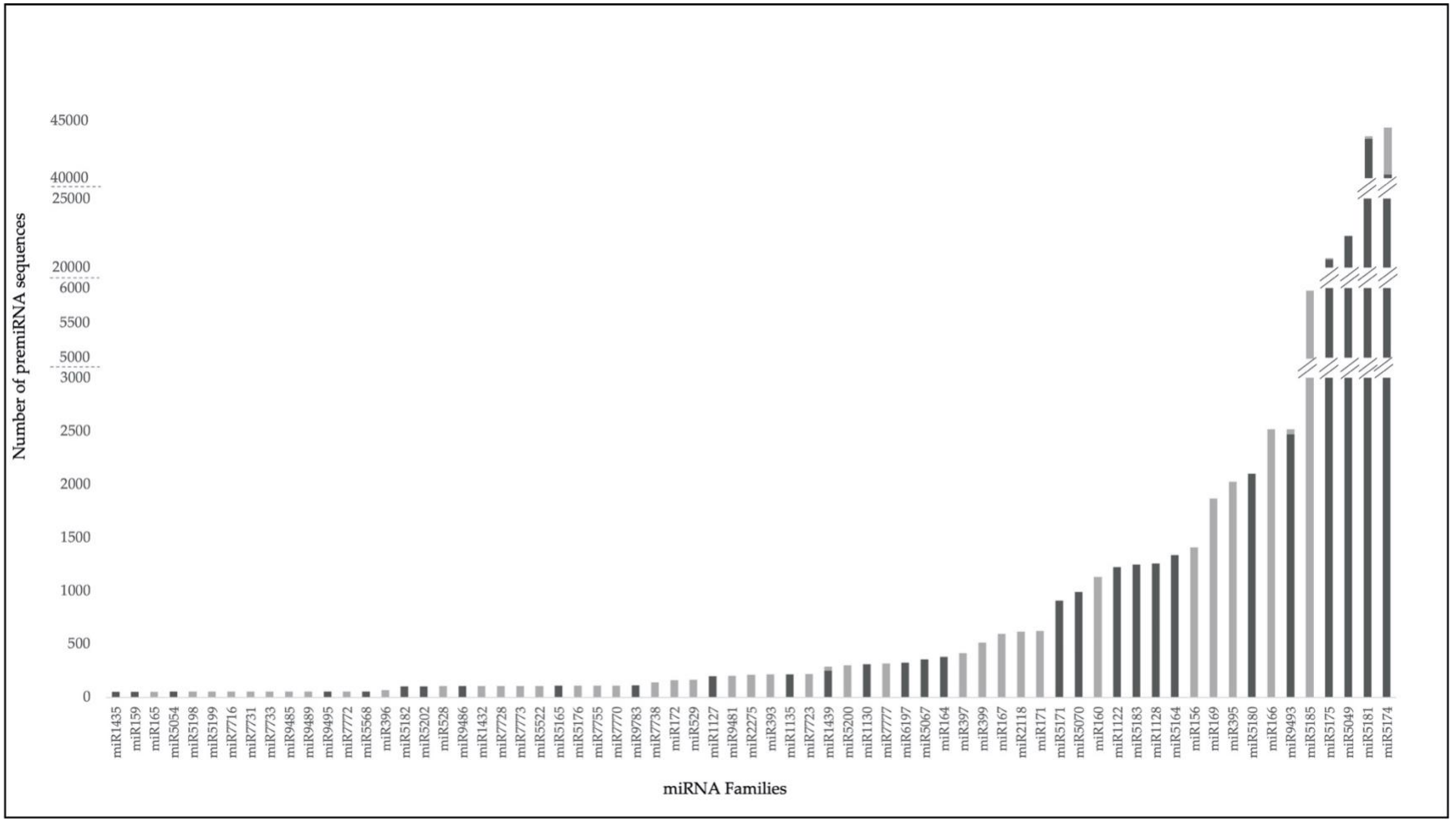
Supplementary Figure 2 The number of precursor miRNA (pre-miRNA) sequences identified for each miRNA family in rare miRNA family group. Light gray bars indicate repetitive pre-miRNAs, defined as containing repetitive elements for more than half of their lengths and non-repetitive pre-miRNAs are shown as the dark shade of gray.



Supplementary Figure 3 The number of precursor miRNA (pre-miRNA) sequences identified for each miRNA family group in moderately conserved miRNA families group. Light gray bars indicate repetitive pre-miRNAs, defined as containing repetitive elements for more than half of their lengths and non-repetitive pre-miRNAs are shown as the dark shade of gray.



Supplementary Figure 4 The number of precursor miRNA (pre-miRNA) sequences identified for each miRNA family group in highly conserved miRNA families group. Light gray bars indicate repetitive pre-miRNAs, defined as containing repetitive elements for more than half of their lengths and non-repetitive pre-miRNAs are shown as the dark shade of gray.



Supplementary Figure 5. The number of precursor miRNA (pre-miRNA) sequences identified for each miRNA family group in common miRNA families group. Light gray bars indicate repetitive pre-miRNAs, defined as containing repetitive elements for more than half of their lengths and non-repetitive pre-miRNAs are shown as the dark shade of gray.

The following supplementary tables and files are available online at https://www.mdpi.com/article/10.3390/plants10050991/s1

Supplementary Table S1: the presence/absence of miRNA families and the number of precursor sequences (pre-miRNAs) for each miRNA family in each *Brachypodium* lineage.

Supplementary Table S2: mature miRNA and pre-miRNA sequences identified in 54 *Brachypodium* lineages.

Supplementary Table S3: potential mRNA target sequences of 54 *Brachypodium* lineages.

Supplementary File S1: representative pre-miRNA sequence alignments for selected miRNA families.

APPENDIX B

The following supplementary tables are available online at: https://www.mdpi.com/article/10.3390/ijms222212349/s1.

Supplementary Table S1. Putative OWBM and WSS resistance loci in cereals, identified by molecular markers and/or transcript analysis.

Supplementary Table S2. Potential annotation of transcripts identified from 3BL-QTL homologous regions

Supplementary Table S3. Proteins identified in homologous QTLs in all species and their potential matches from an iTRAQ-based proteomics study.

Supplementary Table S4. Oat contigs matching the molecular markers defining 1A-QTL, 2B-QTL, 3B-QTL and 4A-QTLs and transcript sequences identified from homologous regions

Supplementary Table S5. All mature miRNA and precursor sequence identified from homologous resistance loci.

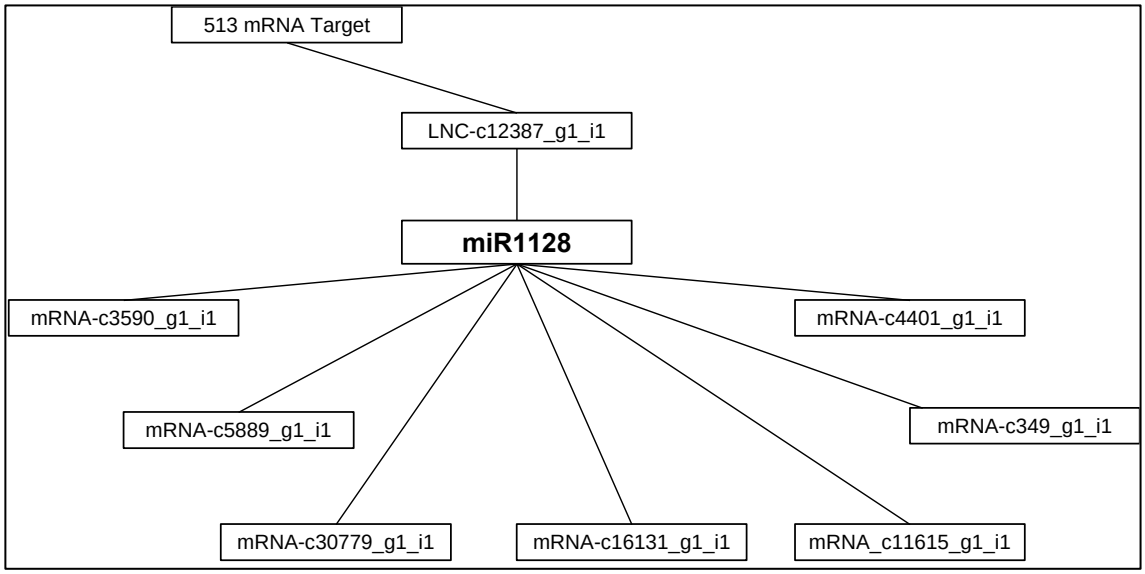
Supplementary Table S6. Presence of miRNA families identified in homologous QTL loci in each cereal.

Supplementary Table S7: miRNA families which have identified to have at least one transcript target in homolog QTL regions.

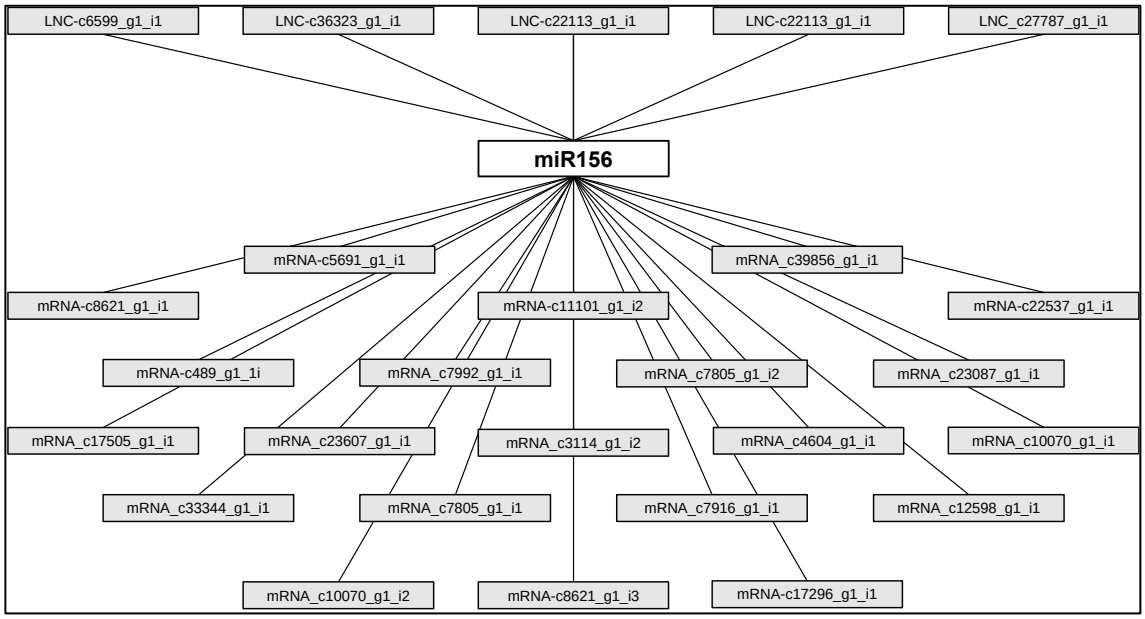
Supplementary Table S8. Potential annotation of transcripts identified from 1A-QTL, 2B-QTL and 4A-QTL homologous regions

APPENDIX C

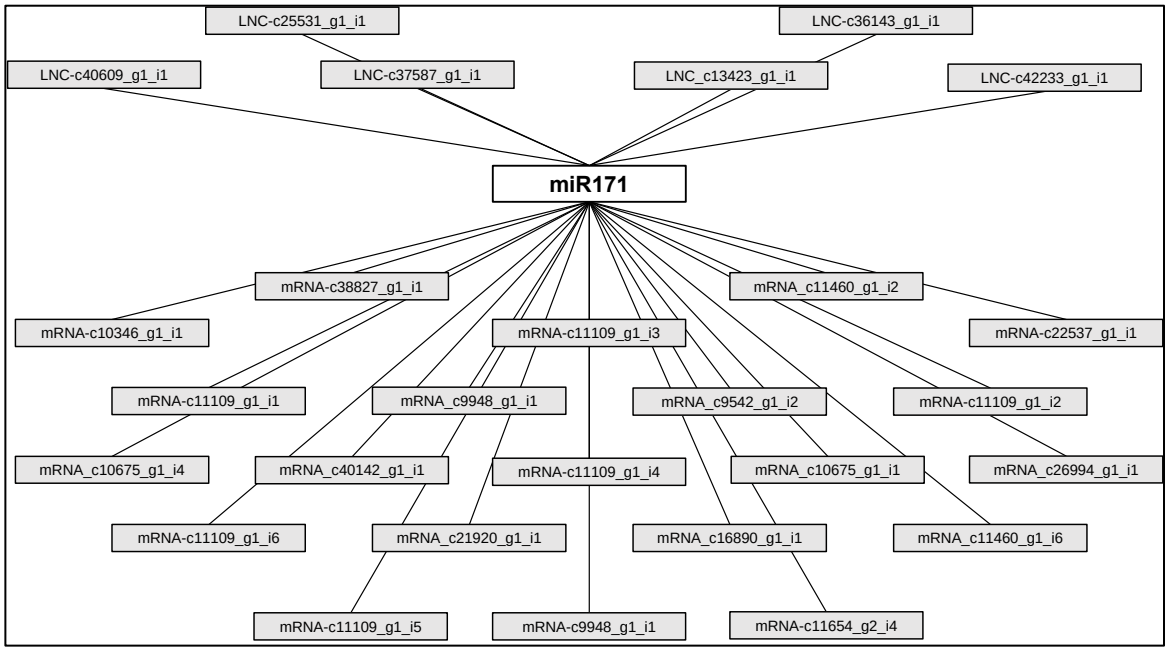
Supplementary Figures



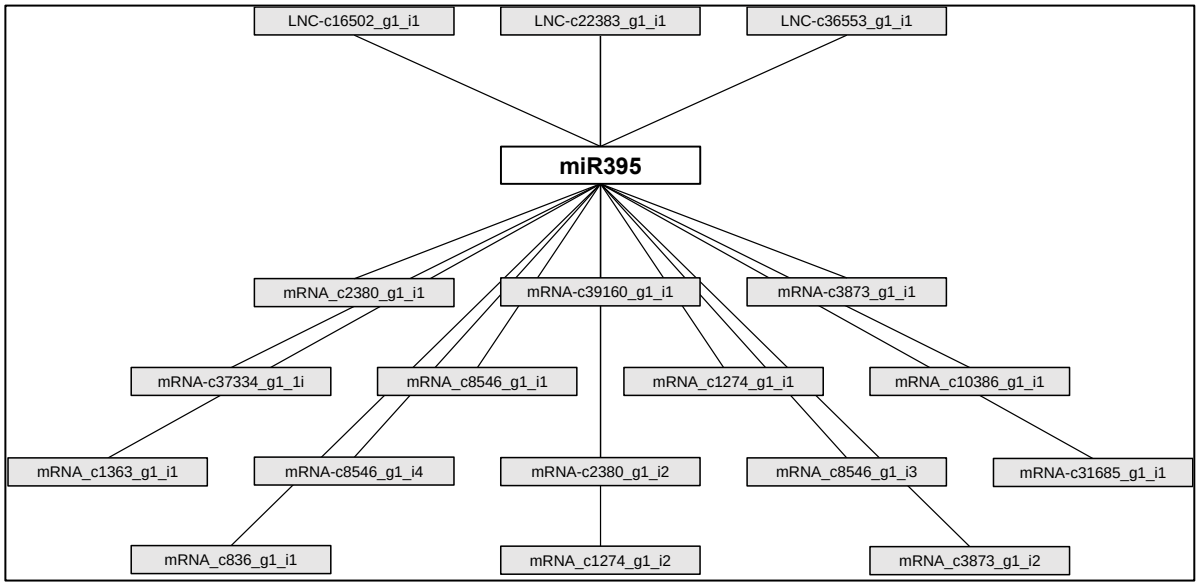
Supplementary Figure 1. LncRNA-miRNA-mRNA regulatory network of miR1128 of Morex transcriptome, the only miRNA is identified to target both lncRNA and mRNA transcripts. The figure shows 7 mRNA transcripts and a lncRNA transcript targeted by miR1128. The lncRNA transcript LNC-c12387_g1_i1 has been found to target 513 mRNA transcripts.



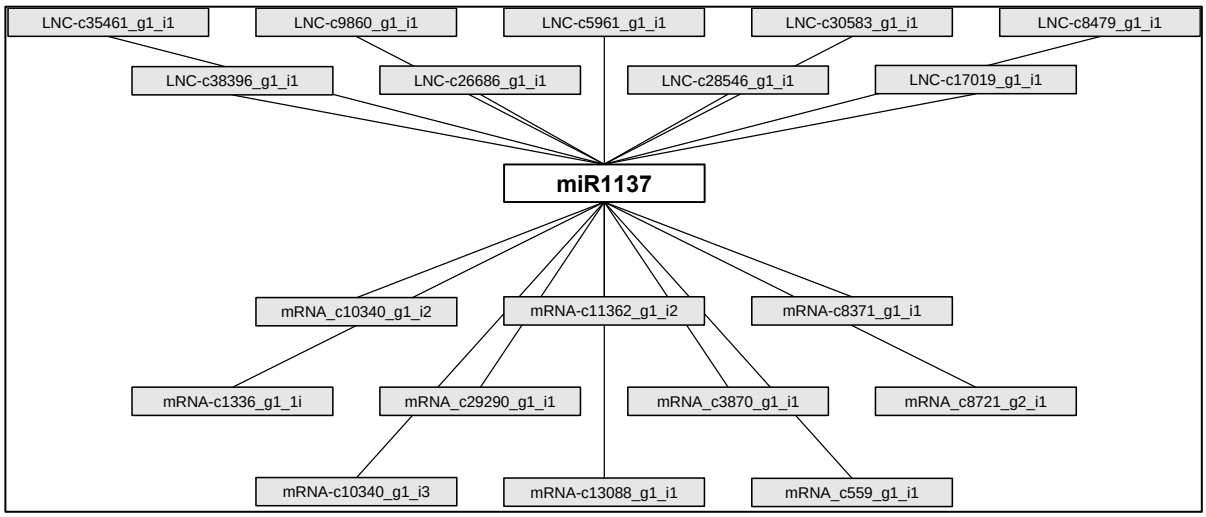
Supplementary Figure 2. LncRNA-miRNA-mRNA regulatory network of miR156 of *Hordeum Spontaneum* transcriptome. miR156 identified from salt-treated transcriptome was shown to interact with 5 lncRNAs and 18 mRNA transcripts.



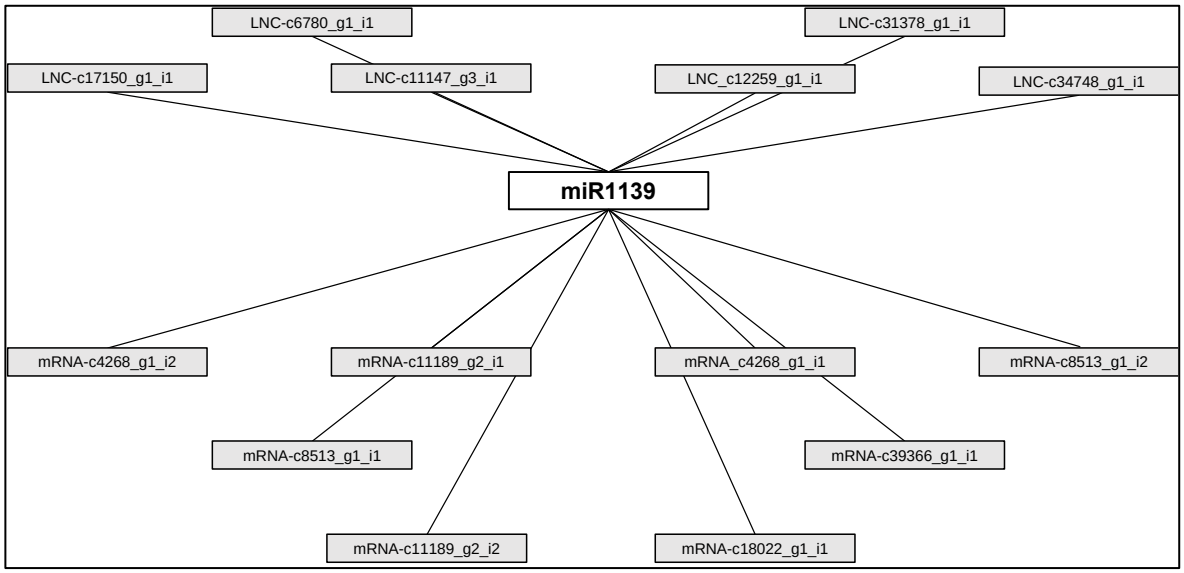
Supplementary Figure 3. LncRNA-miRNA-mRNA regulatory network of miR171 of *Hordeum Spontaneum* transcriptome. miR171 identified from salt-treated transcriptome was shown to interact with 6 lncRNAs and 21 mRNA transcripts.



Supplementary Figure 4. LncRNA-miRNA-mRNA regulatory network of miR395 of *Hordeum Spontaneum* transcriptome. miR395 identified from salt-treated transcriptome was shown to interact with 3 lncRNAs and 15 mRNA transcripts.



Supplementary Figure 5. LncRNA-miRNA-mRNA regulatory network of miR1137 of *Hordeum Spontaneum* transcriptome. miR1137 identified from salt-treated transcriptome was shown to interact with 9 lncRNAs and 10 mRNA transcripts.



Supplementary Figure 6. LncRNA-miRNA-mRNA regulatory network of miR1139 of *Hordeum Spontaneum* transcriptome. miR1139 identified from salt-treated transcriptome was shown to interact with 6 lncRNAs and 8 mRNA transcripts.

Supplementary Tables

The following Supplementary Tables are available online at https://figshare.com/s/42ccd4bf61c118fd3e59

Supplementary Table S1: Mature and premiRNA Sequences of miRNAs identified from Barke Genome

Supplementary Table S2: Mature and premiRNA Sequences of miRNAs identified from Bowman Genome

Supplementary Table S3: Mature and premiRNA Sequences of miRNAs identified from Morex Genome

Supplementary Table S4: miRNA Identification using Morex transcriptome data after Trinity assembly

Supplementary Table S5: miRNA Identification from 0hr salt treatment Hordeum spontaneum transcriptome data (SRR1028011)

Supplementary Table S6: miRNA Identification from 0hr salt treatment Hordeum spontaneum transcriptome data (SRR1028012)

Supplementary Table S7: miRNA Identification from 2 hr salt treatment Hordeum spontaneum transcriptome data (SRR1049587)

Supplementary Table S8: miRNA Identification from 2 hr salt treatment Hordeum spontaneum transcriptome data (SRR1049592)

Supplementary Table S9: miRNA Identification from 12 hr salt treatment Hordeum spontaneum transcriptome data (SRR1049570)

Supplementary Table S10: miRNA Identification from 12 hr salt treatment Hordeum spontaneum transcriptome data (SRR1049609)

Supplementary Table S11: miRNA Identification from 12 hr salt treatment Hordeum spontaneum transcriptome data (SRR1049626)

Supplementary Table S12: miRNA Identification from 24 hr salt treatment Hordeum spontaneum transcriptome data (SRR1049643)

Supplementary Table S13: miRNA Identification from 24 hr salt treatment Hordeum spontaneum transcriptome data (SRR1049648)

Supplementary Table S14: miRNA Identification from 24 hr salt treatment Hordeum spontaneum transcriptome data (SRR1049655)

Supplementary Table S15: psRNATarget tool results showing putative mRNA transcript targets of miRNA families identified from Barke genome assembly

Supplementary Table S16: psRNATarget tool results showing putative mRNA transcript targets of miRNA families identified from Bowman genome assembly

Supplementary Table S17: psRNATarget tool results showing putative mRNA transcript targets of miRNA families identified from Morex genome assembly

Supplementary Table S18: SmallRNA expression of identified miRNAs from Barke genome assembly

Supplementary Table S19: SmallRNA expression of identified miRNAs from Bowman genome assembly

Supplementary Table S20: SmallRNA expression of identified miRNAs from Morex genome assembly

Supplementary Table S21: PremiRNA level expression of identified miRNAs from Barke genome assembly

Supplementary Table S22: PremiRNA level expression of identified miRNAs from Bowman genome assembly

Supplementary Table S23: PremiRNA level expression of identified miRNAs from Morex genome assembly

Supplementary Table S24: Transfer RNA (tRNA) genes determined from unmasked genome assembly of Barke

Supplementary Table S25: Transfer RNA (tRNA) genes determined from unmasked genome assembly of Bowman

Supplementary Table S26: Transfer RNA (tRNA) genes determined from unmasked genome assembly of Morex

Supplementary Table S27: 79 LncRNA transcripts identified from a total of 18851 mRNA transcripts of Morex transcriptome

Supplementary Table S28: miRNA1128 and its lncRNA transcript targets in Morex

Supplementary Table S29: LncRNA transcripts identified from 0hr salt treatment of Hordeum spontaneum transcriptome

Supplementary Table S30: LncRNA transcripts identified from 2hr salt treatment of Hordeum spontaneum transcriptome

Supplementary Table S31: LncRNA transcripts identified from 12hr salt treatment of Hordeum spontaneum transcriptome

Supplementary Table S32: LncRNA transcripts identified from 24 hr salt treatment of Hordeum spontaneum transcriptome

Supplementary Table S33. Predicted functions of target proteins of lncRNA-mRNA targeting six miRNA families identified from *Hordeum spontaneum* transcriptome data

miRNA ID	LncRNA IDs	Functions of targeted proteins
miR156	c14548_g1_i1, c6599_g1_i1 c36323_g1_i1, c22113_g1_i1 c27787_g1_i1	DNA-binding storekeeper protein-related transcriptional regulator, roles in leaf development, vegetative phase change, flower and fruit development, plant architecture, sporogenesis, Gibberelic acid signaling and toxin response, spontaneous calcium independent fusion of synaptic vesicles containing v-SNAREs
miR171	c37578_g1_i1, c25531_g1_i1 c40609_g1_i1, c13423_g1_i1 c42233_g1_i1, c36143_g1_i1	Protein catabolic process, hydrolase activity, ATP binding, roles in maintaining the homeostasis of cellular proteome, roles in lateral root development, mediated protein degradation in regulating auxin accumulation during lateral root primordium development and lateral root meristem emergence
miR395	c22383_g1_i1, c16502_g1_i1 c36553_g1_i1	Sulfate adenylyltransferase (ATP) activity, sulfate assimilation, hydrolase activity, acting on ester bonds, protein serine/threonine kinase activity, signal transduction, ATP binding, embryo development ending in seed dormancy, plastid translation, copper ion binding, chloroplast organization, defense response to fungus, incompatible interaction
miR1128	120 lncRNA targets	Integral component of membrane, regulation of transcription, DNA-templated, association with enzymatic, membrane transporters or DNA-binding domains, key role in pre-mRNA 3'-end formation
miR1137	c35461_g1_i1, c5961_g1_i1 c28546_g1_i1, c30583_g1_i1 c9860_g1_i1, c26686_g1_i1 c8479_g1_i1, c17019_g1_i1 c38396_g1_i1	Integral component of membrane, interaction with ARF exchange factors
miR1139	c12259_g1_i1, c11147_g3_i1 c34748_g1_i1, c31378_g1_i1 c17150_g1_i1, c6780_g1_i1	ADP binding, signal transduction and transcription regulation to cell cycle control, autophagy and apoptosis