

AN ABSTRACT OF THE DISSERTATION OF

Kali M. Brandt for the degree of Doctor of Philosophy in Crop Science presented on June 17, 2019.

Title: Modern Wheat Breeding: Next Generation Sequencing and CRISPR Transformation.

Abstract approved:

Robert S. Zemetra

Wheat (*Triticum aestivum*) is an extremely important crop worldwide. It accounts for almost one quarter of the calories consumed each day by more than one third of the world's population, and is grown over more land area than any other crop. Wheat breeding programs constantly strive to increase or maintain yields while simultaneously improving pest resistance, quality traits, abiotic stress tolerance, nitrogen use efficiency, and more. This must be done while also attempting to address changing environmental factors such as increasing and unpredictable temperatures due to climate change which can affect winter wheat's vernalization requirement and shift severity and duration of disease pressures. While wheat is a highly diverse crop owing to the many environments in which it is planted, within each growing region it is relatively monomorphic and the introgression of new traits is often difficult. In order to keep up with demands for performance and to adapt the crop to changing environmental factors, modern wheat breeders are turning to new and emerging technologies. The advent of reduced representation sequencing and modern computational capabilities has made marker-trait discovery relatively easy and affordable for most wheat breeding programs. These advances make it possible for breeders and researchers to develop markers relevant to regionally-adapted germplasm which can significantly decrease the time it takes to develop a new cultivar via the use of Marker Assisted Selection (MAS). The emerging field of

targeted genetic editing using CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is also being explored for its potential role in wheat breeding through the introduction of traits that cannot not be incorporated using traditional methods.

The work for this dissertation addressed the need for marker-trait identification by utilizing a population of 196 recombinant inbred lines (RILs) derived from a cross between ‘Skiles’ and ‘Goetze,’ two cultivars with diverse phenotypes. The RILs were genotyped using Genotyping by Sequencing (GBS), and Quantitative Trait Loci (QTL) were identified after phenotyping for the desired traits. After field testing in three locations over two years, two QTL for stripe rust (*Puccinia striiformis*) resistance were identified on chromosomes 3B and 3D. A combination of all resistance alleles together outperformed the resistant parent in every treatment. All identified markers are novel and are immediately useful for MAS to integrate extremely high levels of stripe rust resistance into new varieties. The same RILs were also phenotyped for days to heading after varying vernalization treatments. The lines were subjected to either zero, two, four, or six weeks of vernalization at 6°C, then grown in the greenhouse under uniform conditions, and days to heading were recorded. Four QTL on four chromosomes were identified. The QTL on chromosome 5B is likely a facultative allele known to be in Goetze, but not yet characterized in local germplasm. The QTL on chromosome 5D is also likely a facultative allele that was previously unknown to be in this germplasm. Markers from both QTL can be used for MAS to incorporate facultative growth habits into new varieties. The QTL on chromosomes 1D and 3B are potentially associated with poorly-characterized photoperiod genes, and point to the importance of “minor” genes in flowering time that should be tested further.

The work for this dissertation also explored the feasibility of incorporating non-transgenic CRISPR transformation into a public breeding program. The use of CRISPR first requires the *in vivo* testing of desired gene targets to ensure the chromatin state is not inaccessible at the target location. Therefore, a protocol for protoplast isolation and PEG-mediated transformation using CRISPR Ribonucleoproteins (RNPs) was developed to streamline the process and make

implementation in any lab possible. This protocol consistently yields high numbers of protoplasts, does not require specialized equipment, and returns clear positive and negative transformation results for sgRNA targets that are inaccessible *in vivo*. After testing the gene target functionality *in vivo*, immature wheat embryos were biolistically bombarded with gold particles coated in CRISPR RNPs, plantlets were regenerated via tissue culture, and transformed individuals were identified with Sanger sequencing, following published protocols. A transformation efficiency of 6.7 percent was achieved at the T₀ stage, but subsequent sampling of individual tillers from the fully-grown plants showed the presence of chimeric tissue. This reduced the potential transformation efficiency to below one percent and led to the conclusion that this technique requires considerable time and monetary resources that make it infeasible for use in a public breeding program.

It is speculated that successful modern wheat breeding will include a synthesis of all available techniques and technologies for trait improvement such as increased disease resistance, adaptability to unpredictable winter temperatures, and the integration of new traits with genetic editing. This study found novel, useful markers for stripe rust resistance and facultative growth habits in PNW-adapted germplasm that can be used immediately for rapid variety development. While testing CRISPR gene targets *in vivo* using protoplasts is possible with the developed protocol, the non-transgenic transformation of wheat with CRISPR using currently-published protocols is not feasible. Therefore, while it is clear that marker-trait discovery and MAS are useful and important contributors to modern wheat breeding, targeted gene editing with CRISPR technology requires further study and validation before it can be integrated into breeding programs.

©Copyright by Kali M. Brandt
June 17, 2019
All Rights Reserved

Modern Wheat Breeding: Next Generation Sequencing and CRISPR Transformation

by
Kali M. Brandt

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 17, 2019
Commencement June 2020

Doctor of Philosophy dissertation of Kali M. Brandt presented on June 17, 2019

APPROVED:

Major Professor, representing Crop Science

Head of the Department of Crop and Soil Science

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Kali M. Brandt, Author

ACKNOWLEDGEMENTS

I would like to express my gratitude to the members of my committee, Dr. Kelly Vining, Dr. Christopher Mundt, Dr. Patrick Hayes, and Dr. Russell Ingham for their invaluable guidance and help in writing this dissertation. I would also like to thank Dr. Robert Zemetra for giving me the opportunity to come to Corvallis, to work on diverse and amazing projects, and for giving me the advice and encouragement needed to see this through. I must also thank Hilary Gunn for her unending patience, knowledge, and hard work, without which I would not have finished.

I would also like to acknowledge and thank everyone who helped plant, harvest, and process my extensive thesis plots across Oregon. Mark Larson, Tyler Harran, Christy McCarthy, Susi Trittinger, Scott Liu, David Cobertera, Colette Richter, Nathalia Moretti, and all the undergraduate workers over the years, your help was greatly appreciated.

To my husband, Javier Tabima, I say with absolute sincerity that I could not have done this without you. Your emotional and intellectual support was invaluable, and my (mostly) intact sanity thanks you. You are amazing.

To all of my friends and family that supported me and made me feel at home no matter where I was over the last years, thank you.

To Ruth Brandt and Gary Miller, who always made me laugh and helped me look at the world through an empathetic lens, thank you.

To Mom and Dad, who never made me feel anything other than capable of and responsible for achieving my dreams, thank you. I knew I could do it because you knew I could do it. And you are never wrong.

I love you all.

CONTRIBUTION OF AUTHORS

Dr. Robert Zemetra assisted with the design of the studies for Chapters 3, 4, 5, and 6, and with the writing of this dissertation. Adam Heesacker assisted in the design of Chapters 3 and 4. Mark Larson and Tyler Harran assisted in field planting and maintenance for Chapter 3. Dr. Xianming Chen assisted in the field planting and phenotyping for Chapter 3. Dr. Javier F. Tabima assisted in the computational analyses for Chapters 3 and 4 and with the writing of this dissertation. Hilary Gunn, Nathalia Moretti, and Alexander Baasch assisted in the greenhouse work for Chapter 4. Hilary Gunn and Nathalia Moretti assisted in the design and implementation of the studies for Chapters 5 and 6.

TABLE OF CONTENTS

	<u>Page</u>
1 Introduction.....	1
1.1 References.....	8
2 Literature Review.....	12
2.1 Introduction	13
2.2 Growth Habits of Wheat and the Effects of Climate Change.....	13
2.3 Stripe Rust	19
2.4 Breeding and Marker Assisted Selection	23
2.5 Early Markers and First Generation Sequencing	25
2.6 Next Generation Sequencing	27
2.7 SNP Discovery Using Genotyping by Sequencing	30
2.8 SNP Marker Discovery Using Marker-Trait Association Analysis	34
2.9 Genetic Transformation	36
2.10 References	43
3 QTL Analysis of Adult Plant Resistance to Stripe Rust in a Winter Wheat Recombinant Inbred Population	55
3.1 Abstract	56
3.2 Introduction	56
3.3 Materials and Methods	60
3.3.1 Germplasm.....	60
3.3.2 Experimental Design and Phentotyping.....	60
3.3.3 DNA Extraction and Genotyping by Sequencing.....	61
3.3.4 Statistical Analyses.....	62

TABLE OF CONTENTS (Continued)

	<u>Page</u>
3.3.5 Genome Wide Association Study and Candidate Gene Analysis.....	62
3.4 Results	63
3.4.1 Stripe Rust Severity Scores.....	63
3.4.2 SNP Calling.....	64
3.4.3 Statistical Analyses.....	64
3.4.4 Genome Wide Association Study.....	66
3.4.5 Candidate Gene Analysis.....	70
3.5 Discussion	79
3.6 Conclusions	84
3.7 Acknowledgements	84
3.8 References	85
4 Genome-Wide Analysis of Vernalization Requirements in a Winter Wheat Recombinant Inbred Population	88
4.1 Abstract	89
4.2 Introduction	90
4.3 Materials and Methods	93
4.3.1 Germplasm.....	93
4.3.2 Experimental Design and Phenotyping.....	94
4.3.3 DNA Extraction and Genotyping by Sequencing.....	95
4.3.4 Statistical Analyses.....	95

TABLE OF CONTENTS (Continued)

	<u>Page</u>
4.3.5 Genome Wide Association Study.....	96
4.4 Results	96
4.4.1 Heading Date.....	96
4.4.2 SNP Calling.....	98
4.4.3 Statistical Analyses.....	98
4.4.4 Genome Wide Association Study.....	99
4.4.5 Candidate Gene Analysis.....	105
4.5 Discussion	107
4.6 Conclusions	111
4.7 Acknowledgements	112
4.8 References	113
5 A Streamlined Protocol for Wheat (<i>Triticum aestivum</i>) Protoplast Isolation and Transformation with CRISPR Ribonucleoprotein Complexes	118
5.1 Abstract.....	119
5.2 Introduction	119
5.3 Materials.....	123
5.4 Stepwise Procedures	126
5.4.1 <i>In vitro</i> Transcription of sgRNA.....	126
5.4.2 <i>In vitro</i> Cleavage Test Using CRISPR/Cas9 RNPs.....	126
5.4.3 Isolation of Protoplasts.....	127
5.4.4 PEG-Mediated Transformation.....	128

TABLE OF CONTENTS (Continued)

	<u>Page</u>
5.4.5 Extraction of DNA, T7EI Digestion, and Detection of Mutations.....	130
5.5 Results and Discussion	132
5.5.1 Protoplast Isolation.....	132
5.5.2 PEG-mediated Transformation.....	133
5.5.3 Extraction of DNA and T7EI Digestion.....	135
5.5.4 Detection of Mutations.....	136
5.6 Conclusions.....	137
5.7 Acknowledgements	138
5.8 References.....	139
6 Testing Non-Transgenic CRISPR Technology for Wheat Improvement.....	141
6.1 Abstract.....	142
6.2 Introduction	142
6.3 Materials	144
6.3.1 Germplasm.....	144
6.3.2 Genes.....	144
6.4 Methods	144
6.4.1 sgRNA Design.....	144
6.4.2 <i>In vitro</i> Transcription of sgRNA	145
6.4.3 <i>In vitro</i> Cleavage Test Using CRISPR/Cas9 RNPs.....	145
6.4.4 <i>In vivo</i> Cleavage Test Using PEG-Mediated Transformation of Protoplasts with RNPs.....	146

TABLE OF CONTENTS (Continued)

	<u>Page</u>
6.4.5 Tissue Culture	146
6.4.6 Confirming the Efficacy of the Biolistic Delivery System.....	147
6.4.7 Biolistic Bombardment Using RNPs	148
6.4.8 Mutation Detection	148
6.5 Results.....	150
6.5.1 Confirming the Efficacy of the Biolistic Delivery System.....	150
6.5.2 Plantlet Regeneration after Bombardment	151
6.5.3 Sequencing of T ₀ Plantlets in Culture Vessels	151
6.5.4 Sequencing of Individual Tillers from T ₀ Plantlets Transplanted to Soil Pots.....	152
6.6 Discussion and Conclusions	155
6.7 Acknowledgements	161
6.8 References	162
7 Conclusions	165
7.1 References	169
Appendices	171
Appendix A	172
Appendix B	177

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1 Putative winter wheat flowering pathway.....	16
2.2 Putative spring wheat flowering pathway.....	17
2.3 Wheat flag leaves with different levels of stripe rust infection.....	20
2.4 Stripe rust uredia on the heads of severely infected winter wheat plants.....	20
2.5 Genotyping by Sequencing (GBS) workflow.....	32
2.6 The CRISPR bacterial immunity system.....	39
2.7 Modes of action of the CRISPR/Cas9 system.....	41
3.1 Histograms of the number of Taxa scored for each stripe rust severity in each Treatment.....	65
3.2 Skiles x Goetze RILs at Hyslop Research Farm, Corvallis, Oregon in June 2018.....	65
3.3 Principal component analysis representing the genetic diversity of the RIL population used in this study.....	67
3.4 Manhattan plots for each treatment produced by GAPIT.....	68
3.5 Heatmap of linkage disequilibrium (LD) for significant SNPs.....	71
3.6 Location of significant SNPs and their subgroups on the respective chromosomes as determined by the reference genome.....	73
3.7 Chromosome 3B markers and their location on either the GrainGenes linkage map or the reference genome.....	76
3.8 Chromosome 3D markers and their location on either the GrainGenes linkage map or the reference genome.....	78
4.1 Average heading date counts for each Taxa in each treatment.....	97
4.2 Scatterplot of principal component (PC) analysis for data from GBS	99
4.3 Manhattan plots for each treatment produced by GAPIT.....	101

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
4.4 Heatmap of linkage disequilibrium (LD) for significant SNPs.....	103
4.5 Genomic position of significant SNPs and known vernalization and photoperiod genes on the respective wheat chromosomes as determined by the reference genome assembly.....	106
5.1 Overview of the protoplast isolation and transformation protocol.....	122
5.2 Protoplasts stained with Evans Blue dye on a hemacytometer.....	133
5.3 GFP fluorescing protoplast cells.....	134
5.4 A 2% agarose gel with T7EI digested PCR and Negative Controls.....	135
5.5 ICE analysis of the chromatogram sequences.....	137
6.1 Overview of the CRISPR/Cas9 RNP transformation procedure.....	149
6.2 Fluorescent light and green filter image of an immature wheat embryo 24 hours after bombardment with GFP-coated gold particles.....	150
6.3 Callus tissue formed from RNP bombarded embryos showing tight clusters of leaves and leaf tissue.....	151
6.4 Chromatogram sequences from three plantlets grown from RNP bombarded embryos and a negative control plantlet.....	153
6.5 Relative contributions of inferred sequences present in each plantlet PCR sample, as reported by ICE.....	154
6.6 Plantlet 2 taken from a culture vessel in the growth chamber and planted in a one gallon pot in the greenhouse.....	155
6.7 Relative contribution of inferred sequences present in each Tiller's PCR sample from Plantlet 2, as reported by ICE.....	156

LIST OF TABLES

<u>Table</u>	<u>Page</u>
2.1 Examples of the most common molecular marker types with descriptions and examples.....	26
2.2 Most common next Generation Sequencing (NGS) technologies.....	29
3.1 Significantly associated SNPs with stripe rust resistance identified in the Manhattan plots.....	69
3.2 The groupings of significant SNPs on each chromosome as determined by their Linkage Disequilibrium values.....	69
3.3 Significant SNPs from all treatments and their allelic state in the most resistant Taxa and the most susceptible Taxa.....	72
4.1 SNPs significantly associated with heading date variability identified in the Manhattan plots.....	100
4.2 The groupings of significant SNPs on each chromosome as determined by their Linkage Disequilibrium values.....	102
4.3 Taxa with all allelic states for early heading times and for late heading times.....	104
4.4 Taxa with heading times that were the earliest, latest, and most similar to Goetze.....	105
5.1 Sequence of oligonucleotides (sgRNA and primers) used to transcribe the sgRNA and amplify the <i>GW2</i> gene.....	124
5.2 Ingredients for the preparation of Enzyme Solution.....	124
5.3 Ingredients for the preparation of W5 Solution.....	125
5.4 Ingredients for the preparation of W5A Solution.....	125
5.5 Ingredients for the preparation of WI Solution.....	125
5.6 Ingredients for the preparation of PEG Solution.....	126
5.7 PCR reaction for amplifying the desired genomic region.....	130
5.8 PCR reaction parameters.....	131

LIST OF TABLES (Continued)

<u>Table</u>	<u>Page</u>
5.9 Reaction mixture for the formation of heteroduplexes.....	131
5.10 Reaction mixture for the cleavage of heteroduplexes using T7EI endonuclease.....	131
5.11 Ideal ratios of amount of proteins to number of protoplasts, and volume of proteins to volume of protoplasts for GFP control and for CRISPR RNPs...	134
6.1 Sequence of oligonucleotides (sgRNA and primers) used to target the <i>GW2</i> gene.....	145

LIST OF APPENDICES

<u>Appendix</u>	<u>Page</u>
A. QTL Analysis of Adult Plant Resistance to Stripe Rust in a Winter Wheat Recombinant Inbred Population.....	172
B. Genome-Wide Analysis of Vernalization Requirements in a Winter Wheat Recombinant Inbred Population.....	177

LIST OF APPENDIX TABLES

<u>Table</u>	<u>Page</u>
A.1 ANOVA of disease severity as a function of Treatment, Taxa, and Treatment by Taxa Interaction.....	173
A.2 Tukey's HSD (Honest Significant Difference) Test of all Treatments compared to one another.....	173
A.3 Linkage Disequilibrium (LD) analysis of all significant SNPs	173
B.1 ANOVA of Heading Date as a function of Treatment, Taxa, and Treatment by Taxa Interaction.....	178
B.2 Tukey's HSD (Honest Significant Difference) Test of all Treatments compared to one another.....	178
B.3 Linkage Disequilibrium (LD) analysis of all significant SNPs	178

Chapter 1

Introduction

The first plants were domesticated by humans around 10,000 years ago, with this movement from foraging to cultivation likely beginning with the most basic form of plant breeding still practiced today: phenotypic selection. Plants exhibiting certain desirable traits were chosen above other less desirable plants for consumption and re-planting, thus slowly driving the subsequent generations towards the crops we know today[1][2]. It wasn't until the 17th and 18th centuries that botanists began experimenting with artificial pollination and limitations of inter-crossing species. An explosion of targeted breeding began after Darwin's Theory of Evolution took hold and Mendel's laws of heredity were re-discovered in the early 1900s[3][4]. These plant breeding efforts using phenotypic selection increased yields and expanded crops into new areas throughout the first half of the 20th century.

In the 1950s and 1960s agriculture was again transformed by what is called the Green Revolution, which brought about widespread implementation of pesticide use, modern irrigation, and synthetic nitrogen fertilizers as well as high yielding varieties, the most well-known being semi-dwarf grains[5]. Worldwide total factor productivity for agriculture increased by almost a percent from 1970 to 1989, with yield gains for crops like wheat, rice, and maize surpassing 100 percent[5]. In developing countries between 1960 and 1990, Green Revolution technologies increased food supplies by up to 13 percent, preventing food and feed prices from rising up to 65 percent and requiring the cultivation of an additional 62 million acres of land[5][6]. Increase in yield for most major crops has remained fairly linear since 1965, but many areas have begun reaching yield plateaus. Rice, wheat, and maize are the top three food crops in the world, but as of the year 2010, only 12 percent of the production was seeing increased yield, 27 percent was constant, 21 percent had reached its upper yield plateau, and 10 percent was decreasing[7]. The reasons for this plateau and decline are many and can go beyond the crop reaching its biophysical yield ceiling in a particular region. Cyclic and unpredictable weather patterns, land degradation and poor soils, reduction in use of pesticides and fertilizers, and poor investment in research and development can all contribute to stagnating and decreasing yields[8]. Grassini *et al.* (2013) argue that there is no support

for future exponential yield growth in agriculture due to the fact that absolute growth rates are linear, no cases of consistently increasing or steady relative yields are known, and biophysical yield ceilings have already been reached in multiple cases[7]. There is also the problem of global climate change to take into consideration when discussing yield plateaus. An overall increase in temperature, increased occurrence of extreme temperatures, increased drought, and shifts in weed and pest distributions all contribute to the stagnation and sometimes regression of crop yield gains[9][10].

Wheat is one of the three top staple food crops worldwide, providing nearly 25 percent of the calories consumed daily in 89 countries[11][12]. Almost \$50 billion-worth of wheat is traded annually across the globe[13]. In the Pacific Northwest of the United States (PNW) alone, wheat is planted on approximately 4.1 million acres and is a \$1.2 billion per year industry[14]-[16]. Much of the grain produced in this region is shipped overseas as a high value commodity crop, and a disruption in field production would be devastating. Wheat, like most other crops, has been bred over the last century to be extremely well adapted to specific growing regions[17][18]. The duration of cold winter temperatures, the lowest tolerable winter temperatures, planting date, pest and disease resistances, amount of rain, timing of rain, onset of summer heat, tolerance to certain soil types, harvest practices, and seed storage methods are all finely tuned within regions and varieties, and all are predicted to be affected by climate change[19][20]. In the PNW, the average temperature is predicted to increase by as much as 5°C by 2090, reducing snow cover during winter that protects from extreme cold damage, changing amount and timing of rain that affects planting, increasing the overall winter temperature that is needed for winter wheat to vernalize and prepare for growth in the spring, increasing the diversity of diseases and pests that can cause devastating yield losses in fields without resistance and increasing the amount of pesticide use, and increasing the overall spring and summer temperatures that in turn shortens the grain-filling period[21]-[23]. While wheat breeders attempt to mitigate these scenarios, they are simultaneously tasked with increasing yields year after year.

The world's population reached seven billion in 2011 and is expected to increase exponentially to nine billion by 2040, and ten billion by 2055[24]. The United Nations

(UN) has reported that to keep pace with this population growth, the current rate of food production will need to double by the year 2050, while developing countries will begin to experience severe food shortages by 2030 if this is not met[25]. Wheat production specifically will need to increase 60 percent by 2050 to keep pace with current consumption[13]. Coupled with climate change and impending yield plateaus, wheat breeding must once again utilize new knowledge, techniques, and technologies in order to keep pace with these growing demands.

Over the last 40 years, scientists and breeders have been working towards one such solution to overly long breeding times and difficult phenotypic selections called marker assisted selection (MAS). Markers are DNA sequences that are associated with certain traits, usually by a close genetic distance[26]. MAS is the use of those markers to screen plants for the presence or absence of a trait without the need for phenotype testing. This can save space, time, and resources during the development of a new variety, and MAS can cut total breeding time in as much as half[27][28]. For wheat, a new variety takes an average of ten years of extensive field testing before it is released[29]. Being able to significantly reduce this time would allow breeders to better respond to new challenges and help maintain the vitally important wheat industry.

The development of new markers has historically been slow and difficult, as genomes were largely uncharacterized[30][31]. This all changed, however, in the early 2000s with the advent of what is called Next Generation Sequencing (NGS). NGS involves the sequencing of DNA using platforms that can produce millions of 25 to 400 base pair (bp) length reads[32][33]. Soon after the first generation of sequencing technologies were released, others were developed that also utilize high-throughput sequencing but produce extremely long reads of over 100,000bp[34]. Using these technologies in combination with other methods, thousands of genomes were sequenced and millions of sequences from a myriad of organisms and species were made available for public use[35]-[37]. As a result of this explosion of availability and access to sequence data, marker discovery also became much easier and more commonplace. The financial cost of performing NGS for marker discovery has also decreased exponentially, to the point where it is a relatively affordable option[32][38]. NGS has also opened

marker discovery up to the ability to detect associations with genes and genic regions that may not be the major contributor to a trait and thus previously undetected, or genes that were previously unknown[32].

The affordability of the technology also means that these discoveries can be made in adapted germplasm, which is important in crops like wheat that are highly adapted to specific regions and whose phenotypes are determined by unique genes and/or alleles[30]. This opens up an important aspect of modern MAS in which breeders need markers tightly linked to important traits for their particular region. The PNW is a site of constant and heavy stripe rust (*Puccinia striiformis*) infections, and millions of dollars are spent by farmers each year to prevent significant yield loss[39][40]. The best option for reducing pesticide use and cost to farmers is to incorporate multiple stripe rust resistance genes into the varieties grown in this area[41]. A large number of the stripe rust genes in wheat are unknown in PNW germplasm, uncharacterized, and/or often have associated markers that are extremely distant from the gene itself. Breeders in the region would be better able to develop resistant varieties if they had access to high quality markers for genes known to contribute to resistance in their region.

The PNW is also already experiencing unpredictable winter temperatures which threaten to delay wheat flowering due to insufficient vernalization (a sustained period of cold temperatures needed for winter wheat growth)[19][21]. The above-average predicted increase in temperatures in the near future for the most important wheat growing parts of this region make insufficient vernalization of winter wheat a serious threat[22]. In order to overcome this issue, new varieties must have the ability to flower in an acceptable time frame after various amounts of vernalization, rather than rely on a full six weeks of cold temperatures[42][43]. These ‘facultative’ varieties are extremely difficult to test for using phenotypic selection in the field, and as with stripe rust, the genes contributing to this trait are often unknown in PNW germplasm, uncharacterized, and/or have distantly associated markers that are not reliable[44]. In order to successfully and quickly breed for facultative varieties, breeders need access to high quality markers tightly associated with this trait.

MAS is an important complement to traditional breeding methods, and marker-trait discovery using NGS is becoming increasingly common in breeding programs worldwide[28]. A new targeted transformation technology has emerged in the last five years, however, that is still in the early stages and has not yet been used in the development of a released variety, but is also expected to become a significant part of plant breeding in the near future[45][46]. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) transformation uses a sequence specific RNA guide determined by the researcher to target a specific genomic location and create an edit in the DNA sequence. The only requirements are that the targeted sequence is unique in the genome and that the site is located next to a particular three-nucleotide sequence[47][46]. The first use of CRISPR in eukaryotes was in 2013, but within the last six years alone it has been used to transform dozens of crop plants to produce a multitude of traits[48]-[50].

A number of countries around the globe have recently decided to allow CRISPR-transformed crops to be produced without extra oversight and regulation, and the first varieties are expected to be released in the near future[51][52]. Transforming wheat with CRISPR would allow for a large range of traits to be incorporated into wheat varieties that are at present infeasible to incorporate or obtain, either because they do not exist or because the linkage drag would be too great. CRISPR transformation could also allow for the rapid integration of new traits into regionally adapted germplasm, thus significantly reducing the breeding time of new varieties[53][54]. The potential for CRISPR transformation in wheat breeding is substantial, but the wheat industry and consumer positions on transgenically modified wheat is an important consideration. A likely caveat to the use of CRISPR in wheat is that the entire process be completely free of transgenes at every stage. This limits the options for delivering the CRISPR construct into plants, and so far only one research group has reported success with non-transgenic CRISPR transformation of wheat[55][56]. Therefore, the feasibility of using this technique in a wheat breeding program needs to be tested and evaluated.

Wheat breeding is an extremely important aspect of the critically important worldwide wheat industry. Wheat breeders have been tasked with increasing yields in

the face of yield plateaus and climate change causing major setbacks to productivity. Consequently, successful modern wheat breeding will likely need to encompass classical phenotypic selection, marker assisted selection, and targeted transformation. Therefore, the objectives of the proposed research are to address the needs for decreased breeding times and integration of new traits in PNW winter wheat using marker-trait discovery and targeted transformation.

The objectives of this study were:

- 1) To use next generation sequencing for marker-trait discovery of stripe rust resistance genes in winter wheat in order to identify useful markers and quantitative trait loci for future gene characterization and breeding of resistant varieties.
- 2) To use next generation sequencing for marker-trait discovery of reduced vernalization requirements in winter wheat in order to identify useful markers and quantitative trait loci for future gene characterization and breeding of facultative varieties.
- 3) To develop a protocol for testing CRISPR transformation *in vivo* using a wheat protoplast system in order to test the efficiency of gene targets before transformation.
- 4) To transform immature wheat embryos using non-transgenic CRISPR technology in order to determine its viability for use in a breeding program and address potential pitfalls and limitations.

1.1 References

- [1] Z. Peleg, T. Fahima, A. B. Korol, S. Abbo, and Y. Saranga, “Genetic analysis of wheat domestication and evolution under domestication,” *J Exp Bot*, vol. 62, no. 14, pp. 5051–5061, Oct. 2011.
- [2] P. Gustafson, O. Raskina, X. Ma, and E. Nevo, “Wheat Evolution, Domestication, and Improvement,” in *Wheat Science and Trade*, B. F. Carver, Ed. Oxford, UK: Wiley-Blackwell, 2009, pp. 3–30.
- [3] A. R. Hallauer, “History, Contribution, and Future of Quantitative Genetics in Plant Breeding: Lessons From Maize,” *Crop Science*, vol. 47, no. Supplement_3, p. S-4-S-19, Dec. 2007.
- [4] P. Gepts, “A Comparison between Crop Domestication, Classical Plant Breeding, and Genetic Engineering,” *Crop Science*, vol. 42, no. 6, pp. 1780–1790, Nov. 2002.
- [5] P. L. Pingali, “Green Revolution: Impacts, limits, and the path ahead,” *Proc Natl Acad Sci U S A*, vol. 109, no. 31, pp. 12302–12308, Jul. 2012.
- [6] R. E. Evenson and D. Gollin, “Assessing the Impact of the Green Revolution, 1960 to 2000,” *Science*, vol. 300, no. 5620, pp. 758–762, May 2003.
- [7] P. Grassini, K. M. Eskridge, and K. G. Cassman, “Distinguishing between yield advances and yield plateaus in historical crop production trends,” *Nature Communications*, vol. 4, p. 2918, Dec. 2013.
- [8] M. J. Kropff, K. G. Cassman, F. W. T. P. de Vries, and H. H. van Laar, “Increasing the Yield Plateau in Rice and the Role of Global Climate Change,” *J. Agric. Meteorol.*, vol. 48, no. 5, pp. 795–798, 1993.
- [9] D. B. Lobell, W. Schlenker, and J. Costa-Roberts, “Climate Trends and Global Crop Production Since 1980,” *Science*, vol. 333, no. 6042, pp. 616–620, Jul. 2011.
- [10] J. Hatfield *et al.*, “Ch. 6: Agriculture. Climate Change Impacts in the United States: The Third National Climate Assessment”, J. M. Melillo, T.C. Richmond, G. W. Yohe, Eds., U.S. Global Change Research Program, 2014, 150-174.
- [11] A. Kamran, M. Iqbal, and D. Spaner, “Flowering time in wheat (*Triticum aestivum* L.): a key factor for global adaptability,” *Euphytica*, vol. 197, no. 1, pp. 1–26, May 2014.
- [12] W. Bushuk, “Wheat breeding for end-product use,” in *Braun HJ., Altay F., Kronstad W.E., Beniwal S.P.S., McNab A. (eds) Wheat: Prospects for global improvement*, Ankara, Turkey, 1997, vol. 6, pp. 203–211.
- [13] CIMMYT “Wheat Atlas.” [Online]. Available: <http://wheatatlas.org> . [Accessed: 02-May-2019].
- [14] “USDA/NASS 2016 State Agriculture Overview for Oregon.” [Online]. Available: https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=OREGON. [Accessed: 27-Nov-2017].
- [15] “USDA/NASS 2016 State Agriculture Overview for Idaho.” [Online]. Available: https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=IDAHO. [Accessed: 01-Dec-2017].
- [16] “USDA/NASS 2016 State Agriculture Overview for Washington.” [Online]. Available:

- https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=WASHINGTON. [Accessed: 01-Dec-2017].
- [17] M. S. Lopes *et al.*, “Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change,” *Journal of Experimental Botany*, vol. 66, no. 12, pp. 3477–3486, Jun. 2015.
 - [18] T. P. Khumalo, A. Barnard, L. Maphosa, and T. J. Tsilo, “Impact of Growth Habit and Architecture Genes on Adaptation and Performance of Bread Wheat,” in *Wheat Improvement, Management and Utilization*, R. Wanyera and J. Owuoché, Eds. InTech, 2017.
 - [19] Rosenzweig, “Climate Change and Extreme Weather Events; Implications for Food Production, Plant Diseases, and Pests,” *GLOBAL CHANGE*, vol. 2, no. 2, p. 15, 2001.
 - [20] M. A. Semenov, P. Stratonovitch, F. Alghabari, and M. J. Gooding, “Adapting wheat in Europe for climate change,” *Journal of Cereal Science*, vol. 59, no. 3, pp. 245–256, May 2014.
 - [21] D. Z. Skinner and B. S. Bellinger, “Freezing tolerance of winter wheat as influenced by extended growth at low temperatures and exposure to freeze–thaw cycles,” *Canadian Journal of Plant Science*, vol. 97, no. 2, pp. 250–256, Sep. 2016.
 - [22] E. P. Salathé, R. Steed, C. F. Mass, and P. H. Zahn, “A High-Resolution Climate Model for the U.S. Pacific Northwest: Mesoscale Feedbacks and Local Responses to Climate Change,” *J. Climate*, vol. 21, no. 21, pp. 5708–5726, Nov. 2008.
 - [23] S. Asseng *et al.*, “Rising temperatures reduce global wheat production,” *Nature Climate Change*, vol. 5, no. 2, pp. 143–147, Feb. 2015.
 - [24] D. of E. and S. A. United Nations, “World Population Prospects: The 2017 Revision, Key Findings and Advance Tables,” *DemRes*, vol. 30, p. ESA/P/WP/248, 2017.
 - [25] FAO, “FAOSTAT database,” *Agricultural crops: Wheat: area harvested*, 2014. [Online]. Available: <http://www.fao.org/>. [Accessed: 04-Dec-2017].
 - [26] B. C. Y. Collard, M. Z. Z. Jahufer, J. B. Brouwer, and E. C. K. Pang, “An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts,” *Euphytica*, vol. 142, no. 1–2, pp. 169–196, Jan. 2005.
 - [27] Collard Bertrand C.Y and Mackill David J, “Marker-assisted selection: an approach for precision plant breeding in the twenty-first century,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1491, pp. 557–572, Feb. 2008.
 - [28] P. K. Gupta, P. Langridge, and R. R. Mir, “Marker-assisted wheat breeding: present status and future possibilities,” *Mol Breeding*, vol. 26, no. 2, pp. 145–161, Aug. 2010.
 - [29] B. Wang, D. L. Liu, S. Asseng, I. Macadam, and Q. Yu, “Impact of climate change on wheat flowering time in eastern Australia,” *Agricultural and Forest Meteorology*, vol. 209–210, pp. 11–21, Sep. 2015.
 - [30] R. M. D. Koebner and R. W. Summers, “21st century wheat breeding: plot selection or plate detection?,” *Trends in Biotechnology*, vol. 21, no. 2, pp. 59–63, Feb. 2003.

- [31] J. W. Davey, P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter, "Genome-wide genetic marker discovery and genotyping using next-generation sequencing," *Nature Reviews Genetics*, vol. 12, no. 7, pp. 499–510, Jul. 2011.
- [32] D. Edwards, J. Batley, and R. J. Snowdon, "Accessing complex crop genomes with next-generation sequencing," *Theor Appl Genet*, vol. 126, no. 1, pp. 1–11, Jan. 2013.
- [33] J. Dubcovsky, "Marker-assisted selection in public breeding programs: The wheat experience," *CROP SCIENCE*, vol. 44, p. 5, 2004.
- [34] A. N. Egan, J. Schlueter, and D. M. Spooner, "Applications of next-generation sequencing in plant biology," *American Journal of Botany*, vol. 99, no. 2, pp. 175–185, 2012.
- [35] T. P. Michael and S. Jackson, "The First 50 Plant Genomes," *The Plant Genome*, vol. 6, no. 2, Jul. 2013.
- [36] P. J. Kersey, "Plant genome sequences: past, present, future," *Current Opinion in Plant Biology*, vol. 48, pp. 1–8, Apr. 2019.
- [37] K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers, "GenBank," *Nucleic Acids Res*, vol. 44, no. Database issue, pp. D67–D72, Jan. 2016.
- [38] H. Kuchel *et al.*, "The successful application of a marker-assisted wheat breeding strategy," *Mol Breeding*, vol. 20, no. 4, pp. 295–308, Nov. 2007.
- [39] X. M. Chen, "Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat," *Canadian Journal of Plant Pathology*, vol. 27, no. 3, pp. 314–337, Sep. 2005.
- [40] P. Bulli, J. Zhang, S. Chao, X. Chen, and M. Pumphrey, "Genetic Architecture of Resistance to Stripe Rust in a Global Winter Wheat Germplasm Collection," *G3: Genes, Genomes, Genetics*, vol. 6, no. 8, pp. 2237–2253, Aug. 2016.
- [41] R. S. Ren, M. N. Wang, X. M. Chen, and Z. J. Zhang, "Characterization and molecular mapping of Yr52 for high-temperature adult-plant resistance to stripe rust in spring wheat germplasm PI 183527," *Theor Appl Genet*, vol. 125, no. 5, pp. 847–857, Sep. 2012.
- [42] S. M. Grogan *et al.*, "Allelic Variation in Developmental Genes and Effects on Winter Wheat Heading Date in the U.S. Great Plains," *PLOS ONE*, vol. 11, no. 4, pp. 1–23, Apr. 2016.
- [43] D. Gouache, X. Le Bris, M. Bogard, O. Deudon, C. Pagé, and P. Gate, "Evaluating agronomic adaptation options to increasing heat stress under climate change during wheat grain filling in France," *European Journal of Agronomy*, vol. 39, pp. 62–70, May 2012.
- [44] N. Kippes, M. Guedira, L. Lin, M. A. Alvarez, G. L. Brown-Guedira, and J. Dubcovsky, "Single nucleotide polymorphisms in a regulatory site of VRN-A1 first intron are associated with differences in vernalization requirement in winter wheat," *Mol Genet Genomics*, vol. 293, no. 5, pp. 1231–1243, Oct. 2018.
- [45] D. Jaganathan, K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman, "CRISPR for Crop Improvement: An Update Review," *Front Plant Sci*, vol. 9, no. 985, Jul. 2018.

- [46] L. Arora and A. Narula, "Gene Editing and Crop Improvement Using CRISPR-Cas9 System," *Front. Plant Sci.*, vol. 8, no. 1932, Nov. 2017.
- [47] A. Malzahn, L. Lowder, and Y. Qi, "Plant genome editing with TALEN and CRISPR," *Cell & Bioscience*, vol. 7, no. 1, p. 21, Apr. 2017.
- [48] Q. Shan, Y. Wang, J. Li, and C. Gao, "Genome editing in rice and wheat using the CRISPR/Cas system," *Nature Protocols*, vol. 9, no. 10, pp. 2395–2410, Sep. 2014.
- [49] Y. Osakabe *et al.*, "CRISPR–Cas9-mediated genome editing in apple and grapevine," *Nature Protocols*, p. 1, Nov. 2018.
- [50] J. W. Woo *et al.*, "DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins," *Nature Biotechnology*, vol. 33, no. 11, pp. 1162–1164, Oct. 2015.
- [51] D. Normile, Mar. 19, 2019, and 1:15 Pm, "Gene-edited foods are safe, Japanese panel concludes," *Science | AAAS*, 19-Mar-2019. [Online]. Available: <https://www.sciencemag.org/news/2019/03/gene-edited-foods-are-safe-japanese-panel-concludes>. [Accessed: 02-May-2019].
- [52] A. Salleh, "Gene editing of plants and animals gets green light. Is CRISPR bacon next?," *ABC News*, 30-Apr-2019. [Online]. Available: <https://www.abc.net.au/news/science/2019-04-30/crispr-gene-editing-in-the-food-chain/11053622>. [Accessed: 02-May-2019].
- [53] S. Mondal *et al.*, "Harnessing Diversity in Wheat to Enhance Grain Yield, Climate Resilience, Disease and Insect Pest Resistance and Nutrition Through Conventional and Modern Breeding Approaches," *Front Plant Sci*, vol. 7, Jul. 2016.
- [54] W. Wang *et al.*, "Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat," *The CRISPR Journal*, vol. 1, no. 1, pp. 65–74, Feb. 2018.
- [55] Z. Liang *et al.*, "Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes," *Nat Commun*, vol. 8, no. 14261, pp. 1–5, Jan. 2017.
- [56] Z. Liang *et al.*, "Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 *in vitro* transcripts or ribonucleoproteins," *Nature Protocols*, vol. 13, no. 3, pp. 413–430, Mar. 2018.

Chapter 2

Literature Review

2.1 Introduction

The improvement of crop performance via phenotypic selection has been practiced by farmers for thousands of years, and the targeted improvement of crops for specific traits by breeders has been practiced for over 100 years[1]. It is only within the past 40 years that the ability of breeders to select for specific genes using molecular techniques has been available. Within this short time, however, incredible advancements have been made and the technology and techniques have improved by leaps and bounds to the point where breeders are able to test for numerous genes at early breeding stages rather than wait for phenotypic confirmation, as well as directly manipulate genes to achieve traits that were previously infeasible.

The techniques and technologies behind this precipitous change in crop breeding are numerous and constantly changing. Yet it is important for breeders and scientists to understand the concepts, advantages, and limitations in order to utilize them to the full potential. The aim of this literature review is four-fold: first, to outline the growth habits of wheat and the genes that control it; second, to describe stripe rust, the most important disease affecting wheat in the Pacific Northwest of the United States; third, to explore the foundation of molecular breeding techniques using genetic markers and their current use in crop improvement; and fourth, to describe the methods that could be used for the targeted transformation of wheat.

2.2 Growth Habits of Wheat and the Effects of Climate Change

Wheat (*Triticum aestivum*) was domesticated between 8,000 and 10,000 years ago in the Fertile Crescent of what is now the Middle East. Since then it has spread around the world to become one of the most vital food crops providing almost a quarter of the daily calories and protein consumed by people on every continent except Antarctica[2]–[5]. This means wheat is grown in sub-tropical to temperate steppe climate zones, in both rain-fed and dryland conditions, covering more land area and regional climate zones than any other crop[6]–[8]. The main reason for wheat's success in these vastly different environments is its flowering time adaptability[9][10]. Flowering time is determined mainly by genes in the flowering pathway that are sensitive to either day length, called

photoperiod genes, or temperature, called vernalization genes[5][8]. These genes work together to up-regulate or down-regulate one another in order to prevent flowering during cold winters, promote flowering during spring, avoid damaging summer heat, and to take advantage of cool early summers.

The genes that contribute the most to the flowering pathway are the vernalization genes VRN1, VRN2, VRN3, and VRN4; the long-day photoperiod sensitivity gene PPD1; and the short-day sensitivity gene TaFT3 recently characterized by Halliwell *et al.* (2016)[9] and Zikhali *et al.* (2017)[11]. Modern bread wheat is an allohexaploid, meaning there are three copies of most genes on each chromosome (A, B, and D). Homoeologous genes can often express at different levels and can also often fully or partially compensate for the loss of expression in one or two other copies[12][13]. This means that while there are six primary genes in the flowering pathway, there are sixteen copies of those genes, and multiple allelic variations within those copies that all contribute to a wide variety of phenotypes.

VRN1 is a flowering promoter located on all three genomes of chromosome 5 that is up-regulated by the cold temperatures of vernalization. It is also the most well-characterized of the flowering pathway genes[14]. Allelic variation in VRN1 is found on all three genomes and is widely dispersed around the world[15][16]. VRN2 is a flowering repressor that is down-regulated by VRN1 during vernalization and is located on chromosomes 5A, 4B, and 4D[11]. VRN2 does not contain as much allelic diversity as VRN1, although there is some[17][18]. VRN3 integrates the vernalization and photoperiod pathways by up-regulating VRN1 during both vernalization and long-day photoperiods[11][19]. It is located on all three genomes of chromosome 7, but to date allelic variation has only been found on the B genome[10]. VRN4 is thought to have arisen from a retrotransposon insertion that duplicated the A genome copy of VRN1 onto the short arm of chromosome 5D. It has been shown to be up-regulated during vernalization, which in turn down-regulates VRN2 and up-regulates VRN1[18][20][21]. PPD1 up-regulates VRN3 during long-day photoperiods, which thereby promotes flowering. It is located on all three genomes of chromosome 2 and has

been found to contain some allelic diversity[9][11]. TaFT3 is orthologous to the barley (*Hordeum vulgare*) short-day sensitivity gene PPD-H2[22]. It is located on all three genomes of chromosome 1 and has been found to promote flowering during short-days, but where it acts on the flowering pathway and whether there is allelic diversity is currently not known[9][11].

The interactions of these genes within the flowering pathway for the two main growth habits of wheat, winter and spring, are shown in Figures 2.1 and 2.2, respectively. Winter wheat is planted in the fall, germinates, and then requires at least six to eight weeks of temperatures below 8°C (a process called vernalization), before reproductive growth begins in the spring[21]. Winter wheat generally has the following genotype: VRN1 – all functional wild type (WT) recessive alleles[15][16][23], VRN2 – all functional WT dominant alleles[18][19], VRN3 – recessive functional alleles that vary at Vrn-B3 across the world[6][11], Vrn-D4 – presence of gene and alleles varies across the world[20][21], PPD1 – either functional WT recessive alleles or overexpression day neutral alleles[24], and TaFT3 – most likely WT dominant functional alleles[9].

Spring wheat is planted in the spring, after the threat of frost has passed, and grows steadily without the need for vernalization[6]. Spring wheat generally has the following genotype: VRN1 – at least one dominant mutant allele[15][16][23], VRN2 – all recessive null (knock-out/non-functional) alleles[18][19], VRN3 – recessive functional alleles with the possibility of a dominant allele at Vrn-B3[6][11][25], Vrn-D4 – presence of gene and alleles varies across the world[20][21], PPD1 - either functional WT recessive alleles or overexpression day neutral alleles[24], and TaFT3 - most likely WT dominant functional alleles[9].

Genes in the flowering pathway interact with each other in a number of ways to affect flowering date and thereby adaptability to local climates[9][10]. As mentioned previously, the manipulation of the genes, loci, and alleles in the flowering pathway results in a plethora of flowering times that has allowed wheat to become adapted to a multitude of environments by ensuring the plants do not enter floral initiation when temperatures are too hot or too cold for grain production[16][26][27]. This means that while wheat as a species is well adapted to diverse environments, within each regional

environment the locally-adapted cultivars are extremely affected by even slight changes in temperature[28]–[30].

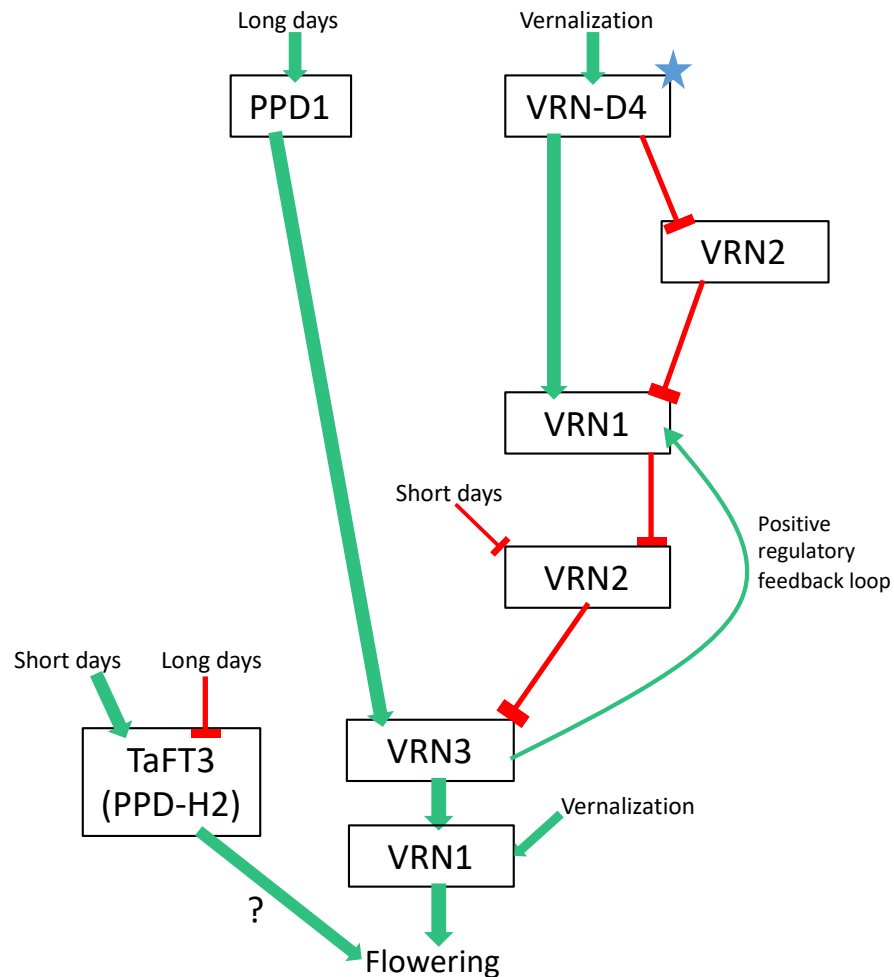


Figure 2.1 Putative winter wheat flowering pathway. The putative flowering pathway of winter wheat is shown with all known and potentially involved genes. Green arrows indicate up-regulation of genes, and red lines indicate down-regulation. TaFT3 and PPD1 are part of the photoperiod pathway and are regulated by day length, while all the VRN genes are part of the vernalization pathway and are regulated by temperature. The ? means it is not yet known where TaFT3 acts on the flowering pathway, but it is known to promote flowering under short days. The blue star indicates that Vrn-D4 has not been characterized in worldwide germplasm. If Vrn-D4 is not present, vernalization will act directly on VRN1.

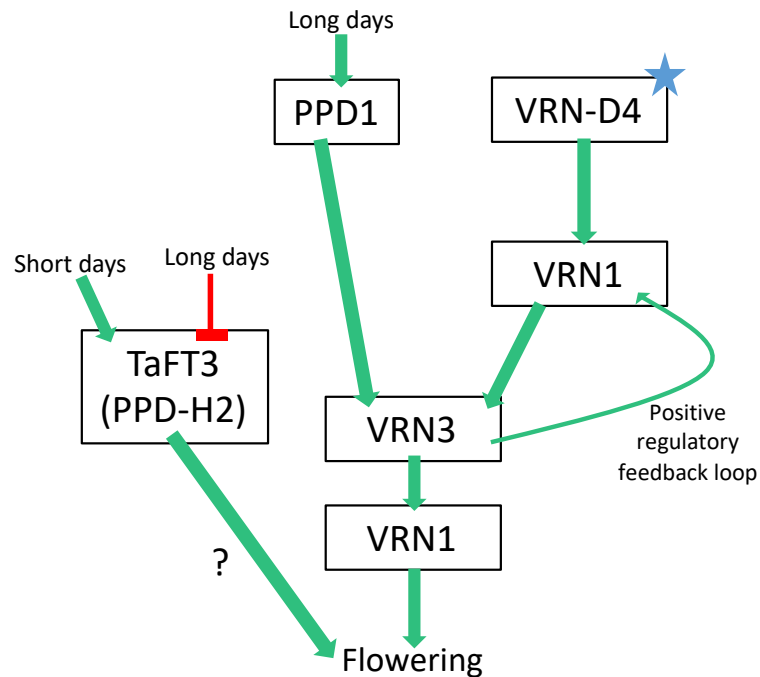


Figure 2.2 Putative spring wheat flowering pathway. The putative flowering pathway of spring wheat is shown with all known and potentially involved genes. Green arrows indicate up-regulation of genes, and red lines indicate down-regulation. Unlike winter wheat, spring wheat does not have functional VRN2 genes and is therefore not regulated by winter temperatures and begins reproductive growth as soon as the photoperiod pathway is activated. The ? means it is not yet known where TaFT3 acts on the flowering pathway, but it is known to promote flowering under short days. The blue star indicates that Vrn-D4 has not been characterized in worldwide germplasm.

Climate change has already altered weather patterns, and consequently yields, in virtually all of the wheat-growing regions of the world, and is only predicted to become worse in the future[27][29]–[32]. Lobell and Field (2007)[31] estimated that warming since 1981 has already resulted in losses of 40Mt or \$5 billion per year in maize, barley, and wheat combined. Wheat growers across the globe can expect overall warmer winters, warmer summers, and more unpredictable extremes in temperatures year-round[33][34]. The Pacific Northwest (PNW) has already seen a reduction in the number of frost days as well as increased winterkill of wheat due to insufficient snow cover and late frosts[33][35]. Climate models predict that the PNW will see annual temperature increases of 1.1°C by the 2020s and 1.8°C by the 2040s, with regional differences having the potential to be even greater[36]. Salathé *et al.* (2008)[37] reported that the flanks of

mountain ranges in Oregon and Washington and the high plains of Eastern/Southern Oregon, Eastern Washington, and Southern Idaho will see average winter temperatures more than 5°C higher by 2090 than at the end of the 20th century. This will also lead to a significant decrease in snow cover in these areas that typically protects winter wheat from damage due to extreme cold[33][37]. A reduction in the number of cumulative vernalization days during winter can delay wheat flowering by up to two weeks, which also increases the chance of yield loss due to exposure to hot, dry conditions during sensitive reproductive stages[26][27][32]. Asseng *et al.* (2015)[38] reported that for each degree Celsius increase in temperature there is a six percent decrease in wheat production, which translates to a loss of approximately one quarter of global wheat trade. Consequently, the increased winter temperatures of the near future pose an extremely alarming situation for wheat growers, as the majority of varieties in this area are dependent upon sustained, low winter temperatures, and a disruption of this cycle can cause complete crop stand losses and puts the entire PNW wheat industry at risk.

As one of the main agricultural exports of the Pacific Northwest, soft white winter wheat from the area is shipped all over the world. In 2016 alone, wheat brought in sales of over \$1.2 billion for Washington, Oregon, and Idaho[39]–[41]. This incredible source of revenue in the PNW is severely threatened by an inconsistent and overall warmer climate. Winter wheat requires a long period of continuous cold temperatures in order to vernalize and activate reproductive growth in the spring. If this vernalization requirement is not met, the wheat takes exponentially longer to reach maturity, or may not set seed at all[19][27][28]. Spring wheat is planted later and avoids the need for vernalization, but yields significantly less due to decreased growth time, and cannot be planted as early as winter wheat due to its lack of cold-hardiness and/or poor planting conditions in the early spring[26][42][43]. Both types require specific conditions be met in order to produce profitable yields. To try and avoid massive yield losses due to unpredictable weather, researchers and growers of winter wheat are looking towards incorporating facultative growth habit into their germplasm[8][27][28][44]. Facultative wheat does not have a strict genotype associated with it, and rather, the only condition is

that it has a reduced or non-existent vernalization requirement, but will not flower during cold winter or early spring temperatures[16][42][43][45].

2.3 Stripe Rust

Stripe rust (*Puccinia striiformis* Westend. f. sp. *Tritici Erikss.* (*Pst*)), also known as yellow rust, is a fungal disease of wheat that causes significant yield losses worldwide[46][47]. It affects wheat crops in 60 countries on every continent where wheat is grown[48][49]. As of 2017, 88 percent of the world's wheat production was susceptible to stripe rust, which causes annual yield losses estimated to be worth \$1billion[50]. The first description of the disease was in Europe in 1777, with research into its epidemiology and control being conducted for over a century[49]. The stripe rust pathogen infects green tissues of wheat at any stage of development, from one-leaf stage to full maturity before dry-down. Symptoms begin to appear one week after infection, with sporulation following in as quickly as two weeks in optimal conditions. The first noticeable signs of stripe rust are tiny yellow-orange pustules called uredia, each of which contain thousands of urediniospores[49]. While the individual urediniospores are too small to see with the naked eye, they form an easily recognizable mass of yellow-orange powder that sticks to hands, clothing, and other leaves if rubbed. The uredia appear first and on young seedling leaves, but do not usually advance past this stage until the plant enters stem elongation[49].

The name 'stripe rust' comes from the distinct parallel lines of either uredia or necrosis that run along the wheat leaves (Figure 2.3). Chlorosis and necrosis of these stripes indicates a hypersensitive response of the plant which confers resistance to the pathogen[48]. Severe infections can also move beyond the leaves and be found on the heads of the developing wheat plants (Figure 2.4). The pathogen diverts water and nutrients from the host, weakening it[49]. Severe infections also significantly reduce the amount of green tissue that is available for photosynthesis[51]. Plants that are infected with stripe rust, but still able to finish seed set, produce seeds with much lower test weight, quality, vigor, and germination rates[49][52]. Dark-colored teliospores can often

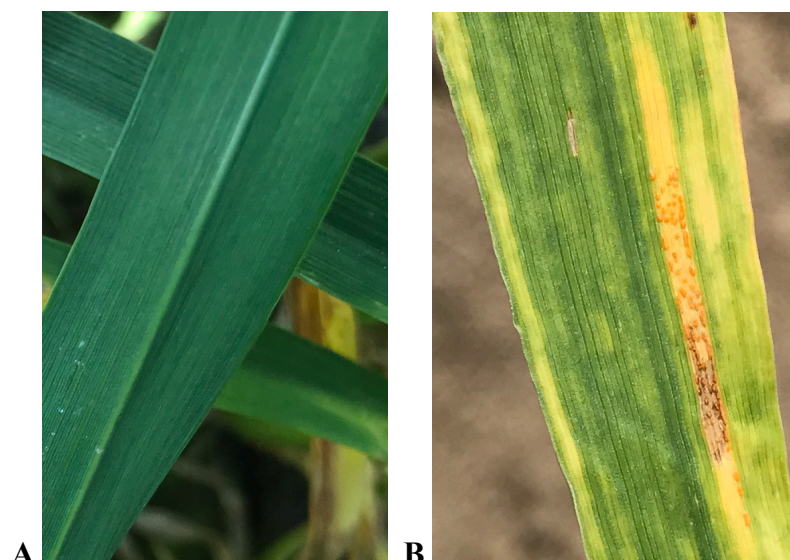


Figure 2.3 Wheat flag leaves with different levels of stripe rust infection. A. Healthy, green flag leaf with no signs of infection. B. Flag leaf infected with stripe rust. Chlorotic flecks and stripes indicate early infection symptoms. Orange uredia in the chlorotic stripe is a classic sign of stripe rust. The fungus is also producing dark brown/black teliospores for over-summering.

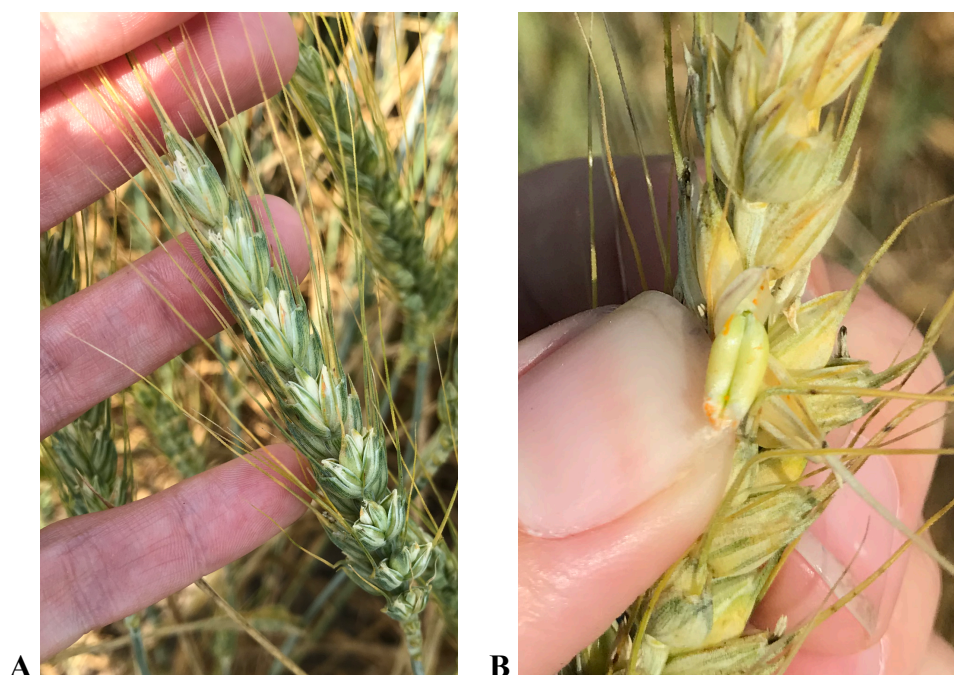


Figure 2.4 Stripe rust uredia found on the heads of severely infected winter wheat plants. A. Orange uredia can be seen on the spikelets and on the awns of a severely infected plant. B. This plant is so severely infected that uredia have developed on the immature seed itself.

develop late in the growing season when the plants begin to senesce or when the temperature begins to cool (Figure 2.3). These teliospores are structures specialized for survival over extended periods of unfavorable conditions, such as over the winter or summer[47]. They remain on the dead plant tissue and produce basidiospores which go on to infect new hosts in the same field or are dispersed to new fields[53].

Initial stripe rust infection is primarily determined by moisture, as the amount of moisture in the environment determines spore germination, infection, and survival[47]. The pathogen requires at least three hours of continuous moisture on the wheat's surface to germinate and then infect the host plant. This requirement is easily met in irrigated crops and in areas with moist weather, like the PNW[49]. High moisture can also have a negative impact on the stripe rust pathogen's urediniospores later in its growth cycle. The spores require a dry environment in order to survive and infect new plants in the fall and winter. In the PNW, however, the dry weather in summer allows the urediniospores to survive and re-infect winter wheat planted in the moist autumn season[49]. Temperature also plays an important role in the stripe rust life cycle, as well as in host resistance. Germination, infection, and survival are inhibited by high temperatures[47]. Extremely low temperatures below -10°C can also decrease survival and limit infections. Stripe rust is mainly wind-dispersed within and between regions but can also be spread mechanically and by splash-dispersal within a small area. For example, stripe rust infections in Canada have been shown to have urediniospores originating from as far away as Mexico[49].

There are two forms of stripe rust resistance: all stage resistance (ASR) and adult plant resistance (APR)[46]. Genes for ASR are generally race-specific, are active during the entire life cycle of the plant, qualitative, and easy to phenotypically select for in the field[52]. Due to their race-specific nature, cultivars containing single ASR genes or low-complexity resistance gene pyramids have been easily overcome by new stripe rust races. This has led to many cultivars losing their resistance shortly after commercial release or even sometimes before release, and therefore many of the characterized ASR genes are no longer effective[46][47][52][54][55]. On the other hand, APR genes are generally quantitative and not race-specific, although there are some exceptions. APR

includes a category called High Temperature Adult Plant resistance (HTAP). HTAP resistance is durable, not race-specific, and is influenced by increasing temperatures and progressing plant growth stages. As the wheat plants mature and the temperature increases, the level of resistance in the plant also increases. The most effective daytime temperatures for HTAP resistance are between 25 and 30°C[56]. A plant with HTAP resistance will be susceptible to stripe rust as a seedling, but the number of new infections will be reduced, the amount of inoculum will decrease, and the spread of the inoculum will be lessened by the resistance genes as the season progresses[50][52][57]. Some cultivars with extremely high levels of HTAP resistance can entirely inhibit sporulation of the pathogen[49][51]. Most of the characterized and named resistance genes are ASR and/or race-specific genes. APR and HTAP resistance genes are less well characterized and therefore new resistance loci are being identified regularly[46].

The PNW has an ideal climate for stripe rust, and consequently, this area sees yearly infections and battles frequent epidemics. Its mild, wet winters and dry hot summers provide the ideal environment for perpetuating the stripe rust life cycle of infection, overwintering, over-summering, and re-infecting[54][55]. The region is often influenced by inoculum from outside the PNW as well[50]. Before resistant cultivars and effective fungicides were introduced, the PNW saw extreme ranges of infection and yield loss[47]. In 1958, Washington lost 2.9 million bushels of wheat totaling 4 percent of the production. In 1960, however, the state experienced an epidemic of stripe rust that resulted in 25 percent of the production being lost. In the 1980s yield losses were significantly reduced by the introduction of stripe rust-resistant cultivars and commercially available fungicides to combat infection[49][54]. However, around the year 2000 new races of stripe rust emerged that overcame most of the cultivar resistances[52][56]. In 2002, the environmental conditions were extremely favorable for stripe rust infection and 70 percent of the wheat acreage in Washington was affected. In order to combat this epidemic, 170,000 acres of wheat had to be sprayed with fungicides, at a cost of more than \$2.5 million. Without the application of fungicides, the yield loss would have been 20 to 25 percent, resulting in losses of \$26 million to \$33 million. Similar situations were seen in the states of Oregon and Idaho due to their geographic

proximity, similar weather conditions, and similar field management practices[49]. In 2010 stripe rust caused a 10 percent loss of winter wheat yield in Oregon, and in 2011 the disease caused a yield loss of 17.5 percent in Oregon, Washington, and Idaho combined[57]. Years like these show that even with timely fungicide applications, severe stripe rust infections can cause devastating yield losses, with up to 70 percent loss in susceptible varieties[58]. Intensive applications of fungicide also come at a huge cost to growers, negatively impact the environment, and can lead to fungicide resistant pathogen strains[51]. In the PNW, foliar fungicides are applied nearly every year on some varieties, and in heavy stripe rust infection years, multiple applications of fungicide are used[54].

2.4 Breeding and Marker Assisted Selection

The first humans to settle in one place and farm crops around 10,000 years ago were undoubtedly also the first plant breeders. They were responsible for the domestication of grasses, vines, and tubers which would eventually evolve to become wheat, rice (*Oryza sativa*), squash (*Cucurbita spp.*), corn (*Zea mays*), potatoes (*Solanum tuberosum*), cassava (*Manihot esculenta*) and countless other crops grown all across the globe today. This was accomplished by the selection for desirable traits such as non-shattering heads, disease and pest resistances, and higher yields. All these traits are controlled by genes which early farmers were selecting, though this was not understood for many thousands of years. Commercial plant breeding, which involves deliberate and artificial cross pollination, took hold in the early 1900s[59]. Conventional (also known as traditional) plant breeding is still practiced all over the world today by both public and private breeders in virtually every commercially produced crop[60]. These breeders choose parents for crosses based on factors such as diversity, disease resistance, yield, quality traits, height, water use efficiency, and any other trait necessary for the improvement of the crop's performance in their target region.

Despite the long history of trait improvement through breeding, the intricacies of allelic control of the traits being selected for (or against) was not fully understood until after rapid DNA sequencing was invented by Frederik Sanger in 1977[61][62]. Soon

after scientists gained the ability to sequence specific parts of DNA, plant breeders began utilizing molecular markers to rapidly improve varieties[59][63][64]. Molecular markers are sequences of DNA that are associated with certain locations in a genome, and thus associated with certain traits in the organism[65]. The use of molecular markers for breeding is called Marker Assisted Selection (MAS). The first examples of markers used for MAS were two wheat protein-based markers discovered in the 1980s. One was from protein storage subunits in seed embryos that were correlated with bread making quality, and the other was the Pchl protein which confers eyespot disease resistance[59][66]. Molecular markers have a decided advantage over traditional morphological selection for traits that are highly influenced by the environment, are recessive, are difficult or costly to phenotype, and that require gene pyramiding (the combination of multiple traits that generally produce the same phenotype)[1][65][67].

Traits that are highly influenced by the environment can be difficult to select for phenotypically if the necessary environmental influence is not present during field testing. If a region requires freezing tolerance in a crop, for example, the breeder would need to have sufficiently cold winters to discard lines that did not survive well. If none of the winters were cold enough to show a phenotypic difference, the breeder may be releasing cultivars that will not survive future harsh winters. Recessive traits can also be difficult to select for phenotypically and often require progeny tests, which take time and money. Some phenotype tests are costly and beyond the scope of every breeding program to perform on every line every year. Knowing the lines that would perform poorly in the phenotype test beforehand would allow the breeder to discard these lines and reduce the final test number, and thus the cost. Gene pyramiding for resistance to pests and diseases can also be extremely difficult to do phenotypically, especially if the breeder does not have access to every possible race or pathotype and an ability to artificially inoculate each line in a controlled environment. A resistance gene needed in the pyramid may be effective against only one race of a pathogen, and that race may not be present in a particular field year, or the environmental conditions for the pathogen may not be ideal and infection rates are low independent of resistance level of the plant. In all

of these cases, plant breeders would greatly benefit from the ability to use MAS over simply conventional breeding and phenotypic selection[1][59][65].

2.5 Early Markers and First Generation Sequencing

Despite the known benefits of MAS, breeding programs by and large did not adopt the use of many markers (if any) during the 30 years since they were first developed. This was primarily due to the difficult and costly processes of marker-trait association, marker development, and marker validation[59]. There was also concern that the developed markers were unreliable and not representative of adapted germplasm[67][68]. The first plant breeding markers were Restriction Fragment Length Polymorphisms (RFLPs)[69]. In RFLP marker discovery, DNA samples are digested with restriction enzymes that cut specific sequences, then these fragments are separated by gel electrophoresis, and the resulting size differences are the markers. RFLPs are codominant (meaning homozygous and heterozygous alleles can be differentiated), robust, and reliable, but they were also time consuming, laborious, expensive, required a large amount of DNA, and the number of polymorphisms in related lines were limited[65]. Thus they were replaced with other, more reliable PCR-based markers such as RAPD, SCAR, CAPS, and AFLPs (Table 2.1)[68].

Amplified Fragment Length Polymorphism (AFLP) markers simplify RFLP discovery by attaching adapters with known sequences to the restriction enzyme-digested fragments of DNA and then PCR amplifying them. This method is easier and more reproducible than RFLPs, but is dominant rather than codominant, meaning the results do not differentiate between homozygous and heterozygous for the trait of interest[65][68]. Cleaved Amplified Polymorphic Sequence (CAPS) markers are also a simplified version of RFLPs. In this technique, specific primers are designed to target and amplify 800 to 2,000 base pair (bp) long regions of DNA which are then digested with restriction enzymes. The resulting fragments are separated via electrophoresis and the size differences are the markers. CAPS markers are codominant, reproducible, and only require a small amount of DNA, but can be difficult to find in some genomes and the researcher must have previous knowledge of the DNA sequence. Random Amplified

Polymorphic DNA (RAPD) markers are discovered by amplifying DNA using short, arbitrary primers. The resulting amplification pattern is used as the marker. RAPD

Table 2.1 Examples of the most common molecular marker types with descriptions and examples. Adapted from Collard *et al.* (2005)[65].

Molecular Marker	Brief Description	Advantages	Disadvantages	Examples of Organism and Trait
RFLP (Restriction Fragment Length Polymorphism) [62][67][70]	DNA samples digested with restriction enzymes → fragments separated by gel electrophoresis → differences in band size are the markers	• Codominant • Robust • Reliable	• Laborious and expensive • Require large amounts of DNA • Limited number of polymorphisms in related lines	• Tomato (<i>Solanum lycopersicum</i>) virus resistance from related wild species[71]
RAPD (Random Amplified Polymorphic DNA) [62][67]	Several short, arbitrary primers are created and used to amplify DNA with PCR → Resulting pattern of amplification is the marker	• Quick and Simple • Inexpensive • Requires small amount of DNA	• Dominant • Difficult to reproduce • Generally not transferrable across populations	• Winter hardiness in apple (<i>Malus spp</i>) • Apomixis genes in Buffelgrass (<i>Cenchrus ciliaris</i>)[76]
SCAR (Sequence Characterized Amplified Region) [62]	RAPD fragments used to develop longer (~20bp) primers for certain regions	• Less sensitive to reaction conditions • More reproducible • More specific	• Dominant • Difficult to reproduce • Generally not transferrable across populations	• Downy mildew resistance in lettuce (<i>Lactuca sativa</i>)[77] • Interspecific hybridization identification between colonial and creeping bentgrass (<i>Agrostis spp.</i>)[78]
CAPS (Cleaved Amplified Polymorphic Sequence) [62][79]	Specific primers designed to target 800-2,000bp region → PCR amplifies DNA → Amplified DNA is digested with restriction enzymes → polymorphisms visualized by electrophoresis	• Codominant • Locus-specific • Reproducible • Low quantity of DNA needed	• Difficult to find in some genomes • Must have previous knowledge of DNA sequence	• Imidazolinone herbicide resistance in wheat • Fertility restorer gene in cotton (<i>Gossypium hirsutum</i>)[80]
AFLP (Amplified Fragment Length Polymorphism) [62][67][70]	DNA digested with restriction enzymes → adapters ligated to sticky ends of fragments → fragments with adapters amplified with PCR	• Multiple loci • High levels of generated polymorphism • High reproducibility	• Dominant • Requires large amount of DNA • Complicated methodology	• Nematode resistance in potato[81] • Leaf rust resistance in Poplar (<i>Populus spp</i>)[82] • Reducing linkage drag in soybean (<i>Glycine max</i>) crosses[83]
Microsatellite/SSR (Simple Sequence Repeat) [62][67][70]	Design primers to amplify 1-10bp repeat sections → PCR amplify regions → run electrophoresis to separate sizes → variable sizes are markers	• Codominant • Technically simple • Robust and reliable • Transferable between populations	• Time and labor intensive to produce primers • Requires knowledge of sequence • Spontaneous mutations may occur	• Photoperiod response genes in wheat • Seed germination gene in <i>Arabidopsis</i>[84] • Genetic mapping markers in kiwifruit (<i>Actinidia deliciosa</i>)[85]
SNPs (Single Nucleotide Polymorphisms) [61][62][84]	PCR validation and/or Sanger sequencing OR Next Generation Sequencing (NGS) of whole or part of the genome → single base difference between DNA segments is marker	• Codominant • Most abundant type of marker • Easy discovery and design with high-throughput automation • Can be 'perfect' markers	• Can be difficult to discover in polymorphic species with no reference genome • Can be difficult to multiplex • Highly sensitive to sequencing coverage	• Grain length and width genes in rice[86] • Leaf, stripe, and stem rust resistance genes in wheat[87] • Late blight resistance genes in potato[88] • Oil quality genes in canola (<i>Brassica napus</i>)[89] • Flowering time and norther leaf blight disease resistance in maize[90][91]

markers are quick and inexpensive to produce, but are dominant and can be difficult to reproduce in other laboratories and populations. Sequence Characterized Amplified Region (SCAR) markers take RAPD markers a step further to improve specificity and reproducibility. These markers are created by amplifying certain RAPD fragments with longer, more specific primers[65]. SCAR markers are still dominant and can be difficult to reproduce across laboratories and populations, however[60].

The most common marker types used for MAS today are microsatellite markers (also known as SSR or simple sequence repeat markers) and SNP (Single Nucleotide Polymorphism) markers[70]. SSRs are tandemly repeated di- to pentanucleotide motifs that are highly abundant in most genomes, extremely variable, and prone to mutations. Therefore, they are highly polymorphic between species, within populations, and even between varieties. In SSR marker discovery, primers are designed to amplify these regions and then the regions are separated by size using electrophoresis. While SSR markers are codominant and highly reproducible, they do require previous knowledge of the DNA sequence and can be prone to spontaneous mutations[68]. SNP markers are the most highly used in MAS today[1][59][71]. SNPs are the most common source of polymorphism in any genome, with the frequency in wheat thought to be at least 1 in every 540 bases[60][71][72]. There are three kinds of SNPs, transitions (C to T and G to A), transversions (C to G, A to T, C to A, and T to G), and indels (single base insertions or deletions), all of which are informative as markers[72]. SNP markers can be discovered by Sanger sequencing any previously known or unknown part of the genome, or by using one of the many Next Generation Sequencing (NGS) technologies available (discussed in detail later). SNP markers are codominant and can often be ‘perfect’ markers, meaning the marker itself is based on the allelic variation that is responsible for the trait of interest, rather than just linked by close genetic distance[66][73].

2.6 Next Generation Sequencing

Sanger sequencing is expensive, low-throughput, and ineffective in highly repetitive regions, yet DNA sequencing and gene discovery were limited to this method

until the 1990s, when the Human Genome Project began[62]. The Human Genome Project was developed in order to sequence every nucleotide in the human genome, but new methods for sequencing needed to be developed to shorten the time to complete genome sequencing, bring down the costs, and increase the ability to sequence GC rich and repetitive regions. It was during this time that second generation sequencing (a part of and commonly referred to as Next Generation Sequencing) technologies were developed (Table 2.2). Traditionally, NGS refers to optical sequencing platforms that can produce millions of short DNA sequence reads typically between 25 and 400bp in length, unlike first generation semi-conductor sequencing technologies that produced read lengths of up to 1000bp but were extremely low throughput[61][66]. The current form of NGS allows for affordable and relatively fast sequencing of whole genomes for model and non-model species. The advent of NGS technologies in the early 2000s caused an exponential increase in the amount of sequencing projects and data generated. The first genome to be sequenced was a bacterium in 1995[90]. Just 24 years later, as of March 2019, over 1,000 species and strains from eukaryotes, prokaryotes, viruses, plasmids, and organelles have had their full genomes sequenced, and more than 40,000 genomes are in the process of being sequenced[91]. In plants alone, the total number of published sequenced genomes increased from 1 in 2000, to 18 in 2010, to 53 in 2013, then up to 328 in 2019[92][93]. In recent years, third and fourth generation sequencing technologies have also been developed (such as Nanopore and PacBio), which utilize the high-throughput production of long sequence reads[66][94].

The explosion of fully sequenced genomes illustrates only a small portion of the total amount of sequencing data available. NCBI's GenBank Release 230.0 reports that its database includes more than 200 million sequences, all of which are searchable and accessible to researchers[95]. This vast amount of data generated by researchers can be used for a wide variety of projects with various goals in mind, such as cancer research, population genetics, and ancestry determination[61]. For plant breeders, the usefulness of NGS data lies in its wide-ranging applicability for trait improvement. Next Generation Sequencing technologies have been successfully used in crops for linkage mapping, wide

Table 2.2 The most common Next Generation Sequencing (NGS) technologies. Not all NGS platforms available are listed, only the most common and well-known platforms are included. The Read Length is the longest possible length of a single read. Reads per Run is how many of those single reads are generated in a single run on each platform. The Cost per Run is given for the highest cost version, some cheaper options may be available, but will usually be smaller read lengths, lower depth, and/or lower quality. The cost can also vary greatly depending on the library preparation, accessibility of the technology, availability of computational resources, and bioinformatics costs. The ideal use of each platform is given, but some technologies can be successfully used for other applications as well.

	Technology	Method	Read length	Accuracy	Reads per Run	Time per Run	Cost per run	Advantages	Disadvantages	Ideal uses
2nd Generation Sequencing - high throughput sequencing of large genomes at a low cost which produce short reads [68]	Roche 454 <i>*Obsolete d, no longer supported</i> [74][70][68]	Pyrosequencing	400bp	99%	1 million	23 hours	\$6,000	- Long reads - Fast	- Expensive - Homopolymer errors (indels)	De novo assembly of BACs, plasmids, and microbial genomes; 16s variable region, and targeted amplicon sequencing
	Illumina HiSeq [74][70][96]	Sequencing by Synthesis	150bp	99%	3 to 6 billion	2-3 days	\$6,000 per lane	- Mid cost instrument - High throughput	- Mid cost per sample - Short reads do not allow for assembly/mapping of highly repetitive regions - Difficult to fill high yield flow cells	Assembly, resequencing, transcriptome sequencing, SNP detection, metagenomics
	Illumina NovaSeq [96][64]	Sequencing by Synthesis	150bp	99%	40 billion	13 to 44 hours	\$10,000	- Increased throughput - Accelerated data processing time	- Not (yet) cheaper than HiSeq - Short reads do not allow for assembly/mapping of highly repetitive regions	Same applications as HiSeq
	Illumina MiSeq [74][96][64]	Sequencing by Synthesis	300bp	99%	25 million	2.7 days	\$2,000 per lane	- Short run time - Long reads	- High error rate - Short reads do not allow for assembly/mapping of highly repetitive regions	Small genome, amplicons, 16s, and low diversity applications
	Ion Torrent [74][96]	Chip-based semiconductor sequencing	400bp	98%	130 million	3 to 21 hours	\$1,000	- Inexpensive equipment - Fast - Long read lengths	- Homopolymer errors - Low throughput - Short reads do not allow for assembly/mapping of highly repetitive regions	Amplicons, small genomes, transcriptomes, microbial, and targeted sequencing
3rd Generation - large fragment sizes from single molecule sequencing [68]	Ion Proton [96][64]	Chip-based semiconductor sequencing	200bp	97%	80 million	4 hours	\$1,000	- Low cost equipment - Fast/high throughput	Indel errors	Transcriptome, exome, and medium sized genome sequencing
	SOLiD 5500 [74][70][68]	Sequencing by ligation	50 or 60bp	99.9%	2.6 billion	6 to 10 days	\$5,000 to \$10,000	Low cost per base	- Slower - Short reads not ideal for <i>de novo</i> assembly - A/T bias errors	<i>De novo</i> sequencing, differential expression, and resequencing
	PacBio Sequel [74][68][96]	Single Molecule Real Time (SMRT) sequencing	15,000 bp	87%	370,000	4 hours	\$850	- Long reads - Sequences high G/C areas - Methylation detection - No PCR amplification	- Low read output - Expensive - High error rate	Small genome assembly (can be combined with other methods for larger genome assembly), sequencing of difficult regions, and epigenetic studies
4th Generation - nanopores produce long reads of single molecules [96]	MinION [96][95][64]	Nanopore sequencing	10,000 bp	90%	160,000	72 hours	\$900	- Long reads - Fast - Cheap - No PCR amplification	- High error rate - High levels of inter-run variability	Bacterial strain identification, viral genome assembly, haplotyping, direct RNA sequencing, and PCR-free cDNA sequencing

crosses, alien introgression, epigenetic modification analysis, transcriptomics, population genetics, *de novo* sequence assembly, association mapping, and the development of molecular markers[61][96]. This last application, as mentioned previously, is of extreme importance in plant breeding as the need for increased yield in the face of exponentially increasing human population, loss of good quality farmable land, increasingly unpredictable weather, lower water availability, new and emerging diseases and pests, and approaching biological yield ceilings simultaneously reduce gains made by traditional breeding and require that improvements be made faster than ever before. MAS is expected to substantially increase the genetic gains in crop species and reduce expenditures in variety development, especially now that sequencing technologies are cheap, easy, and high throughput[61][97].

2.7 SNP Discovery Using Genotyping by Sequencing

As mentioned previously, the most reliable and ubiquitous marker type are SNPs. While SNPs are spread throughout a plant's genome, knowing the location of every SNP is not required for marker development. The most useful SNP markers are those that are closely associated with a trait of interest in adapted germplasm that the breeder will utilize for multiple years. Before NGS became widely available and cost-effective, SNPs were difficult to discover especially in large, polyploid crop species that did not have published reference genomes. Now multiple technologies and techniques exist that allow for affordable, easy SNP discovery. The most commonly used method is a reduction of the genome complexity called Genotyping by Sequencing (GBS)[93].

GBS was developed as an affordable sequencing method for a large number of segregating individuals. As such, it enables the multiplexing and sequencing of up to hundreds of samples in a single Illumina lane, for example. GBS is referred to as a 'reduced representation sequencing' (RRS) technique that reduces the target genome's complexity, and was initially developed as RAD-sequencing (Restriction site Associated DNA)[98][99]. These reduced representation sequencing techniques allow for the sequencing of a large number of markers across the entirety of the genome while simultaneously reducing the overall number of genomic regions being sequenced. This

means each individual region can be sequenced at high coverage in an affordable manner[61][100]. Under the assumption of high sequencing coverage (as variant discovery in RRS techniques are highly sensitive to sequencing coverage), GBS improves the accuracy of genotyping across individuals and can allow for imputation of missing genotypes as long strands of haplotypes can be detected in the population of interest[68][101]. An overview of the GBS pipeline is given in Figure 2.5.

The first step in GBS is to normalize the DNA library so that each individual is contributing the same amount of DNA to the subsequent steps. Next, each individual plant's DNA is cut up with restriction enzymes (RE). The most common REs used are methylation sensitive, meaning they only cut unmethylated DNA. Methylation typically represses gene expression, so a gene that is not being expressed due to methylation will not be cut by the RE. This increases the proportion of SNPs that are associated with expressed genes that will be more likely to contribute to a phenotype and thereby useful as molecular markers[68][102]. The use of one frequent-cutter and one rare-cutter restriction enzyme is generally recommended to further reduce genome complexity and improve fragment library uniformity[60][100]. At this step there are many fragments of highly variable length, and only those between 150 and 350 bases long are selected. Then, adapters are designed to anneal to the restriction enzyme overhang(s) and are subsequently used in PCR amplification. At this stage the library consists of millions of 150 to 350bp long fragments with the structure of "Forward Adapter – Genomic DNA – Reverse Adapter"[94][100].

Most NGS technologies utilize multiplex sequencing, which means that each individual's library of sequences must be tagged for identification and separation in the future[60][103]. This is done using DNA barcodes that are attached to the adapters. The unique sequence of each barcode can be used after sequencing to identify which individual that sample originated from[100]. Once the library is created, all samples are pooled together and run on a NGS technology of choice. The two most commonly used for GBS are Ion and Illumina. Both use slightly different sequencing methods and have different outputs, and the ideal machine and type of run must be determined by the individual project needs[94]. Regardless of which technology is used, the output will be

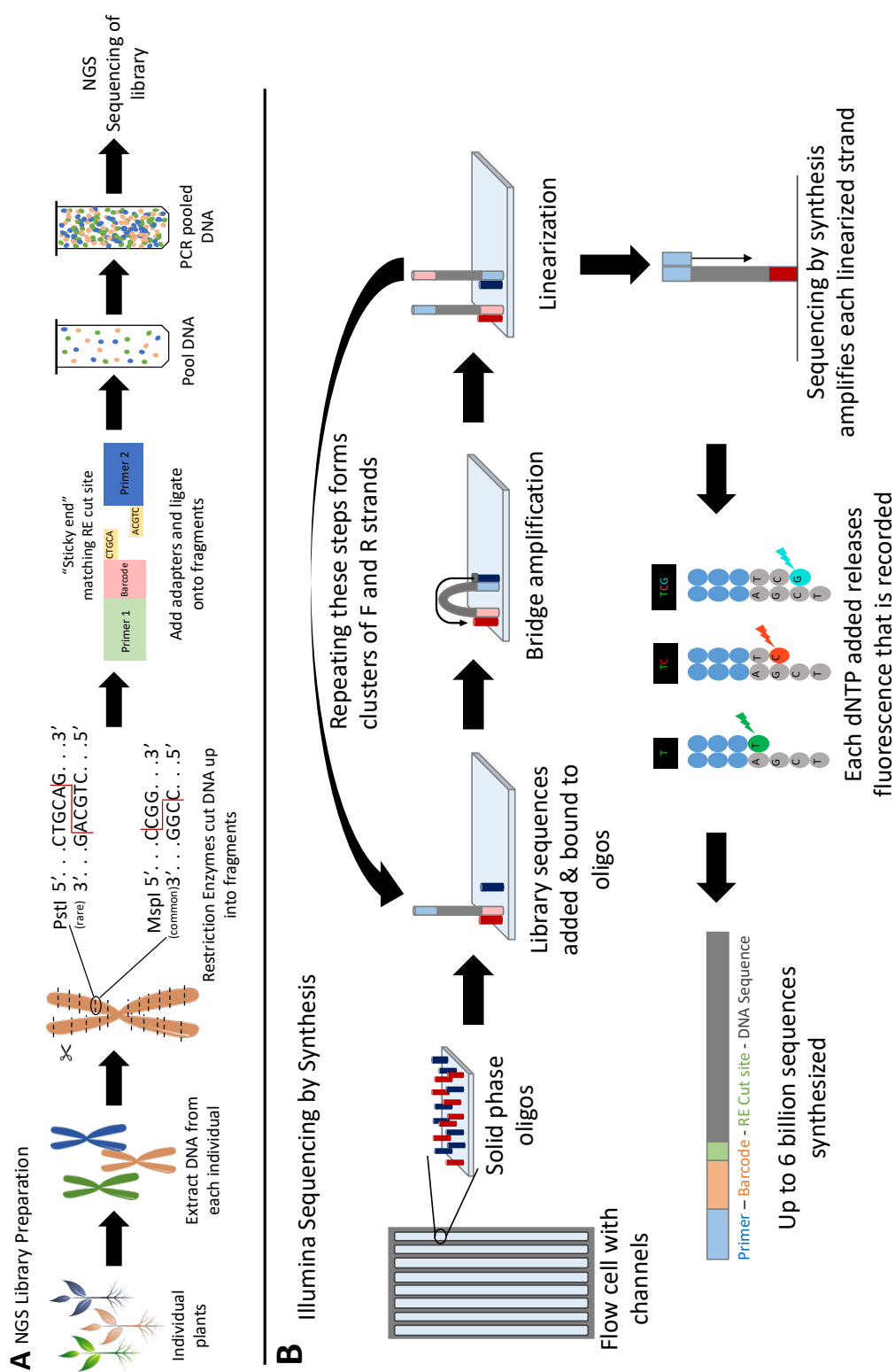


Figure 2.5 Genotyping by Sequencing (GBS) workflow. Adapted in part from Elshire *et al.* (2011), Kohn *et al.* (2013), and Glaubitz *et al.* (2014)[101][107][108]. **A**) NGS Library preparation. **B**) GBS using Illumina technology. **C**) GBS using Ion technology. **D**) SNP analysis pipeline in TASSEL (Trait Analysis by aSSociation Evolution and Linkage)[109].

millions to billions of short reads from all across the genome. This data is then filtered for quality, the sequences aligned to each other, then they are either aligned to a reference genome or *de novo* aligned, and SNPs are identified. This can be done using a number of different methods, which again are ideal for use in different situations and for different end-uses[104]. Once SNPs have been identified and mapped, they are ready to be used in the marker discovery pathway.

2.8 SNP Marker Discovery Using Marker-Trait Association Analysis

Once SNP data is obtained using GBS, there are a few options for marker-trait association analysis, such as genomic diversity studies, genetic linkage analyses, molecular marker discovery, and genomic selection; but one of the most common in plant breeding is a genome-wide association study (GWAS)[60][108]. GWAS utilizes phenotypic analysis combined with genome-wide genetic variants, like SNPs from GBS, to determine whether any of the variants are associated with a trait or traits of interest[60]. The first step to completing such a marker-trait association analysis is to obtain or create a population of plants in which markers can be identified. It is recommended to have 100 to 500 individuals in the population, ideally including the parents. Typically, two parents are selected that are homozygous, but differ either for only the trait of interest or for multiple traits of interest, and a biparental cross is made between them[68]. In self-fertilizing species, the offspring are then allowed to self-fertilize for at least 5 generations in order to obtain homozygous Recombinant Inbred Lines (RILs)[65][77]. To speed up the process, or for species that do not self-fertilize, homozygosity can also be achieved by generating doubled haploids (DHs). Once a segregating population is produced, it must be phenotyped for the trait(s) of interest. It is important that the phenotyping be done accurately with adequate replications to reduce anomalous results. Regardless of how the data is scored originally, it must be quantitative for use in GWAS[70]. The population must also be sampled for GBS using a technology and analysis platform of choice in order to obtain SNPs and their locations. Especially for populations created from adapted germplasm, it is important to

determine the population structure and relatedness present in order to control for false positives[109][110]. The output of a GWAS is SNP locations on a chromosome or linkage map that are significantly associated with a trait of interest. This association can be with a single gene or more commonly with a Quantitative Trait Locus (QTL). These associations can then be used to determine which allelic states at certain SNP locations correspond to the trait of interest and will be useful markers for genotyping other individuals. For species that have a published genome, the locations of the SNPs can be searched for genes in the area or nearby that are involved in the trait of interest. This can make gene discovery and characterization much easier than was ever possible before. While QTL discovery using GWAS is an extremely useful tool for marker discovery and is a common technique employed by scientists, it can be limited in its ability to detect small effect loci and is often only relevant to the germplasm tested. These are important things to keep in mind when using or developing markers identified via GWAS, but the increasing accuracy and decreasing costs of NGS technologies is beginning to resolve these drawbacks.

In the first four months of 2019 alone, studies have been published using GWAS to identify QTL for nutritional elements in foxtail millet (*Setaria italica*)[111], drought tolerance in soybean[112], root architecture in durum wheat (*Triticum durum*)[113], thermotolerance in maize[114], bacterial blight resistance in rice[115], starch content in maize[116], domestication genes in wild chickpea (*Cicer spp.*)[117], flowering time genes in Canola[118], drought tolerance in sesame (*Sesamum inindicum*)[119], yield in barley (*Hordeum vulgare*)[120], low temperature germination tolerance in cucumber (*Cucumis sativus*)[121], stripe rust resistance in wheat[122], fiber length in cotton[123], and many more. The ever-increasing amount of published data and improvements being made in sequencing technologies make the incorporation of MAS into every breeding program a much more practical option than it has ever been before.

2.9 Genetic Transformation

While there are a multitude of wheat cultivars and land races, each with great genetic diversity, these are extremely adapted to local climates, soils, diseases, cultural

practices, and pests[28]–[30]. Therefore, within an area such as a single country, the genetic diversity from one region to another does not lend itself to rapid gene integration using conventional breeding. A variety well adapted to Kansas, for example, will do extremely poorly if planted in Oregon. Therefore, if a specific trait from the Kansas variety is desired in Oregon, the breeder must cross the two and then spend many years backcrossing to locally adapted varieties before the genetic linkage drag has been reduced to acceptable levels. This process of first cross to varietal release can easily take ten years or more to complete, even with the assistance of MAS, and that is the timeline for a trait that already exists[27]. There are many traits and phenotypes that do not exist at all in cultivated wheat, and which can only be found in distantly related species, if at all[124][125]. This can make the integration of traits of interest prohibitive at best and impossible in many situations, such as with the stripe rust resistance genes *Sr32* and *Sr37* found in *Aegilops speltoides*[126] and *Triticum timopheevi*[127], respectively. These genes were introgressed into hexaploid wheat for increased disease resistance, but were never commercially successful because of their associated negative linkage drag that could not be overcome.

For many years now, genetic engineering has been used as an alternative to traditional breeding for traits that cannot be easily introgressed[125][128][129]. RNA interference (RNAi) has been widely utilized as a successful genetic engineering tool, such as with the non-browning Arctic apple[130], and the study of dough properties of wheat with RNAi silenced γ -gliadins[131][132]. In this technique, an RNAi vector that includes a sequence matching the desired gene to be silenced is inserted into the genome of the target plant. When expressed, the vector produces double-stranded RNA which the plant's innate defense mechanisms cut up into short sequences, then incorporate into an RNA-induced silencing complex (RISC). The RISC then recognizes the messenger RNA (mRNA) of the target gene and degrades it, thus silencing expression of the target gene[133][134]. This system, however, involves the stable integration of plant pathogen sequences, and thus any plants produced with this method are considered transgenic. While successfully deregulated RNAi transformed varieties are available for some crops, the deregulation of transgenic wheat has proven financially and socially prohibitive. Due

to this, no Genetically Modified (GM) wheat variety has ever been released, and companies such as Monsanto have slowed or completely stopped their GM wheat projects[135]. Therefore, RNAi is a nonviable option for stable trait integration in wheat varietal releases, and is only useful in gene discovery[132].

A different genetic engineering technology that has been used to improve wheat is TILLING (Targeting Induced Local Lesions In Genomes). TILLING begins with mutating a large number of wheat plants of the same variety with a mutagen such as ethyl methanesulfonate (EMS) to induce point mutations throughout the genome. The population is then screened for nonsense or premature stop mutations in desired genes[12][13]. Obtaining lines with sufficiently reduced protein levels can be a problem in wheat as this species contains a hexaploid (6x) genome, and the desired mutations on each genome often must be combined into one plant using crosses, followed by more breeding to cross the trait into desired germplasm. This once again results in a process that takes many years to complete, as it did for a resistant starch wheat variety that took more than seven years to release after the first publication of mutants combined in all three genomes[12].

A somewhat newer method is available that attempts to overcome the innate difficulties and obstacles of these methods. Targeted genetic editing is the process of making a specific edit or mutation to a known gene in order to achieve a desired phenotype[125][128][129]. The most common early technologies for targeted genetic editing were Transcription Activator-Like Effector Nucleases (TALENs) and Zinc Fingers (ZFNs). Both consist of DNA-binding proteins combined with FokI endonucleases that cut one strand of DNA when successfully bound. When a TALEN or ZFN is designed to target the same place on each DNA strand, the resulting double stranded break (DSB) is then repaired by the cell in one of two ways: Homology Directed Repair (HDR) or Non-Homologous End Joining (NHEJ)[136]–[140]. HDR uses a short template sequence introduced by the researcher at the same time as the TALEN or ZFN to repair the break. This means any desired change can be made to the template sequence and it will be incorporated into the DNA sequence[141]. NHEJ is the most commonly used system when nonsense mutations or premature stops are required, as it is a very

error-prone repair mechanism. The cell's machinery tries to put the two broken ends of the DNA strand back together as quickly as possible, often incorporating or deleting one or many nucleotides in the process. This creates indels (insertions/deletions) in the sequence that can lead to a change in the protein functionality[137][141]. The DNA binding domain of TALENs consists of repeated, highly conserved 33 to 34 amino acid sequences that have a divergent 12th and 13th amino acid that is used to recognize a single base pair in the target sequence[137][138]. The binding domain of ZFNs contains three to six zinc finger repeats that can recognize between nine and 18 base pairs[139][140]. TALENs have successfully been used in wheat to confer powdery mildew resistance[134][136], and ZFNs have been used to create a wheat variety with imidazolinone herbicide resistance[142].

Both TALENs and ZFNs can be designed to target and bind specific DNA sequences, but these technologies never became widespread because they have high production costs, are complex to design and construct, and are time consuming and difficult to implement. They are also restricted to delivery by plasmid DNA, viral vectors, or *in vitro* transcribed mRNA, all of which can result in integration of foreign DNA and therefore restrict their use in certain crop plants[129][143]. In the past five years, however, a new technology called CRISPR has been discovered that can bypass these issues and has thus taken over almost completely from TALENs and ZFNs.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a bacterial adaptive immunity system discovered in *Escherichia coli* and *Streptococcus thermophilus* and named in 2002[144]–[146]. In this system, invading pathogen DNA is cut up into pieces by RNase III:Cas complexes and the pathogen sequence is then attached to the complex and used for future recognition of the same invading pathogen during subsequent infections (Figure 2.6). This system was first adapted for use in eukaryotes in 2013, and has since been used to transform grape (*Vitis vinifera*)[147][148], *Arabidopsis thaliana*[149], rice[144], potato[150][151], multiple filamentous fungi[152], maize[153]–[155], tobacco (*Nicotiana tabacum*)[149], sorghum (*Sorghum bicolor*)[156], apple[147][148], poplar (*Populus tomentosa*)[157], tomato (*Solanum lycopersicum*)[158],

soybean[159], citrus (*Citrus sinensis*)[160], lettuce (*Lactuca sativa*)[149], and wheat[128][144][153][161]–[164].

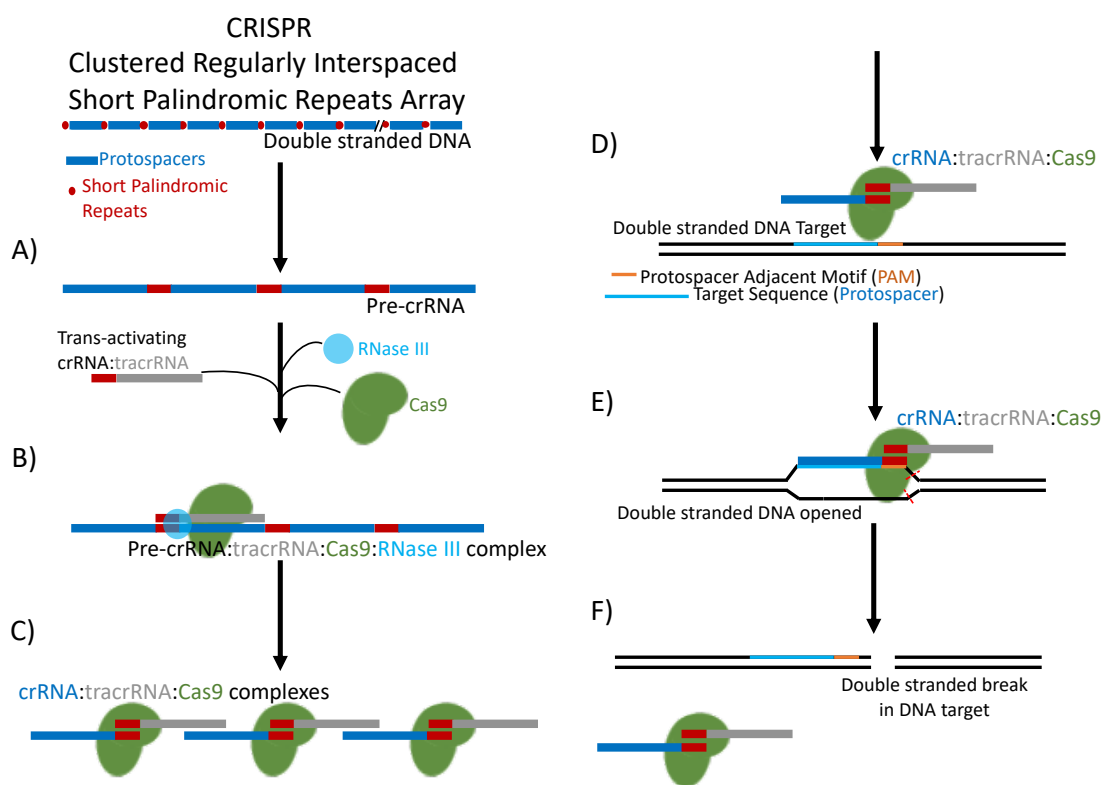


Figure 2.6 The CRISPR bacterial immunity system. A) The CRISPR array is transcribed into pre-crRNA. B) pre-crRNA is processed into individual crRNAs, then the tracrRNA recruits RNase III and Cas9, which separate the individual crRNAs. C) tracrRNA and Cas9 form a complex with each individual, unique crRNA. D) Each complex seeks out complementary DNA sequences with the necessary PAM sequence. E) After binding, Cas9 separates the DNA target and cleaves both strands. F) The complex unbinds from the DNA, leaving the DNA broken and nonfunctional.

All CRISPR systems for genetic editing include a Cas (CRISPR Associated) protein and a single-guide RNA (sgRNA). When expressed together, the Cas protein and the sgRNA form a ribonucleoprotein complex (RNP) that causes Cas to undergo a conformational change, which shifts it from inactive to active and thus able to search the DNA. All Cas proteins, including the most commonly used in plant transformation, Cas9, recognize a specific Protospacer Adjacent Motif (PAM) sequence. Once this PAM sequence has been found, the sgRNA sequence then attempts to bind to the adjacent

DNA sequence. If successful binding occurs, the DNA is cut by the two endonucleases located on the Cas protein[139][143][146]. This cut results in a DSB, which is repaired and exploited as with TALENs and ZFNs[139][143][146]. The only part of the CRISPR system that is designed by the user is the 20-nucleotide long sgRNA sequence. This system can theoretically edit anywhere in a genome that meets two conditions: first, the target sequence must be unique, and second, the target site must be included in the target sequence and be three nucleotides away from the correct PAM sequence (Figure 2.7).

The CRISPR/Cas9 system can be delivered into cells in a multitude of ways, including plasmid expression vectors, viral vectors, Cas9 mRNA vectors, and RNPs. This can be accomplished by *Agrobacterium* mediated transformation, carborundum abrasion, biolistic delivery (gene gun), electroporation, and Gesicles (nanovesicles). The CRISPR/Cas9 complex can be stably integrated into the plant genome or transiently expressed for various amounts of time. The plants themselves can be transformed as whole organisms, callus tissue, embryos, pollen, microspores, or protoplasts. Each of these options, however, comes with caveats and potential problems that depend on the chosen crop's success in tissue culture, the propagation method of the crop, whether integration of foreign DNA is acceptable, costs, and tolerance of off-target effects[146][165]. For example, a researcher who must completely avoid off-target transformation, cannot have any foreign DNA integration, and is using a plant that cannot be grown from protoplasts would likely be limited to RNPs introduced to microspores via electroporation or embryos via a gene gun.

CRISPR also has some further limitations that have come to light as researchers have attempted transformation of more and more genes. Knowledge of the gene sequence and functionality, presence of the PAM sequence in the desired location, and sgRNA sequence specificity are the first obstacles to designing successful gene targets. These issues are much more easily overcome when working with model plants with sequenced genomes and extensive gene annotation than with non-model crops. Organisms without extensive gene annotation may require better gene models for predictive genic annotation, an extensive set of reference sequences from closely related plants, and more extensive primer design and gene confirmation steps before a sgRNA

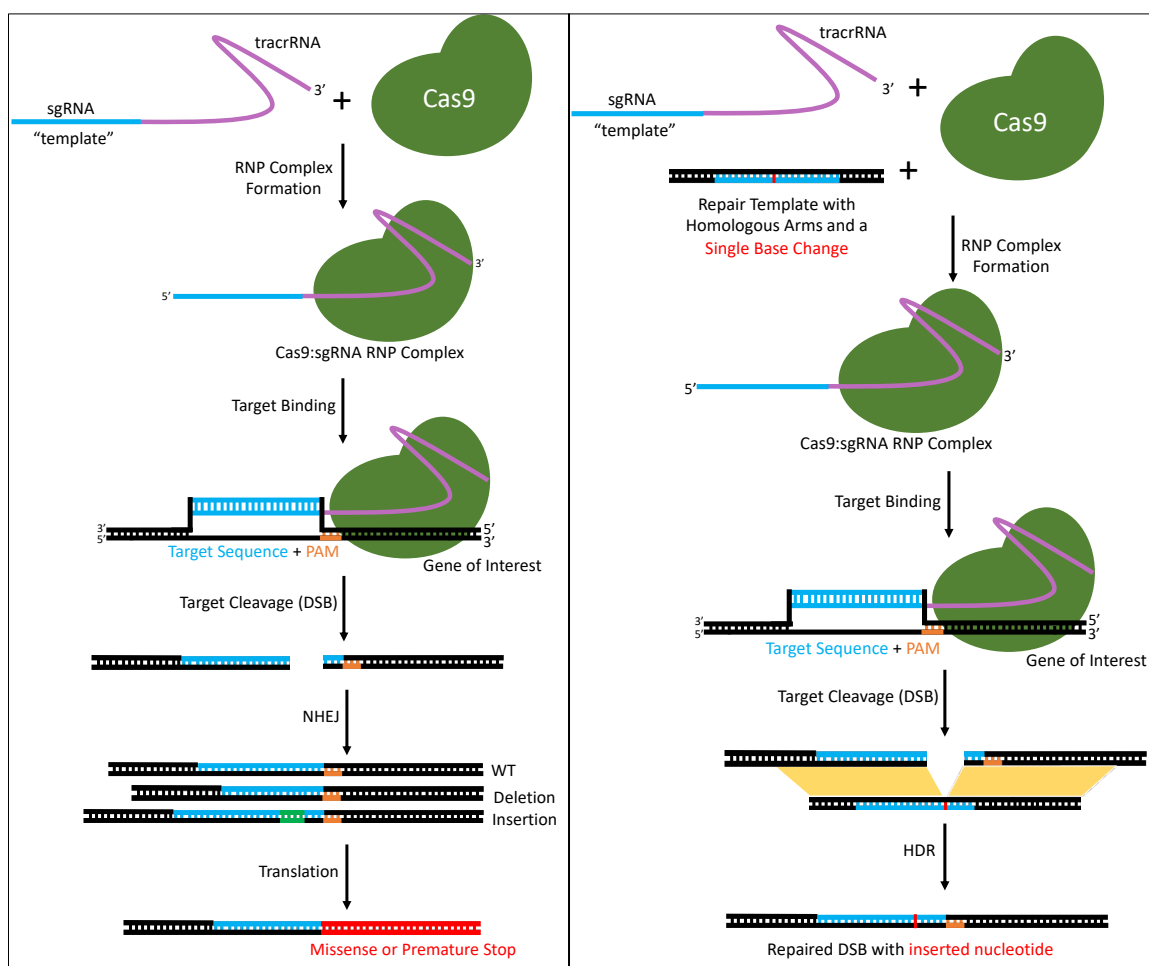


Figure 2.7 Modes of action of the CRISPR/Cas9 system. A) The Cas9:sgRNA ribonucleoprotein complex (RNP) is formed; the RNP searches the DNA sequence for the correct PAM sequence and sgRNA sequence; successful sgRNA binding leads to a double strand break (DSB); the DSB is repaired via non-homologous end joining (NHEJ) which can lead to insertions/deletions (indels); indels can lead to a missense or premature stop codon which results in a nonfunctional protein. B) The Cas9:sgRNA RNP is formed and a repair template is introduced concurrently; the RNP searches the DNA sequence for the correct PAM sequence and sgRNA sequence; successful sgRNA binding leads to a DSB; the DSB is repaired via homology directed repair (HDR) using the repair template; the desired nucleotide is inserted into the repaired DNA sequence.

can be designed. Another recently discovered obstacle is the strand placement of the sgRNA. If the RNP is designed in a way that causes it to bind to the non-template strand, it may remain bound because the polymerase is unable to knock the RNP off the DNA strand. If this happens, it may appear as if the transformation was unsuccessful, without an obvious reason as to why. Therefore, it is important to always design sgRNAs to bind

the template DNA strand[166]. Furthermore, an additional obstacle that is becoming increasingly prevalent due to the increase in CRISPR projects is a lack of functionality due to the chromatin state at the target site[144]. There is currently no way to predict this in advance, and a failure to screen for it will lead to the same result as the previous problem: an unsuccessful experiment with no clear explanation. The only solution to this is to test the chosen sgRNA *in vivo* in single cells before beginning transformations. For many crops, including wheat, protoplasts are the best system to achieve this, and more researchers are beginning to include this step in their experiments[167].

CRISPR/Cas9 technology has often been shown to be cheaper, easier to design and implement, more effective, and not subject to costly governmental regulation when compared to the other options. And because of this, it has quickly overtaken the genetic engineering landscape, and has begun to infiltrate the traditional breeding world as well. For example, CRISPR/Cas9 has already been used in wheat to produce experimental lines with increased powdery mildew resistance[134][136] and lines with increased seed size and thousand grain weight[13][125][162], though no variety has been released to date that involves CRISPR/Cas9 editing in its development.

2.10 References

- [1] Collard Bertrand C.Y and Mackill David J, “Marker-assisted selection: an approach for precision plant breeding in the twenty-first century,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1491, pp. 557–572, Feb. 2008.
- [2] P. Gustafson, O. Raskina, X. Ma, and E. Nevo, “Wheat evolution, domestication, and improvement,” in *Wheat Science and Trade*, Wiley-Blackwell, 2009, pp. 3–30.
- [3] Z. Peleg, T. Fahima, A. B. Korol, S. Abbo, and Y. Saranga, “Genetic analysis of wheat domestication and evolution under domestication,” *J Exp Bot*, vol. 62, no. 14, pp. 5051–5061, Oct. 2011.
- [4] W. Bushuk, “Wheat breeding for end-product use,” in *Braun HJ., Altay F., Kronstad W.E., Beniwal S.P.S., McNab A. (eds) Wheat: Prospects for global improvement*, Ankara, Turkey, 1997, vol. 6, pp. 203–211.
- [5] Z. Ivaničová *et al.*, “Heritable heading time variation in wheat lines with the same number of Ppd-B1 gene copies,” *PLOS ONE*, vol. 12, no. 8, pp. 1–15, Aug. 2017.
- [6] A. Kamran, M. Iqbal, and D. Spaner, “Flowering time in wheat (*Triticum aestivum* L.): a key factor for global adaptability,” *Euphytica*, vol. 197, no. 1, pp. 1–26, May 2014.
- [7] FAO, “FAOSTAT database,” *Agricultural crops: Wheat: area harvested*, 2014. [Online]. Available: <http://www.fao.org/>. [Accessed: 04-Dec-2017].
- [8] S. M. Grogan *et al.*, “Allelic Variation in Developmental Genes and Effects on Winter Wheat Heading Date in the U.S. Great Plains,” *PLOS ONE*, vol. 11, no. 4, pp. 1–23, Apr. 2016.
- [9] J. Halliwell *et al.*, “Systematic Investigation of FLOWERING LOCUS T-Like Poaceae Gene Families Identifies the Short-Day Expressed Flowering Pathway Gene, TaFT3 in Wheat (*Triticum aestivum* L.),” *Front. Plant Sci.*, vol. 7, pp. 1–15, Jun. 2016.
- [10] A. Muterko, R. Kalendar, and E. Salina, “Allelic variation at the VERNALIZATION-A1, VRN-B1, VRN-B3, and PHOTOPERIOD-A1 genes in cultivars of *Triticum durum* Desf.,” *Planta*, vol. 244, no. 6, pp. 1253–1263, Dec. 2016.
- [11] M. Zikhali, L. U. Wingen, M. Leverington-Waite, S. Specel, and S. Griffiths, “The identification of new candidate genes *Triticum aestivum* FLOWERING LOCUS T3-B1 (TaFT3-B1) and TARGET OF EAT1 (TaTOE1-B1) controlling the short-day photoperiod response in bread wheat,” *Plant, Cell & Environment*, vol. 40, no. 11, pp. 2678–2690, Nov. 2017.
- [12] A. J. Slade *et al.*, “Development of high amylose wheat through TILLING,” *BMC Plant Biology*, vol. 12, p. 69, May 2012.
- [13] W. Wang *et al.*, “Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat,” *Theor Appl Genet*, vol. 131, no. 11, pp. 2463–2475, Nov. 2018.
- [14] L. Yan, M. Helguera, K. Kato, S. Fukuyama, J. Sherman, and J. Dubcovsky, “Allelic variation at the VRN-1 promoter region in polyploid wheat,” *Theoretical and Applied Genetics*, vol. 109, no. 8, pp. 1677–1686, Nov. 2004.

- [15] D. K. Santra, M. Santra, R. E. Allan, K. G. Campbell, and K. K. Kidwell, "Genetic and Molecular Characterization of Vernalization Genes *Vrn-A1*, *Vrn-B1*, and *Vrn-D1* in Spring Wheat Germplasm from the Pacific Northwest Region of the U.S.A.," *Plant Breeding*, vol. 128, no. 6, pp. 576–584, Dec. 2009.
- [16] J. Zhang, Y. Wang, S. Wu, J. Yang, H. Liu, and Y. Zhou, "A single nucleotide polymorphism at the *Vrn-D1* promoter region in common wheat is associated with vernalization response," *Theoretical and Applied Genetics*, vol. 125, no. 8, pp. 1697–1704, Dec. 2012.
- [17] N. Kippes, A. Chen, X. Zhang, A. J. Lukaszewski, and J. Dubcovsky, "Development and characterization of a spring hexaploid wheat line with no functional *VRN2* genes," *Theor Appl Genet*, vol. 129, no. 7, pp. 1417–1428, Jul. 2016.
- [18] N. Kippes *et al.*, "Fine mapping and epistatic interactions of the vernalization gene *VRN-D4* in hexaploid wheat," *Mol Genet Genomics*, vol. 289, no. 1, pp. 47–62, Feb. 2014.
- [19] T. Kiss *et al.*, "Effects of ambient temperature in association with photoperiod on phenology and on the expressions of major plant developmental genes in wheat (*Triticum aestivum* L.)," *Plant, Cell & Environment*, vol. 40, no. 8, pp. 1629–1642, Aug. 2017.
- [20] N. Kippes *et al.*, "Identification of the *VERNALIZATION 4* gene reveals the origin of spring growth habit in ancient wheats from South Asia," *PNAS*, vol. 112, no. 39, pp. E5401–E5410, Sep. 2015.
- [21] B. Trevaskis, "Wheat gene for all seasons," *PNAS*, vol. 112, no. 39, pp. 11991–11992, Sep. 2015.
- [22] I. Karsai *et al.*, "Effects of photo and thermo cycles on flowering time in barley: a genetical phenomics approach," *J Exp Bot*, vol. 59, no. 10, pp. 2707–2715, Jul. 2008.
- [23] L. Meng *et al.*, "Effects of the *Vrn-D1b* allele associated with facultative growth habit on agronomic traits in common wheat," *Euphytica*, vol. 211, no. 1, pp. 113–122, Sep. 2016.
- [24] H. Nishida *et al.*, "Structural variation in the 5' upstream region of photoperiod-insensitive alleles *Ppd-A1a* and *Ppd-B1a* identified in hexaploid wheat (*Triticum aestivum* L.), and their effect on heading time," *Mol Breeding*, vol. 31, no. 1, pp. 27–37, Jan. 2013.
- [25] F. Chen, M. Gao, J. Zhang, A. Zuo, X. Shang, and D. Cui, "Molecular characterization of vernalization and response genes in bread wheat from the Yellow and Huai Valley of China," *BMC Plant Biol*, vol. 13, no. 199, pp. 1–12, Dec. 2013.
- [26] U. Steinfort, B. Trevaskis, S. Fukai, K. L. Bell, and M. F. Dreccer, "Vernalisation and photoperiod sensitivity in wheat: Impact on canopy development and yield components," *Field Crops Research*, vol. 201, no. Supplement C, pp. 108–121, Feb. 2017.
- [27] B. Wang, D. L. Liu, S. Asseng, I. Macadam, and Q. Yu, "Impact of climate change on wheat flowering time in eastern Australia," *Agricultural and Forest Meteorology*, vol. 209–210, pp. 11–21, Sep. 2015.
- [28] T. P. Khumalo, A. Barnard, L. Maphosa, and T. J. Tsilo, "Impact of Growth Habit and Architecture Genes on Adaptation and Performance of Bread Wheat," in *Wheat*

- Improvement, Management and Utilization*, R. Wanyera and J. Owuoche, Eds. InTech, 2017, pp. 77–95.
- [29] M. S. Lopes *et al.*, “Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change,” *Journal of Experimental Botany*, vol. 66, no. 12, pp. 3477–3486, Jun. 2015.
- [30] E. I. Teixeira, G. Fischer, H. van Velthuisen, C. Walter, and F. Ewert, “Global hot-spots of heat stress on agricultural crops due to climate change,” *Agricultural and Forest Meteorology*, vol. 170, pp. 206–215, Mar. 2013.
- [31] D. B. Lobell and C. B. Field, “Global scale climate–crop yield relationships and the impacts of recent warming,” *Environmental Research Letters*, vol. 2, no. 014002, Mar. 2007.
- [32] X. Wu, H. Liu, X. Li, Y. Tian, and M. D. Mahecha, “Responses of Winter Wheat Yields to Warming-Mediated Vernalization Variations Across Temperate Europe,” *Front. Ecol. Evol.*, vol. 5, no. 126, pp. 1–10, Oct. 2017.
- [33] C. Rosenzweig, A. Iglesias, X. B. Yang, P. R. Epstein, and E. Chivian, “Climate change and extreme weather events; implications for food production, plant diseases, and pests,” *Global change & human health*, vol. 2, no. 2, pp. 90–104, 2001.
- [34] M. A. Semenov, P. Stratonovitch, F. Alghabari, and M. J. Gooding, “Adapting wheat in Europe for climate change,” *Journal of Cereal Science*, vol. 59, no. 3, pp. 245–256, May 2014.
- [35] D. Z. Skinner and B. S. Bellinger, “Freezing tolerance of winter wheat as influenced by extended growth at low temperatures and exposure to freeze–thaw cycles,” *Canadian Journal of Plant Science*, vol. 97, no. 2, pp. 250–256, Sep. 2016.
- [36] P. W. Mote and E. P. Salathé, “Future climate in the Pacific Northwest,” *Climatic Change*, vol. 102, no. 1–2, pp. 29–50, Sep. 2010.
- [37] E. P. Salathé, R. Steed, C. F. Mass, and P. H. Zahn, “A High-Resolution Climate Model for the U.S. Pacific Northwest: Mesoscale Feedbacks and Local Responses to Climate Change,” *J. Climate*, vol. 21, no. 21, pp. 5708–5726, Nov. 2008.
- [38] S. Asseng *et al.*, “Rising temperatures reduce global wheat production,” *Nature Climate Change*, vol. 5, no. 2, pp. 143–147, Feb. 2015.
- [39] “USDA/NASS 2016 State Agriculture Overview for Idaho.” [Online]. Available: https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=IDAHO. [Accessed: 01-Dec-2017].
- [40] “USDA/NASS 2016 State Agriculture Overview for Oregon.” [Online]. Available: https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=OREGON. [Accessed: 27-Nov-2017].
- [41] “USDA/NASS 2016 State Agriculture Overview for Washington.” [Online]. Available: https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=WASHINGTON. [Accessed: 01-Dec-2017].
- [42] A. Cuesta-Marcos *et al.*, “The Relationships between Development and Low Temperature Tolerance in Barley Near Isogenic Lines Differing for Flowering Behavior,” *Plant and Cell Physiology*, vol. 56, no. 12, pp. 2312–2324, Dec. 2015.

- [43] A. Ozturk, O. Caglar, and S. Bulut, "Growth and Yield Response of Facultative Wheat to Winter Sowing, Freezing Sowing and Spring Sowing at Different Seeding Rates," *Journal of Agronomy & Crop Science*, vol. 192, no. 1, pp. 10–16, Feb. 2006.
- [44] D. Gouache, X. Le Bris, M. Bogard, O. Deudon, C. Pagé, and P. Gate, "Evaluating agronomic adaptation options to increasing heat stress under climate change during wheat grain filling in France," *European Journal of Agronomy*, vol. 39, pp. 62–70, May 2012.
- [45] X. Guo *et al.*, "Distribution of the Vrn-D1b allele associated with facultative growth habit in Chinese wheat accessions," *Euphytica*, vol. 206, no. 1, pp. 1–10, Nov. 2015.
- [46] R. S. Ren, M. N. Wang, X. M. Chen, and Z. J. Zhang, "Characterization and molecular mapping of Yr52 for high-temperature adult-plant resistance to stripe rust in spring wheat germplasm PI 183527," *Theor Appl Genet*, vol. 125, no. 5, pp. 847–857, Sep. 2012.
- [47] X. Chen, "Review Article: High-Temperature Adult-Plant Resistance, Key for Sustainable Control of Stripe Rust," vol. 4, pp. 608–627, Mar. 2013.
- [48] G. M. Rosewarne, S. A. Herrera-Foessel, R. P. Singh, J. Huerta-Espino, C. X. Lan, and Z. H. He, "Quantitative trait loci of stripe rust resistance in wheat," *Theor Appl Genet*, vol. 126, no. 10, pp. 2427–2449, Oct. 2013.
- [49] X. M. Chen, "Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat," *Canadian Journal of Plant Pathology*, vol. 27, no. 3, pp. 314–337, Sep. 2005.
- [50] B. Schwessinger, "Fundamental wheat stripe rust research in the 21st century," *New Phytologist*, vol. 213, no. 4, pp. 1625–1631, Mar. 2017.
- [51] G. M. Rosewarne *et al.*, "Analysis of leaf and stripe rust severities reveals pathotype changes and multiple minor QTLs associated with resistance in an Avocet $\times$ Pastor wheat population," *Theor Appl Genet*, vol. 124, no. 7, pp. 1283–1294, May 2012.
- [52] M. Maccaferri *et al.*, "A Genome-Wide Association Study of Resistance to Stripe Rust (*Puccinia striiformis* f. sp. *tritici*) in a Worldwide Collection of Hexaploid Spring Wheat (*Triticum aestivum* L.)," *G3: Genes, Genomes, Genetics*, vol. 5, no. 3, pp. 449–465, Mar. 2015.
- [53] J. A. Kolmer, M. E. Ordóñez, and J. V. Groth, "The Rust Fungi," in *Encyclopedia of Life Sciences (ELS)*, John Wiley & Sons, Ltd, Ed. Chichester, UK: John Wiley & Sons, Ltd, 2009.
- [54] C. C. Mundt, "Pyramiding for Resistance Durability: Theory and Practice," *Phytopathology*, vol. 108, no. 7, pp. 792–802, Apr. 2018.
- [55] P. Bulli, J. Zhang, S. Chao, X. Chen, and M. Pumphrey, "Genetic Architecture of Resistance to Stripe Rust in a Global Winter Wheat Germplasm Collection," *G3: Genes, Genomes, Genetics*, vol. 6, no. 8, pp. 2237–2253, Aug. 2016.
- [56] J. Chen *et al.*, "Genome-wide identification of QTL conferring high-temperature adult-plant (HTAP) resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in wheat," *Mol Breeding*, vol. 29, no. 3, pp. 791–800, Mar. 2012.

- [57] D. M. Vazquez *et al.*, “Genetic analysis of adult plant, quantitative resistance to stripe rust in wheat cultivar ‘Stephens’ in multi-environment trials,” *Theor Appl Genet*, vol. 124, no. 1, pp. 1–11, Jan. 2012.
- [58] X. M. Chen, “Integration of cultivar resistance and fungicide application for control of wheat stripe rust,” *Canadian Journal of Plant Pathology*, vol. 36, no. 3, pp. 311–326, Jul. 2014.
- [59] Y. Naruoka, K. A. Garland-Campbell, and A. H. Carter, “Genome-wide association mapping for stripe rust (*Puccinia striiformis* F. sp. *tritici*) in US Pacific Northwest winter wheat (*Triticum aestivum* L.),” *Theor Appl Genet*, vol. 128, no. 6, pp. 1083–1101, Jun. 2015.
- [60] USDA-ARS, “Small Grain Losses Due to Rust : USDA ARS.” [Online]. Available: <https://www.ars.usda.gov/midwest-area/stpaul/cereal-disease-lab/docs/small-grain-losses-due-to-rust/small-grain-losses-due-to-rust/>. [Accessed: 12-Mar-2019].
- [61] R. M. D. Koebner and R. W. Summers, “21st century wheat breeding: plot selection or plate detection?,” *Trends in Biotechnology*, vol. 21, no. 2, pp. 59–63, Feb. 2003.
- [62] J. He, X. Zhao, A. Laroche, Z.-X. Lu, H. Liu, and Z. Li, “Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding,” *Front. Plant Sci.*, vol. 5, no. 484, Sep. 2014.
- [63] D. Edwards, J. Batley, and R. J. Snowdon, “Accessing complex crop genomes with next-generation sequencing,” *Theor Appl Genet*, vol. 126, no. 1, pp. 1–11, Jan. 2013.
- [64] B. E. Slatko, A. F. Gardner, and F. M. Ausubel, “Overview of Next-Generation Sequencing Technologies,” *Current Protocols in Molecular Biology*, vol. 122, no. 1, p. e59, 2018.
- [65] H. S. Bariana, G. N. Brown, U. K. Bansal, H. Miah, G. E. Standen, and M. Lu, “Breeding triple rust resistant wheat cultivars for Australia using conventional and marker-assisted selection technologies,” *Aust. J. Agric. Res.*, vol. 58, no. 6, pp. 576–587, Jul. 2007.
- [66] Y. Xu and J. H. Crouch, “Marker-Assisted Selection in Plant Breeding: From Publications to Practice,” *Crop Science*, vol. 48, no. 2, pp. 391–407, Mar. 2008.
- [67] B. C. Y. Collard, M. Z. Z. Jahufer, J. B. Brouwer, and E. C. K. Pang, “An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts,” *Euphytica*, vol. 142, no. 1–2, pp. 169–196, Jan. 2005.
- [68] J. Dubcovsky, “Marker-assisted selection in public breeding programs: The wheat experience,” *CROP SCIENCE*, vol. 44, p. 5, 2004.
- [69] P. K. Gupta, P. Langridge, and R. R. Mir, “Marker-assisted wheat breeding: present status and future possibilities,” *Mol Breeding*, vol. 26, no. 2, pp. 145–161, Aug. 2010.
- [70] J. W. Davey, P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter, “Genome-wide genetic marker discovery and genotyping using next-generation sequencing,” *Nature Reviews Genetics*, vol. 12, no. 7, pp. 499–510, Jul. 2011.

- [71] G. Kochert, "Restriction Fragment Length Polymorphism in Plants and Its Implications," in *Plant Genetic Engineering*, vol. 17, B. B. Biswas and J. R. Harris, Eds. New York: Plenum Press, 1991.
- [72] S. Eltaher *et al.*, "Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing," *Front. Genet.*, vol. 9, 2018.
- [73] H. M. William, R. Trethowan, and E. M. Crosby-Galvan, "Wheat breeding assisted by markers: CIMMYT's experience," *Euphytica*, vol. 157, no. 3, pp. 307–319, Oct. 2007.
- [74] P. J. Berkman, K. Lai, M. T. Lorenc, and D. Edwards, "Next-generation sequencing applications for wheat crop improvement," *American Journal of Botany*, vol. 99, no. 2, pp. 365–371, Feb. 2012.
- [75] T. Miedaner and V. Korzun, "Marker-Assisted Selection for Disease Resistance in Wheat and Barley Breeding," *Phytopathology*, vol. 102, no. 6, pp. 560–566, Jun. 2012.
- [76] S. A. Ranade, N. Farooqui, E. Bhattacharya, and A. Verma, "Gene tagging with random amplified polymorphic DNA (RAPD) markers for molecular breeding in plants," *Critical Reviews in Plant Sciences; Boca Raton*, vol. 20, no. 3, p. 251, 2001.
- [77] I. Paran and R. W. Michelmore, "Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce," *Theoret. Appl. Genetics*, vol. 85, no. 8, pp. 985–993, Feb. 1993.
- [78] E. A. Scheef, M. D. Casler, and G. Jung, "Development of Species-Specific SCAR Markers in Bentgrass," *Crop Science*, vol. 43, no. 1, pp. 345–349, Jan. 2003.
- [79] R. K. Varshney, A. Graner, and M. E. Sorrells, "Genomics-assisted breeding for crop improvement," *Trends in Plant Science*, vol. 10, no. 12, pp. 621–630, Dec. 2005.
- [80] Y. N. Shavrukov, "CAPS markers in plant biology," *Russ J Genet Appl Res*, vol. 6, no. 3, pp. 279–287, May 2016.
- [81] A. Ballvora, J. Hesselbach, J. Niewöhner, D. Leister, F. Salamini, and C. Gebhardt, "Marker enrichment and high-resolution map of the segment of potato chromosome VII harbouring the nematode resistance gene Gro1," *Molec. Gen. Genet.*, vol. 249, no. 1, pp. 82–90, Jan. 1995.
- [82] M. T. Cervera *et al.*, "Identification of AFLP molecular markers for resistance against *Melampsora larici-populina* in *Populus*," *Theor. Appl. Genet.*, vol. 93, no. 5–6, pp. 733–737, Oct. 1996.
- [83] T. Vantoai, "Publication : USDA ARS." [Online]. Available: <https://www.ars.usda.gov/research/publications/publication/?seqNo115=69573>. [Accessed: 16-Mar-2019].
- [84] R. K. Varshney, A. Graner, and M. E. Sorrells, "Genic microsatellite markers in plants: features and applications," *Trends in Biotechnology*, vol. 23, no. 1, pp. 48–55, Jan. 2005.
- [85] L. G. Fraser, C. F. Harvey, R. N. Crowhurst, and H. N. De Silva, "EST-derived microsatellites from *Actinidia* species and their potential for mapping," *Theor Appl Genet*, vol. 108, no. 6, pp. 1010–1016, Apr. 2004.

- [86] H. Yu *et al.*, “Gains in QTL Detection Using an Ultra-High Density SNP Map Based on Population Sequencing Relative to Traditional RFLP/SSR Markers,” *PLoS One*, vol. 6, no. 3, Mar. 2011.
- [87] E. S. Lagudah *et al.*, “Gene-specific markers for the wheat gene Lr34/Yr18/Pm38 which confers resistance to multiple fungal pathogens,” *Theor Appl Genet*, vol. 119, no. 5, pp. 889–898, Sep. 2009.
- [88] K. Pajerowska-Mukhtar *et al.*, “Single Nucleotide Polymorphisms in the Allene Oxide Synthase 2 Gene Are Associated With Field Resistance to Late Blight in Populations of Tetraploid Potato Cultivars,” *Genetics*, vol. 181, no. 3, pp. 1115–1127, Mar. 2009.
- [89] X. Hu, M. Sullivan-Gilbert, M. Gupta, and S. A. Thompson, “Mapping of the loci controlling oleic and linolenic acid contents and development of fad2 and fad3 allele-specific markers in canola (*Brassica napus* L.),” *Theor Appl Genet*, vol. 113, no. 3, pp. 497–507, Aug. 2006.
- [90] J. A. Poland, P. J. Bradbury, E. S. Buckler, and R. J. Nelson, “Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize,” *PNAS*, vol. 108, no. 17, pp. 6893–6898, Apr. 2011.
- [91] E. S. Buckler *et al.*, “The Genetic Architecture of Maize Flowering Time,” *Science*, vol. 325, no. 5941, pp. 714–718, Aug. 2009.
- [92] R. D. Fleischmann *et al.*, “Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd,” *Science*, vol. 269, no. 5223, pp. 496–512, Jul. 1995.
- [93] “Genome List - Genome - NCBI.” [Online]. Available: <https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/>. [Accessed: 20-Mar-2019].
- [94] T. P. Michael and S. Jackson, “The First 50 Plant Genomes,” *The Plant Genome*, vol. 6, no. 2, Jul. 2013.
- [95] P. J. Kersey, “Plant genome sequences: past, present, future,” *Current Opinion in Plant Biology*, vol. 48, pp. 1–8, Apr. 2019.
- [96] A. N. Egan, J. Schlueter, and D. M. Spooner, “Applications of next-generation sequencing in plant biology,” *American Journal of Botany*, vol. 99, no. 2, pp. 175–185, Feb. 2012.
- [97] K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers, “GenBank,” *Nucleic Acids Res*, vol. 44, no. Database issue, pp. D67–D72, Jan. 2016.
- [98] R. K. Varshney, S. N. Nayak, G. D. May, and S. A. Jackson, “Next-generation sequencing technologies and their implications for crop genetics and breeding,” *Trends in Biotechnology*, vol. 27, no. 9, pp. 522–530, Sep. 2009.
- [99] H. Kuchel *et al.*, “The successful application of a marker-assisted wheat breeding strategy,” *Mol Breeding*, vol. 20, no. 4, pp. 295–308, Nov. 2007.
- [100] J. W. Davey and M. L. Blaxter, “RADSeq: next-generation population genetics,” *Brief Funct Genomics*, vol. 9, no. 5–6, pp. 416–423, Dec. 2010.
- [101] R. J. Elshire *et al.*, “A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species,” *PLoS One*, vol. 6, no. 5, May 2011.
- [102] J. A. Poland, P. J. Brown, M. E. Sorrells, and J.-L. Jannink, “Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme

- Genotyping-by-Sequencing Approach,” *PLOS ONE*, vol. 7, no. 2, p. e32253, Feb. 2012.
- [103] K. R. Andrews, P. A. Hohenlohe, M. R. Miller, B. K. Hand, J. E. Seeb, and G. Luikart, “Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al.,” *Molecular Ecology*, vol. 23, no. 24, pp. 5943–5946, Oct. 2014.
 - [104] W. Pootakham, C. Sonthirod, C. Naktang, N. Jomchai, D. Sangsrakru, and S. Tangphatsornruang, “Effects of methylation-sensitive enzymes on the enrichment of genic SNPs and the degree of genome complexity reduction in a two-enzyme genotyping-by-sequencing (GBS) approach: a case study in oil palm (*Elaeis guineensis*),” *Mol Breed*, vol. 36, no. 11, p. 154, Nov. 2016.
 - [105] J. A. Poland and T. W. Rife, “Genotyping-by-Sequencing for Plant Breeding and Genetics,” *The Plant Genome*, vol. 5, no. 3, pp. 92–102, Nov. 2012.
 - [106] D. Torkamaneh, J. Laroche, M. Bastien, A. Abed, and F. Belzile, “Fast-GBS: a new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data,” *BMC Bioinformatics*, vol. 18, no. 1, p. 5, Jan. 2017.
 - [107] J. C. Glaubitz et al., “TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline,” *PLOS ONE*, vol. 9, no. 2, p. e90346, Feb. 2014.
 - [108] A. B. Kohn, T. P. Moroz, J. P. Barnes, M. Netherton, and L. L. Moroz, “Single-Cell Semiconductor Sequencing,” *Methods Mol Biol*, vol. 1048, pp. 247–284, May 2014.
 - [109] P. J. Bradbury, Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler, “TASSEL: software for association mapping of complex traits in diverse samples,” *Bioinformatics*, vol. 23, no. 19, pp. 2633–2635, Oct. 2007.
 - [110] L.-X. Yu, P. Zheng, S. Bhamidimarri, X.-P. Liu, and D. Main, “The Impact of Genotyping-by-Sequencing Pipelines on SNP Discovery and Identification of Markers Associated with Verticillium Wilt Resistance in Autotetraploid Alfalfa (*Medicago sativa* L.),” *Front. Plant Sci.*, vol. 8, no. 89, Feb. 2017.
 - [111] S.-B. Wang et al., “Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology,” *Scientific Reports*, vol. 6, p. 19444, Jan. 2016.
 - [112] Z. Zhang et al., “Mixed linear model approach adapted for genome-wide association studies,” *Nat Genet*, vol. 42, no. 4, pp. 355–360, Apr. 2010.
 - [113] V. Jaiswal et al., “Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (*Setaria italica* L.),” *Journal of Cereal Science*, vol. 85, pp. 48–55, Jan. 2019.
 - [114] M. A. Khan, F. Tong, W. Wang, J. He, T. Zhao, and J. Gai, “Analysis of QTL–allele system conferring drought tolerance at seedling stage in a nested association mapping population of soybean [*Glycine max* (L.) Merr.] using a novel GWAS procedure,” *Planta*, vol. 248, no. 4, pp. 947–962, Oct. 2018.
 - [115] S. Alahmad et al., “A Major Root Architecture QTL Responding to Water Limitation in Durum Wheat,” *Front Plant Sci*, vol. 10, Apr. 2019.
 - [116] J. Gao et al., “Linkage mapping and GWAS reveal candidate genes conferring thermotolerance of seed-set in maize,” *J Exp Bot*, vol. erz171, Apr. 2019.

- [117] S.-M. Kim and R. F. Reinke, "A novel resistance gene for bacterial blight in rice, Xa43(t) identified by GWAS, confirmed by QTL mapping using a bi-parental population," *PLOS ONE*, vol. 14, no. 2, p. e0211775, Feb. 2019.
- [118] F. Lin *et al.*, "QTL mapping for maize starch content and candidate gene prediction combined with co-expression network analysis," *Theor Appl Genet*, Mar. 2019.
- [119] M.-G. Shin *et al.*, "Multi-trait analysis of domestication genes in *Cicer arietinum* – *Cicer reticulatum* hybrids with a multidimensional approach: modeling wide crosses for crop improvement," *Plant Science*, Apr. 2019.
- [120] H. Jian *et al.*, "Joint QTL mapping and transcriptome sequencing analysis reveal candidate flowering time genes in *Brassica napus* L.," *BMC Genomics*, vol. 20, no. 1, p. 21, Jan. 2019.
- [121] K. Dossa *et al.*, "The genetic basis of drought tolerance in the high oil crop *Sesamum indicum*," *Plant Biotechnology Journal*, pp. 1–16, Feb. 2019.
- [122] B. Khahani, E. Tavakol, and V. Shariati J., "Genome-wide meta-analysis on yield and yield-related QTLs in barley (*Hordeum vulgare* L.)," *Mol Breeding*, vol. 39, no. 4, p. 56, Mar. 2019.
- [123] M. Yagcioglu, B. Jiang, P. Wang, Y. Wang, S. S. Ellialtioglu, and Y. Weng, "QTL mapping of low temperature germination ability in cucumber," *Euphytica*, vol. 215, no. 4, p. 84, Apr. 2019.
- [124] L. Liu, C. Y. Yuan, M. N. Wang, D. R. See, R. S. Zemetra, and X. M. Chen, "QTL analysis of durable stripe rust resistance in the North American winter wheat cultivar Skiles," *Theor Appl Genet*, Feb. 2019.
- [125] C. Zhang *et al.*, "Identification of Loci and Candidate Genes Responsible for Fiber Length in Upland Cotton (*Gossypium hirsutum* L.) via Association Mapping and Linkage Analyses," *Front Plant Sci*, vol. 10, Feb. 2019.
- [126] S. Mondal *et al.*, "Harnessing Diversity in Wheat to Enhance Grain Yield, Climate Resilience, Disease and Insect Pest Resistance and Nutrition Through Conventional and Modern Breeding Approaches," *Front Plant Sci*, vol. 7, Jul. 2016.
- [127] W. Wang *et al.*, "Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat," *The CRISPR Journal*, vol. 1, no. 1, pp. 65–74, Feb. 2018.
- [128] R. A. McIntosh, C. R. Wellings, and R. F. Park, *Wheat Rusts: An Atlas of Resistance Genes*. Melbourne, Australia: Csiro Publishing, 1995.
- [129] R. A. McIntosh and J. Gyrfas, "Triticum timopheevi as a source of resistance to wheat stem rust," *Z Pflanzenzucht*, 1971.
- [130] P. Bhowmik *et al.*, "Targeted mutagenesis in wheat microspores using CRISPR/Cas9," *Sci Rep*, vol. 8, pp. 1–10, Apr. 2018.
- [131] E. Haque *et al.*, "Application of CRISPR/Cas9 Genome Editing Technology for the Improvement of Crops Cultivated in Tropical Climates: Recent Progress, Prospects, and Challenges," *Front Plant Sci*, vol. 9, pp. 1–12, May 2018.
- [132] E. Waltz, "Nonbrowning GM apple cleared for market," *Nature Biotechnology*, vol. 33, pp. 326–327, Apr. 2015.

- [133] J. Gil-Humanes *et al.*, “High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9,” *The Plant Journal*, vol. 89, no. 6, pp. 1251–1262, Mar. 2017.
- [134] T. Cowan, “Unapproved Genetically Modified Wheat Discovered in Oregon and Montana: Status and Implications,” Congressional Research Service, CRS Report for Congress R43100, Dec. 2014.
- [135] Z. Zhai, T. Sooksa-nguan, and O. K. Vatamaniuk, “Establishing RNA Interference as a Reverse-Genetic Approach for Gene Functional Analysis in Protoplasts,” *Plant Physiology*, vol. 149, no. 2, pp. 642–652, Feb. 2009.
- [136] Y. Zhang *et al.*, “Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat,” *The Plant Journal*, vol. 91, no. 4, pp. 714–724, Aug. 2017.
- [137] E. Stokstad, “Monsanto Pulls the Plug on Genetically Modified Wheat,” *Science*, vol. 304, no. 5674, pp. 1088–1089, May 2004.
- [138] Y. Wang *et al.*, “Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew,” *Nature Biotechnology*, vol. 32, no. 9, pp. 947–951, Sep. 2014.
- [139] A. Malzahn, L. Lowder, and Y. Qi, “Plant genome editing with TALEN and CRISPR,” *Cell & Bioscience*, vol. 7, no. 1, p. 21, Apr. 2017.
- [140] J. K. Joung and J. D. Sander, “TALENs: a widely applicable technology for targeted genome editing,” *Nature Reviews Molecular Cell Biology*, vol. 14, no. 1, pp. 49–55, Jan. 2013.
- [141] T. Gaj, C. A. Gersbach, and C. F. Barbas, “ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering,” *Trends in Biotechnology*, vol. 31, no. 7, pp. 397–405, Jul. 2013.
- [142] D. Carroll, “Genome Engineering With Zinc-Finger Nucleases,” *Genetics*, vol. 188, no. 4, pp. 773–782, Aug. 2011.
- [143] K. Li, G. Wang, T. Andersen, P. Zhou, and W. T. Pu, “Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System,” *PLOS ONE*, vol. 9, no. 8, p. e105779, Aug. 2014.
- [144] Y. Ran *et al.*, “Zinc finger nuclease-mediated precision genome editing of an endogenous gene in hexaploid bread wheat (*Triticum aestivum*) using a DNA repair template,” *Plant Biotechnology Journal*, vol. 16, no. 12, pp. 2088–2101, Apr. 2018.
- [145] D. Jaganathan, K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman, “CRISPR for Crop Improvement: An Update Review,” *Front Plant Sci*, vol. 9, no. 985, Jul. 2018.
- [146] Q. Shan, Y. Wang, J. Li, and C. Gao, “Genome editing in rice and wheat using the CRISPR/Cas system,” *Nature Protocols*, vol. 9, no. 10, pp. 2395–2410, Sep. 2014.
- [147] S. Kim, D. Kim, S. W. Cho, J. Kim, and J.-S. Kim, “Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins,” *Genome Res.*, vol. 24, no. 6, pp. 1012–1019, Jun. 2014.
- [148] L. Arora and A. Narula, “Gene Editing and Crop Improvement Using CRISPR-Cas9 System,” *Front. Plant Sci.*, vol. 8, no. 1932, Nov. 2017.
- [149] Y. Osakabe *et al.*, “CRISPR–Cas9-mediated genome editing in apple and grapevine,” *Nature Protocols*, p. 1, Nov. 2018.

- [150] M. Malnoy *et al.*, “DNA-Free Genetically Edited Grapevine and Apple Protoplast Using CRISPR/Cas9 Ribonucleoproteins,” *Front. Plant Sci.*, vol. 7, no. 1904, pp. 1–9, Dec. 2016.
- [151] J. W. Woo *et al.*, “DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins,” *Nature Biotechnology*, vol. 33, no. 11, pp. 1162–1164, Oct. 2015.
- [152] M. Andersson *et al.*, “Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery,” *Physiologia Plantarum*, vol. 164, no. 4, pp. 378–384, Dec. 2018.
- [153] M. Andersson, H. Turesson, A. Nicolai, A.-S. Fält, M. Samuelsson, and P. Hofvander, “Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts,” *Plant Cell Rep.*, vol. 36, no. 1, pp. 117–128, Jan. 2017.
- [154] T.-Q. Shi *et al.*, “CRISPR/Cas9-based genome editing of the filamentous fungi: the state of the art,” *Appl Microbiol Biotechnol*, vol. 101, no. 20, pp. 7435–7443, Oct. 2017.
- [155] M. Singh, M. Kumar, M. C. Albertsen, J. K. Young, and A. M. Cigan, “Concurrent modifications in the three homeologs of Ms45 gene with CRISPR-Cas9 lead to rapid generation of male sterile bread wheat (*Triticum aestivum* L.),” *Plant Mol Biol*, vol. 97, no. 4, pp. 371–383, Jul. 2018.
- [156] C. Feng, J. Yuan, R. Wang, Y. Liu, J. A. Birchler, and F. Han, “Efficient Targeted Genome Modification in Maize Using CRISPR/Cas9 System,” *Journal of Genetics and Genomics*, vol. 43, no. 1, pp. 37–43, Jan. 2016.
- [157] C. Feng *et al.*, “High-efficiency genome editing using a dmc1 promoter-controlled CRISPR/Cas9 system in maize,” *Plant Biotechnology Journal*, vol. 16, no. 11, pp. 1–10, Mar. 2018.
- [158] W. Jiang, H. Zhou, H. Bi, M. Fromm, B. Yang, and D. P. Weeks, “Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice,” *Nucleic Acids Res.*, vol. 41, no. 20, pp. e188–e188, Nov. 2013.
- [159] D. Fan *et al.*, “Efficient CRISPR/Cas9-mediated Targeted Mutagenesis in Populus in the First Generation,” *Scientific Reports*, vol. 5, p. 12217, Jul. 2015.
- [160] Y. Ito, A. Nishizawa-Yokoi, M. Endo, M. Mikami, and S. Toki, “CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening,” *Biochemical and Biophysical Research Communications*, vol. 467, no. 1, pp. 76–82, Nov. 2015.
- [161] H. Kim, S.-T. Kim, J. Ryu, B.-C. Kang, J.-S. Kim, and S.-G. Kim, “CRISPR/Cpf1-mediated DNA-free plant genome editing,” *Nature Communications*, vol. 8, p. 14406, Feb. 2017.
- [162] A. Peng *et al.*, “Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus,” *Plant Biotechnology Journal*, vol. 15, no. 12, pp. 1509–1519, May 2017.
- [163] Y. Zhang *et al.*, “Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA,” *Nature Communications*, vol. 7, p. 12617, Aug. 2016.

- [164] Z. Liang *et al.*, “Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes,” *Nat Commun*, vol. 8, no. 14261, pp. 1–5, Jan. 2017.
- [165] H. Hamada, Y. Liu, Y. Nagira, R. Miki, N. Taoka, and R. Imai, “Biolistic-delivery-based transient CRISPR/Cas9 expression enables in planta genome editing in wheat,” *Scientific Reports*, vol. 8, no. 1, p. 14422, Sep. 2018.
- [166] X. Cui, “Targeted Gene Editing Using CRISPR/Cas9 in a Wheat Protoplast System,” University of Ottawa, 2017.
- [167] A. Scheben, F. Wolter, J. Batley, H. Puchta, and D. Edwards, “Towards CRISPR/Cas crops – bringing together genomics and genome editing,” *New Phytologist*, vol. 216, no. 3, pp. 682–698, Nov. 2017.

Chapter 3

QTL Analysis of Adult Plant Resistance to Stripe Rust in a Winter Wheat Recombinant Inbred Population

3.1 Abstract

Stripe rust, caused by the basidiomycete fungus *Puccinia striiformis*, is a worldwide pathogen of wheat that causes devastating crop losses. Resistant cultivars have been developed over the last 40 years that have significantly reduced the economic impact of the disease on growers, but in heavy infection years it is mostly controlled through the intensive application of fungicides. The Pacific Northwest of the United States has an ideal climate for stripe rust and also has one of the most diverse race compositions in the country. This has resulted in many waves of epidemics that have overcome most of the resistance genes traditionally used in elite germplasm. In 2002, \$2.5million worth of fungicides was sprayed in the state of Washington to combat the pathogen and prevent tens of millions of dollars in yield losses due to a lack of durable resistance genes. The best way to prevent high yield losses, reduce production costs to growers, and reduce the heavy application of fungicides is to pyramid multiple stripe rust resistance genes into new cultivars. Using Genotyping by Sequencing (GBS), we identified 458,150 variant positions in a Recombinant Inbred Line (RIL) population of 196 individuals derived from a cross between Skiles, an extremely resistant winter wheat variety, and Goetze, a moderately to highly susceptible winter wheat variety, both developed at Oregon State University. A subsequent Genome Wide Association Study (GWAS) identified two Quantitative Trait Loci (QTL) on two chromosomes within the predicted locations of stripe rust resistance genes. Resistance QTL from both parents, when combined together, conferred extremely high levels of stripe rust resistance, indicating that these QTL would be important additions to any future breeding efforts of PNW winter wheat varieties.

3.2 Introduction

Stripe rust (*Puccinia striiformis* Westend. f. sp. *Tritici* Erikss. (*Pst*)), also known as yellow rust, is a basidiomycete fungal disease of wheat (*Triticum aestivum*) that causes significant yield losses around the world[1][2]. It can be found affecting wheat crops in 60 countries across every continent except Antarctica[3][4]. As of 2017, 88 percent of the world's wheat production was susceptible to stripe rust, causing annual yield losses

estimated to be worth \$1billion[5]. Research into its epidemiology and control has been conducted for over a century[4].

In the Pacific Northwest (PNW), the climate is ideal for stripe rust, and this region has been battling yearly infections and frequent epidemics since the mid-twentieth century. Its mild, wet winters and dry hot summers provide the ideal environment for stripe rust to infect, overwinter, over-summer, and re-infect[6][7]. Yield losses were mitigated beginning in the 1980s by the introduction of resistant cultivars and commercially available fungicides[6][4]. However, around the year 2000, new races of stripe rust emerged that overcame much of the genetic resistance[8][9]. Therefore, foliar fungicides must be applied nearly every year on some varieties, and in heavy stripe rust infection years multiple applications of fungicide are used in an attempt to save the crop[6]. Current resistance present in commercial cultivars has saved PNW growers hundreds of millions of dollars, but it is still not sufficient during heavy stripe rust years[6]. The best way to reduce management costs, lower the negative environmental impacts, and improve food safety is to incorporate more durable sources of stripe rust resistance into commercial cultivars[1].

For this reason, most breeders have focused their efforts on incorporating genetic resistance to stripe rust into their germplasm[3][10][11]. Stripe rust resistance comes in two forms, all stage resistance (ASR) and adult plant resistance (APR)[1]. ASR genes are usually race-specific and are active during the entire life cycle of the plant. They are usually qualitative and easily selected for phenotypically in the field[8]. However, because ASR genes are race-specific, cultivars containing single genes or in low-complexity pyramids have been easily overcome by new pathogen races within a very short time period after commercial release, sometimes even just before release, resulting in many of the characterized ASR genes no longer being effective against the pathogen[1][3][8][11][12]. Conversely, APR is typically quantitative and not race-specific, although there are a few exceptions. Adult plant resistance includes a specific category designated High Temperature Adult Plant resistance, or HTAP resistance. This type of resistance is durable, non-race-specific, and is influenced by temperature and plant growth stage. As the wheat plant matures and the temperature increases, the level

of resistance increases, with the most effective daytime temperatures being 25 to 30°C[13]. Seedlings with HTAP are still susceptible to stripe rust, but as the season progresses, the resistance genes protect the plant by reducing the number of new infections, the amount of inoculum, and the spread of inoculum[5][8][9]. Cultivars with the highest level of HTAP can even entirely inhibit sporulation of the stripe rust pathogen[4][10]. While most of the named genes are all-stage and race-specific, the APR genes are less well characterized, with new QTL being identified regularly[1].

In the PNW, new stripe rust races are identified each year, with a race composition that is more diverse than anywhere else in the country[7][14]. With the combination of favorable conditions for the pathogen and susceptible cultivars, complete yield loss has been documented in this region[4]. The use of resistant cultivars is able to reduce the potential yield loss down to just over 8 percent, with chemical control reducing it even further[6][14]. However, chemical control still costs the growers millions of dollars each year, and ASR genes only have an average life-span of 3.5 years when deployed as single-genes in commercial cultivars[7]. Due to this, breeders in the PNW have been focusing on what is called pyramiding resistance genes, or combining multiple resistance genes and types of resistance into one cultivar[6][13][15]. This reduces the chance of the resistance being overcome by the stripe rust pathogen because it would have to overcome multiple resistance mechanisms at once[10]. In 2010 and 2011 the region's conditions were highly favorable for stripe rust and cultivars that had low to moderate HTAP resistance saw 10 to 40 percent yield losses, even with fungicide applications[1]. HTAP resistance genes can have an additive effect in wheat plants, and especially when combined with effective ASR genes, can confer extremely high levels of durable stripe rust resistance[3][9]. Some studies have found that combining four or five HTAP genes can lead to near perfect immunity[10]. One such cultivar that is highly resistant to stripe rust and commonly grown in the PNW is Skiles, a soft white winter wheat released in 2008 that can be virtually free of infection in the field[15][16].

Climate change, end-use quality demands, changing pathogen and pest pressures, and increasing population size all drive the demand for new cultivars[12][17]. This means that while Skiles has performed extremely well for stripe rust resistance in the

PNW, growers and producers regularly require new varieties. Therefore, it is important for breeders to be able to incorporate effective stripe rust resistance like that of Skiles into the next generation of commercial cultivars[1][4][6][18]. Traditionally, breeders used phenotypic selection in the field to determine disease resistance. Levels of resistance and stripe rust infection, however, are extremely sensitive to environmental factors such as moisture, temperature, and wind[7][13]. One year may see a completely different race of stripe rust from the next, and one year may see a completely different inoculum load from the next. This sort of variation from year to year makes phenotypic selection difficult[19]. Therefore, the most efficient and effective way to incorporate these genes is by the use of Marker Assisted Selection (MAS)[2][7][11][20].

MAS overcomes the need for consistent disease conditions and the testing of multiple races by allowing for direct selection on the genotype itself. Markers are DNA sequences that are linked to a trait of interest by a close genetic distance or are a part of the gene causing that trait. In the recent past, marker discovery was done using techniques such as Polymerase Chain Reaction (PCR), size selection with gel electrophoresis, and/or Sanger sequencing. These techniques were cumbersome, low throughput, provided low densities of markers, and were often quite far from the actual gene of interest. Developing markers with these techniques in bread wheat can also be difficult due to its hexaploid nature[17]. The advent of Next Generation Sequencing (NGS) has changed this, however, by providing a high-throughput process that generates Single Nucleotide Polymorphisms (SNPs) as markers. SNPs are generally more tightly linked to the trait of interest or are the actual allele of interest and therefore much more reliable than the previous marker types as they are less likely to be lost due to homologous recombination or neutral mutations[19]. It is important for modern breeders to have access to SNP markers for stripe rust resistance genes if they are to successfully integrate them into new germplasm[17][21].

One of the most efficient ways to identify SNPs in a population is Genotyping by Sequencing (GBS) coupled with a Genome-Wide Association Study (GWAS)[12][22]. GWAS is a common analysis done to determine whether any SNPs are associated with a certain trait, such as disease resistance[23][24]. SNP identification and GWAS have been

performed on many winter wheat cultivars in order to identify stripe rust resistance genes and QTLs, but many of them have not been replicated or validated, and new loci are being discovered regularly, meaning there are likely new sources of resistance that have yet to be characterized[7][17][22].

As mentioned previously, the commercial winter wheat variety Skiles is highly resistant to stripe rust throughout its life cycle, even at somewhat low temperatures, but the genes responsible for this are unknown[15]. The cultivar ‘Goetze’, also a soft white winter wheat released by Oregon State University in 2007, was resistant to stripe rust at its release, but expected to have different resistant genes than Skiles[25]. It has become moderately to highly susceptible in the years since, though still expresses a low level of HTAP in some years. Identifying the sources of resistance in these lines will aid in efforts to pyramid resistance genes and greatly increase the durability of resistance in future PNW winter wheat cultivars. While there are hundreds of QTL associated with stripe rust resistance, they are widely dispersed across the world and most have not had reliable markers developed. Therefore, the goal of this study was to use GBS coupled with GWAS on a Skiles x Goetze recombinant inbred line (RIL) population of these two varieties to characterize the genes and/or QTL associated with their stripe rust resistances to be used for the development of markers for future breeding efforts in the PNW and beyond.

3.3 Materials and Methods

3.3.1 *Germplasm*

The population for this study consisted of 196 F₅ recombinant inbred lines (RILs) from a cross between the cultivars Skiles (high levels of adult plant resistance) and Goetze (moderately to highly susceptible). Skiles is a soft white winter wheat developed by Oregon State University and released in 2008. Goetze is also a soft white winter wheat developed by Oregon State University and released in 2007.

3.3.2 *Experimental Design and Phenotyping*

The 196 progeny lines plus the two parental lines were evaluated for stripe rust

resistance in naturally-inoculated field nurseries in Pullman, WA in 2017 and 2018; Mount Vernon, WA in 2017 and 2018; and in Corvallis, OR in 2018. Mount Vernon and Pullman are approximately 500km apart and have different weather conditions and races of stripe rust[14][15]. Corvallis is approximately 500km from Mount Vernon and over 600km from Pullman, and commonly has different weather conditions and races of stripe rust, although sometimes the race will be the same as one of the other locations (C. Mundt, personal communication). Each trial was planted in the fall of the year before the year indicated following common practices. A single replicate of each of the 196 RILs and two replicates of each parent were planted in Mount Vernon and Pullman in 2017 and 2018 in a randomized design as single 50cm rows with 20cm between rows. The nurseries in Washington were planted, maintained, and phenotyped by Xianming Chen and the Wheat Health, Genetics, and Quality Research team at the United States Department of Agriculture-Agricultural Research Service at Washington State University. Two replicates of each RIL and four replicates of each parent were planted in full 1.5 x 3 meter plots in Corvallis in 2018 in randomized complete blocks. Fertilization and weed control for all sites and years followed common practices. Stripe rust severity was assessed visually as the percentage of the leaf area infected at flowering (Zadoks GS60) in all locations in the spring of the year indicated[26].

3.3.3 DNA Extraction and Genotyping By Sequencing

Leaf tissue from both parents and all 196 RILs was sampled in 2014 and the DNA prepared for sequencing by the Center for Genome Research and Biocomputing at Oregon State University. The DNA was then sent to the Western Regional Small Grains Genotyping Laboratory at Washington State University in Pullman, WA for library preparation and sequencing on the Ion Proton (Thermo Fisher Scientific, Waltham, MA) for SNP variant calling as described by Merriman *et al.* (2012)[27] and Kohn *et al.* (2014)[28]. Raw reads with a length less than 64 bp were discarded. Variants in the population were discovered using the TASSEL 5.0 (Trait Analysis by aSSociation Evolution and Linkage) GBS2 pipeline[29]. A minimum of 10 raw reads per tag was used as a threshold for quality control. The resultant tags were mapped against the wheat

reference genome sequence v1.0 (obtained from the International Wheat Genome Sequencing Consortium[30]) using the Burrows-Wheeler Aligner (BWA) and default parameters[31]. For the prediction of each variant, filtering was performed on the raw variants based on a minimum mapping quality (MQ) of five and a minimum base quality of ten. Post-variant prediction filtering was performed using TASSEL 5.0 to retain high-quality variants. Filtering was performed per variant site with a maximum threshold of 20 percent missing data per position, a minimum minimum allele frequency (MAF) of 0.05, a maximum MAF of 0.5, filtering out of minor SNPs (third allele SNPs), and indels marked as missing. Missing SNP calls were first imputed using the FILLIN option, using parental SNP calls and all options set to default parameters. Post imputation, all heterozygous SNP calls were marked as missing. The resulting dataset was used for all subsequent analyses.

3.3.4 *Statistical Analyses*

All statistical analyses were performed using R version 3.5 in RStudio version 1.1.463[32]. An ANOVA (Analysis of Variance) was performed first to determine the significance of the locations and years (hereby referred to as ‘Treatments’) and the Treatment by genotype (also referred to as ‘Taxa’) interaction. This was done using the ‘aov’ function on a linear model of the form:

$$\text{Disease Severity} = \text{Treatment} + \text{Taxa} + (\text{Treatment} \times \text{Taxa})$$

A Tukey’s HSD (Honest Significant Difference) test was then performed on the ANOVA data in order to determine whether the disease severity scores between each Treatment were significantly different. A Principle Component Analysis (PCA) was also performed to determine the number of principle components necessary for the model to account for population structure. This was done using the ‘prcomp’ function and visualized with ‘ggbiplot’.

3.3.5 *Genome Wide Association Study and Candidate Gene Analysis*

The GWAS was performed using the R package GAPIT[33]. This package

utilizes a compressed mixed linear model (CMLM) that accounts for population structure and kinship[34]. Population structure was accounted for using the first two principal components in order to reduce the occurrence of spurious associations.

The linkage disequilibrium (LD) of each significant SNP was determined using TASSEL 5.0. The locations of the significant SNPs were then extracted from the GAPIT results. Two-thousand base pairs were added to each end of the SNP location for use in Ensembl Plants to search for candidate genes[35]. The region was then explored and annotations of any genes within that region or nearby were determined using the predicted functions from the UniProt database[36] and the NCBI Conserved Domain Search[37].

The United States Department of Agriculture-Agriculture Research Service keeps a repository for wheat data called GrainGenes[38]. This site includes a linkage map of each wheat chromosome with known stripe rust genes and markers placed onto it. The wheat reference genome was published recently in August of 2018, and the vast majority of data in GrainGenes precedes this[30]. This means that the data in GrainGenes is based on linkage mapping and relative distances rather than precise physical locations. In order to more accurately determine the placement of the QTL and significant SNPs found in this study with other markers and stripe rust genes, all available marker sequences from the ‘Wheat, Yr genes and QTL 3B’ linkage map between 0cM and 12.2cM, and 14.6cM and 17.9cM were used in a BLAST (Basic Local Alignment Search Tool) against the RefSeq v1.0 chromosome 3B (urgi.versailles.inra.fr) to determine their physical location. Any result that was below a 98 percent match was not included. The resulting linkage map locations were then aligned by physical location on the reference sequence. The linkage map locations of stripe rust genes on chromosome 3B were also noted and subsequently changed to reflect the new marker placements on the reference genome. This procedure was then repeated for chromosome 3D with the ‘Wheat, Yr genes and QTL 3D’ map.

3.4 Results

3.4.1 *Stripe Rust Severity Scores*

Each location and year show highly variable responses to stripe rust infection throughout the RIL population. Histograms of the number of Taxa (lines) scored at each value in each treatment are shown in Figure 3.1. Examples of a highly resistant line that would receive a score of 0 and a highly susceptible line that would receive a score of 100 are shown in Figure 3.2. Skiles received disease severity ratings of 0%, 0%, 10%, and 5% for Pullman 2017 (Pull17), Pullman 2018 (Pull18), Mount Vernon 2017 (MV17), and Mount Vernon 2018 + Corvallis 2018 (MVC18), respectively. Goetze received ratings of 30%, 10%, 45%, and 75% for Pull17, Pull18, MV17, and MVC18, respectively. In Pull17, 39 of the RILs were also rated at 0% infection like Skiles, and 155 lines had 30% or less infection, like Goetze. In Pull18, only 25 lines had higher infection ratings than Skiles and Goetze. In MV17, five lines were scored at only 5% infection, which is less than the resistant parent Skiles. 33 lines scored the same 10% as Skiles. Overall in MV17, 111 lines scored lower than Goetze and 85 scored higher. In MVC18, 54 lines received disease ratings equal to or better than Skiles, and 53 lines received disease ratings equal to or worse than Goetze. In Pull18, no lines were more than 50% infected by stripe rust. Conversely, in MV17, no lines were completely free of the pathogen.

3.4.2 *SNP Calling*

A total of 2,251,531 variants were identified by TASSEL GBS2. Two lines were removed after filtering due to poor data quality and a high proportion of missing SNP calls. Post-variant call filtering per position resulted in 458,150 SNPs with a minimum allele frequency (MAF) between 0.05 and 0.5, proportion missing between 0 and 0.82, and no heterozygous SNPs.

3.4.3 *Statistical Analyses*

The ANOVA results showed that in terms of stripe rust severity, the location and year Treatments are significantly different (p-value $<2e-16$), the Taxa are significantly different (p-value $<2e-16$), and the Treatment by Taxa interaction is significant (p-value $<2e-16$) (Appendix A: Table A.1). Due to the significant p-values of the Treatments, a Tukey's HSD was performed to determine which Treatments should be analyzed

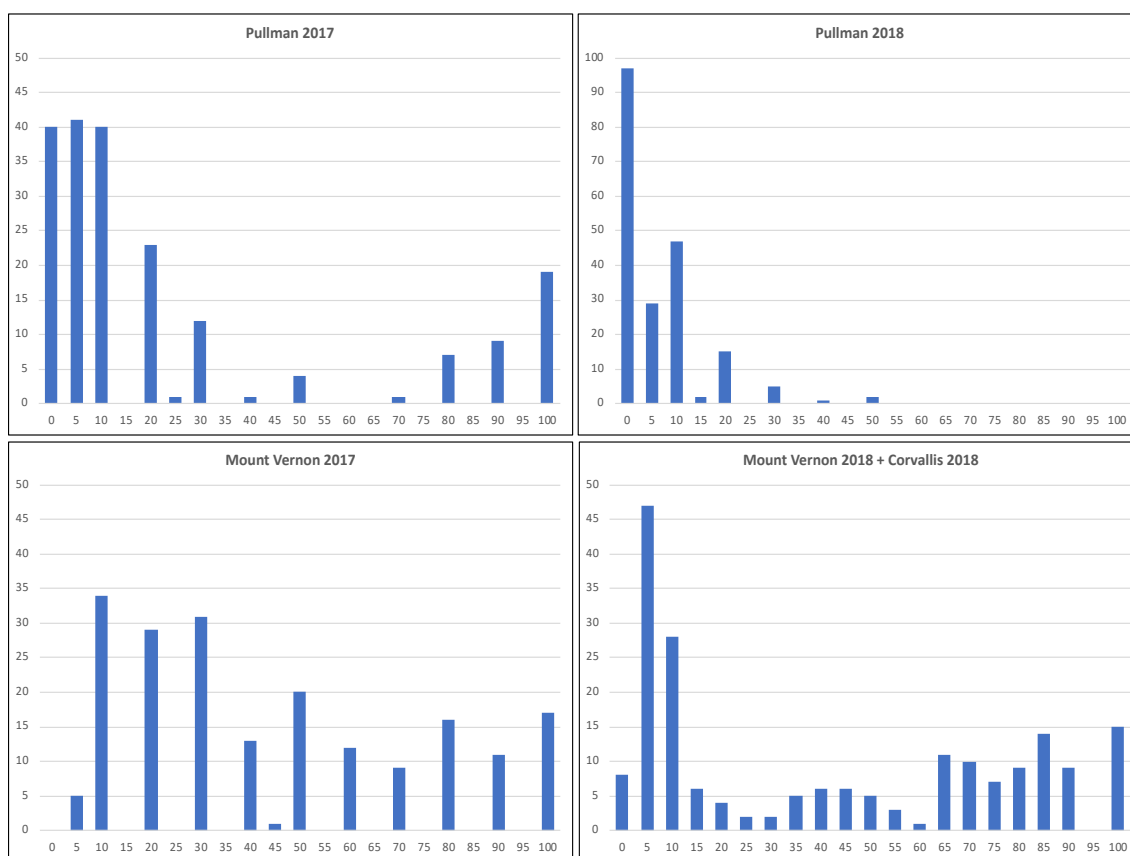


Figure 3.1 Histograms of the number of Taxa scored for each stripe rust severity in each Treatment. Severity is on the horizontal axis and is a measure of the percentage of leaf area affected by the pathogen. The number of Taxa is on the vertical axis.

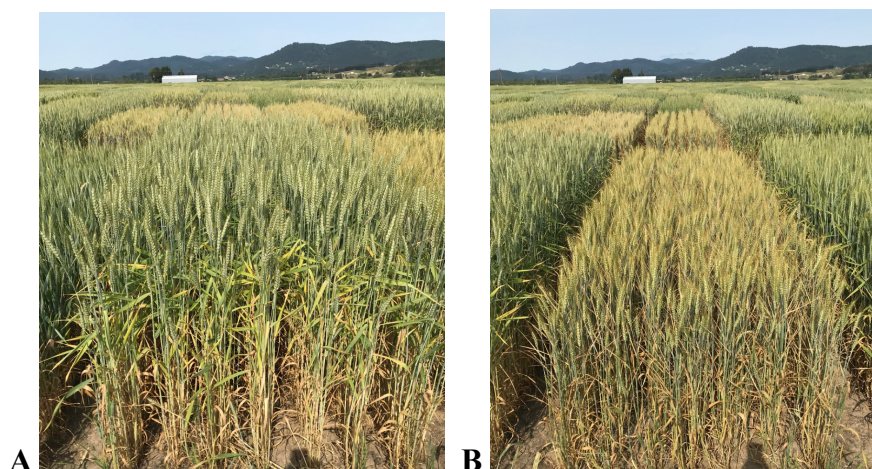


Figure 3.2 Skiles x Goetze RILs at Hyslop Research Farm, Corvallis, Oregon in June 2018. A. A healthy plot with green heads unaffected by stripe rust that received a severity rating of 0. B. A highly stripe rust-infected plot with yellow heads and awns and necrotic, curled leaves that received a score of 100.

separately by GAPIT. The results showed that each treatment is significantly different from the others (p -value $< 1e-7$), except for Corvallis 2018 and Mount Vernon 2018 (p -value = 0.99) (Appendix A: Table A.2). Therefore, each significantly different Treatment was analyzed separately in GAPIT, and Corvallis 2018 (Cor18) was combined with Mount Vernon 2018 (MV18) and analyzed as a single dataset.

The PCA for the first two principle components is shown in Figure 3.3. Goetze is indicated by the blue square, and Skiles is indicated by the green triangle. The contributions of the five Treatments to the PCs is shown by the arrows. The tight cluster to the right that includes Skiles is comprised of lines that are highly resistant to stripe rust in all Treatments. The rest of the lines are spread out based on varying resistance or susceptibility for each Treatment, with the most overall susceptible lines being in the top left corner of the graph. PC1 explains 79.9 percent of the variation and PC2 explains 9.0 percent of the variation. Together they explain 88.9 percent of the variation and are sufficient for use in the GAPIT model for QTL discovery.

3.4.4 *Genome Wide Association Study*

The Manhattan plots produced by the GWAS for each treatment are shown in Figure 3.4, with the green horizontal line indicating the threshold for significance. All dots above the threshold line represent SNPs that are significantly associated with increased stripe rust resistance. Pullman 2017 returned 10 significant SNPs: Two on chromosome 3B, one on chromosome 3D, and seven on the Unidentified chromosome (denoting a genomic scaffold that could not be mapped to any particular chromosome in the reference genome). Pullman 2018 returned 18 significant SNPs: Five on chromosome 3B, four on chromosome 3D, and nine on the Unidentified chromosome. Mount Vernon 2017 returned 13 significant SNPs: Four on chromosome 3B, three on chromosome 3D, and six on the Unidentified chromosome. Mount Vernon 2018 and Corvallis 2018 returned 17 significant SNPs: Four on chromosome 3B, four on chromosome 3D, and nine on the Unidentified chromosome. A list of all unique significant SNPs identified by GAPIT, their location on the reference genome assembly, p -value, R^2 value, and associated treatment are reported in Table 3.1.

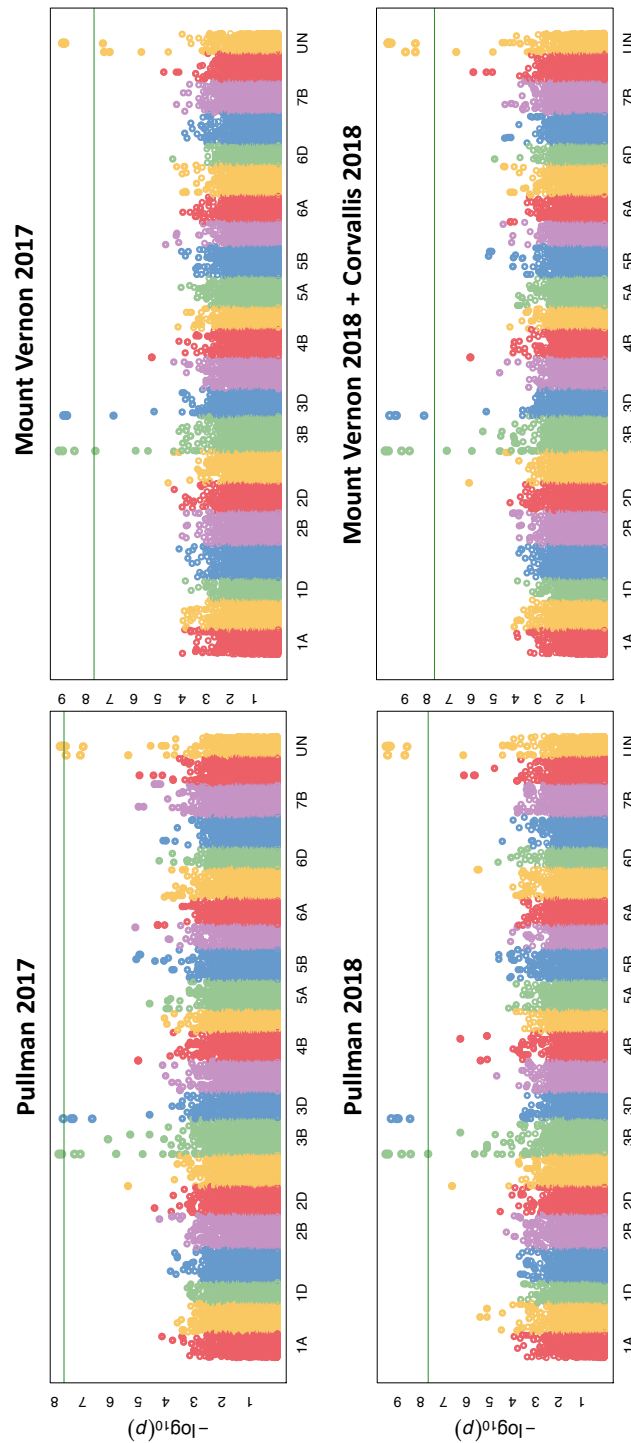


Figure 3.4 Manhattan plots for each treatment produced by GAPIT. The horizontal axis is the chromosome, with ‘UN’ denoting a genomic scaffold that could not be mapped to any particular chromosome in the reference genome. The vertical axis is the p-value for each SNP’s association with the trait of interest. The significant p-value threshold is shown as the green horizontal line. All colored dots are SNPs predicted by TASSEL, and any dot above the green line is a significant SNP.

Table 3.1 Significantly associated SNPs with stripe rust resistance identified in the Manhattan plots. The ‘Chromosome’ and ‘Position’ columns correspond to the chromosome and position of the particular SNP in the reference sequence assembly in base pairs (bp). ‘p-value’ is the significance level used for the Manhattan plots in Figure 3.4. ‘MAF’ is the minor allele frequency at that SNP. The ‘R² of Model without SNP’ is the value of the coefficient of determination (R²) if the SNP were not present. The ‘R² of Model with SNP’ is the R² value with that SNP included and indicates the effect of the QTL. ‘Treatment’ indicates in which treatment type the SNP was significantly associated with stripe rust resistance.

SNP	Chromosome	Position	p-value	MAF	R ² of Model without SNP	R ² of Model with SNP	Treatment In Which SNP is Significant			
S3B_10644041	3B	10644041	2.15E-08	0.49484536	0.004192961	0.182417858		P18		
S3B_5601689	3B	5601689	1.53E-09	0.48969072	0.004192961	0.21453956		P18	MV17	MVC18
S3B_6309966	3B	6309966	1.44E-08	0.49484536	0.025363202	0.204534284	P17	P18	MV17	MVC18
S3B_6309968	3B	6309968	1.80E-08	0.49484536	0.025363202	0.201901283	P17	P18	MV17	MVC18
S3B_6309973	3B	6309973	3.77E-09	0.49226804	0.004192961	0.203497949		P18	MV17	MVC18
S3D_4068757	3D	4068757	1.10E-09	0.49226804	0.004192961	0.21864089		P18	MV17	MVC18
S3D_4068759	3D	4068759	2.04E-08	0.49484536	0.025363202	0.200442713	P17	P18	MV17	MVC18
S3D_4068764	3D	4068764	9.04E-10	0.5	0.004192961	0.221088638		P18	MV17	MVC18
S3D_909572	3D	909572	3.56E-09	0.4871134	0.004192961	0.20419397		P18		MVC18
SUN_234960006	UN	234960006	2.64E-09	0.4871134	0.004192961	0.207848839		P18		MVC18
SUN_242439365	UN	242439365	1.60E-08	0.49742268	0.025363202	0.203244259	P17	P18	MV17	MVC18
SUN_242439370	UN	242439370	1.88E-08	0.49226804	0.025363202	0.20136551	P17	P18	MV17	MVC18
SUN_242439372	UN	242439372	1.80E-08	0.49484536	0.025363202	0.201901283	P17	P18	MV17	MVC18
SUN_242452400	UN	242452400	1.80E-08	0.49484536	0.025363202	0.201901283	P17	P18	MV17	MVC18
SUN_242452405	UN	242452405	2.44E-08	0.49226804	0.025363202	0.19831623	P17	P18	MV17	MVC18
SUN_242452407	UN	242452407	1.80E-08	0.49484536	0.025363202	0.201901283	P17	P18	MV17	MVC18
SUN_34103779	UN	34103779	2.11E-09	0.4871134	0.004192961	0.21062752		P18		MVC18
SUN_36153637	UN	36153637	2.61E-08	0.49226804	0.025363202	0.197552981	P17	P18		MVC18

Table 3.2 The groupings of significant SNPs on each chromosome as determined by their Linkage Disequilibrium values. UN refers to the Unidentified Chromosome. An R² of 1.0 indicates that the SNPs are completely correlated with one another in this population, meaning the presence of one allele can predict the allelic state of the other(s).

Chromosome	Subgroup	Significant SNP	R ² Within Subgroup	R ² of 1.0 With Other Subgroups
3B	Subgroup1	S3B_6309966	1.0	3D Subgroup1 UN Subgroup1
		S3B_6309968		
		S3B_6309973		
	Subgroup2	S3B_5601689		3D Subgroup2 UN Subgroup2
3D	Subgroup1	S3D_10644041	1.0	3B Subgroup1 UN Subgroup1
		S3D_4068757		
		S3D_4068759		
	Subgroup2	S3D_4068764		3B Subgroup2 UN Subgroup2
UN	Subgroup1	S3D_909572	1.0	3B Subgroup1 3D Subgroup1
		SUN_242439365		
		SUN_242439370		
		SUN_242439372		
		SUN_242452400		
		SUN_242452405		
	Subgroup2	SUN_242452407		
		SUN_34103779	1.0	3B Subgroup2 3D Subgroup2
		SUN_234960006		
	Subgroup3	SUN_36153637		

unlinked from one another in this population. The LD values show that there are subgroups within the QTL that are highly linked with one another, but not with other QTL on the same chromosome. There are three subgroups of SNPs on chromosome 3B: Subgroup1 (S3B_6309966, S3B_6309968, and S3B_6309973), Subgroup2 (S3B_5601689 by itself), and Subgroup3 (S3B_10644041 by itself). There are two subgroups of SNPs on chromosome 3D: Subgroup1 (S3D_4068757, S3D_4068759, and S3D_4068764) and Subgroup2 (S3D_909572 by itself). There are three subgroups on the Unidentified chromosome: Subgroup1 (SUN_242439365, SUN_242439370, SUN_242439372, SUN_242452400, SUN_242452405, and SUN_242452407), Subgroup2 (SUN_34103779 and SUN_234960006), and Subgroup3 (SUN_36153637 by itself). Subgroup1 of 3B is highly correlated with Subgroup1 of 3D and Subgroup1 of the Unidentified chromosome (all with R^2 of 1.0). Subgroup1 of 3D is also highly correlated with Subgroup1 of the Unidentified chromosome (R^2 of 1.0). Subgroup2 of 3B is highly correlated with Subgroup2 of 3D and Subgroup2 of the Unidentified chromosome (all with R^2 of 1.0). Subgroup3 of 3B is not highly correlated with any other subgroup or individual SNP. Subgroup3 of the Unidentified chromosome is somewhat highly correlated with Subgroup1 of 3B and 3D (R^2 of between 0.95 and 0.96), and somewhat correlated with Subgroup2 of 3B and 3D (R^2 of 0.89).

The top three Taxa that showed the highest level of resistance in every location and the bottom six Taxa that showed the lowest level of resistance in every location are shown in Table 3.3. The allelic state of each individual at each significant SNP is given, with “N” denoting an unknown allele.

3.4.5 *Candidate Gene Analysis*

The placement of all significant SNPs in relation to one another along with known stripe rust genes are shown in Figure 3.6. The QTL identified on chromosome 3B are all located on the distal end of the short arm, between 5.6Mb and 12.3Mb. SNP S3B_5601689 (Subgroup2) is located 216bp downstream of Gene TraesCS3B02G012400 at 5.6Mb, which is predicted to code for a Knottin scorpion toxin-like superfamily that includes some plant defensins and antifungal proteins. SNPs S3B_6309966,

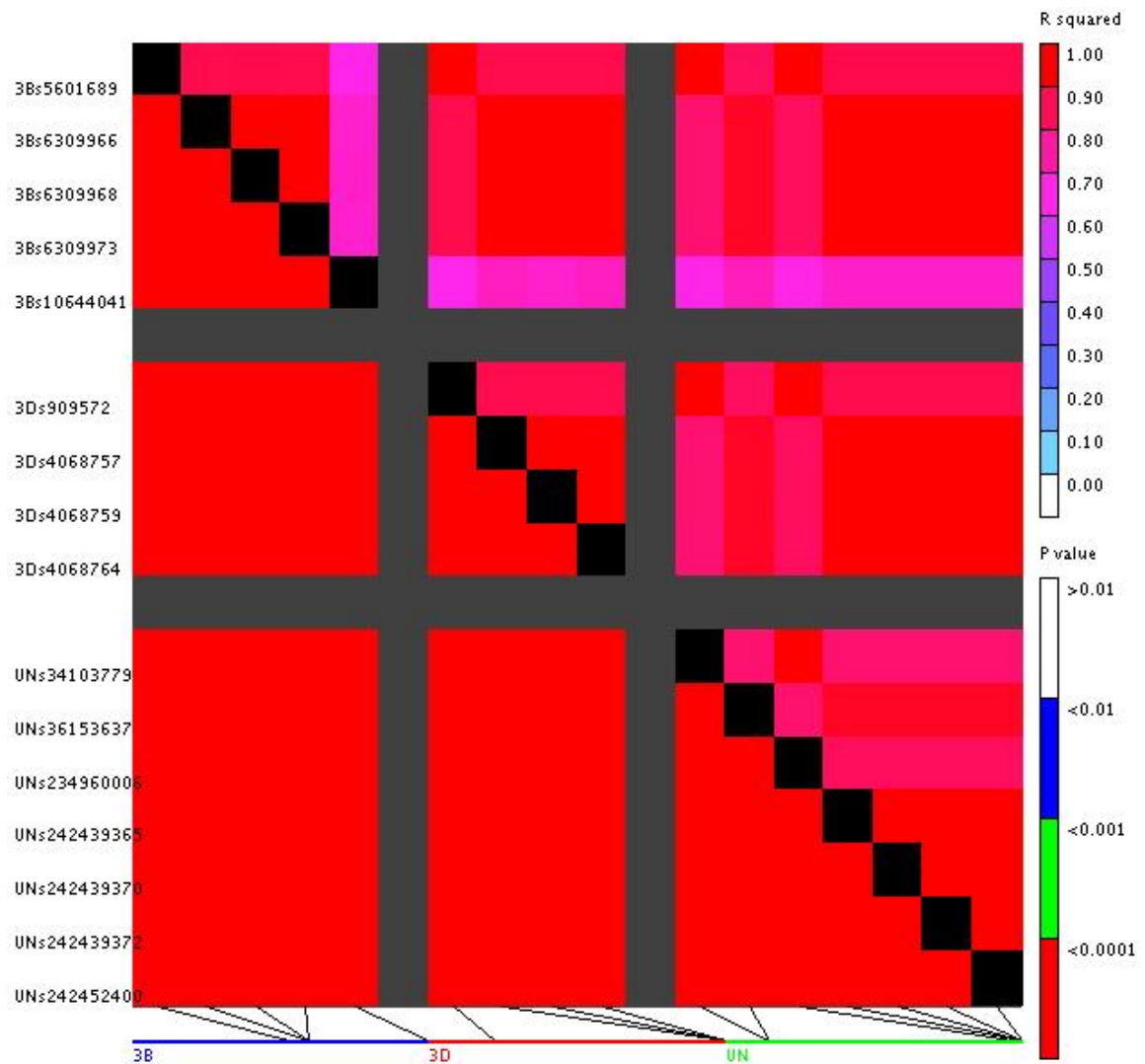


Figure 3.5 Heatmap of linkage disequilibrium (LD) for significant SNPs. Tile colors represent the R^2 value for each pairwise SNP (top right half), where R^2 represents the degree of association between two SNPs. Significance of association between pairs of SNPs is represented in the bottom left half of the figure (The p-value of the R^2 value, legend in the bottom right). The horizontal line at the bottom is each chromosome with the identified SNPs being shown as black lines pointing to the corresponding placement on the plot.

Table 3.3 Significant SNPs from all treatments and their allelic state in the most resistant Taxa and the most susceptible Taxa. The stripe rust severity of each Taxa for each Treatment is given in the columns to the right. 'N' means the allele at that location is unknown.

Taxa Name	Significant SNPs from all Treatments															Treatments								
	3B_56 01689	3B_63 09966	3B_63 09968	3B_63 09973	UN_24 243936	UN_24 243937	UN_24 243937	UN_24 243937	UN_24 245240	UN_24 245240	UN_24 245240	UN_24 245240	UN_24 245240	UN_24 245240	UN_23 496000	3D_40 68757	3D_40 68759	3D_40 8764	Pull17 Severity		Pull18 Severity	MV17 Severity	MV18 Severity	
SG8_163A	C	G	A	G	C	T	C	C	C	T	C	C	G	G	C	G	A	G	0	0	5	5	Top 3 Resistant in every location	
SG8_179B	C	G	A	G	C	T	C	C	C	T	C	C	G	G	C	G	A	G	0	0	5	5		
SG8_195A	N	G	A	G	C	T	C	C	C	T	C	C	N	N	N	G	A	G	0	0	5	5		
SG8_156A	A	C	G	C	G	C	G	G	G	C	C	G	A	T	A	T	C	G	C	100	40	100	100	Bottom 6 most susceptible in every location
SG8_172A	A	C	G	C	G	C	G	G	G	C	C	G	A	T	A	T	C	G	C	100	30	100	100	
SG8_180A	A	C	G	C	G	C	G	C	G	G	C	G	A	T	A	T	C	G	C	100	20	100	100	
SG8_7A	A	C	G	C	G	C	G	C	G	C	G	C	A	T	A	T	C	G	C	100	20	100	100	

S3B_6309968, and S3B_6309973 from chromosome 3B Subgroup1 are located approximately 260bp from gene TraesCS3B02G015100 at 6.3Mb, which is a Pentatricopeptide repeat (PPR) gene in the PLN03218 superfamily. SNP S3B_10644041 (Subgroup3) is located in an untranslated region (UTR). The closest gene, TraesCS3B02G024600, is approximately 10,000 nucleotides away at 10.6Mb. This gene is in the PLN02930 superfamily and is predicted to encode a serine exchange enzyme. Three stripe rust resistance genes, Yr4, Yr30, and Yr57, along with many QTL have also been mapped to this region of chromosome 3B.

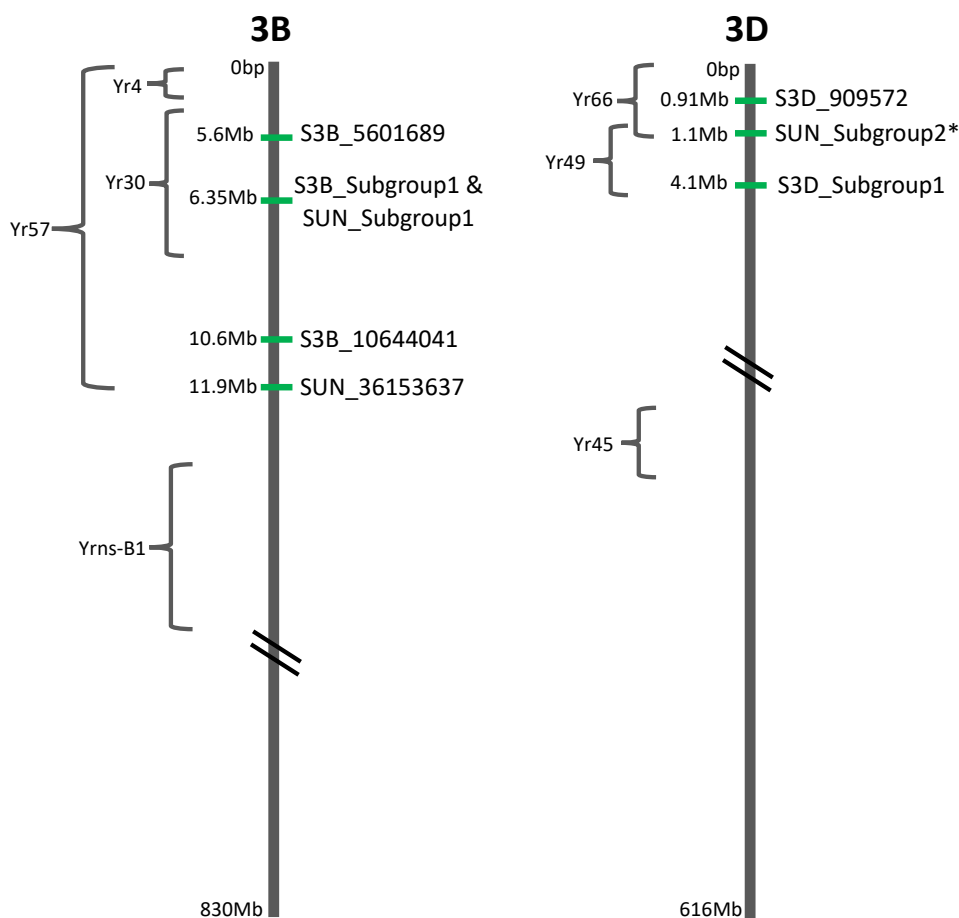


Figure 3.6 Location of the significant SNPs and their subgroups on the respective chromosomes as determined by the reference genome. The known stripe rust genes on each chromosome are also shown in their relative locations based on the GrainGenes maps. The double slash indicates the centromere. Chromosomes are not to scale.

The QTL identified on chromosome 3D are all located on the distal end of the short arm, between 0.91Mb and 4.1Mb. The SNPs in Subgroup1 of chromosome 3D (S3D_4068757, S3D_4068759, and S3D_4068764) are located in an intron of gene TraesCS3D02G011200, which is also a PPR Repeat protein in the PLN03218 superfamily. The SNP in Subgroup2 of 3D, S3D_909572, is located in a UTR. The closest gene, TraesCS3D02G002000, is located approximately 16,000 nucleotides downstream at 0.93Mb. It is in the Paf1 superfamily, which is an RNA polymerase II associated factor. The stripe rust resistance genes Yr66 and Yr49 have also been mapped to this region of chromosome 3D.

The QTL from the Unidentified chromosome segregate onto named chromosomes according to their subgroups. The SNPs from Subgroup2 (SUN_34103779 and SUN_234960006) are most likely located on chromosome 3D at approximately 1.1MB. SNP SUN_34103779 is located within gene TraesCSU02G041500, which is characterized as a leucine-rich repeat domain. Performing a BLAST search of the gene's protein sequence in Ensembl Plants returned a match of 88 percent for gene TraesCS3D02G002400 located at 1.1Mb on chromosome 3D. SNP SUN_234960006 is located at gene TraesCSU02G162800, which is also a leucine-rich repeat domain. Performing a BLAST search in Ensembl Plants also returned an 88 percent match for gene TraesCS3D02G002400 on chromosome 3D. The other two subgroups from the Unidentified chromosome are most likely located on chromosome 3B. SNP SUN_36153637 (Subgroup3) is located in a UTR, with the closest gene, TraesCSU02G046100 located approximately 150,000 nucleotides away. This gene encodes a SINAT5 protein that functions in ubiquitin-mediated degradation resulting in the downregulation of auxin. Performing a BLAST of the protein sequence in Ensembl Plants returned a 99 percent match to gene TraesCS3B02G027400, located at 11.9Mb on chromosome 3B. The SNPs in Subgroup1 (SUN_242439365, SUN_242439370, SUN_242439372, SUN_242452400, SUN_242452405, and SUN_242452407) are located close to genes TraesCSU02G166200 and TraesCSU02G166300, both of which are PPR repeat proteins in the PLN03218 superfamily. Performing a BLAST search for

each gene returned the same result: a 100 percent match to gene TraesCS3B02G015100 located at 6.3MB on chromosome 3B.

The placement of markers and stripe rust genes from the GrainGenes linkage map for chromosome 3B is shown in Figure 3.7. In image A, the markers are ordered according to their location on the linkage map with the physical location to the right of the marker name. It is clear that while the general trend of the physical locations matches the linkage map, there are many markers that are not in order, and some that are extremely distant to their original mapped location. The markers have been reordered in image B to match the reference genome locations, and the locations of the stripe rust genes have also been modified to match. For example, marker IWB63385 is the first marker on the linkage map and is supposed to be the only marker within Yr4. Its location on the physical map, however, is much further away and most likely around 10Mb downstream of Yrns-B1, quite far from Yr4. According to Liu *et al.* (2019), Yr30 is tightly linked with Xgwm533, which is further downstream on the physical map than most of the surrounding markers on the linkage map[15]. SNP S3B_5601689 is located between markers IWB56857 and IWB8756 and within the potential location of Yr57 and Yr30. The SNPs in Subgroup1 of both 3B and the Unidentified chromosome are all located between markers IWB1837 and IWB23378, also within the potential location of Yr57 and Yr30. SNP S3B_10644041 is located between markers IWA758 and IWB13827 within the potential location of Yr30. SNP SUN_36153637 was mapped between markers IWB57993 and IWB39411, also within the potential location of Yr30.

The same process was repeated for the linkage map of chromosome 3D and is shown in Figure 3.8. This chromosome is less well-characterized than 3B and therefore contains significantly fewer markers. Image A shows the markers ordered according to their location on the linkage map, with the physical location on the reference genome to the right, and the locations of stripe rust genes on the linkage map to the left. Upon reordering based on physical location (Image B), most markers remain in the same order, with the exceptions of IWA3573, IWA5030, and IWA6225. SNP S3D_909572 is located more than 32Mb upstream of the first marker in GrainGenes and is within the potential region of both Yr66 and Yr49. The SNPs in Subgroup1 of 3D are located at 4.1Mb,

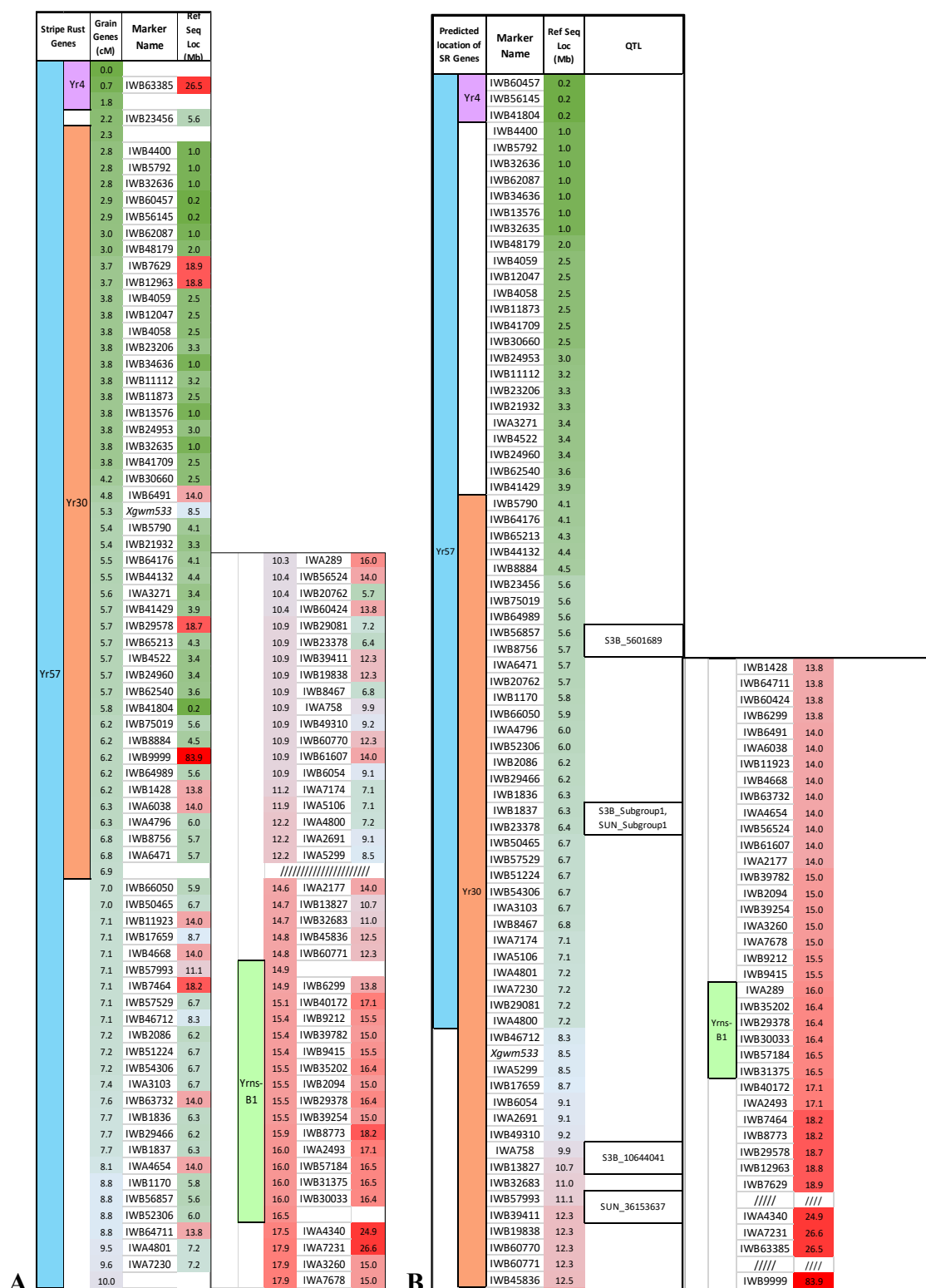


Figure 3.7 Chromosome 3B markers and their location on either the GrainGenes linkage map or the reference genome. (Continued)

C

Predicted location of SR Genes	Marker Name	Ref Seq Loc (Mb)	QTL
Yr57	IWB5790	4.1	
	IWB64176	4.1	
	IWB65213	4.3	
	IWB44132	4.4	
	IWB8884	4.5	
	IWB23456	5.6	
	IWB75019	5.6	
	IWB64989	5.6	
	IWB56857	5.6	
	IWB8756	5.7	S3B_5601689
	IWA6471	5.7	
	IWB20762	5.7	
	IWB1170	5.8	
	IWB66050	5.9	
	IWA4796	6.0	
	IWB52306	6.0	
	IWB2086	6.2	
	IWB29466	6.2	
	IWB1836	6.3	
	IWB1837	6.3	S3B_Subgroup1, SUN_Subgroup1
	IWB23378	6.4	
Yr30	IWB50465	6.7	
	IWB57529	6.7	
	IWB51224	6.7	
	IWB54306	6.7	
	IWA3103	6.7	
	IWB8467	6.8	
	IWA7174	7.1	
	IWA5106	7.1	
	IWA4801	7.2	
	IWA7230	7.2	
	IWB29081	7.2	
	IWA4800	7.2	
	IWB46712	8.3	
	Xgwm533	8.5	
	IWA5299	8.5	
	IWB17659	8.7	
	IWB6054	9.1	
	IWA2691	9.1	
	IWB49310	9.2	
	IWA758	9.9	S3B_10644041
	IWB13827	10.7	
	IWB32683	11.0	
	IWB57993	11.1	SUN_36153637
	IWB39411	12.3	
	IWB19838	12.3	
	IWB60770	12.3	
	IWB60771	12.3	
	IWB45836	12.5	

(Continued) Figure 3.7 Chromosome 3B markers and their location on either the GrainGenes linkage map or the reference genome. Known stripe rust genes and QTL found in this study are marked. **A)** Markers from the GrainGenes ‘Wheat, Yr genes and QTL 3B’ linkage map arranged in map order with the linkage map location of the marker and stripe rust genes (cM) to the left of the marker name. The location of that marker on the wheat reference genome (Mb) is on the right. The locations are color coded with green being the most distal markers and red being the most proximal. **B)** The same markers as in A but arranged only by their location on the reference genome. The stripe rust gene locations have been updated but are still speculative. The SNPs and/or Subgroups found in this study are to the right of the location. **C)** A zoomed-in view of image B covering the SNPs found in this study.

Stripe Rust Genes	Grain Genes (cM)	Marker Name	Ref Seq Loc (Mb)
Yr66	0.4		
	2.4		
Yr49	7.1		
	8.6		
	23.7	IWA1123	33.2
	29.1	IWA1082	48.4
	36	IWA6225	286.7
Yr45	60.2		
	66.9	IWA8610	552
	72		
	89.5	IWA1312	600.2
	89.7	IWA1367	591.9
	89.7	IWA6485	600.3
	89.7	IWA7157	610.3
	90.1	IWA5030	448.3
	90.5	IWA1796	609.2
	90.5	IWA6777	610.3
	90.5	IWA8038	609.2
	90.9	IWA8203	612.9
	92.1	IWA6725	611.3
	92.3	IWA8059	604.4
	92.3	IWA1624	604.4
	92.5	IWA3573	94.5
	92.7	IWA1715	612.9
	92.7	IWA7274	612.9
	92.8	IWA1616	609.1
	92.8	IWA7902	606.1
	92.8	IWA5223	606.9
	92.8	IWA3177	610.3
	93.7	IWA1847	604.7
	93.8	IWA153	604.4
	94.3	IWA7997	603.3
	97.3	IWA643	613.4
	97.3	IWA4081	613.7

A

Stripe Rust Genes	Marker Name	Ref Seq Loc (Mb)	QTL
		0	
Yr49	Yr66	0.91	S3D_909572
		1.1	SUN_Subgroup2*
		4.1	S3D_Subgroup1
		33.2	
		48.4	
		94.5	
		286.7	
	Yr45	448.3	
		552	
		600.2	
		591.9	
		600.3	
		610.3	
		609.2	
		610.3	
		610.3	
		609.2	
		612.9	
		611.3	
		604.4	
		604.4	
		612.9	
		612.9	
		610.3	
		609.1	
		606.1	
		606.9	
		610.3	
		604.7	
		604.4	
		603.3	
		613.4	
		613.7	

B

Figure 3.8 Chromosome 3D markers and their location on either the GrainGenes linkage map or the reference genome. Known stripe rust genes and QTL found in this study are marked. **A)** Markers from the GrainGenes ‘Wheat, Yr genes and QTL 3D’ linkage map arranged in map order with the map location of the marker and stripe rust genes (cM) to the left of the marker name. The location of that marker on the wheat reference genome (Mb) is on the right. The locations are color coded with green being the closest to the distal end of the short arm and red being the furthest away from this point. **B)** The same markers as A but arranged by their location on the reference genome. The stripe rust gene locations have been updated but are still speculative. The SNPs and/or Subgroups found in this study are to the right of the location. *SUN_Subgroup2 is not a highly confident placement, as the BLAST search of the related genes only matched this reference sequence location at 88%.

approximately 29Mb upstream of the closest marker in GrainGenes. The SNPs in Subgroup2 of the Unidentified chromosome are potentially located between these loci at 1.1Mb, but the homology to this location was not as strong as the others in the Unidentified chromosome, so the location is not considered highly confident. Yr45 is located across the centromere from these QTL.

3.5 Discussion

Based on initial response to stripe rust in the field, it was thought that Skiles potentially carried the following resistance genes: Yr4 (3B), Yr5 (2B), Yr15 (1B), Yr24 (1B), Yr29 (1B), and/or Yr32 (2A). Also based on field responses, it was thought that Goetze potentially carried the following resistance genes: Yr4 (3B), Yr5 (2B), Yr10 (1B), Yr16 (2D), Yr17 (2A), Yr29 (1B), and/or Yr32 (2A). Yr4 is a seedling resistance gene, while the disease severity data in this study was taken at the adult stage only, and therefore it is unlikely that the QTL found on 3B is due to Yr4. Yr29 was thought to be the main contributor to Goetze's resistance, but this gene is located on chromosome 1B which was not identified as a source of resistance in this study. As mentioned previously, it was not known which gene or genes were contributing to the resistance in Skiles, and no potential gene identified previously, besides Yr4, is located on a chromosome detected in this study.

In the field, the recombinant inbred lines showed a wide range of resistance and susceptibility to the stripe rust pathogen. They also exhibited transgressive segregation when compared to the parental lines, with some exhibiting extreme susceptibility and others almost complete resistance across locations and years. Pullman 2018 was an anomalous year in that no line showed a stripe rust severity score higher than 50% and most lines were scored within the resistant ranges. Mount Vernon and Corvallis in the same year did not exhibit this unusually low infection rate, however. Therefore, it is possible that in 2018 Pullman had a different stripe rust race than the other two locations that is not as virulent against Goetze's resistance genes. At the time of its release, Goetze was considered resistant to stripe rust, but this resistance quickly waned. Goetze is now considered a susceptible variety, although it sometimes behaves as a moderately susceptible variety with some HTAP resistance as seen in Mount Vernon and Pullman in

2017 where it received scores of 45% and 30% respectively. The fact that Goetze only scored at 10% in Pullman in 2018 combined with the presence of more susceptible RILs in the same Treatment indicate that some of Goetze's resistance genes are likely still effective against some races of the pathogen and can be useful for pyramiding.

Five of the RILs showed the same high level of resistance as Skiles in every location and year. Seventeen RILs showed even higher resistance compared to Skiles in at least one treatment, including five lines that performed better in the most disease heavy Treatment of Mount Vernon 2017. Skiles is considered to be one of the most durable, highly resistant lines in the PNW. The fact that so many lines from this one cross with a moderately susceptible to susceptible variety consistently performed as well as or better than Skiles in a multitude of environments means that there is still room for improvement. This study also shows that the development of varieties that are completely resistant to stripe rust without the need for any chemical treatments in average infection years is realistic. Pyramiding more resistance genes on top of those that Skiles already contains will give growers an extremely reliable, resistant variety that will significantly reduce the amount of pesticides needed and improve yields and profits.

The publication of the wheat reference sequence in August of 2018 makes the ability to map SNPs and QTL more accurate than ever before[30]. However, the map data generated before this are based on linkage maps and relative distances rather than physical distances. In order to accurately map the QTL found in this study with relevant markers, the marker locations from GrainGenes were mapped to the reference genome where possible. Due to the nature of linkage mapping and the large LD decay of hexaploid wheat, many of the markers had been placed incorrectly in the GrainGenes maps, especially on chromosome 3B. The markers were subsequently reordered according to their physical locations and the QTL from this study were added. The results of this are shown in Figures 3.7 and 3.8 for chromosomes 3B and 3D respectively. While the mapping of marker sequences to the reference is a relatively simple process, mapping the locations of the stripe rust genes is not. The sequences of the genes themselves is not known, and some of them are likely QTL rather than single genes, making their placement on the reference genome difficult. Therefore, their locations

have been updated based on associated markers, but is still speculative rather than definite. Further characterization is needed to narrow down the actual locations and find highly reliable markers for each individual gene. This can be done with inoculations of RILs using stripe rust races with known interactions to specific stripe rust resistance genes. Targeted mutagenesis can also be used to knock out proteins within the suspected regions and then analyze the plants for stripe rust susceptibility changes.

Another advantage the publication of the wheat reference genome gives is the ability to search the location of significant SNPs for associated genes and then determine their function. In this study, five of the genes associated with QTLs also have functional annotations with disease resistance potential. Subgroup1 of chromosome 3D and Subgroup1 of chromosome 3B (which was determined to include Subgroup1 of the Unidentified chromosome) are characterized as Pentatricopeptide repeat (PPR) genes. PPRs are tandem repeats of degenerate 35 amino acid motifs that are highly sequence-specific. They are involved in many aspects of RNA metabolism and are extremely common in plant genomes, numbering in the hundreds[39]. This sequence specificity makes PPR genes good candidates to be used as reliable markers. The SNP S3B_5601689 is located just downstream of a gene in chromosome 3B that is predicted to be in the Knottin scorpion toxin-like superfamily, which includes plant defensin and antifungal proteins. Therefore, this gene may be a part of the Yr57 or Yr30 QTL and should be looked into in more detail. The SNPs from Subgroup2 of the Unidentified chromosome were tentatively mapped onto the distal end of chromosome 3D and are located within genes that possess six leucine-rich repeats (LRR) each. A leucine repeat is a conserved eleven-residue sequence motif of the form LxxLxLxxN/CxL that is usually involved in protein-protein interactions and often associated with innate plant immunity to pathogens. These LRR domains are involved in detecting pathogen invasion either on the cell-surface or intracellularly and then triggering downstream defenses. The majority of disease resistance genes encode leucine-rich repeat genes[40]. Therefore, these SNPs are extremely likely to be directly associated with stripe rust resistance genes in this population and would be useful markers for breeding resistant cultivars. While these SNPs are potentially associated with either Yr66 or Yr49, their exact location is not

certain, and this should be resolved so that the resistance gene can be identified and characterized. These two SNPs were only significant in Pull18 and MVC18 but were close to the significance threshold in the other two Treatments and therefore are likely important contributors to overall stripe rust resistance in this population.

A recent study by Liu *et al.* (2019) used Skiles-derived RILs and breeding lines to identify stripe rust resistance QTL on chromosomes 3B, 4B, 1B, 5A, 6B, and 7D[15]. Their ability to detect more QTL than this study is likely due to their wider range of germplasm that included Doubled Haploid lines, PNW cultivars, and breeding lines that involved Skiles in their lineages. The large number of contributing varieties besides Skiles are the likely reason many of their QTL were not identified in this study. Another probable reason for the difference in identified QTL is the difference in Treatments. Liu *et al.* tested their lines in Pullman in 2016, 2017, and 2018; Mount Vernon in 2017 and 2018; and in the greenhouse under controlled dust-inoculation. The increased number of treatments and alternative type of inoculation likely also contributed to an increased ability to detect QTL. Another difference is that Liu *et al.* did not identify a QTL on chromosome 3D as was seen in this study, which suggests that this QTL could be coming from Goetze, as this variety was not an important part of their tested lines. The genotyping by Liu *et al.* was performed on a 90 K Illumina iSelect wheat SNP chip, which restricts the resulting loci to previously reported genomic locations[41]. Genotyping by Sequencing done in this study allows for the discovery of novel SNPs, which could also explain the QTL on chromosome 3D. While both studies identified a QTL in the same general region of chromosome 3B, the SNPs in this area are most likely unique due to the different genotyping methods used.

Liu *et al.* concluded that their QTL on chromosome 3B was most likely a novel stripe rust resistance gene, as Yr4 and Yr57 are seedling resistance genes, Yrns-B1 is associated with marker Xgwm533 which is not linked to the resistance seen in their lines, and Yr30 has been associated with a pseudo black chaff trait not seen in Skiles[42]. The QTL on 3B in this study are in the same general area as those in Liu *et al.*, but the possibility of Yr30 no longer being associated with pseudo-black chaff should not be discounted, as all four QTL span a large portion of the potential location of Yr30.

Therefore, the QTL found on chromosome 3B in this study are most likely associated with Yr30 and/or a novel gene.

The QTL found on chromosome 3D is located in a small region associated with Yr49 and Yr66. Yr66 has not been characterized and the earliest reference and subsequently cited sources reference the Catalogue of Gene Symbols for Wheat: 2013-2014 Supplement, which only lists ‘Bansal U 2014 Personal communication’ as the source of the gene information[43]-[45]. As a result, it is not possible at this time to determine the likelihood of association between the QTL on chromosome 3D and this gene. Yr49, on the other hand, was characterized by Ellis *et al.* (2014) as a race-specific, adult plant resistance gene effective against every Australian stripe rust isolate[46]. As our QTL is associated only with adult plant resistance and is located entirely within the predicted location of this gene, the QTL covering 0.91 to 4.1Mb is likely Yr49. Since Yr49 is known to be susceptible to Chinese races of stripe rust and resistant to Australian races, a greenhouse inoculation test can be performed to determine the validity of this hypothesis.

Unfortunately, when the original cross between Skiles and Goetze was made in 2009, the parents were not kept and propagated alongside the RILs. The seed used in the field tests and in the DNA extraction for the parent lines was from foundation seed stocks. This resulted in both Skiles and Goetze being completely heterozygous at each SNP shown in Table 3.3. Therefore, it is not possible at this time to directly determine which QTL was contributed by which parent. However, there is a clear distinction between the allelic states at each locus and the associated stripe rust severity. Extremely high levels of resistance appear to require the designated alleles at all SNP locations on both chromosomes 3B and 3D. The most susceptible lines from the least severe stripe rust Treatment, Pullman 2018, also have the genotype of the most susceptible lines in Table 3.3; and the most resistant lines from the most severe Treatment, Mount Vernon 2017, have the genotype of the most resistant lines. This result clearly indicates that these SNPs confer extremely high levels of resistance to stripe rust.

3.6 Conclusions

The use of GBS and the publication of the wheat reference genome have allowed for the discovery of novel QTL and SNPs in this population that are strongly associated with highly desirable stripe rust resistance packages that surpass even Skiles. The availability of the reference genome also allowed for the re-ordering of markers on chromosomes 3D and especially 3B that will help future efforts to pinpoint important stripe rust gene locations and narrow down the relevant marker distances in order to improve the efficacy of gene pyramiding. All QTL identified in this study were shown to be important for full resistance in each year and location, and the allelic state at each QTL segregated for resistance and susceptibility. These new markers combined with previously identified markers in the GrainGenes database can be used to further reduce the number of potential locations of important stripe rust resistance genes that have yet to be properly annotated so that they can be used in pyramiding efforts. The SNPs identified in this study can also be used immediately to incorporate elevated levels of resistance to stripe rust into elite germplasm which would greatly benefit growers in this region. These varieties would require significantly less input of chemicals for controlling stripe rust, which would reduce costs, benefit the environment, and reduce the risk of creating pesticide-resistant stripe rust races.

3.7 Acknowledgements

This work was supported by the Oregon Agricultural Experiment Station, Warren E. Kronstad Wheat Research Endowment, the Oregon Wheat Commission, and the ARCS Oregon Chapter Scholar Award.

3.8 References

- [1] R. S. Ren, M. N. Wang, X. M. Chen, and Z. J. Zhang, "Characterization and molecular mapping of Yr52 for high-temperature adult-plant resistance to stripe rust in spring wheat germplasm PI 183527," *Theor Appl Genet*, vol. 125, no. 5, pp. 847–857, Sep. 2012.
- [2] X. Chen, "Review Article: High-Temperature Adult-Plant Resistance, Key for Sustainable Control of Stripe Rust," vol. 4, pp. 608–627, Mar. 2013.
- [3] G. M. Rosewarne, S. A. Herrera-Foessel, R. P. Singh, J. Huerta-Espino, C. X. Lan, and Z. H. He, "Quantitative trait loci of stripe rust resistance in wheat," *Theor Appl Genet*, vol. 126, no. 10, pp. 2427–2449, Oct. 2013.
- [4] X. M. Chen, "Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat," *Canadian Journal of Plant Pathology*, vol. 27, no. 3, pp. 314–337, Sep. 2005.
- [5] B. Schwessinger, "Fundamental wheat stripe rust research in the 21st century," *New Phytologist*, vol. 213, no. 4, pp. 1625–1631, Mar. 2017.
- [6] X. M. Chen, "Integration of cultivar resistance and fungicide application for control of wheat stripe rust," *Canadian Journal of Plant Pathology*, vol. 36, no. 3, pp. 311–326, Jul. 2014.
- [7] Y. Naruoka, K. A. Garland-Campbell, and A. H. Carter, "Genome-wide association mapping for stripe rust (*Puccinia striiformis* F. sp. *tritici*) in US Pacific Northwest winter wheat (*Triticum aestivum* L.)," *Theor Appl Genet*, vol. 128, no. 6, pp. 1083–1101, Jun. 2015.
- [8] M. Maccaferri *et al.*, "A Genome-Wide Association Study of Resistance to Stripe Rust (*Puccinia striiformis* f. sp. *tritici*) in a Worldwide Collection of Hexaploid Spring Wheat (*Triticum aestivum* L.)," *G3: Genes, Genomes, Genetics*, vol. 5, no. 3, pp. 449–465, Mar. 2015.
- [9] M. Dolores Vazquez *et al.*, "Genetic analysis of adult plant, quantitative resistance to stripe rust in wheat cultivar 'Stephens' in multi-environment trials," *Theor Appl Genet*, vol. 124, no. 1, pp. 1–11, Jan. 2012.
- [10] G. M. Rosewarne *et al.*, "Analysis of leaf and stripe rust severities reveals pathotype changes and multiple minor QTLs associated with resistance in an Avocet $\times$ Pastor wheat population," *Theor Appl Genet*, vol. 124, no. 7, pp. 1283–1294, May 2012.
- [11] C. C. Mundt, "Pyramiding for Resistance Durability: Theory and Practice," *Phytopathology*, vol. 108, no. 7, pp. 792–802, Apr. 2018.
- [12] P. Bulli, J. Zhang, S. Chao, X. Chen, and M. Pumphrey, "Genetic Architecture of Resistance to Stripe Rust in a Global Winter Wheat Germplasm Collection," *G3: Genes, Genomes, Genetics*, vol. 6, no. 8, pp. 2237–2253, Aug. 2016.
- [13] J. Chen *et al.*, "Genome-wide identification of QTL conferring high-temperature adult-plant (HTAP) resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in wheat," *Mol Breeding*, vol. 29, no. 3, pp. 791–800, Mar. 2012.
- [14] L. Liu, M. Wang, J. Feng, D. R. See, and X. Chen, "Whole Genome Mapping of Stripe Rust Resistance QTL and Race-specificity Related to Resistance Reduction in Winter Wheat Cultivar Eltan," *Phytopathology*, Feb. 2019.

- [15] L. Liu, C. Y. Yuan, M. N. Wang, D. R. See, R. S. Zemetra, and X. M. Chen, "QTL analysis of durable stripe rust resistance in the North American winter wheat cultivar Skiles," *Theor Appl Genet*, Feb. 2019.
- [16] M. Flowers, C. J. Peterson, A. Hulting, John Burns, S. Guy, and J. Kuehner, "Skiles Soft White Winter Wheat," Corvallis, Or.: Extension Service, Oregon State University, Apr. 2010.
- [17] Q. Zeng *et al.*, "A major QTL co-localized on chromosome 6BL and its epistatic interaction for enhanced wheat stripe rust resistance," *Theor Appl Genet*, Feb. 2019.
- [18] Z. Dong *et al.*, "Validation and characterization of a QTL for adult plant resistance to stripe rust on wheat chromosome arm 6BS (Yr78)," *Theor Appl Genet*, vol. 130, no. 10, pp. 2127–2137, Oct. 2017.
- [19] Q. Zeng *et al.*, "Genome-wide Mapping for Stripe Rust Resistance Loci in Common Wheat Cultivar Qinnong 142," *Plant Disease*, Aug. 2018.
- [20] J. Mu *et al.*, "Genetic architecture of wheat stripe rust resistance revealed by combining QTL mapping using SNP-based genetic maps and bulked segregant analysis," *Theor Appl Genet*, vol. 132, no. 2, pp. 443–455, Feb. 2019.
- [21] J. Ma *et al.*, "Identification and validation of a novel major QTL for all-stage stripe rust resistance on 1BL in the winter wheat line 20828," *Theor Appl Genet*, Jan. 2019.
- [22] P. Juliana *et al.*, "Genome-wide association mapping for resistance to leaf rust, stripe rust and tan spot in wheat reveals potential candidate genes," *Theoretical and Applied Genetics*, vol. 131, no. 7, pp. 1405–1422, Jul. 2018.
- [23] Collard Bertrand C.Y and Mackill David J, "Marker-assisted selection: an approach for precision plant breeding in the twenty-first century," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1491, pp. 557–572, Feb. 2008.
- [24] P. J. Berkman, K. Lai, M. T. Lorenc, and D. Edwards, "Next-generation sequencing applications for wheat crop improvement," *American Journal of Botany*, vol. 99, no. 2, pp. 365–371, Feb. 2012.
- [25] M. D. Flowers and C. J. Peterson, "Goetze, soft winter wheat," Corvallis, Or.: Extension Service, Oregon State University, Apr. 2008.
- [26] J. C. Zadoks, T. T. Chang, and C. F. Konzak, "A decimal code for the growth stages of cereals," *Weed Research*, vol. 14, no. 6, pp. 415–421, Dec. 1974.
- [27] B. Merriman, I. T. R. Team, and J. M. Rothberg, "Progress in Ion Torrent semiconductor chip based sequencing," *ELECTROPHORESIS*, vol. 33, no. 23, pp. 3397–3417, Sep. 2012.
- [28] A. B. Kohn, T. P. Moroz, J. P. Barnes, M. Netherton, and L. L. Moroz, "Single-Cell Semiconductor Sequencing," *Methods Mol Biol*, vol. 1048, pp. 247–284, May 2014.
- [29] P. J. Bradbury, Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler, "TASSEL: software for association mapping of complex traits in diverse samples," *Bioinformatics*, vol. 23, no. 19, pp. 2633–2635, Oct. 2007.
- [30] The International Wheat Genome Sequencing Consortium (IWGSC) *et al.*, "Shifting the limits in wheat research and breeding using a fully annotated reference genome," *Science*, vol. 361, no. 6403, p. eaar7191, Aug. 2018.

- [31] H. Li and R. Durbin, “Fast and accurate short read alignment with Burrows–Wheeler transform,” *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, Jul. 2009.
- [32] RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.
- [33] A. E. Lipka *et al.*, “GAPIT: genome association and prediction integrated tool,” *Bioinformatics*, vol. 28, no. 18, pp. 2397–2399, Sep. 2012.
- [34] Z. Zhang *et al.*, “Mixed linear model approach adapted for genome-wide association studies,” *Nat Genet*, vol. 42, no. 4, pp. 355–360, Apr. 2010.
- [35] S. E. Hunt *et al.*, “Ensembl variation resources,” *Database (Oxford)*, vol. 2018, pp. 1–12, Jan. 2018.
- [36] “UniProt: a worldwide hub of protein knowledge,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D506–D515, Jan. 2019.
- [37] A. Marchler-Bauer *et al.*, “CDD/SPARCLE: functional classification of proteins via subfamily domain architectures,” *Nucleic Acids Res*, vol. 45, no. Database issue, pp. D200–D203, Jan. 2017.
- [38] USDA-ARS, Wheat, Yr genes and QTL. In *_GrainGenes_*. Available: <https://wheat.pw.usda.gov/cgi-bin/GG3/browse.cgi?class=map>. [Accessed: 15-Apr-2019].
- [39] I. D. Small and N. Peeters, “The PPR motif – a TPR-related motif prevalent in plant organellar proteins,” *Trends in Biochemical Sciences*, vol. 25, no. 2, pp. 45–47, Feb. 2000.
- [40] M. Padmanabhan, P. Cournoyer, and S. P. Dinesh-Kumar, “The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities,” *Cell Microbiol*, vol. 11, no. 2, pp. 191–198, Feb. 2009.
- [41] S. Wang *et al.*, “Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array,” *Plant Biotechnology Journal*, vol. 12, no. 6, pp. 787–796, 2014.
- [42] W. Spielmeyer, P. J. Sharp, and E. S. Lagudah, “Identification and Validation of Markers Linked to Broad-Spectrum Stem Rust Resistance Gene Sr2 in Wheat (*Triticum aestivum* L.),” *Crop Science*, vol. 43, no. 1, pp. 333–336, Jan. 2003.
- [43] S. Zheng *et al.*, “Evaluating the contribution of Yr genes to stripe rust resistance breeding through marker-assisted detection in wheat,” *Euphytica*, vol. 213, no. 2, p. 50, Feb. 2017.
- [44] H. Bariana, K. Forrest, N. Qureshi, H. Miah, M. Hayden, and U. Bansal, “Adult plant stripe rust resistance gene Yr71 maps close to Lr24 in chromosome 3D of common wheat,” *Mol Breeding*, vol. 36, no. 7, p. 98, Jul. 2016.
- [45] R. A. McIntosh, J. Dubcovsky, W. J. Rogers, C. Morris, R. Appels, and X. C. Xia, “CATALOGUE OF GENE SYMBOLS FOR WHEAT: 2013-2014 SUPPLEMENT.”
- [46] J. G. Ellis, E. S. Lagudah, W. Spielmeyer, and P. N. Dodds, “The past, present and future of breeding rust resistant wheat,” *Front Plant Sci*, vol. 5, no. 641, pp. 1–13, Nov. 2014.

Chapter 4

Genome-Wide Analysis of Vernalization Requirements in a Winter Wheat Recombinant Inbred Population

4.1 Abstract

The process of vernalization is essential for flowering in winter wheat, but may not be satisfied in years with warmer winters brought on by climate change. Insufficient vernalization in full winter varieties leads to delayed or no flowering, resulting in significant yield loss. The most promising way to combat this problem while retaining the high yield potential of winter wheat is to use facultative varieties with reduced vernalization requirements that are adapted to local environments and cultivation practices. This study identified quantitative trait loci (QTL) conferring reduced vernalization requirements in a recombinant inbred line (RIL) population of 196 individuals derived from a cross between two elite cultivars developed at Oregon State University: ‘Skiles’, a winter line requiring full vernalization for timely heading, and ‘Goetze’, a facultative winter line with reduced vernalization requirements. Days to heading were recorded after growth in a greenhouse following four different vernalization treatments: zero weeks, two weeks, four weeks, and six weeks. The RIL population was analyzed using Genotyping by Sequencing (GBS) with Ion Proton technology and aligned to the wheat reference genome, resulting in 458,150 single nucleotide polymorphisms (SNPs). Subsequent QTL analysis detected 13 significant SNPs on four different chromosomes. The significant QTL on chromosome 1D was close to a photoperiod gene that may contribute to day-neutral phenotypes and may be useful in breeding for facultative varieties. The QTL found on chromosome 3B may be associated with a poorly characterized phytochrome gene and should be examined further to elucidate the genes/loci responsible for reduced vernalization requirements in these lines. The SNP-gene association on chromosome 5B reinforced the significance of a large, important QTL in this region contributed by Goetze that will be useful for immediate breeding efforts. The QTL identified on chromosome 5D revealed the presence of a facultative vernalization allele previously unknown to be in this germplasm that will also be useful for future breeding efforts to obtain facultative varieties.

4.2 Introduction

Wheat (*Triticum aestivum*) was domesticated around 8,000 years ago in the Fertile Crescent, and since then has spread around the world to become one of its most vital food crops[1][2]. Currently, it supplies more than 20 percent of the daily calories and protein consumed by almost 35 percent of the world's population[3]–[5]. Wheat is cultivated on every continent except Antarctica, and as of 2009 it was grown on 225 million hectares, meaning it was grown over more land area and across more regional climate zones than any other crop[5]–[7]. The main reason for wheat's success in this multitude of environments is its flowering time adaptability[8][9]. It has been bred to produce high yields in sub-tropical to temperate steppe climate zones, and from rain-fed to dryland conditions. All of this is possible due to modifications in flowering time via the alteration of genes in the flowering pathway that are sensitive to day length (photoperiod) or temperature (vernalization)[4][7]. Both types of genes work in concert to prevent flowering during cold winters, promote flowering during spring warmth, speed up seed filling to avoid summer heat, or slow seed filling to take advantage of cooler early summers.

The main flowering pathway involves four characterized vernalization genes (VRN1, VRN2, VRN3, and VRN4), a long-day photoperiod sensitivity gene (PPD1), and a putative short-day sensitivity gene (TaFT3)[8][10]. The hexaploid nature and diverse parentage of modern wheat varieties means there are multiple genes that can be involved in the process of flowering, as well as an array of possible alleles, each with the ability to affect the flowering pathway in different ways. The functions of these genes have been studied to varying degrees, with the major genes such as VRN1, a flowering promoter that is up-regulated by cold temperatures, being the most well-characterized[11]. Allelic variation for VRN1 is found on all three genomes and is widely dispersed around the world[12][13]. The interactions of vernalization genes often regulate expression, such as VRN2 being a flowering repressor that is down-regulated by VRN1 during the cold days of winter[10]. VRN2 is located on all three genomes as well but has not been found to contain much allelic diversity[14][15]. VRN3 is the integrator of the vernalization and photoperiod pathways and acts by up-regulating VRN1 during vernalization and also

during long days, thus promoting flowering[10][16]. This gene is located on all three genomes, but allelic variation has only been found on the B genome[9]. VRN4 has been shown to be up-regulated during vernalization, when it down-regulates VRN2 and up-regulates VRN1. VRN4 is only located on the D genome and is thought to have arisen from a retrotransposon insertion event that duplicated Vrn-A1 onto the short arm of chromosome 5D[15][17][18]. PPD1 is a long-day sensitivity gene that up-regulates VRN3, and thereby flowering, during long days, and is also found on all three homoeologous chromosomes[8][10]. Finally, the gene TaFT3 is a putative short-day sensitivity gene orthologous to the barley (*Hordeum vulgare*) short-day sensitivity gene PPD-H2[19]. It has been proposed that TaTF3 promotes flowering during short days via up-regulation, but is down-regulated during long days. Where TaFT3 acts within the flowering pathway is currently unknown[8][10]. Other genes are known to be involved in the flowering pathway, but these are generally newly-discovered, contribute at lower levels to flowering, and are therefore often poorly characterized and understood[20].

Genes in the flowering pathway interact with each other in a number of ways to affect flowering date and thereby adaptability to local climates[8][9]. The manipulation of these genes and their allelic states results in a plethora of flowering times and cold tolerances that have been used to adapt wheat to a multitude of environments by ensuring the plants do not enter floral initiation when temperatures are too hot or too cold for grain production[13][21][22]. This means that while wheat as a species is adapted to multiple environments, within each regional environment the local cultivars are extremely susceptible to even slight changes in temperature[23]–[25]. Global climate change has already begun to affect the duration and stability of winter temperatures, and this change in stable weather patterns is only expected to become more extreme in the near future[22][24]–[27]. In the Pacific Northwest (PNW), the winter temperatures are predicted to increase more dramatically than the worldwide average in some of the most important wheat growing regions[28][29]. This reduction in the number of cumulative vernalization days can have a profound effect on winter wheat yields[21][22][27].

Therefore, researchers and growers have come to the conclusion that the best way to avoid massive yield losses due to unpredictable weather is to further manipulate the

flowering pathway and create what are called facultative wheats[7][22][23][30]. This growth habit does not have a strict definition or genotype, and its only condition is that it has a reduced or non-existent vernalization requirement, but will not enter reproductive growth during the cold of winter[13][31]-[33]. Some breeding programs have created facultative lines using only, or mostly, phenotypic selection. However, the genotypes of these wheat varieties are extremely variable, poorly understood, and often linked with poor winter-hardiness. The genotypes of these varieties are also highly dependent upon parentage as well as local environmental requirements[34][35]. Due to this, the genetic factors contributing to differences in vernalization requirement are not well understood[36]. Breeders would greatly benefit from the ability to use marker-assisted selection (MAS) in the development of facultative varieties, but the markers need to be specific for their particular region and germplasm, as the genetic contributors to vernalization requirements differ greatly from region to region.

A widely-used technique for identifying single nucleotide polymorphisms (SNPs) in an adapted population is combining Genotyping by Sequencing (GBS) with a Genome-Wide Association Study (GWAS)[37]-[39]. GBS allows for the genotyping of hundreds of samples using a restriction-enzyme guided reduced representation of the genome approach that is suitable for species with very large genomes. GWAS allows for the discovery of quantitative trait loci (QTL) that are associated with traits of interest. Once QTL have been identified, the genotyping data from GBS, in the form of SNPs, can be used in combination with the newly-published wheat reference genome to find nearby genes associated with that trait and to develop markers using the SNP locations that can be used in MAS[40]-[42].

It has been shown in four independent studies that allelic variation in the VRN1 genes are responsible for a large proportion of the vernalization requirement differences in winter wheat varieties[17][36][43][44]. However, the four studies concluded that different polymorphisms were responsible for each of the observed phenological differences. Therefore, the polymorphisms found by each study are unreliable as markers for use in other breeding programs, especially those that do not use the same genetic material as the respective studies. While it is clear that allelic variation in the VRN1

genes is important for the development of facultative winter wheat varieties, it is insufficient for use in MAS in the PNW[8][32]. As a result, breeders of PNW wheat would greatly benefit from the discovery of markers in adapted germplasm that can easily be used for the development of facultative varieties in their region without the need for extensive validation. Therefore, this study was conducted using two elite cultivars from the PNW: ‘Skiles’ and ‘Goetze’. Skiles is a cold tolerant, full winter cultivar, meaning it requires the full six weeks of vernalization to flower at the appropriate time. Goetze is a facultative winter cultivar that does not possess much cold tolerance. Both lines were released by Oregon State University and have agronomic and quality traits important for the PNW. The majority of the genes in both cultivars contributing to their diverse flowering times are unknown and identifying the genetic sources of the facultative growth habit and developing markers for those QTL would directly benefit future breeding efforts in this region. Therefore, the objective of this study was to use GBS coupled with GWAS on a Skiles x Goetze recombinant inbred line (RIL) population to characterize the genes and/or QTL associated with variable vernalization requirements.

4.3 Materials and Methods

4.3.1 *Germplasm*

A recombinant inbred line (RIL) population was developed from a cross between the winter wheat varieties Skiles and Goetze. The cross was made in 2009 and the progeny were then propagated by single seed descent for five generations in order to create a homozygous, segregating RIL population. The RIL population was comprised of 196 F₅ derived F₆ lines. Skiles is a soft white winter wheat developed by Oregon State University and released in 2008. It is best adapted to Northeastern Oregon and Southeastern Washington, requires full vernalization, and has good winter-hardiness[45]. Goetze is also a soft white winter wheat developed by Oregon State University and released in 2007. Goetze can be grown in Eastern Oregon, but is best adapted to North Central and Northwestern Oregon. Unlike Skiles, Goetze is an early-maturing, facultative variety with little winter-hardiness[46]. It is known that at least some of Goetze’s facultative growth habit comes from a Vrn-B1 allele, but the other vernalization

genes are not well characterized in either variety. ‘Bobtail’, a soft white winter wheat developed by Oregon State University, was used as a winter check as this variety requires full vernalization. ‘Louise’, a soft white spring wheat developed by Washington State University, was used as a spring check[47].

4.3.2 *Experimental Design and Phenotyping*

The 196 progeny lines, the parental varieties, Bobtail, and Louise were planted in 8x16 well flats in a randomized complete block design with one plant per well. The seeds were allowed to germinate and grow for two weeks in the greenhouse at 21-24°C under 16 hours of light per day before being either directly transplanted into cones or moved to the vernalization chamber at 6°C under 10 hours of light per day. The treatments used were four different vernalization times: zero weeks, two weeks, four weeks, and six weeks. Zero weeks of vernalization reproduces the conditions needed for spring wheat growth, and six weeks of vernalization fully satisfies the vernalization needs of full winter lines in the PNW. The zero week treatment included three replicates of each line, parents, and checks. The other three treatments only included two replicates of each due to space constraints. The planting date of each treatment was staggered so that all flats were moved from the vernalization chamber to the greenhouse on the same day. For growth in the greenhouse, the plantlets were transplanted into 164ml cones (Stuewe & Sons, Inc., Tangent, OR, Cat. #RLC4) placed in Ray Leach Cone-tainers (Cat. # RL98), with six flats for each block of the vernalized treatments and nine flats for the zero week treatment. Each flat was placed in a flow tray (Cat. # FLSSLITE) filled with water and was fertilized with one tablespoon of Peters Professional 20-20-20 fertilizer (J.R. Peters Inc, Allentown, PA) to two gallons of water once during growth. Heading dates were recorded as days since planting, minus any time spent in the vernalization chamber. Heading date was recorded when the first tiller reached Zadoks stage GS59, meaning the head had completely emerged above the flag leaf ligule[48]. Any line that did not head after 135 days was recorded as having a heading date of 200 for use in subsequent analyses, as a mark of 0 would have skewed the data.

4.3.3 DNA Extraction and Genotyping by Sequencing

Leaf tissue was sampled from all 196 lines plus both parents in 2014 and the DNA was prepared for GBS analysis by the Center for Genome Research and Biocomputing at Oregon State University. The DNA was then sent to the Western Regional Small Grains Genotyping Laboratory at Washington State University in Pullman, WA for library preparation and sequencing on the Ion Proton (Thermo Fisher Scientific, Waltham, MA) for SNP variant calling as described by Merriman *et al.* (2012)[49] and Kohn *et al.* (2013)[50]. Raw reads with a length less than 64bp were discarded. Variants in the population were discovered using the TASSEL 5.0 (Trait Analysis by aSSociation Evolution and Linkage) GBS2 pipeline[51]. A minimum of 10 raw reads per tag was used as a quality control threshold, and the identified tags were then mapped against the wheat genome reference sequence v1.0 (obtained from the International Wheat Genome Sequencing Consortium[42]) using the Burrows-Wheeler Aligner (BWA)[52] and default parameters. Variants were predicted after quality control for a minimum mapping quality (MQ) of 5 and a minimum base quality of 10 for each variant. Post-variant call filtering was performed using TASSEL 5.0. Filtering was performed per variant site with a maximum threshold of 20 percent missing data per position, a minimum minimum allele frequency (MAF) of 0.05, a maximum MAF of 0.5, removing minor SNPs (third allele SNPs), and marking indels as missing. Missing SNP calls were first imputed using the FILLIN option, using parental SNP calls and all options set to default parameters. Post imputation, all heterozygous SNP calls were marked as missing. The resulting dataset was used for all subsequent analyses.

4.3.4 Statistical Analyses

All statistical analyses were performed using R version 3.5 in RStudio version 1.1.463[53]. First, an ANOVA (Analysis of Variance) was performed to determine the significance of the vernalization treatment and the treatment by genotype (also referred to as ‘Taxa’) interaction using the ‘aov’ function on a linear model of the form:

$$\text{Heading Date} = \text{Treatment} + \text{Taxa} + (\text{Treatment} \times \text{Taxa})$$

A Tukey's HSD (Honest Significant Difference) test was performed on the ANOVA data to determine whether the treatments were significantly different from one another. A Principle Component Analysis (PCA) was also performed for all data using the 'precomp' function and visualized with 'ggbiplot' in order to provide information on population structure.

4.3.5 *Genome Wide Association Study*

The Genome Wide Association Study was performed using the R package GAPIT[54], which utilizes a compressed mixed linear model (CMLM) that accounts for population structure and kinship[55]. In order to reduce the occurrence of spurious associations, population structure was accounted for using the first two principal components.

The linkage disequilibrium (LD) of each significant SNP was determined using TASSEL 5.0. The locations of the significant SNPs were then extracted from the GAPIT results and two thousand base pairs were added to each end of the SNP location for use in Ensembl Plants to search for candidate genes[56]. The region was then explored and annotations of any genes within the region or nearby were determined using the predicted functions from the UniProt database[57].

4.4 **Results**

4.4.1 *Heading Date*

The counts of the number of lines (taxa) in each heading date bin for each treatment is presented in Figure 4.1. All heading dates are averages of the replicates in each treatment. For Figure 4.1, if no replicate in a treatment reached heading before 135 days, the heading date remained 0. If at least one replicate reached heading before 135 days, but another replicate did not, the 0 was changed to a 200 in order to not skew the average to the earlier direction when it should be later. All 0 heading dates from Figure 4.1 were changed to 200 in subsequent analyses.

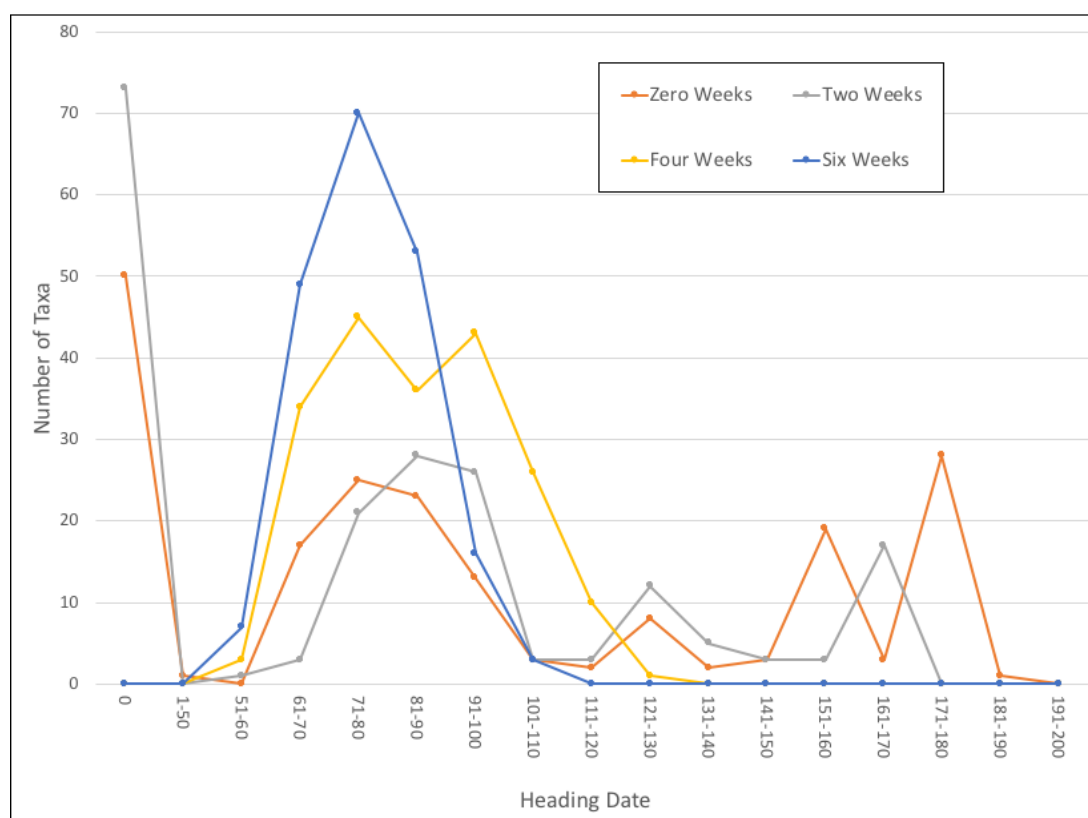


Figure 4.1 Average heading date counts for each Taxa in each treatment. A heading date of '0' indicates that none of the replicates reached heading before 135 days after planting. In order to not skew the data the wrong way, if one replicate headed before 135 days, but the other replicate did not, the heading date of 0 was changed to 200 and used for averaging the replicates.

The winter check variety Bobtail had heading dates of 139, 0, 109, and 92 days for zero, two, four, and six weeks of vernalization, respectively. The winter parent Skiles had dates of 187, 0, 104, and 76 days for zero, two, four, and six weeks of vernalization, respectively. The facultative parent Goetze had heading dates of 72, 75, 69, and 64 days for zero, two, four, and six weeks of vernalization, respectively. And the spring variety Louise had heading dates of 41, 56, 55, and 55 days for zero, two, four, and six weeks of vernalization, respectively. One hundred forty-four of the 196 RILs headed in the zero weeks of vernalization treatment. Of those, 91 lines reached heading before Bobtail, but only 18 lines reached heading faster than Goetze. The shortest time to heading in the RILs for zero weeks was 61 days, 20 days later than Louise. After two weeks of vernalization, 71 lines behaved like Bobtail and Skiles and did not head at all. No lines

reached heading faster than Louise. Nineteen lines reached heading at a similar time to Goetze, and 47 lines took 100 days or more to reach heading. All lines reached heading by day 135 after four weeks of vernalization. While no lines reached heading as fast as Louise, 30 lines reached heading at the same time as Goetze or faster. Overall, 159 lines reached heading in 100 days or less, with the slowest line reaching heading at 126 days. Again, all lines reached heading by day 135 after six weeks of vernalization. One line reached heading in 55 days, as did Louise, and 18 lines reached heading at the same time or faster than Goetze. Ninety-seven lines reached heading after Skiles, and 13 of those reached heading after Bobtail.

4.4.2 *SNP Calling*

A total of 2,251,531 raw variants were identified by TASSEL GBS2. After filtering, two lines were removed due to poor data quality and high a proportion of missing SNP calls. Post-variant call filtering per position resulted in 458,150 SNPs with minimum allele frequency (MAF) between 0.05 and 0.5, proportion missing between 0 and 0.82, and no heterozygous SNPs.

4.4.3 *Statistical Analyses*

The ANOVA results showed that in terms of heading date, the vernalization treatments are significantly different (p-value $<2e-16$), the taxa are significantly different (p-value $<2e-16$), and the treatment by taxa interaction is significant (p-value $<2e-16$) (Appendix B: Table B.1). Due to the significant p-values of the treatments, a Tukey's HSD was performed on the ANOVA results to determine which treatments should be analyzed separately in GAPIT. The results showed that each treatment was significantly different (p-value $<2e-4$), and therefore the QTL analysis was performed on each treatment individually (Appendix B: Table B.2).

The PCA for the first two principle components is shown in Figure 4.2. Louise is indicated by the green triangle, Goetze by the pink star, Bobtail by the orange circle, and Skiles by the blue square. The four vernalization treatments' contributions to the PCs is shown by the arrows. The clustering of Skiles and Bobtail together, and Louise and

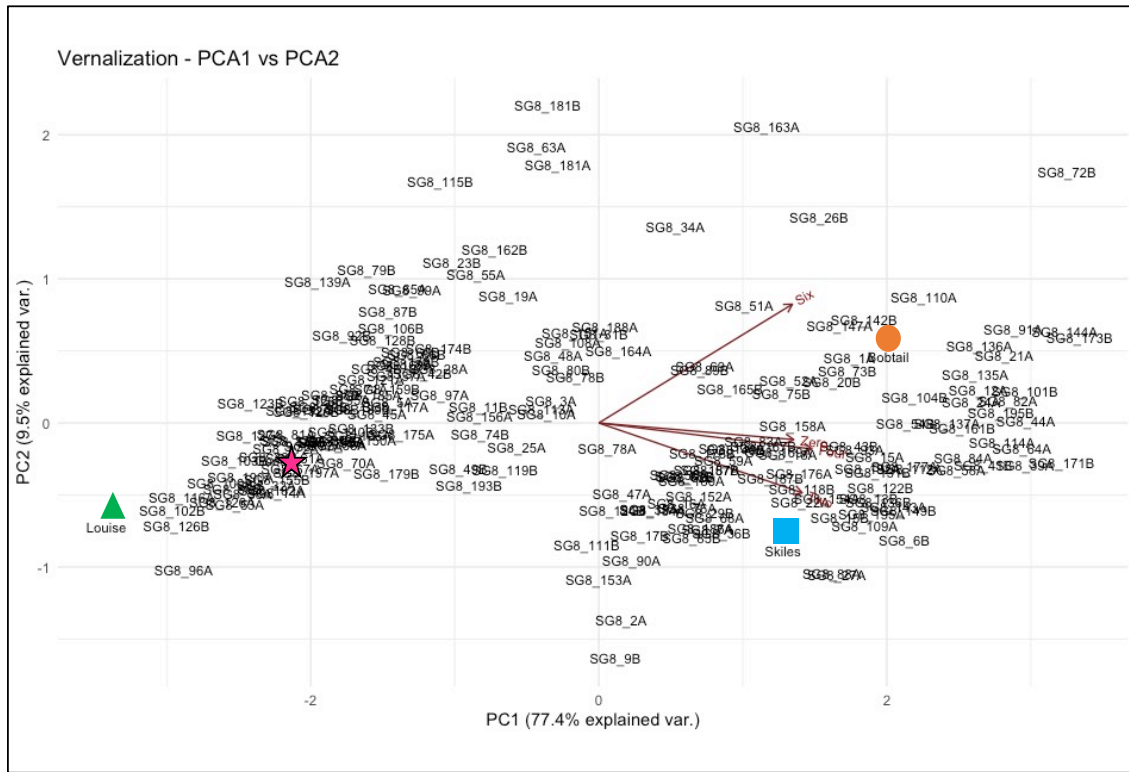


Figure 4.2 Scatterplot of a principal component (PC) analysis for data from GBS. Each axis represents a principal component. PC1 explains 77.4% of the variation and PC2 explains 9.5% of the variation. The green triangle is the spring check Louise, located at -3, -0.75 (PC1, PC2). The pink star is the facultative parent Goetze and is located at -2, 0.5. The blue square is the full winter parent Skiles and is located at 1, -1. The orange circle is the winter check Bobtail and is located at 2, 0.5. The separation of Louise and Goetze from Skiles and Bobtail shows a grouping of lines requiring full vernalization versus those requiring little or no vernalization on PC1.

Goetze together, indicated a separation of lines requiring full vernalization (*ie.* winter lines) versus lines that require none or a reduced amount of vernalization (*ie.* spring and facultative lines). PC1 explains 77.4 percent of the variation and PC2 explains 9.5 percent of the variation. Together they explain 86.9 percent of the variation and are sufficient for use in the GAPIT model for QTL discovery.

4.4.4 Genome Wide Association Study

The Manhattan plots produced by the GWAS for each treatment are shown in Figure 4.3, with the green horizontal line indicating the threshold for significance. All dots above the threshold line represent SNPs that are significantly associated with either

increased or decreased heading dates. Zero weeks of vernalization returned 11 significant SNPs: One on chromosome 1D, nine on chromosome 5B, and one on chromosome 5D. Two weeks of vernalization returned nine significant SNPs: One on 1D, one on 3B, six on 5B, and one on 5D. Four weeks of vernalization returned four significant SNPs: One on 1D and three on 5B. Six weeks of vernalization returned no significant SNPs, but had a similar profile to the other treatments. The list of all significant SNPs identified by GAPIT, their location, p-value, R^2 value, associated treatment, and alleles are reported in Table 4.1.

Linkage disequilibrium (LD) values among all significant SNPs can be found in Table B.3 of Appendix B. The subgroupings of SNPs as determined by these values is given here in Table 4.2. The graphical interpretation of the LD values among all significant SNPs is shown in Figure 4.4. An R^2 of 1.0 indicated the two SNPs completely correspond to one another, hence the state of one allele can predict the state of the second

Table 4.1 SNPs significantly associated with heading date variability identified in the Manhattan plots. The ‘Chromosome’ and ‘Position’ columns correspond to the chromosome and bp (base pair) location of the particular SNP in the reference sequence assembly. ‘p-value’ is the significance level used for the Manhattan plots in Figure 4.3. ‘MAF’ is the minor allele frequency at that SNP. The ‘ R^2 of Model without SNP’ is the value of the coefficient of determination (R^2) if the SNP were not present. The ‘ R^2 of Model with SNP’ is the R^2 value with that SNP included and should be higher than the model without the SNP for significant SNPs. ‘Treatment’ indicates in which treatment type the SNP was significantly associated with heading date. The bold ‘Allele’ indicates the allelic state associated with a faster heading time for all treatments in which the SNP was significant.

SNP	Chromosome	Position	p-value	MAF	R^2 of Model without SNP	R^2 of Model with SNP	Treatment	Allele (Major/Minor)
S1D_439252922	1D	439252922	1.08E-11	0.461340206	0.044912807	0.307009425	Zero, Two, Four	G/A
S3B_184436222	3B	184436222	9.19E-09	0.255154639	0.014963791	0.201382768	Two	A/G
S5B_563716653	5B	563716653	2.39E-08	0.448453608	0.014963791	0.190016469	Two	A/C
S5B_565507633	5B	565507633	4.86E-11	0.394329897	0.044912807	0.288381499	Zero, Two	G/A
S5B_565507635	5B	565507635	2.51E-11	0.389175258	0.044912807	0.296518704	Zero, Two, Four	G/A
S5B_569083742	5B	569083742	4.92E-12	0.456185567	0.044912807	0.316956386	Zero, Two, Four	C/T
S5B_569141895	5B	569141895	8.42E-10	0.451030928	0.044912807	0.253786366	Zero	G/A
S5B_569243058	5B	569243058	7.88E-10	0.425257732	0.044912807	0.2545846	Zero	C/G
S5B_581028391	5B	581028391	4.51E-12	0.476804124	0.044912807	0.318068679	Zero, Two, Four	G/A
S5B_585398903	5B	585398903	1.68E-09	0.443298969	0.044912807	0.24554564	Zero, Two	T/C
S5B_588093888	5B	588093888	2.48E-08	0.471649485	0.044912807	0.214239391	Zero	T/C
S5B_590557005	5B	590557005	3.56E-09	0.471649485	0.044912807	0.236727234	Zero	G/A
S5D_477960042	5D	477960042	8.58E-10	0.443298969	0.044912807	0.253560409	Zero, Two	T/C

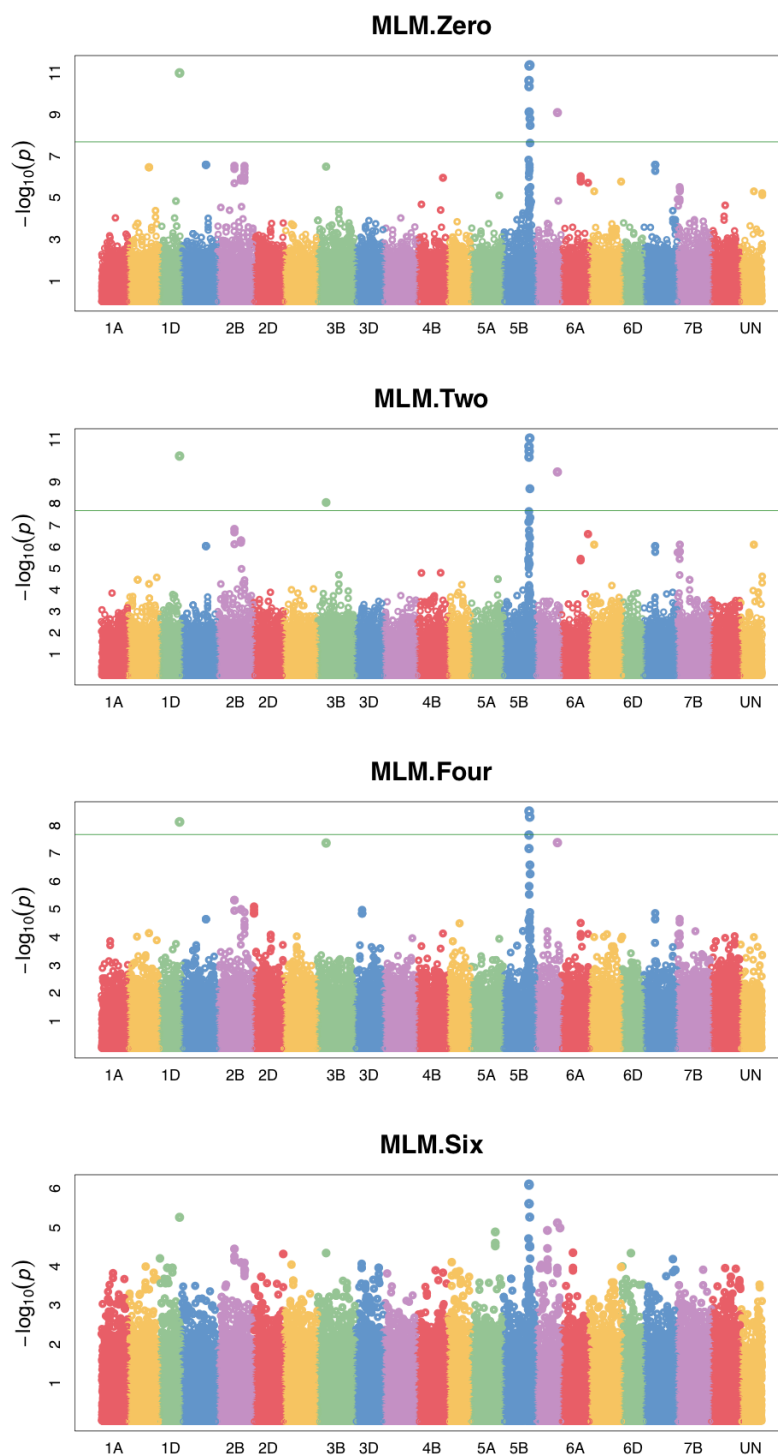


Figure 4.3 Manhattan plots for each treatment produced by GAPIT. The horizontal axis is the chromosome, with ‘UN’ denoting a genomic scaffold that could not be mapped to any particular chromosome in the reference genome. The vertical axis is the p-value for each SNP’s association with the trait of interest. All colored dots are SNPs identified by TASSEL, and any dot above the green line is a significant SNP.

allele within this population. An R^2 of 0.0 indicated that the two SNPs are completely unlinked from one another in this population. SNP S5B_565507633 with S5B_565507635 had an R^2 of 1.0, as do S5B_569243058 with S5B_569141895 and S5D_477960042 with S5B_585398903, meaning they are completely correlated in this dataset. SNP S3B_184436222 had very low R^2 values with all other SNPs, meaning it was not correlated with any other SNP. There were approximately 5 sub-groupings of SNPs on chromosome 5B based on the R^2 values of each. The first is S5B_563716653 by itself; second is S5B_565507633, S5B_565507635, S5B_569083742, S5B_569141895, and S5B_569243058; third is S5B_581028391 by itself; fourth is S5B_585398903 and S5B_588093888; and fifth is S5B_590557005 by itself. SNP S1D_439252922 was somewhat highly correlated with Subgroup2 of chromosome 5B and not with any of the others. SNP S5D_477960042 was somewhat highly correlated with Subgroup4 of 5B.

Table 4.2 The groupings of significant SNPs on each chromosome as determined by their Linkage Disequilibrium values. An R^2 of 1.0 indicates that the SNPs are completely correlated with one another in this population, meaning the presence of one allele can predict the allelic state of the other(s).

Chromosome	Subgroup	Significant SNP	R^2 Within Subgroup	High R^2 With Other Subgroups
5B	Subgroup1	S5B_563716653		None
	Subgroup2	S5B_565507633	Between 0.83 and 0.96	S1D_439252922 (between 0.79 and 0.94)
		S5B_565507635		
		S5B_569083742		
		S5B_569141895		
		S5B_569243058		
	Subgroup3	S5B_581028391		None
	Subgroup4	S5B_585398903	0.94	S5D_477960042 (0.91)
		S5B_588093888		
	Subgroup5	S5B_590557005		None
5D		S5D_477960042		5B Subgroup4 (0.91)
1D		S1D_439252922		5B Subgroup2 (between 0.79 and 0.94)
3B		S3B_184436222		None

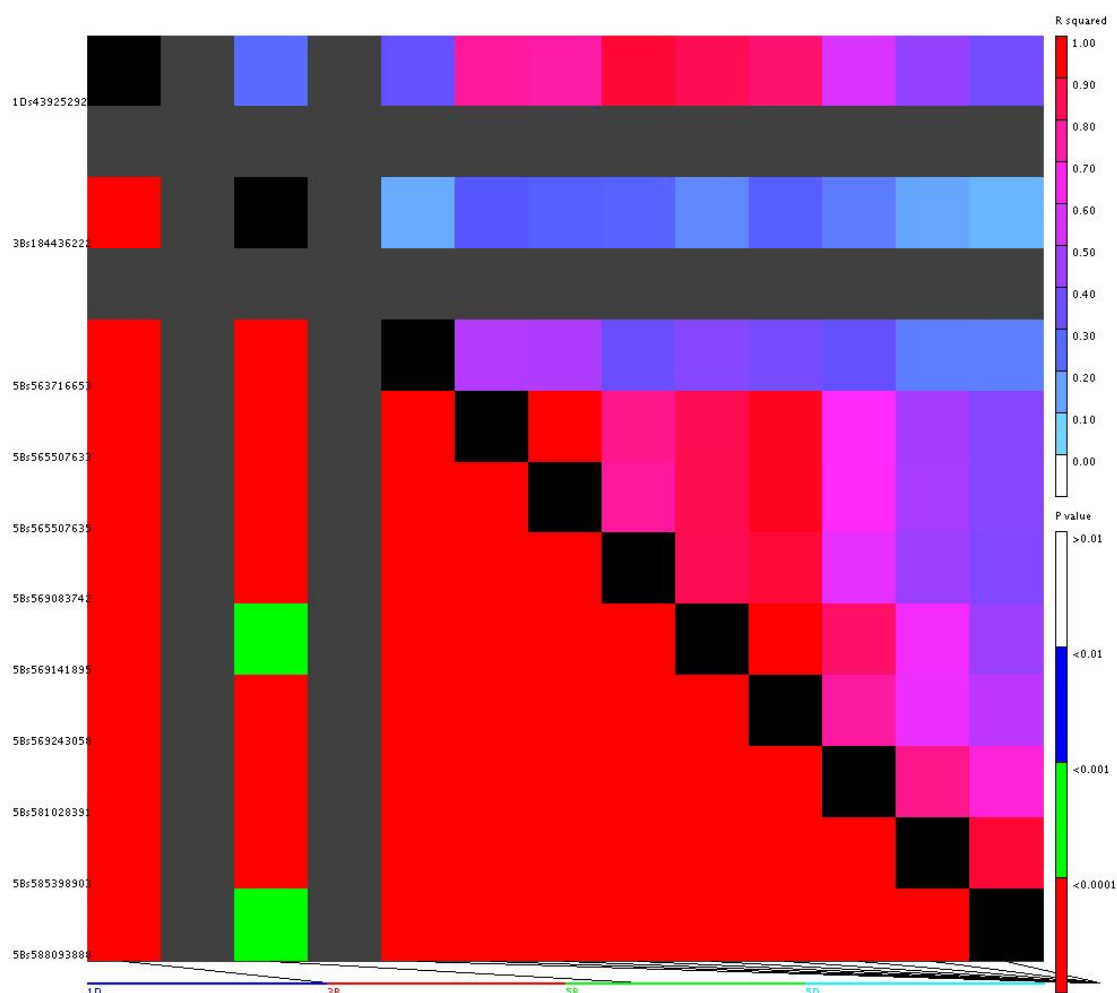


Figure 4.4 Heatmap of linkage disequilibrium (LD) for significant SNPs. Tile colors represent the R^2 value for each pairwise SNP (top right half), where R^2 represents the degree of association between two SNPs. Significance of association between pairs of SNPs is represented in the bottom left half of the figure as the p-value of the R^2 value. The horizontal line at the bottom represents each SNP position (black diagonal lines) per chromosome (colored horizontal lines).

The Taxa with all allelic states for earlier heading dates and those with all allelic states for later heading dates along with their corresponding heading dates per treatment are shown in Table 4.3. Unknown alleles (N) were left in the table when a corresponding SNP with an R^2 of 1.0 was known. The Taxa showing heading dates that were the earliest, latest, and most similar to Goetze along with their corresponding allelic states at

each significant SNP are shown in Table 4.4. There were no taxa that closely resembled the heading dates of Skiles in more than two Treatments.

Table 4.3 Taxa with all allelic states for early heading times and for late heading times. The SNPs and alleles are color-coded according to their LD grouping. Under ‘Treatments, Days to Heading’, any Taxa with a heading date that does not correspond to early (Goetze or earlier) in the first group, or late (Skiles or later) in the second group, is colored green. ‘N’ means the allele at that location is unknown.

Taxa	Significant SNPs													Treatments, Days to Heading				
	1D_4392 52922	3B_1844 36222	5B_5637 16653	5B_5655 07633	5B_5655 07635	5B_5690 83742	5B_5691 41895	5B_5692 43058	5B_5810 28391	5B_5853 98903	5B_5880 93888	5B_5905 57005	5D_4779 60042	Zero	Two	Four	Six	
SG8_92B	A	G	A	A	A	T	G	G	A	C	C	A	C	63	72	70	74	All Allelic States for earlier heading dates
SG8_126A	A	G	A	A	A	T	N	N	A	C	C	A	C	66	80	64	60	
SG8_115B	A	G	A	A	A	T	G	G	A	C	N	A	C	73	79	69	65	
SG8_196B	A	G	A	A	A	T	N	N	A	N	C	A	N	75	86	70	69	
SG8_38A	A	G	A	A	A	T	G	N	A	C	N	A	C	78	79	65	61	
SG8_87B	A	G	A	A	A	T	G	G	A	C	N	A	C	78	84	71	77	
SG8_79B	A	G	A	A	A	T	N	N	A	C	C	A	C	81	73	65	79	
SG8_38B	A	G	A	A	N	T	N	N	A	C	C	A	C	81	92	66	64	
SG8_189B	A	G	A	A	A	T	N	N	A	C	C	A	C	85	96	72	70	
SG8_113A	A	G	A	A	A	T	G	G	A	C	C	A	C	176	129	89	75	
SG8_132A	G	A	C	G	G	C	N	C	G	T	T	G	T	70	77	68	68	All Allelic States for later heading dates
SG8_129A	G	A	C	G	G	C	N	C	G	T	N	G	T	80	84	67	66	
SG8_128B	G	A	C	G	G	C	N	C	G	T	T	G	T	93	87	70	75	
SG8_99A	G	A	C	G	G	C	N	C	G	T	T	G	T	101	87	69	79	
SG8_156A	G	A	C	G	G	C	A	C	G	T	T	G	T	121	122	83	73	
SG8_76A	G	A	C	G	G	C	N	C	G	T	T	G	T	122	163	94	77	
SG8_187B	G	A	C	G	G	C	N	N	G	T	T	G	T	129	200	104	81	
SG8_124B	G	A	C	G	G	C	N	C	G	T	N	G	T	151	167	81	75	
SG8_153A	G	A	C	G	G	C	A	N	G	T	T	G	T	152	166	85	70	
SG8_118B	G	A	C	G	G	C	N	N	G	T	T	G	T	154	200	110	76	
SG8_176B	G	A	C	G	G	C	A	C	G	T	T	G	T	156	200	89	81	
SG8_29B	G	A	C	G	G	C	N	N	G	T	T	G	T	156	200	95	102	
SG8_75B	G	A	C	G	G	C	N	C	G	T	T	G	T	156	200	97	87	
SG8_180A	G	A	C	G	G	C	N	N	G	T	T	G	T	157	200	92	79	
SG8_62B	G	A	C	G	G	C	N	C	G	T	T	G	T	174	167	89	79	
SG8_22A	G	A	C	G	G	C	N	C	G	T	T	G	T	174	200	110	80	
SG8_187A	G	A	C	G	G	C	N	C	G	T	T	G	T	177	200	91	80	
SG8_59A	G	A	C	G	G	C	A	N	G	T	T	G	T	177	200	93	81	
SG8_104B	G	A	C	G	G	C	N	N	G	T	N	G	T	177	200	95	89	
SG8_80B	G	A	C	G	G	C	N	N	G	T	T	G	T	179	123	89	78	
SG8_176A	G	A	C	G	G	C	N	N	G	T	T	G	T	179	200	83	81	
SG8_90A	G	A	C	G	G	C	N	C	G	T	T	G	T	179	200	86	72	
SG8_108A	G	A	C	G	G	C	A	C	G	T	T	G	T	200	85	71	78	
SG8_51A	G	A	C	G	G	C	N	C	G	T	T	G	N	200	129	89	86	
SG8_78B	G	A	C	G	G	C	N	C	G	T	T	G	T	200	131	89	78	
SG8_54B	G	A	C	G	G	C	A	C	G	T	T	G	T	200	164	96	87	
SG8_112A	G	A	C	G	G	C	A	C	G	T	T	G	T	200	166	101	84	
SG8_111B	G	A	C	G	G	C	N	C	G	T	T	G	T	200	200	77	72	
SG8_192A	G	A	C	G	G	C	N	C	G	T	T	G	T	200	200	94	83	
SG8_13B	G	A	C	G	G	C	N	C	G	T	T	G	T	200	200	97	81	
SG8_177A	G	A	C	G	G	C	N	C	G	T	T	G	T	200	200	100	84	
SG8_1A	G	A	C	G	G	C	N	C	G	T	T	G	T	200	200	103	91	
SG8_149B	G	A	C	G	G	C	N	C	G	N	T	G	T	200	200	104	81	
SG8_161B	G	A	C	G	G	C	N	C	G	T	N	G	T	200	200	106	88	
SG8_21A	G	A	C	G	G	C	N	N	G	T	T	G	T	200	200	106	94	
SG8_84A	G	A	C	G	G	C	A	C	G	T	T	G	T	200	200	109	86	
SG8_58A	G	A	C	G	G	C	N	C	G	T	T	G	T	200	200	109	85	
SG8_41B	G	A	C	G	G	C	N	N	G	T	T	G	T	200	200	113	86	
Goetze	A	A/G	A/G	A	A	T	G	G	A	C/T	C/T	A	C/T	72	75	69	64	Parents and Checks
Skiles	A/G	A/G	A/C	G	G	C/T	A	C	G	C/T	C/T	A/G	C/T	187	200	104	76	
Bobtail														139	200	109	92	
Louise														41	56	55	55	

Table 4.4 Taxa with heading times that were the earliest, latest, and most similar to Goetze. The Significant SNPs are color coded according to their LD grouping. The Taxa, allelic states, and Heading Dates are color-coded based on their grouping to the far right. If both parents are heterozygous at a SNP, the allele associated with earlier heading date is in bold. Any allele that does not match the expected allele based on heading date is colored red. Heading times of the top four Taxa are the earliest, the next three are the latest, and the next six are the most similar to Goetze in at least three of the four treatments. Any Treatment that does not follow this expected pattern is noted on the far right. ‘N’ means the allele at that location is unknown.

Taxa	Significant SNPs													Treatments, Days to Heading				
	1D_4392 52922	3B_1844 36222	5B_5637 16653	5B_5655 07633	5B_5655 07635	5B_5690 83742	5B_5691 41895	5B_5692 43058	5B_5810 28391	5B_5853 98903	5B_5880 93888	5B_5905 57005	5D_4779 60042	Zero	Two	Four	Six	
SG8_116A	A	G	A	N	N	T	N	N	A	N	C	N	N	61	66	60	59	Overall Earliest
SG8_126A	A	G	A	A	A	T	N	N	A	C	C	A	C	66	80	64	60	
SG8_96A	G	A	C	G	G	C	N	N	N	T	T	G	T	65	97	61	55	Earliest in all but Two
SG8_102B	A	N	C	G	G	T	A	N	G	T	T	G	N	66	71	58	58	Earliest in all but Zero
SG8_101B	A	N	A	A	A	T	N	N	A	C	C	A	C	200	200	113	92	Overall Latest
SG8_44A	G	A	A	G	G	C	A	N	N	T	T	G	T	200	200	116	90	
SG8_173B	G	A	C	G	G	C	A	C	N	T	N	G	T	200	200	117	97	Like Goetze in all but Zero
SG8_172A	A	N	A	A	A	T	N	N	A	C	C	A	C	69	75	70	65	
SG8_155B	A	A	A	A	A	T	N	N	N	T	T	G	T	73	85	69	63	Like Goetze in all but Two
SG8_115B	A	G	A	A	A	T	G	G	A	C	N	A	C	73	79	69	65	Like Goetze in all but Two
SG8_127A	A	N	A	A	A	T	G	G	A	C	C	A	C	73	75	62	65	Like Goetze in all but Four
SG8_103B	A	N	A	A	A	T	G	G	A	C	C	A	C	72	75	62	63	Like Goetze in all but Four
SG8_81A	A	N	N	A	A	T	G	G	A	N	C	A	C	73	75	68	66	Like Goetze in all but Six
Goetze	A	A/G	A	A	A	T	G	G	A	C/T	C/T	A	C/T	72	75	69	64	Parents
Skiles	A/G	A/G	A/C	G	G	C/T	A	C	G	C/T	C/T	A/G	C/T	187	200	104	76	

4.4.5 Candidate Gene Analysis

Six of the significant SNPs found on chromosome 5B were upstream of the *Vrn-B1* gene, which is located between 573,802,883bp and 573,816,070bp, and four of them were downstream. SNP S5B_569243058 was 4.6Mb upstream of the start site of *Vrn-B1*. SNP S5B_581028391 is 7.2Mb downstream of the end of the *Vrn-B1* gene. One of the three orthologs of the *Vrn1* gene, *Vrn-D1*, is located on chromosome 5D between 467,176,609bp and 467,184,508bp. SNP S5D_477960042 was located 10.78Mb downstream of the *Vrn-D1* gene. *Vrn-D4* is also located on chromosome 5D on the proximal end of the short chromosome arm. However, *Vrn-D4* has not been found in all germplasm and it is not known if this gene is present in Skiles or Goetze. The putative short-day sensitivity gene *TaFT3* is located on chromosome 1D in the region between 420,469,366bp and 430,468,679bp. SNP S1D_439252922 was located 18.8Mb downstream of *TaFT3-D*. The gene *TaGI*, orthologous to the *Arabidopsis* gene *GIGANTEA* (*GI*), and implicated as a part of the photoperiod side of the flowering pathway, is located on chromosome 3B between 117,928,502bp and 117,930,007bp. SNP S3B_184436222 was located 66.5Mb downstream of this gene. No significant SNP was directly in or adjacent to a gene known to be involved in, or predicted to be involved

in, vernalization or photoperiod activity. The placement of these SNPs with the known genes are shown in Figure 4.5.

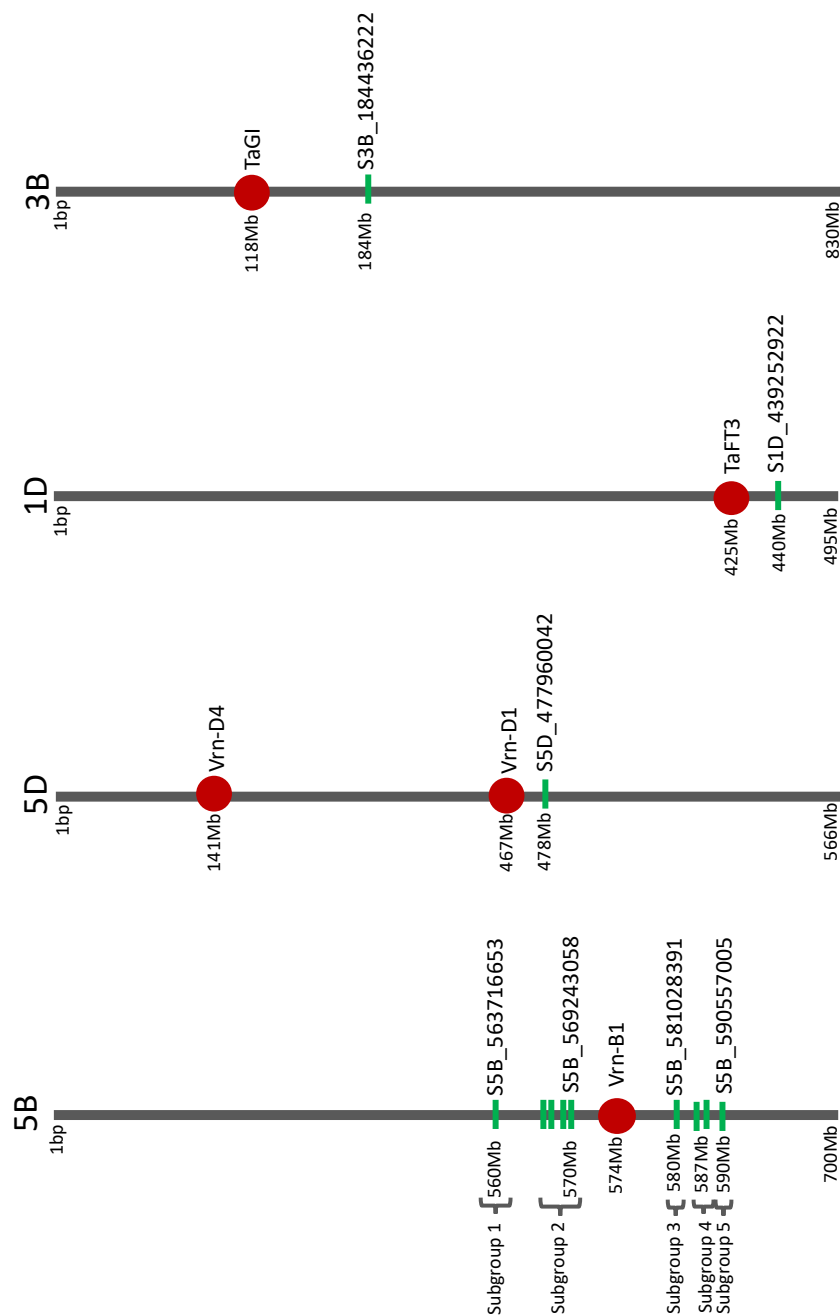


Figure 4.5 Genomic position of significant SNPs (green bars) and known vernalization and photoperiod genes (red circles) on the respective wheat chromosomes (dark grey horizontal lines) as determined by the reference genome assembly. Chromosomes are not to scale.

4.5 Discussion

Sufficient vernalization is fundamental for the production of winter wheat around the world. In the PNW, facultative varieties are needed to address warming winter temperatures and prevent devastating crop losses. This study identified 13 SNPs on four chromosomes that are significantly associated with reduced vernalization requirements and are linked to or likely linked to genes involved in the molecular mechanisms of flowering.

The heading dates of the RILs compared to the parents and checks showed a large variability within each treatment. Six weeks of vernalization resulted in the narrowest, most normally distributed set of heading dates of the four treatments. Most lines headed at 70 to 80 days, which was earlier than the winter check Bobtail (92 days) and slightly later than the facultative Goetze (64 days). This was expected since Skiles is a winter line and Goetze is a facultative line derived from winter lines, and there were no spring lines involved in the production of the RILs. This lack of diversity and narrow range of phenotypes is also the most likely reason that no SNPs were significant in the six week treatment, as there was not enough phenotypic diversity to assess genotypic linkage. Four weeks of vernalization was the second most normally-distributed, with most lines heading somewhere between 60 and 100 days. This treatment also only had four significant SNPs, all of which were also found in the two and zero week treatments. Again, this is likely due to the narrow heading date range.

Two and zero weeks of vernalization were much more spread out and uneven. Goetze headed in roughly the same amount of time under both treatments, similar to Louise, suggesting that some of Goetze's flowering is determined by photoperiod rather than only by vernalization. Vernalization time above two weeks helped to speed up the time to heading for Goetze due to the up-regulation of the VRN1 genes. Skiles and Bobtail both flowered with no vernalization, due to their day neutral alleles at PPD1 that promote flowering regardless of day length. Skiles has a facultative allele at Ppd-D1 on chromosome 2D, and Bobtail has a facultative allele at Ppd-A1 on chromosome 2A. Without this contribution of photoperiod to flowering, they would never have reached heading without vernalization. Although it is also possible that they have overexpressing

alleles at TaFT3 in addition to those at PPD1, which could explain the significant SNP on chromosome 1D where TaFT3-D is found. Despite flowering with no vernalization, both lines failed to flower within the given timeframe after two weeks of vernalization. This is potentially because two weeks of vernalization is insufficient to promote flowering, and instead only acts to slow the growth of the plant. This would explain the increase in time to heading seen in Louise at two, four, and six weeks of vernalization. Louise took 14 days longer to head after being vernalized, but did not show any difference in heading date within the vernalization treatments, further suggesting that vernalization slows growth in plants that either do not require vernalization or that require at least four weeks of vernalization to head in fewer than 135 days.

Unfortunately, when the original cross between Skiles and Goetze was made in 2009, the parental lines were not kept and propagated along with the progeny lines. The seed that was used to grow the plants sampled for GBS was from foundation seed stocks, which resulted in both Skiles and Goetze being heterozygous for multiple SNPs identified in this study (Tables 4.3 and 4.4). Therefore, it is not possible to determine the parental lines' contributions at SNPs S3B_184436222, S5B_585398903, S5B_588093888, or S5D_477960042. It is possible, however, to determine which allelic state is associated with earlier heading dates and which is associated with later heading dates. For the majority of Taxa, a genotype with all early alleles corresponds to early heading dates that are much more similar to Goetze than Skiles, and vice versa for a genotype with all late alleles (Table 4.3).

One early genotype line (SG8_113A) and four late genotype lines (SG8_132A, SG8_129A, SG8_128A, and SG8_99A) do not appear to follow this pattern, however. This is most likely not due to sequencing or SNP-calling errors, since all the alleles that are called are of the same type, but is more likely due to a sampling error or a planting error that caused the genotype and phenotype to be unmatched. Although there is also the possibility that the genotypes and phenotypes are correct, and there is another aspect of these lines that is contributing to the unexpected results. Therefore, these lines should have each SNP location Sanger sequenced to determine whether the results are due to a mistake or are illuminating the presence of something else contributing to heading date

that was not identified in this study. Other lines that should be genotyped using Sanger sequencing are SG8_96A, SG8_102B, SG8_101B, and SG8_155B (Table 4.4). Their genotypes also do not strictly follow the expected associated phenotypes. Despite these anomalies, it is clear that in the majority of lines, those with all early alleles together show early heading dates in all four Treatments.

As mentioned earlier, the presence of the significant SNP S1D_439252922 on chromosome 1D 18.8Mb from TaFT3 may indicate that the experiment was also picking up some photoperiod sensitivity rather than only vernalization response-associated genes. However, the detection of this SNP is most likely due to the selective breeding performed on the parents causing them to maintain a close association between facultative vernalization alleles and short-day photoperiod sensitivity alleles, rather than as a response to photoperiod differences, as each vernalization treatment was under the same day length regime. The average LD decay across the wheat genome has been reported to be between 23Mb and 50Mb, making it difficult to determine whether this SNP was associated with TaFT3-D or another gene, as this distance is too large to analyze gene-by-gene[58][59]. In addition to this, the exact location of TaFT3 within the reported 1Mb region is not currently known, as the discovery of this gene in wheat occurred recently and few studies have been conducted into its contribution and allelic diversity in cultivated wheat[8][10]. Therefore, since it is clear that there is a significant association with facultative growth habit and a QTL on chromosome 1D in this RIL population, further studies should be performed to narrow down the QTL range, determine the true nature of the gene(s) contributing to this phenotype, and obtain markers for use in MAS.

Goetze has a facultative allele at or near the gene Vrn-B1, which is located on chromosome 5B, but the exact allele is unknown. The extremely wide distribution of significant SNPs on chromosome 5B in this study indicate the presence of a large QTL, rather than only a single gene controlling the decreased vernalization requirement. Located only 0.6Mb upstream of Vrn-B1 is the photoreceptor gene TaPHYC, which is a contributor to flowering via the photoperiod pathway[60]. The close proximity of these two genes and their location between the significant SNPs found in this study further emphasize the importance of the entire QTL, rather than only the single vernalization

gene, to flowering time. Therefore, some of the widely spaced markers surrounding Vrn-B1, such as those in subgroup two, three, or four, could be useful in future breeding efforts to retain the facultative growth habit seen in many of the RILs and Goetze. The strong presence of these significant SNPs on chromosome 5B in three of the four treatments also reinforces the strength of the facultative allele contributed by Goetze at this location. Use of this QTL in a cross with any full winter line should produce many progenies with significantly reduced vernalization requirements.

A homoeolog of Vrn-B1, Vrn-D1, has also been shown to have allelic variation in diverse germplasm, although it has been reported to contribute less to vernalization requirements compared to Vrn-A1 and Vrn-B1[61]. This study reinforced its lower level of contribution compared to Vrn-B1, but it also indicated that Vrn-D1 can significantly contribute to highly facultative varieties as evidenced by the significance of SNP S5D_477960042 in the zero and two week treatments. Therefore, it should not be overlooked when breeding for facultative varieties, especially if more fine-tuned variation in vernalization requirement is needed. There is also the possibility that this SNP is associated with Vrn-D4, as this gene is also on chromosome 5D. Vrn-D4 has not, however, been characterized in PNW winter wheat at this time, is most common in Southeast Asian varieties, and is likely located on the other chromosome arm from SNP S5D_477960042 and Vrn-D1[17]. Therefore, the most likely association is with Vrn-D1.

The TaGI gene in wheat is induced by photoperiod and expressed in leaves. To date, three research papers have found that TaGI contributes to flowering induction during long day photoperiods, although the exact mechanisms and magnitude of its contribution are unknown[61]-[63]. Any naturally occurring allelic variation is also unknown at this time. A study done in 1988 by Zemetra *et al.* found that a substitution of chromosome 3B from the winter line ‘Wichita’ into the winter line ‘Cheyenne’ resulted in a spring line[64]. They determined that the gene was contributing to an extremely early flowering phenotype but had been suppressed in Wichita. It is possible that the substituted chromosome contained an allele of TaGI that promoted flowering without vernalization, and that this allele may also exist in the RILs used in this study. SNP S3B_184436222 is located 66.5Mb downstream of the B genome copy of TaGI. While

the LD decay in hexaploid wheat is quite large, 66.5Mb is a significant distance. Therefore, it is not possible to determine from this study whether SNP S3B_184436222 is associated with TaGI or some other unknown gene or QTL on this chromosome. This SNP was also close to significance in the four week vernalization treatment and should therefore be studied further as it appears to be an important contributor to heading date.

One possible way to narrow the search range and increase SNP depth for chromosomes 1D, 3B, and 5D would be to redo this experiment including lines that are more diverse for their vernalization requirement. The number of weeks of vernalization could also be subdivided further to include one, three, and five weeks. A different genotyping technology and/or different restriction enzymes could be used as well. Another study that would help to determine whether genes on 1D and 3B are photoperiod-related is to subject the RILs to different photoperiod regimes such as long versus short days, and then record days to heading. The results of these studies should increase the ability to detect SNPs in the important regions of chromosomes 1D, 3B, and 5D, as well as increase the chances of identifying more significant SNP/trait associations.

4.6 Conclusions

This study illustrates the usefulness of GBS and GWAS as a tool for marker-trait discovery. It reinforces the previous knowledge of an important facultative allele at *Vrn-B1* from Goetze and the presence of a wide QTL in this region that is integral in facultative growth habits. There is also potential for the presence of another facultative allele at *Vrn-D1* that is significantly contributing to reduced vernalization requirements. The genes themselves can be further characterized using Sanger sequencing to amplify the regions of interest in the parents and selected individuals from the RIL population showing discrete and wide vernalization requirements. This will elucidate the alleles responsible for decreased vernalization requirements which can also be used as markers for future crosses. Breeders can then be confident in the facultative ability of the lines being produced without the need for extensive vernalization testing. This study also indicates an important association with what is potentially a photoperiod insensitivity allele on chromosome 1D that may be contributing to heading date variability. There is

also a potential association on chromosome 3B with the photoperiod gene TaGI, a QTL that includes TaGI, or a previously uncharacterized gene involved in the flowering pathway of wheat. Due to the importance of breeding for facultative varieties, breeders would benefit from further investigations at these locations to determine the gene or QTL associated with these two SNPs. Breeders can, however, immediately begin using the knowledge of the vernalization QTLs on chromosomes 5B and 5D for the creation of locally-adapted facultative varieties.

4.7 Acknowledgements

This work was supported by the Oregon Agricultural Experiment Station, Warren E. Kronstad Wheat Research Endowment, the Oregon Wheat Commission, and the ARCS Oregon Chapter Scholar Award.

4.8 References

- [1] P. Gustafson, O. Raskina, X. Ma, and E. Nevo, “Wheat evolution, domestication, and improvement,” in *Wheat Science and Trade*, Wiley-Blackwell, 2009, pp. 3–30.
- [2] Z. Peleg, T. Fahima, A. B. Korol, S. Abbo, and Y. Saranga, “Genetic analysis of wheat domestication and evolution under domestication,” *J Exp Bot*, vol. 62, no. 14, pp. 5051–5061, Oct. 2011.
- [3] W. Bushuk, “Wheat breeding for end-product use,” in *Braun HJ., Altay F., Kronstad W.E., Beniwal S.P.S., McNab A. (eds) Wheat: Prospects for global improvement*, Ankara, Turkey, 1997, vol. 6, pp. 203–211.
- [4] Z. Ivaničová *et al.*, “Heritable heading time variation in wheat lines with the same number of Ppd-B1 gene copies,” *PLOS ONE*, vol. 12, no. 8, pp. 1–15, Aug. 2017.
- [5] A. Kamran, M. Iqbal, and D. Spaner, “Flowering time in wheat (*Triticum aestivum* L.): a key factor for global adaptability,” *Euphytica*, vol. 197, no. 1, pp. 1–26, May 2014.
- [6] FAO, “FAOSTAT database,” *Agricultural crops: Wheat: area harvested*, 2014. [Online]. Available: <http://www.fao.org/>. [Accessed: 04-Dec-2017].
- [7] S. M. Grogan *et al.*, “Allelic Variation in Developmental Genes and Effects on Winter Wheat Heading Date in the U.S. Great Plains,” *PLOS ONE*, vol. 11, no. 4, pp. 1–23, Apr. 2016.
- [8] J. Halliwell *et al.*, “Systematic Investigation of FLOWERING LOCUS T-Like Poaceae Gene Families Identifies the Short-Day Expressed Flowering Pathway Gene, TaFT3 in Wheat (*Triticum aestivum* L.),” *Front. Plant Sci.*, vol. 7, pp. 1–15, Jun. 2016.
- [9] A. Muterko, R. Kalendar, and E. Salina, “Allelic variation at the VERNALIZATION-A1, VRN-B1, VRN-B3, and PHOTOPERIOD-A1 genes in cultivars of *Triticum durum* Desf.,” *Planta*, vol. 244, no. 6, pp. 1253–1263, Dec. 2016.
- [10] M. Zikhali, L. U. Wingen, M. Leverington-Waite, S. Specel, and S. Griffiths, “The identification of new candidate genes *Triticum aestivum* FLOWERING LOCUS T3-B1 (TaFT3-B1) and TARGET OF EAT1 (TaTOE1-B1) controlling the short-day photoperiod response in bread wheat,” *Plant, Cell & Environment*, vol. 40, no. 11, pp. 2678–2690, Nov. 2017.
- [11] L. Yan, M. Helguera, K. Kato, S. Fukuyama, J. Sherman, and J. Dubcovsky, “Allelic variation at the VRN-1 promoter region in polyploid wheat,” *Theoretical and Applied Genetics*, vol. 109, no. 8, pp. 1677–1686, Nov. 2004.
- [12] D. K. Santra, M. Santra, R. E. Allan, K. G. Campbell, and K. K. Kidwell, “Genetic and Molecular Characterization of Vernalization Genes *Vrn-A1*, *Vrn-B1*, and *Vrn-D1* in Spring Wheat Germplasm from the Pacific Northwest Region of the U.S.A.,” *Plant Breeding*, vol. 128, no. 6, pp. 576–584, Dec. 2009.
- [13] J. Zhang, Y. Wang, S. Wu, J. Yang, H. Liu, and Y. Zhou, “A single nucleotide polymorphism at the *Vrn-D1* promoter region in common wheat is associated with vernalization response,” *Theoretical and Applied Genetics*, vol. 125, no. 8, pp. 1697–1704, Dec. 2012.
- [14] N. Kippes, A. Chen, X. Zhang, A. J. Lukaszewski, and J. Dubcovsky, “Development and characterization of a spring hexaploid wheat line with no

- functional VRN2 genes,” *Theor Appl Genet*, vol. 129, no. 7, pp. 1417–1428, Jul. 2016.
- [15] N. Kippes *et al.*, “Fine mapping and epistatic interactions of the vernalization gene VRN-D4 in hexaploid wheat,” *Mol Genet Genomics*, vol. 289, no. 1, pp. 47–62, Feb. 2014.
- [16] T. Kiss *et al.*, “Effects of ambient temperature in association with photoperiod on phenology and on the expressions of major plant developmental genes in wheat (*Triticum aestivum* L.),” *Plant, Cell & Environment*, vol. 40, no. 8, pp. 1629–1642, Aug. 2017.
- [17] N. Kippes *et al.*, “Identification of the VERNALIZATION 4 gene reveals the origin of spring growth habit in ancient wheats from South Asia,” *PNAS*, vol. 112, no. 39, pp. E5401–E5410, Sep. 2015.
- [18] B. Trevaskis, “Wheat gene for all seasons,” *PNAS*, vol. 112, no. 39, pp. 11991–11992, Sep. 2015.
- [19] I. Karsai *et al.*, “Effects of photo and thermo cycles on flowering time in barley: a genetical phenomics approach,” *J Exp Bot*, vol. 59, no. 10, pp. 2707–2715, Jul. 2008.
- [20] J. L. Gouis *et al.*, “Genome-wide association analysis to identify chromosomal regions determining components of earliness in wheat,” *Theor Appl Genet*, vol. 124, no. 3, pp. 597–611, Feb. 2012.
- [21] U. Steinfort, B. Trevaskis, S. Fukai, K. L. Bell, and M. F. Dreccer, “Vernalisation and photoperiod sensitivity in wheat: Impact on canopy development and yield components,” *Field Crops Research*, vol. 201, no. Supplement C, pp. 108–121, Feb. 2017.
- [22] B. Wang, D. L. Liu, S. Asseng, I. Macadam, and Q. Yu, “Impact of climate change on wheat flowering time in eastern Australia,” *Agricultural and Forest Meteorology*, vol. 209–210, pp. 11–21, Sep. 2015.
- [23] T. P. Khumalo, A. Barnard, L. Maphosa, and T. J. Tsilo, “Impact of Growth Habit and Architecture Genes on Adaptation and Performance of Bread Wheat,” in *Wheat Improvement, Management and Utilization*, R. Wanyera and J. Owuoche, Eds. InTech, 2017.
- [24] M. S. Lopes *et al.*, “Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change,” *Journal of Experimental Botany*, vol. 66, no. 12, pp. 3477–3486, Jun. 2015.
- [25] E. I. Teixeira, G. Fischer, H. van Velthuizen, C. Walter, and F. Ewert, “Global hot-spots of heat stress on agricultural crops due to climate change,” *Agricultural and Forest Meteorology*, vol. 170, pp. 206–215, Mar. 2013.
- [26] D. B. Lobell and C. B. Field, “Global scale climate–crop yield relationships and the impacts of recent warming,” *Environmental Research Letters*, vol. 2, no. 014002, Mar. 2007.
- [27] X. Wu, H. Liu, X. Li, Y. Tian, and M. D. Mahecha, “Responses of Winter Wheat Yields to Warming-Mediated Vernalization Variations Across Temperate Europe,” *Front. Ecol. Evol.*, vol. 5, no. 126, pp. 1–10, Oct. 2017.
- [28] P. W. Mote and E. P. Salathé, “Future climate in the Pacific Northwest,” *Climatic Change*, vol. 102, no. 1–2, pp. 29–50, Sep. 2010.

- [29] E. P. Salathé, R. Steed, C. F. Mass, and P. H. Zahn, "A High-Resolution Climate Model for the U.S. Pacific Northwest: Mesoscale Feedbacks and Local Responses to Climate Change," *J. Climate*, vol. 21, no. 21, pp. 5708–5726, Nov. 2008.
- [30] D. Gouache, X. Le Bris, M. Bogard, O. Deudon, C. Pagé, and P. Gate, "Evaluating agronomic adaptation options to increasing heat stress under climate change during wheat grain filling in France," *European Journal of Agronomy*, vol. 39, pp. 62–70, May 2012.
- [31] A. Cuesta-Marcos *et al.*, "The Relationships between Development and Low Temperature Tolerance in Barley Near Isogenic Lines Differing for Flowering Behavior," *Plant and Cell Physiology*, vol. 56, no. 12, pp. 2312–2324, Dec. 2015.
- [32] A. Ozturk, O. Caglar, and S. Bulut, "Growth and Yield Response of Facultative Wheat to Winter Sowing, Freezing Sowing and Spring Sowing at Different Seeding Rates," *Journal of Agronomy & Crop Science*, vol. 192, no. 1, pp. 10–16, Feb. 2006.
- [33] X. Guo *et al.*, "Distribution of the Vrn-D1b allele associated with facultative growth habit in Chinese wheat accessions," *Euphytica*, vol. 206, no. 1, pp. 1–10, Nov. 2015.
- [34] H. A. Eagles, K. Cane, and N. Vallance, "The flow of alleles of important photoperiod and vernalisation genes through Australian wheat," *Crop Pasture Sci.*, vol. 60, no. 7, pp. 646–657, Aug. 2009.
- [35] M. Ishaq, "Genetic potential and variability for phenological and yield attributes in facultative winter wheat," *Pure and Applied Biology*, vol. 6, no. 2, pp. 619–626, Jun. 2017.
- [36] N. Kippes, M. Guedira, L. Lin, M. A. Alvarez, G. L. Brown-Guedira, and J. Dubcovsky, "Single nucleotide polymorphisms in a regulatory site of VRN-A1 first intron are associated with differences in vernalization requirement in winter wheat," *Mol Genet Genomics*, vol. 293, no. 5, pp. 1231–1243, Oct. 2018.
- [37] P. J. Kersey, "Plant genome sequences: past, present, future," *Current Opinion in Plant Biology*, vol. 48, pp. 1–8, Apr. 2019.
- [38] J. He, X. Zhao, A. Laroche, Z.-X. Lu, H. Liu, and Z. Li, "Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding," *Front. Plant Sci.*, vol. 5, no. 484, Sep. 2014.
- [39] L.-X. Yu, P. Zheng, S. Bhamidimarri, X.-P. Liu, and D. Main, "The Impact of Genotyping-by-Sequencing Pipelines on SNP Discovery and Identification of Markers Associated with Verticillium Wilt Resistance in Autotetraploid Alfalfa (*Medicago sativa* L.)," *Front. Plant Sci.*, vol. 8, no. 89, Feb. 2017.
- [40] Collard Bertrand C.Y and Mackill David J, "Marker-assisted selection: an approach for precision plant breeding in the twenty-first century," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1491, pp. 557–572, Feb. 2008.
- [41] P. J. Berkman, K. Lai, M. T. Lorenc, and D. Edwards, "Next-generation sequencing applications for wheat crop improvement," *American Journal of Botany*, vol. 99, no. 2, pp. 365–371, Feb. 2012.
- [42] The International Wheat Genome Sequencing Consortium (IWGSC) *et al.*, "Shifting the limits in wheat research and breeding using a fully annotated reference genome," *Science*, vol. 361, no. 6403, p. eaar7191, Aug. 2018.

- [43] A. Díaz, M. Zikhali, A. S. Turner, P. Isaac, and D. A. Laurie, “Copy Number Variation Affecting the Photoperiod-B1 and Vernalization-A1 Genes Is Associated with Altered Flowering Time in Wheat (*Triticum aestivum*),” *PLOS ONE*, vol. 7, no. 3, p. e33234, Mar. 2012.
- [44] G. Li, M. Yu, T. Fang, S. Cao, B. F. Carver, and L. Yan, “Vernalization requirement duration in winter wheat is controlled by TaVRN-A1 at the protein level,” *The Plant Journal*, vol. 76, no. 5, pp. 742–753, Aug. 2013.
- [45] M. Flowers, C. J. Peterson, A. Hulting, John Burns, S. Guy, and J. Kuehner, “Skiles Soft White Winter Wheat,” Corvallis, Or.: Extension Service, Oregon State University, Apr. 2010.
- [46] M. D. Flowers and C. J. Peterson, “Goetze, soft winter wheat,” Corvallis, Or.: Extension Service, Oregon State University, Apr. 2008.
- [47] K.K. Kidwell *et al.*, “Registration of ‘Louise’ Wheat,” *Crop Science*, vol. 46, no. 3, pp. 1384–1386, Jun. 2006.
- [48] J. C. Zadoks, T. T. Chang, and C. F. Konzak, “A decimal code for the growth stages of cereals,” *Weed Research*, vol. 14, no. 6, pp. 415–421, Dec. 1974.
- [49] B. Merriman, I. T. R. Team, and J. M. Rothberg, “Progress in Ion Torrent semiconductor chip based sequencing,” *ELECTROPHORESIS*, vol. 33, no. 23, pp. 3397–3417, Sep. 2012.
- [50] A. B. Kohn, T. P. Moroz, J. P. Barnes, M. Netherton, and L. L. Moroz, “Single-Cell Semiconductor Sequencing,” *Methods Mol Biol*, vol. 1048, pp. 247–284, May 2014.
- [51] P. J. Bradbury, Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler, “TASSEL: software for association mapping of complex traits in diverse samples,” *Bioinformatics*, vol. 23, no. 19, pp. 2633–2635, Oct. 2007.
- [52] H. Li and R. Durbin, “Fast and accurate short read alignment with Burrows–Wheeler transform,” *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, Jul. 2009.
- [53] RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.
- [54] A. E. Lipka *et al.*, “GAPIT: genome association and prediction integrated tool,” *Bioinformatics*, vol. 28, no. 18, pp. 2397–2399, Sep. 2012.
- [55] Z. Zhang *et al.*, “Mixed linear model approach adapted for genome-wide association studies,” *Nat Genet*, vol. 42, no. 4, pp. 355–360, Apr. 2010.
- [56] S. E. Hunt *et al.*, “Ensembl variation resources,” *Database (Oxford)*, vol. 2018, pp. 1–12, Jan. 2018.
- [57] “UniProt: a worldwide hub of protein knowledge,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D506–D515, Jan. 2019.
- [58] G. Li *et al.*, “Identification of powdery mildew resistance loci in wheat by integrating genome-wide association study (GWAS) and linkage mapping,” *The Crop Journal*, pp. 1–13, Apr. 2019.
- [59] P. Juliana *et al.*, “Genome-wide association mapping for resistance to leaf rust, stripe rust and tan spot in wheat reveals potential candidate genes,” *Theoretical and Applied Genetics*, vol. 131, no. 7, pp. 1405–1422, Jul. 2018.

- [60] C. Shi, L. Zhao, X. Zhang, G. Lv, Y. Pan, and F. Chen, “Gene regulatory network and abundant genetic variation play critical roles in heading stage of polyploidy wheat,” *BMC Plant Biol*, vol. 19, no. 6, pp. 1–16, Jan. 2019.
- [61] A. Loukoianov, L. Yan, A. Blechl, A. Sanchez, and J. Dubcovsky, “Regulation of VRN-1 Vernalization Genes in Normal and Transgenic Polyploid Wheat,” *Plant Physiology*, vol. 138, no. 4, pp. 2364–2373, Aug. 2005.
- [62] N. Mizuno *et al.*, “Loss-of-Function Mutations in Three Homoeologous PHYTOCLOCK 1 Genes in Common Wheat Are Associated with the Extra-Early Flowering Phenotype,” *PLOS ONE*, vol. 11, no. 10, p. e0165618, Oct. 2016.
- [63] X. Y. Zhao, M. S. Liu, J. R. Li, C. M. Guan, and X. S. Zhang, “The wheat TaGI1, involved in photoperiodic flowering, encodes an Arabidopsis GI ortholog,” *Plant Molecular Biology*, vol. 58, no. 1, pp. 53–64, May 2005.
- [64] Zemetra, R.S., and R. Morris. “Effects of an intercultural chromosome substitution on winter hardiness and vernalization in wheat,” *Genetics*, vol. 119, pp. 453–456, Feb. 1988.

Chapter 5

A Streamlined Protocol for Wheat (*Triticum aestivum*) Protoplast Isolation and Transformation with CRISPR Ribonucleoprotein Complexes

5.1 Abstract

The genetic engineering method CRISPR has been touted as an efficient, inexpensive, easily used, and targeted genetic modification technology that is widely suggested as having the potential to solve many of the problems facing agriculture now and in the future. CRISPR technology can theoretically reduce many of the problems associated with traditional breeding such as non-existent traits, prohibitive breeding times, and detrimental linkage drag. Like all new technologies, however, it is not without problems. One of the most difficult problems to anticipate and detect is gene targets that are inaccessible due to the chromatin state at their specific location. There is currently no way to predict this during the process of designing a sgRNA target, as the only evidence of an inaccessible gene is a failed experiment. At this time, the only way to detect this issue before spending time and resources on full transformations is to test the cleavage ability of the sgRNA *in vivo*. In wheat, this is possible using protoplast isolation and PEG transformation with Cas9 Ribonucleoproteins (RNPs). Therefore, we have developed a simplified protocol for the isolation and transformation of wheat protoplast cells using CRISPR/Cas9 RNPs which will assist future researchers in streamlining protoplast isolation and CRISPR transformation projects.

5.2 Introduction

Wheat is one of the staple food crops in the world, and currently feeds more than a quarter of the global population[1][2]. Multiple public and private entities work towards improving wheat for each growing region. Currently, no transgenic wheat varieties have been deregulated by any government, and therefore the majority of breeders rely heavily on traditional breeding to improve traits and release new varieties[3][4]. However, the process from first cross to varietal release can easily take ten years or more[5]. Thus, many researchers are turning to genetic transformation via targeted gene editing as a potential alternative to traditional breeding[6]–[8].

Gene editing is the process of making specific modifications to a known DNA sequence within a cell. These modifications can consist of insertions, deletions, or changes in the gene sequence that cause a desired change in the produced protein[9][10].

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a relatively new technique that facilitates gene editing while increasing the efficiency of transformation for model and non-model plant species, and has quickly replaced most uses of other transformation methods[11][12]. CRISPR has recently been engineered to have the ability to target any desired sequence in eukaryotic cells, and was first used to transform crop plants in 2013[13]–[19]. The CRISPR system is composed of a CRISPR associated (Cas) protein bound to a tracrRNA (trans-activating CRISPR RNA) sequence, which is then linked to a crRNA (CRISPR RNA) sequence forming a Ribonucleoprotein (RNP) complex. This crRNA sequence is a 20 nucleotide long single-guide RNA (sgRNA) that is designed by the user to match a specific target sequence. The most widely used Cas protein is Cas9, which recognizes the protospacer adjacent motif (PAM) NGG (any nucleotide followed by two Guanines). The Cas9/sgRNA complex runs along the target organism's genome until it encounters a correct PAM sequence. Once the Cas9 has bound to the PAM, it undergoes a conformational change that allows the sgRNA sequence to attempt to bind to the DNA sequence. If the sgRNA successfully binds, Cas9 undergoes a second conformational change which activates the two endonuclease domains. These domains cut the double stranded DNA, causing a double stranded break (DSB) to occur. The cell's machinery will repair the break using non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ is a quick process that often results in indels (insertions/deletions), which can change the protein functionality. HDR requires the simultaneous introduction of a template sequence that the cell's machinery uses to repair the break. This template sequence can include any desired change[10][19]–[21].

While CRISPR has been touted as having the ability to target any unique sequence in a genome, this comes with a caveat of its own. There are some regions of the genome that may be inaccessible, most likely due to chromatin being tightly packed around histones in that location[22][23]. If the CRISPR/Cas9 complex cannot reach the intended gene target, no sequence modification will occur. Targeting an inaccessible gene in a whole plant, embryo, callus, or single cell will appear the same as a failed experiment. It is virtually impossible to know if some part of the process failed or if the

gene target is inaccessible. Therefore, it is critical to test every potential sgRNA target *in vivo* at the single cell level before attempting to transform an entire organism. Although it is important to note that the chromatin state of genes is not static and may change over a cell's or tissue's life span and may also differ from tissue to tissue, and this should be kept in mind when performing an *in vivo* test[23].

While protoplast regeneration is not currently possible in wheat, protoplasts are the best tool to test transformation *in vivo* and avoid wasting time and resources on failed sgRNA targets[24][25]. Although many protoplast isolation methods have been reported, there are few that also involve subsequent CRISPR RNP transformation. For instance, there are only three wheat protoplast isolation and CRISPR transformation protocols published at this time[22][24][26]. There are also some differences between protocols that can be contradictory which would benefit from standardization. For example, the PEG-mediated transformation steps reported in each published protocol show deviations in total volume of over 200 μ l and Cas9:sgRNA final amounts of 40 μ g. This disparity is emphasized in the total amount of cells used in the PEG transformation step: Wang *et al.* (2016)[24] utilizes 1×10^6 cells in 250 μ l (4×10^3 cells/ μ l), Shan *et al.* (2014)[22] utilizes 5×10^5 cells in 460 μ l ($\sim 1 \times 10^3$ cells/ μ l), and Cui (2017)[26] utilizes 2×10^4 cells in 220 μ l (~ 90 cells/ μ l). These disparate numbers illustrate a need for consensus in the PEG-mediated CRISPR transformation of wheat protoplasts that would benefit future researchers. In response to this, we have developed a defined protocol for the entire procedure, from isolation to CRISPR/Cas9 transformation, to assist other researchers in quickly implementing protoplast transformation tests to validate sgRNAs (Figure 5.1).

Before beginning the *in vivo* transformation steps, a gene target must be selected and an appropriate sgRNA designed. In the case of bread wheat, an allohexaploid, most genes are present in three copies, one for each genome (A, B, and D). Homoeologs may be expressed at different levels from each other and are often compensatory, meaning if one copy is knocked out, expression in another copy will increase to make up for it[27][28]. This is important to keep in mind when deciding which gene and/or copies of that gene to target. For this study, we chose to target the gene *GW2*, which is a negative regulator of wheat grain size and thousand grain weight and has homoeologs on each

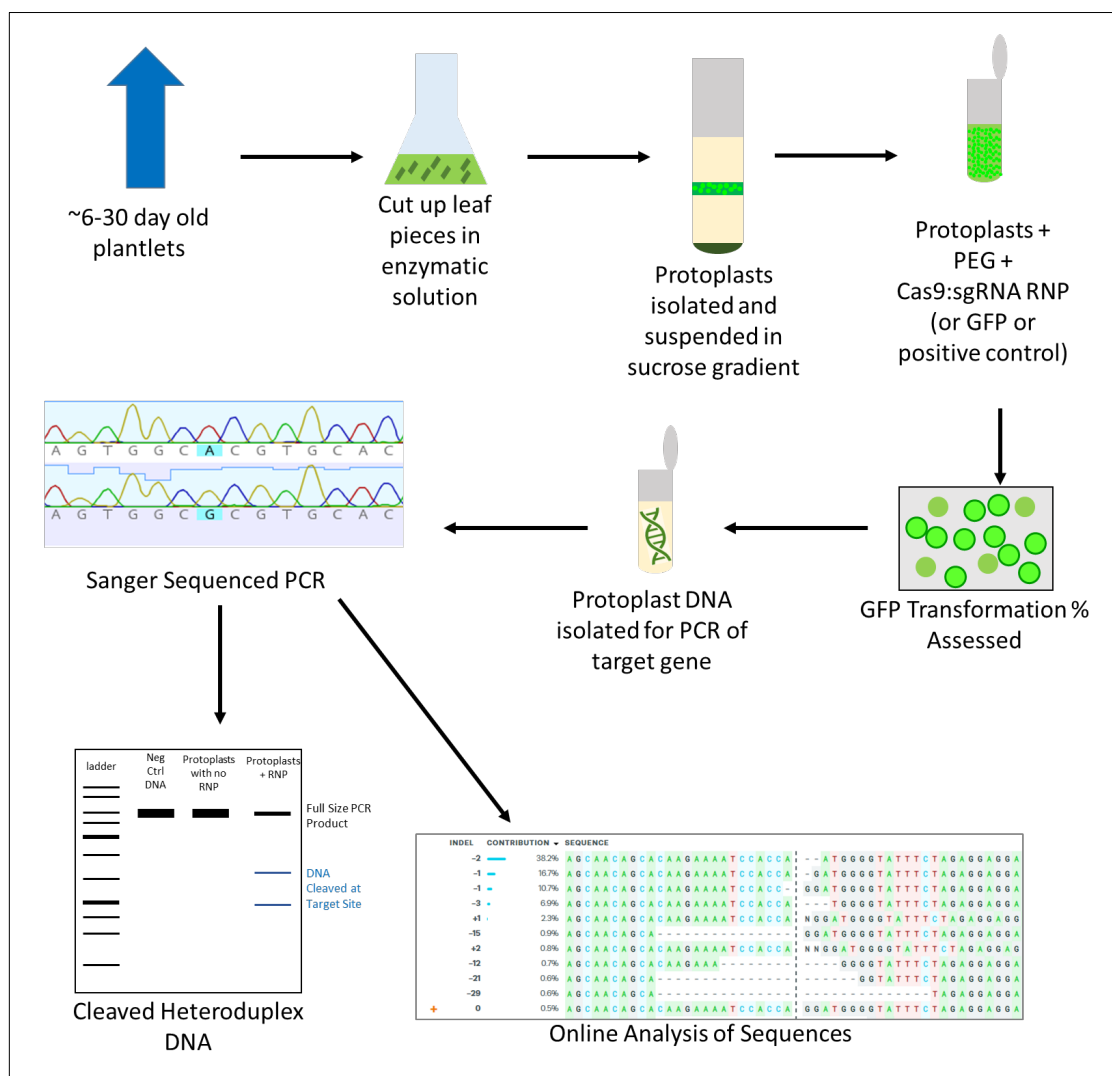


Figure 5.1 Overview of the protoplast isolation and transformation protocol. Leaf tissue is cut into 2mm pieces and digested in an enzymatic solution on a shaker. The protoplasts are then filtered and isolated using a sucrose gradient. A PEG-mediated transformation is performed in microcentrifuge tubes, using GFP as a positive control. After 24-48 hours, the transformation efficiency is assessed by counting GFP fluorescing cells in the positive control. DNA is isolated from CRISPR-transformed protoplasts and PCR is performed. The PCR is then used in a T7EI digestion assay as well as Sanger Sequenced and analyzed using an online program to assess the transformation efficiency of the chosen sgRNA.

genome. The B genome copy of *GW2* has been shown to express at high levels and has been successfully transformed in other CRISPR/Cas9 studies. Therefore, an sgRNA was designed to target *GW2-B* and its functionality was confirmed *in vitro*[5][25][28][29].

The proposed method for the use of protoplast isolation and transformation to determine the potential for trait modification by CRISPR/Cas9 begins with digesting young leaf tissue in an enzymatic solution to remove the cell walls. The protoplasts are then isolated, concentrated, and assessed for viability. Then the protoplasts are transformed with Cas9:sgRNA RNPs via PEG-mediated transformation, using a GFP (Green Fluorescent Protein) construct as a positive control. After incubation with the RNPs, DNA is extracted from the protoplasts and the target sequence is amplified via PCR. The PCR product can then be tested for transformation using a T7EI digestion and/or Sanger sequencing and analysis with a program such as TIDE (Tracking of Indels by Decomposition)[30] or ICE (Inference of CRISPR Edits)[31]. TIDE and ICE are bioinformatic software tools developed to analyze pooled CRISPR-transformed DNA. They use the sequence traces from Sanger sequencing to reconstruct all indels, their frequencies, and quality control metrics, thus providing detailed information about the types and sizes of indels, as well as specific nucleotide changes. After confirmation of the sgRNA *in vivo*, the researcher can move on to transforming a whole organism using any method.

While this protocol was specifically adapted for the assessment of RNP transformation in wheat leaf protoplast cells, there is the potential for it to be applied to other plant species and CRISPR delivery methods, with some modifications. The protoplast isolation steps can also be followed for use with different applications besides transformation with PEG and RNPs.

5.3 Materials

1. Wheat seed (*Triticum aestivum* L. cv. Bobwhite)
2. Oligonucleotides for *GW2* gene (Table 5.1)

Table 5.1 Sequence of oligonucleotides (sgRNA and primers) used to transcribe the sgRNA and amplify the *GW2* gene. Both the sgRNA target sequence and primers were ordered from Eurofins Genomics (Louisville, KY).

sgRNA Name	Sequence (5'-3') (PAM)		
GW2-B	CTAGAAATACCCCATCCTGG(TGG)		
Primer Name	Sequence (5'-3')	T _m °C	Distance from sgRNA Cut Site (base pairs)
GW2-B-F	CTGCCATTACTTTGTATTTTGTACTC	55	500
GW2-B-R	TCCTTCCTCTCTTACCACTTCCC	57	220

3. Guide-it sgRNA *In Vitro* Transcription Kit (Takara Bio, Mountain View, CA, cat. no. 632635)
4. Guide-it IVT RNA Clean-Up Kit (Takara Bio, cat. no. 632638)
5. DNeasy Plant Mini Kit (Qiagen, Germantown, MD, cat. no. 69104)
6. High Fidelity PCR EcoDry™ Premix (Takara Bio, cat. no. 639280)
7. NucleoSpin Gel and PCR Clean-Up kit (Macherey-Nagel, Bethlehem, PA, cat. no. 740609)
8. EnGen™ Spy Cas9 NLS & 10x Cas9 Nuclease Reaction Buffer (New England BioLabs, Ipswich, MA, cat. no. M0386M)
9. Bovine Serum Albumin (Millipore Sigma, St. Louis, MO, cat. no. A2153)
10. Agar (Phytotech Labs, Shawnee Mission, KS, cat. no. A111)
11. 1kb ladder (ThermoFisher Scientific, Waltham, MA, cat. no. 10787018)
12. TBE Buffer (ThermoFisher Scientific, cat. no. B52)
13. Enzyme Solution (Table 5.2)

Table 5.2 Ingredients for the preparation of Enzyme Solution. Solution should either be filter sterilized or autoclaved.

Chemical	Manufacturer	Catalog Number	Quantities
MES, Free Acid	GoldBio, St. Louis, MO	M-095-100	20mM
Cellulase R-10	Duchefa Biochemie, Haarlem, The Netherlands	C8001	1% (wt/vol)
Macerozyme R-10	Duchefa Biochemie	M8002	0.25% (wt/vol)
D-Mannitol	Millipore Sigma	M1902	0.4M
KCl	Millipore Sigma	1.04935	10mM
NaOH	Fisher Scientific, Waltham, MA	S318-100	pH to between 5.7 and 6.0

14. W5 Solution (Table 5.3)

Table 5.3 Ingredients for the preparation of W5 Solution. Solution should either be filter sterilized or autoclaved.

Chemical	Manufacturer	Catalog Number	Quantities
MES, Free Acid	GoldBio	M-095-100	2mM
NaCl	Fisher Scientific	S271-500	154mM
CaCl ₂	Millipore Sigma	C8106	125mM
KCl	Millipore Sigma	1.04935	5mM
NaOH	Fisher Scientific	S318-100	pH to between 5.7 and 6.0

15. W5A Solution (Table 5.4)

Table 5.4 Ingredients for the preparation of W5A Solution. Solution should either be filter sterilized or autoclaved.

Chemical	Manufacturer	Catalog Number	Quantities
Glucose	Millipore Sigma	G8270	5mM
NaCl	Fisher Scientific	S271-500	154mM
CaCl ₂	Millipore Sigma	C8106	125mM
KCl	Millipore Sigma	1.04935	5mM
MES, Free Acid	GoldBio	M-095-100	0.1%
NaOH	Fisher Scientific	S318-100	pH to between 5.7 and 6.0

16. Sucrose (Fisher Scientific, cat. no. S5-500)

17. WI Solution (Table 5.5)

Table 5.5 Ingredients for the preparation of WI Solution. Solution should either be filter sterilized or autoclaved.

Chemical	Manufacturer	Catalog Number	Quantities
D-Mannitol	Millipore Sigma	M1902	0.5M
KCl	Millipore Sigma	1.04935	20mM
MES, Free Acid	GoldBio	M-095-100	4mM
NaOH	Fisher Scientific	S318-100	pH to between 5.7 and 6.0

18. Evans Blue (Millipore Sigma, cat. no. E2129)

19. GFP reporter plasmid (obtained from Cathleen Ma, Oregon State University, via personal communication)

20. PEG Solution (Table 5.6)

Table 5.6 Ingredients for the preparation of PEG Solution. The solution should be freshly prepared but done at least one hour before transformation to completely dissolve the PEG.

Chemical	Manufacturer	Catalog Number	Quantities
PEG 4000	Millipore Sigma	1546569	40% (wt/vol)
D-Mannitol	Millipore Sigma	M1902	0.2M
CaCl ₂	Millipore Sigma	C8106	100mM
ddH ₂ O			Up to final volume

21. Phusion HF Polymerase & Phusion HF buffer 5x (ThermoFisher Scientific, cat. no. F531S)
22. dNTPs (ThermoFisher Scientific, cat. no. R0182)
23. QIAquick PCR Purification Kit (Qiagen, cat. no. 28104)
24. 10x NEBuffer 2 (New England BioLabs, cat. no. B7003S)
25. T7 Endonuclease I (New England BioLabs, cat. no. M0302S)
26. EDTA (Millipore Sigma, cat. no. 819040)

5.4 Stepwise Procedures

5.4.1 In vitro *Transcription of sgRNA*

1. Transcription of the sgRNA is performed with the Guide-it sgRNA *In Vitro* Transcription Kit following the manufacturer's instructions.
2. Purification of the sgRNA is performed with the Guide-it IVT RNA Clean-Up Kit following the manufacturer's instructions.

5.4.2 In vitro *Cleavage Test Using CRISPR/Cas9 RNPs*

1. RNase-free DNA is prepared from Bobwhite plantlets using the Dneasy Plant Mini Kit and following the manufacturer's instructions, omitting the addition of Rnase A to prevent digestion of the sgRNA transcript.
2. PCR amplification of the *GW2* gene target is performed using the gene target primers (Table 5.1) and the High Fidelity PCR EcoDry™ Premix following the manufacturer's instructions.
3. The PCR product is then purified with the NucleoSpin Gel and PCR Clean-Up kit following the manufacturer's instructions.
4. The RNPs are formed by mixing 70ng of sgRNA, 250ng of Cas9 nuclease, and Rnase-Free H₂O to a volume of 1.23μl per reaction.

5. The RNPs are then incubated at 37°C for 5 minutes.
6. The RNP complex is then added to 100ng of PCR product, 1µl of 10x Cas9 buffer, 0.5µl of 10x BSA, and Rnase-Free H₂O to 15µl total.
7. The cleavage reaction is incubated at 37°C for 1 hour, followed by 80°C for 5 minutes.
8. The cleaved PCR product is then run on a 2% agarose gel in TBE buffer with a 1kb ladder. Cleavage activity is assessed based on the amount of digested product compared to the amount of total input DNA

5.4.3 Isolation of Protoplasts

Growth of plantlets takes approximately 3 weeks. Digestion of leaf tissue and isolation of protoplasts takes 1 day.

1. Cut 2g of wheat leaves from young (approximately 6-28 day old) wheat plants and leave to soak in sterile water in the refrigerator while preparing the solutions
2. Add enzymes to Enzyme Solution, warm to 55°C for 10min, cool to room temperature, add CaCl₂ and BSA
**Note: 2g of tissue requires 100ml of Enzyme Solution*
3. Remove wheat leaves from water and cut into 2mm pieces using scissors, put cut pieces into 0.4M D-Mannitol
4. Once Enzyme Solution is ready and all leaf pieces have been cut, filter leaves through a mesh snap ball tea strainer
5. Put leaves into a flask containing the Enzyme Solution, cover with tin foil, and place on an incubating shaker at 27°C, 100rpm, for 3 hours
6. Gently shake the flask to release the protoplasts
7. Filter through a mesh snap ball tea strainer into a small beaker. Rinse the flask and leaf pieces with 20ml of W5 Solution
8. Filter the liquid through a 100µm mesh into a 50ml tube. Rinse the mesh with 1-2ml of W5 Solution
9. Distribute evenly into 4 round-bottomed centrifuge tubes

10. Centrifuge at 100g for 5 minutes
11. Be very careful not to disturb the pellet, and remove the supernatant by pipetting
12. Resuspend the pellet in 4ml of W5A Solution
13. Add 8ml of filter sterilized 21% Sucrose Solution to a new round-bottomed centrifuge tube
14. Very slowly and carefully, layer the protoplast solution on top of the sucrose
**Note: Always cut the ends off of the pipette tips in order to prevent shearing the protoplasts*
15. Centrifuge at 720g for 13 minutes
16. There should be a layer of clear W5A Solution on top, followed by a small layer of green protoplasts, then a large layer of sucrose, and at the bottom are dead/broken protoplasts
17. Slowly harvest the living protoplasts by pipetting, and place in 2.0ml microcentrifuge tubes
18. Bring total volume to 2.0ml with WI Solution
**Note: Without WI solution, the protoplasts may not settle to the bottom, but remain suspended in solution*
19. Cover the tubes with tin foil and leave at 4°C overnight to let the protoplasts settle
20. Once settled, pipette off the supernatant
21. Check concentration with a hemocytometer
22. Check viability with Evans Blue dye by adding 3µl of dye per 100µl of protoplasts and incubating for 10 minutes at room temperature. Any dead tissue will be dyed blue.
23. Make up to desired concentration with WI Solution (ideal concentration is 0.7-1.0 x 10⁶ cells/ml)
 - a. Save some protoplasts at 4°C to use as a negative control in 5.4.5 step 1

5.4.4 PEG-Mediated Transformation

PEG-mediated transformation takes 1 to 2 days.

1. Make 40-60µg of RNPs with a 3:1 ratio of Cas9:sgRNA in a total volume of <20µl by mixing Cas9 + sgRNA + Cas9 Nuclease Reaction Buffer (if needed) and incubating for 10 minutes at room temperature
 - a. Add 15µg of GFP reporter plasmid to positive control instead of RNPs (15ug of GFP in 40µl or less + 50µl of protoplasts at $0.7-1 \times 10^6$ cells/ml + 50% volume PEG)
2. Add RNPs to 50µl of protoplasts (at $0.7-1.0 \times 10^6$ cells/ml) and mix them gently by tapping the tube
3. Add the same volume in µl of freshly prepared PEG solution as there is of RNPs and protoplasts (so that the final volume is approximately 50% PEG solution), and mix thoroughly by gently inverting the tube until homogenous

**Note: DNA, protoplasts, and PEG solution should be added to the microcentrifuge tube in that order. Don't delay in between adding protoplasts and PEG. Once protoplasts and DNA have been mixed together, add PEG immediately*
4. Incubate mixture for 15-20min at room temperature in the dark. Add 2 times the volume of WI Solution to the tube (eg. 400 µl RNP/protoplast mix gets 800 µl WI solution) and mix well by inverting the tube to stop the transformation process
5. Centrifuge tubes at 150g for 3min at room temperature. Remove the supernatant by pipetting
6. Resuspend protoplasts gently to 500µl with WI Solution
7. Coat 2ml microcentrifuge tubes with 5% BSA (5mg BSA in 1ml sterile water) to prevent protoplasts from sticking to the plastic and each other. Discard excess BSA
8. Transfer all 500µl of the protoplast solution into the BSA-coated tubes
9. Wrap the tubes with aluminum foil, lay them on their sides, and incubate at 23°C for 24 to 48 hours
 - b. GFP expression peaks after 24 hours if using a transient plasmid
 - c. RNPs should be allowed to incubate for 48 hours
 - d. Check the transformation efficiency by counting the number of GFP-fluorescing cells in the positive control using a fluorescent microscope and hemacytometer

10. Collect the protoplasts by centrifuging at 12,000g for 2min at room temperature.

Remove the supernatant by pipetting

**Note: Remember to also collect the protoplasts from the negative control sample from step A23*

5.4.5 Extraction of DNA, T7EI Digestion, and Detection of Mutations

The extraction of protoplast DNA takes approximately 4 hours. PCR of the extracted DNA and heteroduplex formation takes approximately 5 hours. The T7EI digestion takes approximately 2 hours.

1. Extract genomic DNA with a method of your choice
2. Determine the DNA concentration with a spectrophotometer. The usual concentration is around 30ng/μl in a total volume of 30μl

**Pause Point: Extracted DNA can be stored at -20°C for several months, or 4°C for several weeks*

3. Set up the PCR reaction to amplify the genomic region targeted for mutagenesis (Table 5.7)

**It is important to use high-fidelity polymerase to reduce the error rate and to use the protoplasts saved in step A23 as a negative control*

Table 5.7 PCR reaction for amplifying the desired genomic region

Component	Amount (μl)	Final concentration
Phusion HF buffer, 5x	5	1x
dNTP, 2mM (0.5mM each)	2.5	0.2mM
PCR-Fwd primer, 10μM	1.25	0.5μM
PCR-Rev primer, 10μM	1.25	0.5μM
Phusion HF polymerase, 2U/μl	0.25	0.5 units
DNA template	8	3-6 ng/μl
ddH ₂ O	To 25	

4. Perform the PCR for amplifying the desired genomic region (Table 5.8)

Table 5.8 PCR reaction parameters

Step Number	Denature	Anneal	Extend
1	98°C, 30s		
2-4	98°C, 10s	62°C, 30s, -0.5°C per cycle	72°C, 30s
5	Go to 2, 8 times		
6-8	98°C, 10s	58°C, 30s	72°C, 30s
9	Go to 6, 32 times		
10			72°C, 5min

6. Clean up the PCR with a kit of your choice following the manufacturer's instructions
 - a. At this point, Sanger sequence each sample and use CRISPR analysis software to quantify the transformation rate and determine the type of change

**Pause Point: PCR products can be stored at 4°C for several weeks.*

7. Using the PCR product from above, set up an annealing reaction to form DNA heteroduplexes (Table 5.9)

Table 5.9 Reaction mixture for the formation of heteroduplexes

Component	Amount (μl)
DNA	200ng
10x NEB buffer2	2
Nuclease-free H ₂ O	X
Total	19

8. Run the annealing reaction in a thermocycler using the following conditions:
95°C for 5min, 95-85°C (-2°C/sec), 85-25°C (-10°C/min), 4°C forever
9. Use the annealed heteroduplex to set up a T7EI nuclease digestion (Table 5.10)

Table 5.10 Reaction mixture for the cleavage of heteroduplexes using T7EI endonuclease

Component	Amount (μl)
PCR	19
T7EI	1
Total	20

10. Mix well and spin the mixture down briefly. Incubate the reaction at 37°C for 15min, then stop reaction with 1.5µl of 0.25M EDTA
 - a. The incubation time can be increased if necessary
11. Run digestion products on a 2% agarose gel in TBE buffer using standard protocols. Include a DNA ladder and negative controls on the same gel. If digested bands are only observed in the CRISPR/Cas-treated sample, and not in the negative control, an indel mutation has occurred

5.5 Results and Discussion

5.5.1 Protoplast Isolation

Stem tissue and older leaves that had been growing more than one month and lost their light green color and pliability consistently gave lower protoplast yields than young- to relatively-young leaf tissue (below 2.0×10^5 cells/ml). Cutting the plantlet's leaves back to approximately 12.5mm from the stem is an alternative to planting more seeds, as the new growth is acceptable for use in protoplast isolation. This is likely due to levels of cellulose and lignin increasing as the tissue ages, making digestion more difficult[32].

Despite keeping the ratio of plant tissue to enzymatic solution the same, increasing or decreasing the total volume in the flask at 5.4.3 step 2 resulted in decreased yields. Two grams of leaf tissue per 100 ml of enzymatic solution in a 250ml flask gave the highest total viable protoplast yields. Using this protocol, 2g of leaf tissue consistently yielded an average of 1.3×10^6 total viable protoplasts at 5.4.3 step 21, which is easily made up to a concentration of 0.7 to 1.0×10^6 cells/ml for use in transformation. On average, Evans Blue assays showed approximately 15 percent dead protoplasts at this stage (Figure 5.2). The protoplasts survived storage in 2ml microcentrifuge tubes in the refrigerator for up to 3 days after isolation and concentration at 5.4.3 step 23 before viability decreased to unacceptable levels.

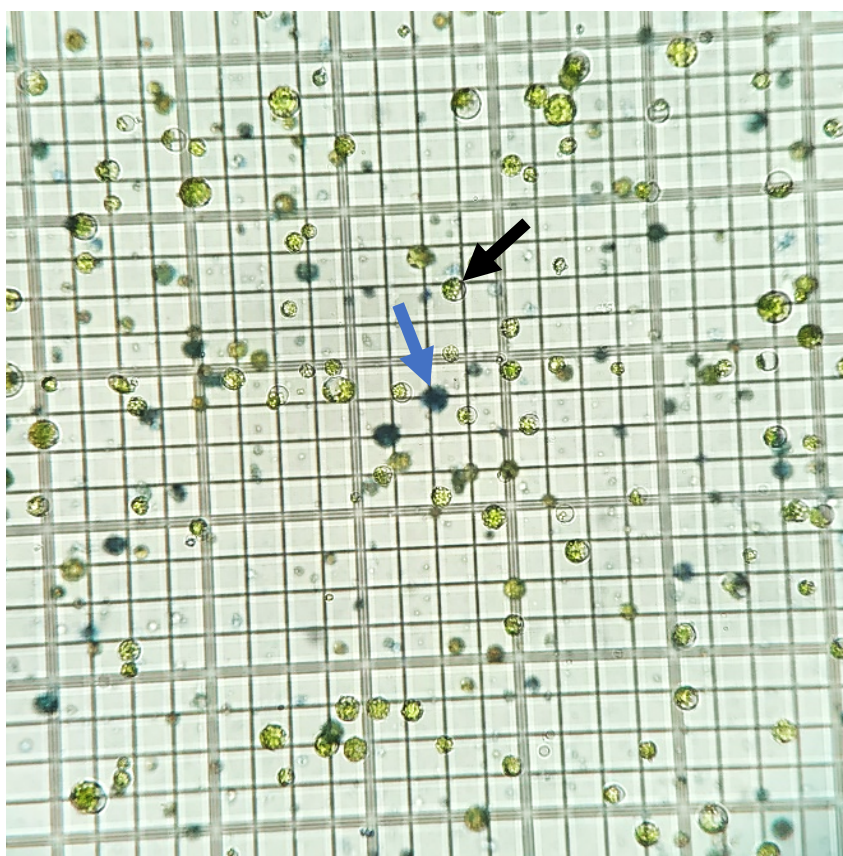


Figure 5.2 Protoplasts stained with Evans Blue dye on a hemacytometer. Viable cells are bright green and indicated by the black arrow. Nonviable, dead cells are blue and indicated by the blue arrow.

5.5.2 PEG-Mediated Transformation

Low concentrations of sgRNA lead to low transformation percentages due to the high volumes of reagents required, and this low percentage did not reach the minimum threshold for use in CRISPR applications. We found that any sgRNA should have a concentration of least 2000ng/ μ l for use in PEG-mediated transformation so that the total volume of RNP solution can be kept low. We also found that a Cas9:sgRNA ratio of 3:1 works better than a ratio of 1:1. Other protoplast transformation protocols reported success with a 1:1 ratio, and we found that it can be used if needed, although the transformation efficiencies will be slightly lower. Protoplast concentrations of 0.7 to 1.0 $\times 10^6$ cells/ml are ideal for transformation in order to keep the volume of cells at around 50 μ l, while at the same time ensuring the concentration of protoplasts will be high

enough to isolate sufficient DNA, but not so high as to overwhelm the assay causing low transformation efficiencies (Table 5.11).

Table 5.11 Ideal ratios of amount of proteins to number of protoplasts, and volume of proteins to volume of protoplasts for GFP control and for CRISPR RNPs.

For GFP control:	μg of GFP/total protoplast number	0.0002 to 0.0004
	μl of GFP/total μl protoplasts	0.4 to 0.6
For RNPs:	μg of RNPs/total protoplast number	0.001 to 0.002
	μl of RNPs/total μl protoplasts	0.2 to 0.4

While higher transformation efficiencies are easier to visualize on the microscope and on the T7EI gel, it is still possible to visualize positive results at as low as 20% GFP transformation (Figure 5.3). An sgRNA construct with associated transformation efficiencies of $\frac{1}{5}$ (or greater) that of the GFP transformation efficiency can be used for the transformation of plants.

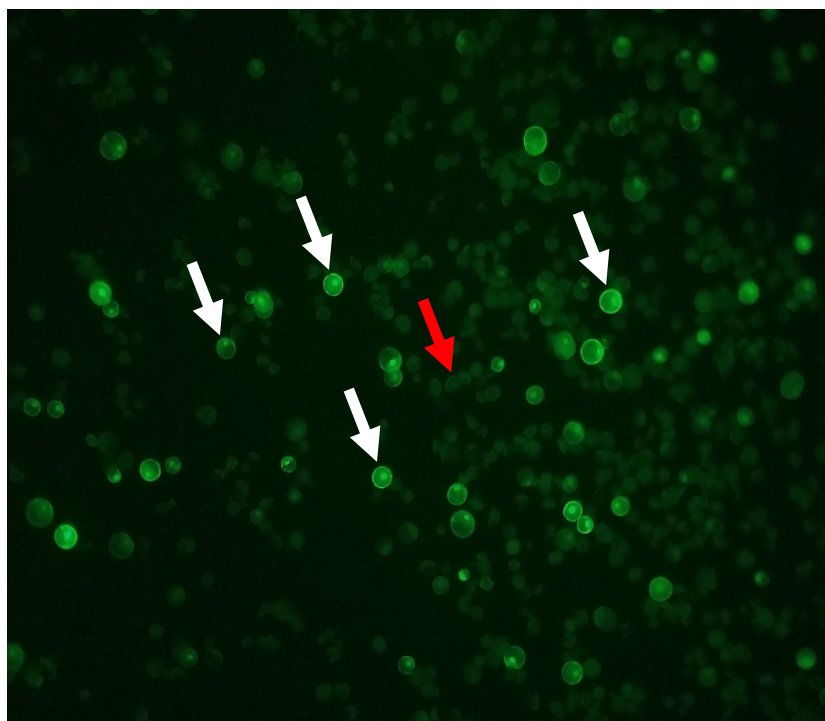


Figure 5.3 GFP fluorescing protoplast cells. Cells with bright green borders and nuclei are expressing the GFP plasmid. A selection of GFP fluorescing cells are indicated with white arrows. Solid, light green cells are untransformed. An example of an untransformed cell is indicated with the red arrow.

5.5.3 Extraction of DNA and T7EI Digestion

DNA yields after extraction with the Qiagen DNeasy Plant Mini Kit were consistently around 30ng/μl in a total volume of 30μl when using 50μl of cells (~50,000 cells total) for transformation. These yields increased to 60ng/μl when 200μl of cells (~200,000 cells total) were used. At 5.4.5 step 10, we found that the T7EI digestion time should be altered to match the expected transformation efficiency based off of the GFP transformation percentage. If GFP transformation is low, around 10 percent, the digestion time should be increased up to one hour. If GFP transformation is high, around 60 percent, the digestion time can be left at 15 minutes. This ensures that even with low transformation rates, the digested bands will become concentrated enough to be visible on the electrophoresis gel. The *GW2* gene target in this study showed successful digestion with T7EI endonuclease after 15 minutes (Figure 5.4).

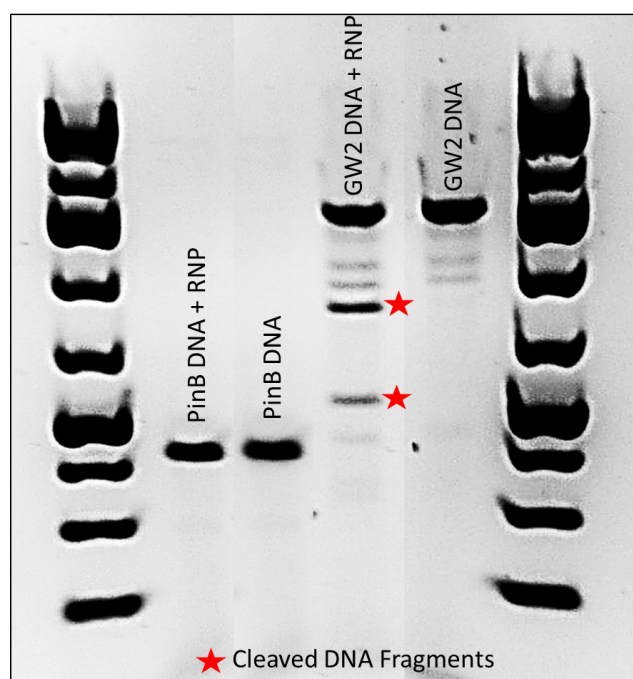


Figure 5.4 A 2% agarose gel with T7EI digested PCR and Negative Controls. A 1Kb ladder is on each end. *PinB* is a gene target that is inaccessible due to its chromatin state. *GW2* is the positive control gene target. '*PinB* DNA + RNP' shows no difference from the '*PinB* DNA' alone, meaning that no transformation occurred in the protoplasts. '*GW2* DNA + RNP' shows two bands of DNA that are the expected size, and together they add up to the full *GW2* PCR fragment size seen at the top of the lane and in the '*GW2* DNA' negative control lane.

5.5.4 Detection of Mutations

The PCR product that showed successful T7EI digested *GW2* product in Figure 5.4 was also sequenced using Sanger Sequencing. Sanger Sequencing results are usually given as chromatograms with the occurrence of nucleotides represented as colored peaks. The chromatograms of transformed protoplast DNA do not typically give any indication of the transformation efficiency or type of transformation due to the variation in types of edits in each protoplast cell and due to the presence of untransformed protoplast cell DNA in each sample. Therefore, the changes must be visualized using software developed for deconvoluting pooled sequence data from CRISPR transformations. The visualization of the *GW2* DNA in this study was done using ICE. A representative sample of protoplast DNA sequencing analysis is shown in Figure 5.5. For this sample, ICE reported that out of all the protoplast DNA sequences present, 85.8 percent were unchanged, 3.4 percent had a single nucleotide insertion after the expected cut site, 2.4 percent had six deleted nucleotides, 1.2 percent had four deleted nucleotides, one percent had five deleted nucleotides, and fewer than one percent of protoplast sequences had three, 20, one, and 21 deleted nucleotides. The total percent of protoplast sequences that were transformed with the CRISPR RNP was 14.2 percent, with nine percent of those causing frameshift mutations that would significantly impact protein function of the *GW2* gene. All protoplast transformation assays using this protocol with the *GW2-B* sgRNA had total transformation percentages above 13. Protoplast transformations with *GW2* also showed a penchant for deletions rather than insertions, with approximately 70 percent of the deletions leading to premature stop codons. Based on previous *in vivo* transformation studies, an sgRNA with a protoplast transformation rate of at least 10 percent is a viable candidate for recovering transformed plants using other methods[22][24]. The 13+ percent transformation achieved in this study is above this threshold, which reinforces the use of *GW2-B* as a positive control gene target as well as shows the ability of this protocol to screen potential sgRNA candidates *in vivo*.

Indel	Contribution	Sequence
0	85.8%	CCTCCTCTAGAAATACCCCATCC <u>T</u> GGTGGATTTTCTTGTGCTGTTGCTG
+1	3.4%	CCTCCTCTAGAAATACCCCATCC <u>N</u> GGTGGATTTTCTTGTGCTGTTGCTG
-6	2.4%	CCTCCTCTAGAAATACCC- - - - -GGTGGATTTTCTTGTGCTGTTGCTG
-4	1.2%	CCTCCTCTAGAAATACCC- - - - -TGGTGGATTTTCTTGTGCTGTTGCTG
-5	1.0%	CCTCCTCTAGAAATACCC- - - - -GGTGGATTTTCTTGTGCTGTTGCTG
-3	0.9%	CCTCCTCTAGAAATACCCCA- - -TGGTGGATTTTCTTGTGCTGTTGCTG
-20	0.7%	CCTCCTCT- - - - -GATTTTCTTGTGCTGTTGCTG
-1	0.6%	CCTCCTCTAGAAATACCCCATCC- GGTGGATTTTCTTGTGCTGTTGCTG
-21	0.6%	CCTCCTCT- - - - -ATTTTCTTGTGCTGTTGCTG

Figure 5.5 ICE Analysis of the chromatogram sequences. The sgRNA target is underlined in the first sequence. This mutant protoplast DNA sample has a 14.2% total transformation rate, with a single nucleotide addition at the cut site being the most common edit at 3.4%.

5.6 Conclusions

This research demonstrates the effectiveness of our simplified protocol for protoplast isolation from wheat leaf tissue and subsequent transformation with CRISPR/Cas9 RNPs and GFP. While other protocols are available for protoplast isolation and transformation, there is a need for consensus in some steps. Methods including peeling the abaxial epidermis tissue and vacuum infiltration in the enzyme solution were found to be unnecessary for isolation of sufficient, viable protoplasts. Researchers attempting to integrate *in vivo* testing of CRISPR sgRNA gene targets will also benefit from information presented in this study on ideal weight of leaf tissue used, ideal ratio of leaf tissue to enzyme solution, expected viability of isolated protoplasts, and ideal amount and volume ratios of RNPs to cells. This study shows that we have developed a simplified protocol for use in any lab that wishes to isolate protoplasts and/or implement CRISPR transformation, which gives comparable results to the other available methods. This protocol requires less specialized equipment than others and provides easily replicable steps and results, thus making implementation in any lab straightforward.

5.7 Acknowledgements

Our methods were modified from Shan *et al.* (2014)[22] and Yoo *et al.* (2007)[33], and we wish to acknowledge their important contributions. We would also like to thank Taj Arndell of The University of Adelaide for his willingness to share protocols and knowledge, and Cathleen Ma of Oregon State University for providing the GFP construct used as a positive control. This work was supported by the Oregon Agricultural Experiment Station, Warren E. Kronstad Wheat Research Endowment, the Oregon Wheat Commission, and the ARCS Oregon Chapter Scholar Award.

5.8 References

- [1] W. Bushuk, "Wheat breeding for end-product use," in *Braun H.J., Altay F., Konstad W.E., Beniwal S.P.S., McNab A. (eds) Wheat: Prospects for global improvement*, Ankara, Turkey, 1997, vol. 6, pp. 203–211.
- [2] P. Gustafson, O. Raskina, X. Ma, and E. Nevo, "Wheat evolution, domestication, and improvement," in *Wheat Science and Trade*, Wiley-Blackwell, 2009, pp. 3–30.
- [3] T. Cowan, "Unapproved Genetically Modified Wheat Discovered in Oregon and Montana: Status and Implications," Congressional Research Service, CRS Report for Congress R43100, Dec. 2014.
- [4] E. Stokstad, "Monsanto Pulls the Plug on Genetically Modified Wheat," *Science*, vol. 304, no. 5674, pp. 1088–1089, May 2004.
- [5] B. Wang, D. L. Liu, S. Asseng, I. Macadam, and Q. Yu, "Impact of climate change on wheat flowering time in eastern Australia," *Agricultural and Forest Meteorology*, vol. 209–210, no. Supplement C, pp. 11–21, Sep. 2015.
- [6] W. Wang *et al.*, "Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat," *The CRISPR Journal*, vol. 1, no. 1, pp. 65–74, Feb. 2018.
- [7] P. Bhowmik *et al.*, "Targeted mutagenesis in wheat microspores using CRISPR/Cas9," *Sci Rep*, vol. 8, pp. 1–10, Apr. 2018.
- [8] E. Haque *et al.*, "Application of CRISPR/Cas9 Genome Editing Technology for the Improvement of Crops Cultivated in Tropical Climates: Recent Progress, Prospects, and Challenges," *Front Plant Sci*, vol. 9, pp. 1–12, May 2018.
- [9] K. Li, G. Wang, T. Andersen, P. Zhou, and W. T. Pu, "Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System," *PLOS ONE*, vol. 9, no. 8, p. e105779, Aug. 2014.
- [10] A. Malzahn, L. Lowder, and Y. Qi, "Plant genome editing with TALEN and CRISPR," *Cell & Bioscience*, vol. 7, no. 1, p. 21, Apr. 2017.
- [11] D. Jaganathan, K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman, "CRISPR for Crop Improvement: An Update Review," *Front Plant Sci*, vol. 9, no. 985, Jul. 2018.
- [12] A. Scheben, F. Wolter, J. Batley, H. Puchta, and D. Edwards, "Towards CRISPR/Cas crops – bringing together genomics and genome editing," *New Phytologist*, vol. 216, no. 3, pp. 682–698, Nov. 2017.
- [13] K. Xie and Y. Yang, "RNA-Guided Genome Editing in Plants Using a CRISPR–Cas System," *Molecular Plant*, vol. 6, no. 6, pp. 1975–1983, Nov. 2013.
- [14] Q. Shan *et al.*, "Targeted genome modification of crop plants using a CRISPR–Cas system," *Nature Biotechnology*, vol. 31, pp. 686–688, Aug. 2013.
- [15] V. Nekrasov, B. Staskawicz, D. Weigel, J. D. G. Jones, and S. Kamoun, "Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease," *Nature Biotechnology*, vol. 31, no. 8, pp. 691–693, Aug. 2013.
- [16] J.-F. Li *et al.*, "Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9," *Nature Biotechnology*, vol. 31, no. 8, pp. 688–691, Aug. 2013.
- [17] Z. Feng *et al.*, "Efficient genome editing in plants using a CRISPR/Cas system," *Cell Research*, vol. 23, no. 10, pp. 1229–1232, Oct. 2013.

- [18] L. Bortesi and R. Fischer, "The CRISPR/Cas9 system for plant genome editing and beyond," *Biotechnology Advances*, vol. 33, no. 1, pp. 41–52, Jan. 2015.
- [19] T. Gaj, C. A. Gersbach, and C. F. Barbas, "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering," *Trends in Biotechnology*, vol. 31, no. 7, pp. 397–405, Jul. 2013.
- [20] D. Carroll, "Genome Engineering With Zinc-Finger Nucleases," *Genetics*, vol. 188, no. 4, pp. 773–782, Aug. 2011.
- [21] J. K. Joung and J. D. Sander, "TALENs: a widely applicable technology for targeted genome editing," *Nature Reviews Molecular Cell Biology*, vol. 14, no. 1, pp. 49–55, Jan. 2013.
- [22] Q. Shan, Y. Wang, J. Li, and C. Gao, "Genome editing in rice and wheat using the CRISPR/Cas system," *Nature Protocols*, vol. 9, no. 10, pp. 2395–2410, Sep. 2014.
- [23] V. Ondřej *et al.*, "Chromatin structural rearrangement during dedifferentiation of protoplasts of *Cucumis sativus* L.," *Molec Cells*, vol. 27, no. 4, pp. 443–447, April 2009.
- [24] W. Wang, A. Akhunova, S. Chao, and E. Akhunov, "Optimizing multiplex CRISPR/Cas9-based genome editing for wheat," *bioRxiv* doi: 10.1101/051342, May 2016.
- [25] Z. Liang *et al.*, "Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 *in vitro* transcripts or ribonucleoproteins," *Nature Protocols*, vol. 13, no. 3, pp. 413–430, Mar. 2018.
- [26] X. Cui, "Targeted Gene Editing Using CRISPR/Cas9 in a Wheat Protoplast System," University of Ottawa, 2017.
- [27] A. Slade *et al.*, "Development of high amylose wheat through TILLING," *BMC Plant Bio*, vol. 12, pp. 69, May 2012.
- [28] W. Wang *et al.*, "Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat," *Theor Appl Genet*, vol. 131, no. 11, pp. 2463–2475, Nov. 2018.
- [29] Z. Liang *et al.*, "Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes," *Nat Commun*, vol. 8, no. 14261, pp. 1–5, Jan. 2017.
- [30] E. K. Brinkman, T. Chen, M. Amendola, and B. van Steensel, "Easy quantitative assessment of genome editing by sequence trace decomposition," *Nucleic Acids Res*, vol. 42, no. 22, pp. e168–e168, Dec. 2014.
- [31] T. Hsiao *et al.*, "Inference of CRISPR Edits from Sanger Trace Data," *bioRxiv* doi: 10.1101/251082, Jan. 2019.
- [32] C. Bi, F. Chen, L. Jackson, B. S. Gill, and W. Li, "Expression of Lignin Biosynthetic Genes in Wheat during Development and upon Infection by Fungal Pathogens," *Plant Mol Biol Rep*, vol. 29, no. 1, pp. 149–161, Mar. 2011.
- [33] S.-D. Yoo, Y.-H. Cho, and J. Sheen, "Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis," *Nature Protocols*, vol. 2, no. 7, pp. 1565–1572, Jul. 2007.

Chapter 6
Testing Non-Transgenic CRISPR Technology for Wheat Improvement

6.1 Abstract

CRISPR/Cas9 transformation is being widely used in many crop science disciplines to modify gene expression because it is relatively easy, often less expensive than other methods, requires only a moderate level of expertise, and does not currently fall under the United States Department of Agriculture's regulation of genetically modified plant material. CRISPR has the potential to contribute to crop improvements that were previously infeasible due to a lack of traits in existing germplasm, tightly linked undesirable traits, and/or extensive breeding times. A caveat for its use in wheat, however, is the need to achieve editing with no off-target effects and no incorporation of foreign DNA using a robust method of introduction of the CRISPR/Cas9 system. Here, we describe a test conducted to determine the feasibility of using non-transgenic CRISPR gene editing in a wheat breeding program modeled after the two published reports of successful editing using an entirely non-transgenic method. To achieve editing, the Cas9:sgRNA construct was delivered as a ribonucleoprotein (RNP) complex via particle bombardment of wheat immature embryos. A total of 34 T₀ transformed plantlets out of 510 were recovered, for a transformation efficiency of 6.7%. However, upon subsequent sampling of individual tillers from the fully-grown transformed plantlets, the presence of chimeric tissue became apparent. This reduced the potential transformation efficiency to less than 1%. Therefore, due to the large amounts of input required to achieve successful editing, this technique is not considered feasible for use in a breeding program, and alternative methods that do not result in the formation of chimeric tissue are needed.

6.2 Introduction

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a bacterial adaptive immunity system that was first adapted for targeted gene editing in eukaryotes in 2013[1]-[3]. The most common system used in plants today consists of the nonspecific Cas9 protein bound to a tracrRNA (trans-activating crRNA) which is then bound to a 20 nucleotide-long sgRNA (target-specific crRNA) sequence designed by the user. This system can theoretically target any sequence in a genome that meets two conditions: first, the target sequence must be unique, and second, the target sequence

must be adjacent to the correct Protospacer Adjacent Motif (PAM) sequence which is determined by the Cas protein used[3][4].

Since its introduction, CRISPR/Cas9 has been used to introduce targeted genetic modifications to a multitude of crops[5]-[8]. Normally, the CRISPR/Cas9 constructs are delivered into cells via plasmids and *Agrobacterium*-mediated transformation. This often results in the intentional or unintentional incorporation of these constructs into the target plant's genome, resulting in the creation of a transgenic organism[9]. This integration can also significantly increase the chances of off-target mutations occurring due to the increased duration of the construct's expression. At present, the only way to remove these transgenes is by crossing the transformed plant to an untransformed plant[10]. This can be difficult to do in certain species, sometimes requires multiple generations, and increases the amount of time needed to incorporate a trait into the adapted germplasm. The creation of transgenic material and unintended mutations are also of concern to regulatory agencies, growers, and the public[11]. In response to these issues, efforts have been made to find transient transformation systems that do not become incorporated into the target genome and are expressed for only a short time before being degraded, thus reducing the occurrence of off-target mutations[12][13].

One such system for transient transformation is utilizing CRISPR/Cas9 Ribonucleoproteins (RNPs). RNPs have been shown to be functional almost immediately after transfection and then quickly degrade, thus significantly decreasing off-target mutations and avoiding incorporation into the target organism's DNA, unlike plasmids[12]. They are also simple to construct as they do not require the creation of plasmids or cell culture and are cost efficient compared to some of the other transformation methods[14].

Recently, CRISPR/Cas9 RNPs were used by Liang *et al.* (2017 and 2018) to successfully transform wheat[14][15]. Their methods reported the use of particle bombardment of immature embryos with *in vitro* transcribed sgRNAs, which were then propagated via callus formation and tissue culture. They found no off-target mutations, determined that the mutants were completely transgene-free, and reported a transformation rate of 4.4%, which is comparable to other particle bombardment

methods. Thus, Liang *et al.* present a plausible approach for the successful transformation of wheat using CRISPR/Cas9.

Wheat is one of the most important crops in the world, and therefore many researchers are beginning to use CRISPR transformation to study gene function for disease resistance, water use efficiency, salt tolerance, pesticide resistance, yield, and quality traits[3][16]-[18]. While CRISPR has been touted as having the ability to significantly change crop breeding, the feasibility of its use in a wheat breeding program for commercialization, rather than only for genetic studies, needs to be evaluated. Therefore, we followed the protocol published by Liang *et al.* (2018) for transforming immature embryos with RNPs in order to test the feasibility of implementing non-transgenic CRISPR/Cas9 transformation in a public wheat breeding program.

6.3 Materials

6.3.1 Germplasm

Triticum aestivum L. cv. Bobwhite, a spring wheat, was grown in one-gallon pots in the greenhouse at 70-75°C under 16 hours of light per day.

6.3.2 Genes

GW2 is a negative regulator of wheat grain size and thousand grain weight, with homoeologues on each of the three genomes of wheat. The B genome copy showed high expression and transformation efficiency in previous studies, and therefore was chosen as the target for this study[14][15][19][20].

6.4 Methods

6.4.1 sgRNA Design

Using BLAST (Basic Local Alignment Search Tool), we designed primers and a sgRNA to target the B genome copy of the *GW2* gene (Table 6.1). Both the sgRNA target sequence and primers were ordered from Eurofins Genomics (Louisville, KY). PCR was performed on RNase-free DNA and the sequence was confirmed using Sanger sequencing.

Table 6.1 Sequence of oligonucleotides (sgRNA and primers) used to target the *GW2* gene. These oligos were used for all subsequent experiments. 'T_m' is the melting temperature.

sgRNA Name	Sequence (5'-3') (PAM)		
GW2-B	CTAGAAATACCCCATCCTGG(TGG)		
Primer Name	Sequence (5'-3')	T _m °C	Distance from sgRNA Cut Site (base pairs)
GW2-B-F	CTGCCATTACTTTGTATTTTGTACTC	55	500
GW2-B-R	TCCTTCCTCTCTTACCACTCCC	57	220

6.4.2 In vitro Transcription of sgRNA

Transcription of sgRNA was performed with the Guide-it sgRNA *In Vitro* Transcription Kit (cat. no. 632635) from Takara Bio USA, Inc (Mountain View, CA) following the manufacturer's instructions. Purification of the sgRNA was then performed using the Guide-it IVT RNA Clean-Up Kit (Takara Bio, cat. no. 632638), following the manufacturer's instructions.

6.4.3 In vitro Cleavage Test Using CRISPR/Cas9 RNPs

RNase-free DNA was prepared using the DNeasy Plant Mini Kit from Qiagen (Germantown, MD, cat. no. 69104), omitting the addition of RNase A to prevent digestion of the sgRNA transcript. PCR amplification was performed using the gene target primers and the High Fidelity PCR EcoDry™ Premix (cat. no. 639280) from Takara Bio USA, Inc. and purified with the NucleoSpin Gel and PCR Clean-Up kit (cat. no. 740609) from Macherey-Nagel (Bethlehem, PA). EnGen™ Cas9 NLS, *S. pyogenes*, (cat. no. M0646T) was obtained from New England Biolabs (Ipswich, MA).

The cleavage reaction was as follows: First, the RNP complex was formed by mixing 70ng of sgRNA, 250ng of Cas9 nuclease, and RNase-Free H₂O to a volume of 1.23μl per reaction. The mix was incubated at 37°C for 5 minutes. Then the RNP complex was added to 100ng of PCR product, 1μl of 10x Cas9 buffer, 0.5μl of 10x BSA, and RNase-Free H₂O to 15μl total. The reaction was incubated at 37°C for 1 hour, followed by 80°C for 5 minutes. The product was then run on a 2% agarose gel in TBE

buffer. Cleavage activity was assessed based on the amount of digested product compared to the amount of total input DNA.

6.4.4 *In vivo Cleavage Test Using PEG-Mediated Transformation of Protoplasts with RNPs*

The protoplasts were isolated and transformed following Brandt *et al.* 2019 (in prep). Briefly, wheat seeds were germinated and grown in soil to approximately six to 28 days old, then leaf tissue was cut into 2mm pieces and used for digestion in an enzymatic solution to remove the cell walls. The protoplasts were then isolated, concentrated, and assessed for viability. Then the protoplasts were transformed using RNP complexes via PEG-mediated transformation, using a Green Fluorescent Protein (GFP) construct as a positive control (obtained from Cathleen Ma, Oregon State University via personal communication). After incubation with the RNPs, DNA was extracted from the protoplasts and the target sequence was amplified via PCR. Heteroduplexes were then created and subsequently digested with T7EI endonuclease. At this stage, an agarose gel image was analyzed and the pre-heteroduplexed PCR was Sanger sequenced and analyzed using the online program ICE (Inference of CRISPR Edits)[21]. ICE is a bioinformatic software tool developed to analyze CRISPR-transformed DNA that has been pooled or comes from chimeric tissue. ICE uses the sequence trace files from Sanger sequencing to reconstruct all indels, their frequencies, and produce quality control metrics, thus providing detailed information about the types and sizes of indels as well as specific nucleotide changes.

6.4.5 *Tissue Culture*

Wheat callus was produced following Hansen *et al.* (1998)[22]. In short, spikes from Bobwhite plants were collected eleven days post-anthesis and stored in water at 4°C for up to one week. The heads were sterilized in 80% Ethanol for 30 seconds, followed by 5% bleach plus 1ml of Tween per 100ml of solution for 10 minutes. The heads were then rinsed twice in distilled water and allowed to air dry for two to five minutes in a laminar flow hood. Using aseptic technique, immature seeds were collected from the primary and secondary florets of each spikelet. The outer layer of seed coat was cut and

peeled away using a sterile scalpel to expose the immature embryo. The embryo was gently removed using the scalpel and placed scutellum side up (flat side up, radicle side down) on a 15mm x 100mm petri dish of Callus Induction Medium. Once a 32mm diameter circle in the middle of the petri dish was filled with embryos (approximately 8 wheat heads), the plate was wrapped with parafilm and placed in a dark growth chamber at 20-25°C for up to one week prior to particle bombardment. Callus Induction Medium is made with 4.43g of Murashige and Skoog Basal medium (Sigma-Aldrich, St. Louis, MO, cat. no. M0404), 2ml of 1mg/ml 2-4, D Solution (PhytoTech Labs, Shawnee Mission, KS, cat. no. D295), 30g of sucrose (Fisher Scientific, Waltham, MA, cat. no. S5-500), 100mg/L of *myo*-Inositol (PhytoTech Labs, cat. no. 1703), 0.5ml/L of Plant Preservative Mixture (PPM) (Plant Cell Technology, Washington, D.C., cat. no. 250), DI H₂O to 1L, pH between 5.7 and 5.8, and 2g of Gelrite (Gelzan CM) (Millipore Sigma, St. Louis, MO, cat. no. G1910), then autoclaved.

Two to three weeks after bombardment, the callus was transferred to Shoot Induction Medium and placed in a lighted growth chamber with a 12 hour light/dark cycle at 20-25°C for approximately four weeks, or once the shoots were large enough to be transferred. Shoot Induction Medium is made as Callus Induction Medium, but with 0.5ml of 1mg/ml Dicamba (Phytotech Labs, cat. no. D165) instead of 2-4, D.

When the shoots were green and approximately 25 to 50mm long, they were transferred to Root Induction Medium in 98mm tall culture vessels (PhytoTech Labs, cat. no. C2100) in the same lighted growth chamber, removing as much of the callus as possible without damaging the plantlet. Leaf samples were taken from the plants for DNA extraction once they were close to touching the top of the culture vessel. Root Induction Medium is made as Callus Induction Medium, but with only half strength Murashige and Skoog Basal medium (2.215g) and no hormones.

6.4.6 *Confirming the Efficacy of the Biolistic Delivery System*

As a positive control for the bombardment process, immature embryos were bombarded with a Green Fluorescent Protein (GFP) construct. Ten µl of gold particles were coated with 1µg of GFP and delivered into immature embryos following the same

procedure as with biolistic bombardment of RNP complexes. After 24 hours the embryos were visualized on a fluorescent microscope to confirm GFP integration and expression.

6.4.7 Biolistic Bombardment Using RNPs

The GW2-B RNPs were delivered to the wheat embryos following Liang *et al.* (2017)[14] with some modifications from Svitashev *et al.* (2015)[23], using a Bio-Rad PDS-1000 HeTM gene gun (Hercules, CA, cat no. 1652257) and accessories. In short, 15mg of 0.6µm gold particles (Bio-Rad, cat. no. 1652262) were washed with 500µl of cold 100% EtOH. They were then vortexed for five minutes, followed by centrifugation for one minute at 5,000rpm, and then the supernatant was discarded. The gold particles were then washed three times with 500µl of RNase-free water and centrifuged for two minutes at 8,000rpm. The supernatant was removed with a pipette and discarded. Then 250µl of RNase-free water was added and the solution vortexed for five minutes. The RNPs were formed by combining 2µg Cas9 + 2µg sgRNA + 0.14µl RNase Inhibitor + Xµl Cas9 reaction buffer for a total volume of 10µl and incubating at 25°C for 10 minutes. Then 150µg of gold nanoparticles in RNase-free water were added to the RNPs and gently homogenized. This mixture was spread onto a microcarrier (Bio-Rad, cat. no. 1652335) and allowed to air dry for approximately 15 minutes. The immature embryos on Callus Induction Medium were then transformed at a distance of 6.0cm and a rupture pressure of 650psi (Bio-Rad, cat. no. 1652327) following the manufacturer's instructions.

6.4.8 Mutation Detection

Once the plantlets grew to reach the top of the culture vessels, approximately 25mm of leaf tissue total was sampled from multiple leaves on the same plantlet and disrupted with a tungsten bead in a Retsch Mixer Mill (Haan, Germany, cat. no. 20.746.0001). The DNA of the disrupted leaf tissue was extracted using the DNeasy Plant Mini Kit (cat. no. 69104) from Qiagen. PCR amplification was performed using the GW2-B-F and GW2-B-R primers and Phusion® High-Fidelity DNA Polymerase (cat. no. M0530) from New England Biolabs following the manufacturer's protocol. The PCR

was then cleaned with the QIAquick PCR Purification Kit (cat. no. 28104) from Qiagen and sent to Oregon State University's Center for Genome Research and Biocomputing for Sanger sequencing. Any transformed plantlets were transplanted from culture vessels into soil pots in the greenhouse at 70-75°C under 16 hours of light per day. An overview of the procedure is given in Figure 6.1.

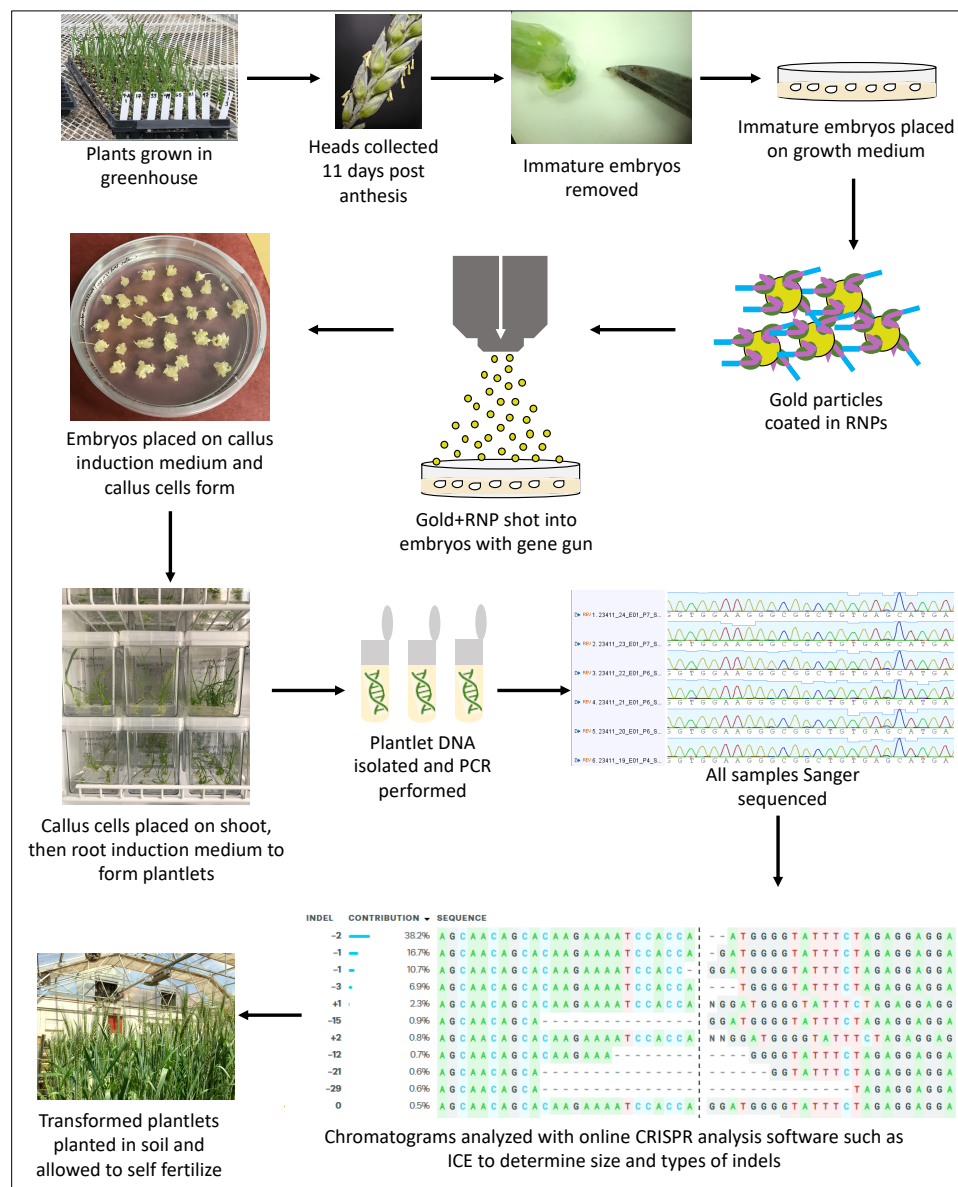


Figure 6.1 Overview of the CRISPR/Cas9 RNP transformation procedure. Plantlets are grown in greenhouse, heads are harvested 11 days post-anthesis, cleaned in bleach and ethanol solutions, then seeds are removed from the heads. Immature embryos are then removed with a scalpel and placed scutellum-side up on callus induction medium. 0.6µm

gold particles are mixed with Cas9 and sgRNA to form RNP complexes, which coat the gold particles. The gold is then placed on a microcarrier in a gene gun and delivered at 6.0cm and a rupture pressure of 650psi into the embryos, which are then allowed to grow in a dark growth chamber at 20-25°C for two to three weeks, placed on shoot induction medium, then root induction medium in culture vessels in a lighted growth chamber at 20-25°C with a 12 hour light/dark cycle. Leaves of each plantlet are sampled, DNA is isolated, and PCR is performed for the gene target. This PCR is then Sanger sequenced and analyzed using a CRISPR analysis software of choice. Any transformed plantlets are transplanted into soil pots in the greenhouse

6.5 Results

6.5.1 *Confirming the Efficacy of the Biolistic Delivery System*

The positive control assay for the biolistic delivery system showed bright green fluorescing cells within the bombarded embryo, meaning that the cells had incorporated and begun expressing the GFP construct (Figure 6.2). This indicated that the biolistic delivery of constructs was successful.

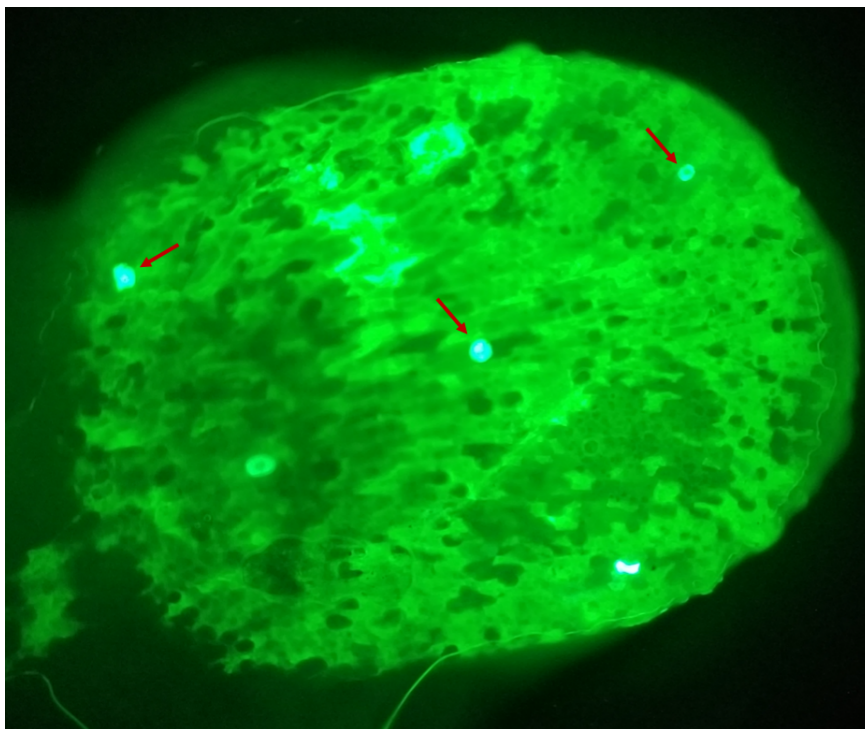


Figure 6.2 Fluorescent light and green filter image of an immature wheat embryo 24 hours after bombardment with GFP-coated gold particles. Bright green cells expressing the GFP construct are indicated with red arrows. Lighter colored cells represent non-transformed cells.

6.5.2 Plantlet Regeneration After Bombardment

The regeneration assays showed that a single plate of embryos (approximately 8 heads) regenerated upwards of 65 plantlets, occasionally above 100 plantlets. Callus induction and subsequent plantlet formation often led to the formation of tight clusters of sprouting plantlets that could not be easily separated into true individuals (Figure 6.3). Therefore, we continued to name these clusters ‘plantlets’, but excised a piece of leaf tissue from each tiller during the sampling process to ensure all potential transformants were sampled and analyzed. In total, 510 plantlets representing almost 60 heads of wheat were successfully recovered and sampled from culture vessels and Sanger sequenced to test for transformation events.

6.5.3 Sequencing of T_0 Plantlets in Culture Vessels

The evaluation of chromatograms produced by Sanger sequencing of plantlet PCR products did not give definitive indication of whether or not there were nucleotide insertions or deletions (indels), nor the type of indels that may have been present (Figure

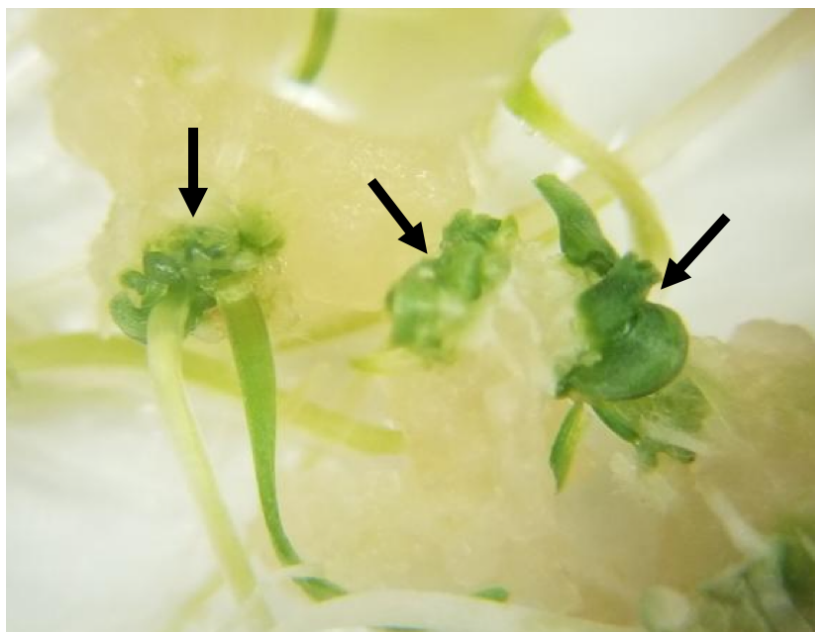


Figure 6.3 Callus tissue formed from RNP bombarded embryos showing tight clusters of leaves and leaf tissue (indicated by black arrows). The leaf tissue from each tiller was sampled and pooled into a single ‘plantlet’ DNA sample that was deconvoluted at a later date if transformation was detected.

6.4). Analysis with ICE deconvoluted those pooled samples and gave individual indel frequencies and types, reported as the relative contributions of inferred sequences present in each plantlet's PCR sample (Figure 6.5). The sequences in Figures 6.4 and 6.5 are a representative subset of the sequencing done for this project. In summary, 96.3% of the inferred sequences were untransformed in Plantlet 1. In Plantlet 2, 9.3% of the inferred sequences were untransformed. The other 90.7% of inferred sequences in Plantlet 2 were transformed, with 80.1% of the total inferred sequences consisting of a single deletion of the cysteine directly upstream of the expected cut site. In Plantlet 3, 34.6% of inferred sequences consisted of a two-nucleotide deletion, 22.3% consisted of a single nucleotide deletion, and 8.6% consisted of a different single nucleotide deletion. The average transformation efficiency for negative control plantlets that were taken through the tissue culture process but never bombarded with RNP complexes was 4.6%, with a standard deviation of 2.8%. Therefore, the 3.7% transformation efficiency from Plantlet 1 is within this error range and the plantlet is considered untransformed. Plantlet 2 and Plantlet 3 had transformation efficiencies of 80.1% and 8.6-34.6%, respectively, and are therefore both considered to be transformed. In total, there were 34 plantlets out of 510 that showed greater than 7.4% transformation efficiency, for a total of 6.7% plantlet transformation efficiency.

6.5.4 Sequencing of Individual Tillers from T_0 Plantlets Transplanted to Soil Pots

Since the plantlets regenerated from callus tissue had the potential to be tightly-clustered individuals and some showed indications of being pooled samples (Figure 6.5), any plantlet that showed transformation above the negative control limit of 7.4% was transplanted into gallon pots to grow in the greenhouse. Once large enough, the individual tillers and offshoots were tagged and leaf tissue from each was sampled independently in order to deconvolute any pooled results (Figure 6.6). Here, Plantlet 2 is used as the representative sample. In summary, of the inferred sequences from Tiller 1, 96.2% were Wt and the rest were below 1%. For Tiller 2, 99.4% were Wt, and the rest were below 1%. For Tiller 3, 95.1% were Wt and the rest were below 1%. For Tiller 4, 99.3% were Wt and the rest were below 1%. For Tiller 5, 90.1% were Wt, 2.3% were a

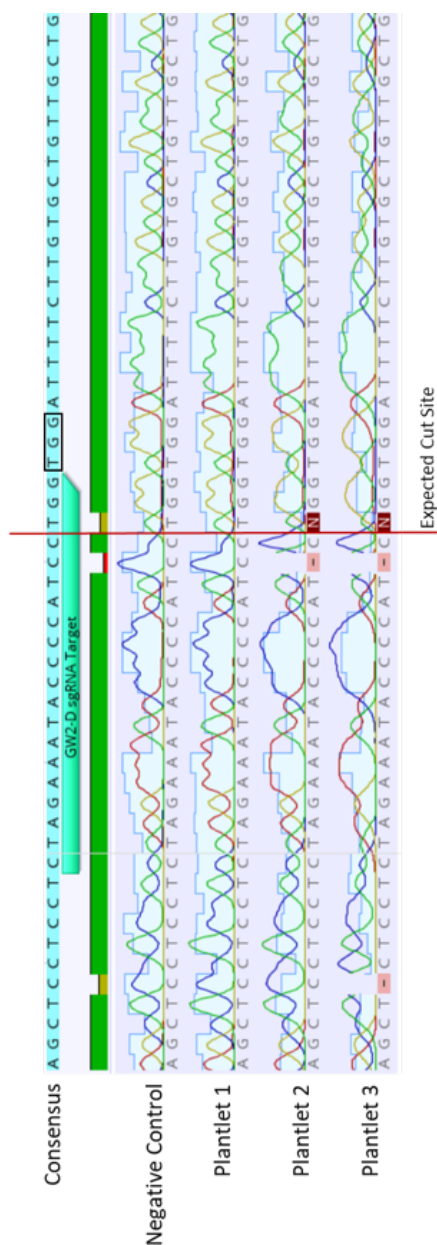


Figure 6.4 Chromatogram sequences from three plantlets grown from RNP bombarded embryos and a negative control plantlet. The negative control was propagated through tissue culture but never transformed with RNPs. The black box represents the PAM sequence recognized by Cas9, and the red vertical line indicates the expected cut site. Plantlet 1 does not differ from the negative control chromatogram. Plantlet 2 appears to have a deleted nucleotide near the cut site. Plantlet 3 also appears to have a deleted nucleotide near the cut site with the possibility of another deletion further upstream. Visualizing chromatograms is not sufficient to determine the indel frequency nor type of indel if the sample was taken from pooled DNA of individuals and/or chimeras since the chromatogram shows the most highly abundant nucleotide at each position, and therefore cannot accurately reflect multiple types of sequences.

Plantlet 1		
Indel	Contribution	Sequence
0	96.3%	CCTCCTCTAGAAATACCCCATCC TGGTGGATTTTCTTGCTGTTGCTG
+15	1.2%	CCTCCTCTAGAAATACCCCATCC NNNNNNNNNNNNNNNNNTGGTGGATTTT
+14	1.1%	CCTCCTCTAGAAATACCCCATCC NNNNNNNNNNNNNNNTGGTGGATTTTC

Plantlet 2		
Indel	Contribution	Sequence
-1	80.1%	CCTCCTCTAGAAATACCCCATC- TGGTGGATTTTCTTGCTGTTGCTG
0	9.3%	CCTCCTCTAGAAATACCCCATCC TGGTGGATTTTCTTGCTGTTGCTG
-2	5.3%	CCTCCTCTAGAAATACCCCAT-- TGGTGGATTTTCTTGCTGTTGCTG

Plantlet 3		
Indel	Contribution	Sequence
-2	34.6%	CCTCCTCTAGAAATACCCCAT-- TGGTGGATTTTCTTGCTGTTGCTG
-1	22.3%	CCTCCTCTAGAAATACCCCATC- TGGTGGATTTTCTTGCTGTTGCTG
-1	8.6%	CCTCCTCTAGAAATACCCCATCC- GGTGGATTTTCTTGCTGTTGCTG

Figure 6.5 Relative contributions of inferred sequences present in each plantlet PCR sample, as reported by ICE. Only the top three inferred sequences from each sample are reported here. ‘Indel’ is the type of insertion or deletion, with ‘0’ being a wild type (Wt) (or untransformed) sequence. ‘Contribution’ is the percentage of the total number of inferred sequences that are of that indel type. The dashed vertical line indicates the expected cut site of Cas9 for the *GW2-B* sgRNA used. N’s indicate unknown nucleotides. In Plantlet 1, 96.3% of the inferred sequences are Wt, with 1.2% and 1.1% being a 15 nucleotide insertion and a 14 nucleotide insertion, respectively. In Plantlet 2, 80.1% of inferred sequences show a single deletion at the expected cut site, with only 9.3% of inferred sequences being Wt. In Plantlet 3, 34.6% of inferred sequences show a two-nucleotide deletion, 22.3% show a single nucleotide deletion upstream of the expected cut site, and 8.6% show a single nucleotide deletion downstream of the expected cut site.



Figure 6.6 Plantlet 2 taken from a culture vessel in the growth chamber and planted in a one gallon pot in the greenhouse. The white tags were numbered and placed on individual tillers. Leaf tissue from each tiller was then sampled for DNA extraction and Sanger sequencing as in the culture vessels, and then analyzed for transformation efficiency.

three-nucleotide deletion, and 2.1% were a single nucleotide deletion. For Tiller 6, 90.2% were Wt, 3.6% were a single nucleotide deletion, and 1.4% were a three-nucleotide deletion (Figure 6.7).

6.6 Discussion and Conclusions

The results we obtained from sequencing the individual tillers of putatively transformed plantlets do not match the results from the culture vessel sequencing, as shown in the transformation efficiencies dropping from 80.1% in Plantlet 2's pooled sample to 3.6% at the highest in the individual tillers. Each of the culture vessel samples with greater than 7.4% transformation efficiency were significantly different from the

			Tiller 1	
Indel	Contribution	Sequence		
0	96.2%	CCTCCTCTAGAAATACCCCATCC	TGGTGGATTTTCTTGCTGTTGCTG	
-6	0.6%	CCTCCTCTAGAAATACCCCATCC	-----ATTTTCTTGCTGTTGCTG	
-20	0.6%	CCTCCTCTA	-----TTTTCTTGCTGTTGCTG	
			Tiller 2	
Indel	Contribution	Sequence		
0	99.4%	CCTCCTCTAGAAATACCCCATCC	TGGTGGATTTTCTTGCTGTTGCTG	
-3	0.2%	CCTCCTCTAGAAATACCCCATCC	---TGGATTTTCTTGCTGTTGCTG	
-1	0.1%	CCTCCTCTAGAAATACCCCATCC	-GGTGGATTTTCTTGCTGTTGCTG	
			Tiller 3	
Indel	Contribution	Sequence		
0	95.1%	CCTCCTCTAGAAATACCCCATCC	TGGTGGATTTTCTTGCTGTTGCTG	
-1	0.9%	CCTCCTCTAGAAATACCCCATC	-TGGTGGATTTTCTTGCTGTTGCTG	
-1	0.6%	CCTCCTCTAGAAATACCCCATCC	-GGTGGATTTTCTTGCTGTTGCTG	
			Tiller 4	
Indel	Contribution	Sequence		
0	99.3%	CCTCCTCTAGAAATACCCCATCC	TGGTGGATTTTCTTGCTGTTGCTG	
+7	0.2%	CCTCCTCTAGAAATACCCCATCC	NNNNNNNTGGTGGATTTTCTTGCTG	
-9	0.1%	CCTCCTCTAGAAATA	-----GGTGGATTTTCTTGCTGTTGCTG	
			Tiller 5	
Indel	Contribution	Sequence		
0	90.1%	CCTCCTCTAGAAATACCCCATCC	TGGTGGATTTTCTTGCTGTTGCTG	
-3	2.3%	CCTCCTCTAGAAATACCCCAT	---GGTGGATTTTCTTGCTGTTGCTG	
-1	2.1%	CCTCCTCTAGAAATACCCCATC	-TGGTGGATTTTCTTGCTGTTGCTG	
			Tiller 6	
Indel	Contribution	Sequence		
0	90.2%	CCTCCTCTAGAAATACCCCATCC	TGGTGGATTTTCTTGCTGTTGCTG	
-1	3.6%	CCTCCTCTAGAAATACCCCATCC	-GGTGGATTTTCTTGCTGTTGCTG	
-3	1.4%	CCTCCTCTAGAAATACCCCATCC	---TGGATTTTCTTGCTGTTGCTG	

Figure 6.7 Relative contributions of inferred sequences present in each Tiller's PCR sample from Plantlet 2, as reported by ICE. Only the top three inferred sequences are reported here. 'Indel' is the type of insertion or deletion, with '0' being the wild type (Wt) (or untransformed) sequence. 'Contribution' is the percentage of the total number of inferred sequences that are of that indel type. The dashed vertical line indicates the expected cut site of Cas9 for the *GW2-B* sgRNA used. N's indicate unknown nucleotides. In Tiller 1, 96.2% of the inferred sequences are Wt, with a six-nucleotide deletion and a 20 nucleotide deletion each occurring at 0.6%. In Tiller 2, 99.4% of inferred sequences are Wt, 0.2% are a three-nucleotide deletion, and 0.1% are a single nucleotide deletion. In Tiller 3, 95.1% of inferred sequences are Wt, with two different single nucleotide deletions each occurring at 0.9 and 0.6%. In Tiller 4, 99.3% of inferred sequences are Wt, 0.2% are a seven-nucleotide insertion, and 0.1% are a nine-nucleotide deletion. In Tiller 5, 90.1% of inferred sequences are Wt, 2.3% are a three-nucleotide deletion, and 2.1% are a single nucleotide deletion. In Tiller 6, 90.2% of inferred sequences are Wt, 3.6% are a single nucleotide deletion, and 1.4% are a three-nucleotide deletion.

untransformed negative control samples (p-value <0.001). However, none of the individual tillers from these plantlets sampled in the greenhouse pots showed a transformation efficiency above 7.4%. We can rule out that the inability to detect transformation in individual tillers is caused by the sgRNA being nonfunctional due to the successful *in vitro* validation tests performed on *GW2* PCR products, as well as the *in vivo* protoplast transformations. Both showed positive results consistent with the literature using the same *GW2* sgRNA construct[14][15][19][20]. We can also rule out a problem with the biolistic delivery system since the GFP construct expressed in bombarded embryo cells.

Therefore, the most likely explanation for this drop in transformation efficiency is that we are creating chimeric plants due largely to the lack of a selection method that helps to eliminate non-modified plantlets during the tissue culture process. Chimeras can either be formed through the production of chimeric clusters where it is impossible to physically separate one plant from the other, or through chimeric individuals, where some cells are transformed, but others are not[24]. Chimeric plants have been reported in other transformation protocols that utilize tissue culture and is likely what we are seeing here since sampling individual tillers did not resolve the issue[25][26]. Usually, the addition of a selectable marker gene such as herbicide resistance during transformation assists in reducing the number of chimeras produced by eliminating the cells that have not been transformed[27]. However, this does not always succeed in killing chimeric plants and chimeric tissue, as evidenced in Zhang *et al.* (2019)[28]. They transformed wheat embryos with CRISPR/Cas9 using *Agrobacterium* and included a selectable marker, but still had chimeric transformed individuals. At the T₀ stage, equivalent to our individual tiller sampling stage, they were unable to detect any mutations using T7EI digestions or Sanger sequencing. Detection of mutations required deep amplicon sequencing and showed efficiencies of only 3.3 to 5.2%. In the next generation, they were able to use Sanger sequencing to detect mutations in 25 out of 226 T₁ plants that had successful transformation in gametic cells in the T₀ generation. This study shows that even with selection during tissue culture, the occurrence of chimeric individuals is highly likely. Therefore, we hypothesize that sampling of culture vessel plantlet leaves was done on

young and therefore small plants, leading to a positive transformation result. When the plantlets were allowed to grow in the greenhouse in gallon pots, however, the proportion of transformed to untransformed cells decreased significantly and we were no longer able to detect transformation, similar to Zhang *et al.* (2019).

Up through the culture vessel sampling stage (Figure 6.5) we obtained similar results to those of Liang *et al.* In Liang *et al.* 2017, they describe transformation efficiencies of 4.4% in T₀ plantlet DNA[14]. Their second report, Liang *et al.* 2018, is a protocol extension in which they obtained similar results using similar procedures to the first report[15]. We obtained a comparable overall transformation efficiency at the T₀ culture vessel stage of 6.7%, although we did not see evidence of only homozygous or heterozygous mutations as Liang *et al.* reported. They also did not include any indication of obtaining chimeric plants after tissue culture nor did they report sampling individual tillers after subsequent growth.

In order to further address these discrepancies and evaluate the hypothesis that we have chimeric individuals, we need to deep amplicon sequence the DNA. Deep amplicon sequencing involves the targeted amplification of a region of interest using next generation sequencing techniques and can be used to quantify the amount of transformed versus untransformed tissue based on the number of reads with and without evidence of CRISPR/Cas9 guided mutations[28][29]. Deep amplicon sequencing should be performed on the transformants obtained from the culture vessel samples that showed positive transformation, the individual tillers from these putative transformants, and the negative control samples that went through tissue culture but were never bombarded with RNP complexes. This will allow us to see if the transformations reported by ICE were accurate, if the culture vessel samples truly were pools of mutated and un-mutated DNA, and if the individual tillers were chimeric as well. We will also need to harvest seed from the self-fertilized T₀ plantlets, germinate them, and sample them for Sanger sequencing analysis as in Zhang *et al.* (2019) to attempt to recover fully mutated T₁ plants. While the T₁ plants may not show any mutations if the transformed cells in the T₀ plants were never in the gametic tissue, the results should be a clear zero percent (or comparable to the negative controls) if this is the case. Both of these tests are in the process of being

completed.

Our results also indicate that the GFP constructs used as a control resulted in transformed cells. The GFP that was shot into the immature wheat embryos in Figure 6.4 is a stable construct, and we will be following these embryos throughout the tissue culture process. This will allow us to visualize the number of cells that are initially transformed, how many of them proliferate during callus induction, how many of the transformed cells become plantlets, how common chimeric clusters and individual plantlets are, and whether the proportion of transformed to untransformed cells decreases after growth in greenhouse gallon pots.

Overall, the primary aim of this work to test the feasibility of transforming immature wheat embryos using RNP complexes was successful, and is therefore valuable for any future work on the non-transgenic transformation of wheat with CRISPR. The results presented here suggest that despite the fact that transformation likely occurred, this protocol cannot feasibly be implemented in a public breeding program. Without the ability to include a selectable marker during tissue culture, there are too many plantlets to test making the costs for identifying modified plants prohibitive. If the problem we are experiencing is in fact chimeric plants, then every plantlet generated through tissue culture should have its gene target deep sequenced after initial detection in culture vessels, and that is financially prohibitive for a breeding program. The alternative option is to plant, grow, and sample every T_0 seed from putative transformants. If the average number of tillers per plantlet is five, and the average number of seeds produced per tiller is 10, this would result in the need to sample and sequence approximately 50 plantlets per potential transformant. In this study the number would be around 1,700 samples that would need to be Sanger sequenced and analyzed. Therefore, this protocol being used to transform wheat using CRISPR with the aim of integrating it as a supplement to other breeding methods does not appear to be economically feasible.

While this particular protocol is not practical, there are other potential options for non-transgenic transformation of wheat with CRISPR for use in a breeding program. The main finding of this experiment is the need to avoid the callus stage of tissue culture due to the high rate of chimeras. Therefore, we propose three future project directions: i)

Electroporation of wheat pollen grains with RNP complexes, ii) Electroporation of wheat microspores with RNP complexes, and iii) Bombarding Shoot Apical Meristems of immature wheat embryos with RNP complexes. Option one (i) is based off work done by Liu (2016) and Kumar (2004) where pollen grains were harvested and transformed via electroporation with DNA, then used to pollenate emasculated wheat heads[30][31]. This method would produce heterozygous mutant plants that could then be self-fertilized to generate homozygous mutant offspring. Option two (ii) is based off work by Bhowmik *et al.* (2018) in which wheat microspores were electroporated with Cas9:sgRNA expressing plasmids and then induced to form embryos[32]. The regeneration of wheat plantlets from microspores is a difficult process and would take some optimization, but this method utilizes direct somatic embryogenesis and does not require a callus stage, thus potentially bypassing many of the problems caused by chimeras. This method was done using transgenic techniques and has not yet been performed using RNPs[33]. Option three (iii) is based off work by Hamada *et al.* (2018) where the shoot apical meristems of imbibed seeds were biolistically transformed with plasmids expressing Cas9 and sgRNA[13]. The transformation efficiency was 5.2% in the T₀ generation and 1.4% in the T₁ generation, this technique does not require callus induction, and the authors state that it should also work with RNP complexes rather than transgenic-inducing plasmids, but this has not yet been done.

While there are still potential options for a CRISPR transformation method in wheat without the incorporation of transgenes, this report shows that the biolistic delivery of RNP complexes to wheat immature embryos is not a feasible approach in its current state. We demonstrated that while this method can result in plants with modifications at the gene target site, the challenges and costs associated with identifying fully modified plants prevent this approach being practical for use in a public wheat breeding program. This research shows that it is possible to modify wheat genes using transient transformation, but the key for success is the need to reduce the potential for creating chimeric plants by developing protocols that result in transformation and regeneration of modified plants from single cells. If plants can be developed from single transformed cells such as pollen or microspores rather than chimeric callus tissue, CRISPR could be a

valuable tool to rapidly modify genes in wheat.

6.7 Acknowledgements

This work was supported by the Oregon Agricultural Experiment Station, Warren E. Kronstad Wheat Research Endowment, the Oregon Wheat Commission, and the ARCS Oregon Chapter Scholar Award.

6.8 References

- [1] Q. Shan, Y. Wang, J. Li, and C. Gao, “Genome editing in rice and wheat using the CRISPR/Cas system,” *Nature Protocols*, vol. 9, no. 10, pp. 2395–2410, Sep. 2014.
- [2] S. Kim, D. Kim, S. W. Cho, J. Kim, and J.-S. Kim, “Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins,” *Genome Res.*, vol. 24, no. 6, pp. 1012–1019, Jun. 2014.
- [3] L. Arora and A. Narula, “Gene Editing and Crop Improvement Using CRISPR-Cas9 System,” *Front. Plant Sci.*, vol. 8, no. 1932, Nov. 2017.
- [4] A. Malzahn, L. Lowder, and Y. Qi, “Plant genome editing with TALEN and CRISPR,” *Cell & Bioscience*, vol. 7, no. 1, p. 21, Apr. 2017.
- [5] M. Malnoy *et al.*, “DNA-Free Genetically Edited Grapevine and Apple Protoplast Using CRISPR/Cas9 Ribonucleoproteins,” *Front. Plant Sci.*, vol. 7, no. 1904, pp. 1–9, Dec. 2016.
- [6] T.-Q. Shi *et al.*, “CRISPR/Cas9-based genome editing of the filamentous fungi: the state of the art,” *Appl Microbiol Biotechnol*, vol. 101, no. 20, pp. 7435–7443, Oct. 2017.
- [7] W. Jiang, H. Zhou, H. Bi, M. Fromm, B. Yang, and D. P. Weeks, “Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice,” *Nucleic Acids Res*, vol. 41, no. 20, pp. e188–e188, Nov. 2013.
- [8] A. Peng *et al.*, “Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus,” *Plant Biotechnology Journal*, vol. 15, no. 12, pp. 1509–1519, May 2017.
- [9] J. Gil-Humanes *et al.*, “High-efficiency gene targeting in hexaploidy wheat using DNA replicons and CRISPR/Cas9,” *The Plant Journal*, vol. 89, no. 6, pp. 1251–1262, Mar. 2017.
- [10] L. Bortesi and R. Fischer, “The CRISPR/Cas9 system for plant genome editing and beyond,” *Biotechnology Advances*, vol. 33, no. 1, pp. 41–52, Jan. 2015.
- [11] D. Jaganathan, K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman, “CRISPR for Crop Improvement: An Update Review,” *Front Plant Sci.*, vol. 9, no. 985, Jul. 2018.
- [12] M. Andersson *et al.*, “Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery,” *Physiologia Plantarum*, vol. 164, no. 4, pp. 378–384, Dec. 2018.
- [13] H. Hamada, Y. Liu, Y. Nagira, R. Miki, N. Taoka, and R. Imai, “Biolistic-delivery-based transient CRISPR/Cas9 expression enables in planta genome editing in wheat,” *Scientific Reports*, vol. 8, no. 1, p. 14422, Sep. 2018.
- [14] Z. Liang *et al.*, “Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes,” *Nat Commun*, vol. 8, no. 14261, pp. 1–5, Jan. 2017.
- [15] Z. Liang *et al.*, “Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 *in vitro* transcripts or ribonucleoproteins,” *Nature Protocols*, vol. 13, no. 3, pp. 413–430, Mar. 2018.

- [16] T. Langner, S. Kamoun, and K. Belhaj, "CRISPR Crops: Plant Genome Editing Toward Disease Resistance," *Annual Review of Phytopathology*, vol. 56, no. 1, pp. 479–512, Jul. 2018.
- [17] P. M. Rogowsky, "CRISPR-Cas Technology in Plant Science," *Potato Res.*, vol. 60, no. 3, pp. 353–360, Dec. 2017.
- [18] J. Parsons, "The theory and applications of CRISPR in plant and tree improvement.," *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, vol. 13, no. 007, Jun. 2018.
- [19] W. Wang *et al.*, "Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat," *Theor Appl Genet*, vol. 131, no. 11, pp. 2463–2475, Nov. 2018.
- [20] W. Wang *et al.*, "Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat," *The CRISPR Journal*, vol. 1, no. 1, pp. 65–74, Feb. 2018.
- [21] T. Hsiao *et al.*, "Inference of CRISPR Edits from Sanger Trace Data," *bioRxiv* doi: 10.1101/251082, Jan. 2019.
- [22] J. Hansen, P.J. Shiel, P. McCarthy, P.H. Berger, and R.S. Zemetra, "Transformation of soft white winter wheat (*Triticum aestivum*) for virus resistance," in *Proc. 9th Intern. Wheat Genetics Symp.*, Saskatoon Saskatchewan Canada, vol. 3, Aug. 1998, pp. 186-188.
- [23] S. Svitashhev, J. K. Young, C. Schwartz, H. Gao, S. C. Falco, and A. M. Cigan, "Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA," *Plant Physiology*, vol. 169, no. 2, pp. 931–945, Oct. 2015.
- [24] M. R. Janga, L. M. Campbell, and K. S. Rathore, "CRISPR/Cas9-mediated targeted mutagenesis in upland cotton (*Gossypium hirsutum* L.)," *Plant Mol Biol*, vol. 94, no. 4, pp. 349–360, Jul. 2017.
- [25] M. H. Frank and D. H. Chitwood, "Plant chimeras: The good, the bad, and the 'Bizzaria,'" *Developmental Biology*, vol. 419, no. 1, pp. 41–53, Nov. 2016.
- [26] G. Q. Chen, "Effective Reduction of Chimeric Tissue in Transgenics for the Stable Genetic Transformation of *Lesquerella fendleri*," *HortScience*, vol. 46, no. 1, pp. 86–90, Jan. 2011.
- [27] M. C. Jordan, "Green fluorescent protein as a visual marker for wheat transformation," *Plant Cell Reports*, vol. 19, no. 11, pp. 1069–1075, Nov. 2000.
- [28] Z. Zhang *et al.*, "Development of an Agrobacterium-delivered CRISPR/Cas9 system for wheat genome editing," *Plant Biotechnology Journal*, vol. doi: 10.1111/pbi.13088, Feb. 2019.
- [29] N. J. Hathaway, C. M. Parobek, J. J. Juliano, and J. A. Bailey, "SeekDeep: single-base resolution *de novo* clustering for deep amplicon sequencing," *Nucleic Acids Research*, vol. 46, no. 4, pp. e21, Nov. 2017.
- [30] J. Liu, "Identifying the Parameters for Wheat Pollen Electroporation Transformation," Oregon State University, Corvallis, Oregon, 2016.
- [31] M. Kumar, "Genetic Modification of Wheat (*Triticum aestivum* L.) Straw for Potential Use as a Biofuel," University of Idaho, Moscow, Idaho, 2004.

- [32] P. Bhowmik *et al.*, “Targeted mutagenesis in wheat microspores using CRISPR/Cas9,” *Sci Rep*, vol. 8, pp. 1–10, Apr. 2018.
- [33] F. Eudes, S. Acharya, A. Laroche, and L. B. Selinger, “A novel method to induce direct somatic embryogenesis, secondary embryogenesis and regeneration of fertile green cereal plants,” *Plant Cell, Tissue and Organ Culture*, vol. 73, pp. 147–157, May 2003.

Chapter 7

Conclusions

Wheat has been cultivated for many thousands of years and is grown across the globe in a vast range of environments. It is a vitally important crop for human consumption and for the global economy[1][2]. Increasing population size, decreasing arable land area, and growing concern for environmental impacts have pushed breeders to develop varieties that yield higher with fewer inputs[3]. Modern wheat breeding is now faced with the additional task of adapting wheat to the many and varied impacts of climate change. This means that breeders in each region must develop regionally adapted varieties with a multitude of important traits, some of which are extremely difficult or impossible to integrate using traditional breeding methods[4][5]. This dissertation addresses two different techniques that are expected to become integral to modern wheat breeding.

Next generation sequencing (NGS) techniques were used in combination with a Genome Wide Association Study (GWAS) to discover novel Quantitative Trait Loci (QTL) and markers for stripe rust resistance in a winter wheat Recombinant Inbred Line (RIL) population produced from a cross of the varieties ‘Skiles’ and ‘Goetze’. This study found significant QTL on chromosomes 3B and 3D that when deployed together show extremely high levels of stripe rust resistance, even above that of the resistant parent. All markers found are novel, and the markers on chromosome 3D are located almost 30Mb away from the closest known marker[6]. The markers on chromosome 3B will aid in efforts to determine which resistance gene is present on this chromosome, and the markers on 3D will be extremely useful in marker assisted selection for integration of this gene into new varieties. Together, the markers discovered in this study can be used to immediately breed new varieties with extremely high levels of stripe rust resistance.

The same techniques of NGS and GWAS were used to discover novel QTL and markers for reduced vernalization requirements (also known as facultative growth habit) in the Skiles x Goetze RIL population. This study found four QTL on four different chromosomes that contribute to flowering time under various vernalization regimes. The QTL on chromosome 5B represents an allele known to contribute to facultative ability that has yet to be characterized in local germplasm[7]. The QTL on chromosome 5D is also likely a facultative allele that was previously unknown to be contributing to the

reduced vernalization requirements of the parental variety Goetze[8]. It can now be used for breeding this trait into new varieties. The QTL on chromosomes 1D and 3B are likely associated with photoperiod genes previously unknown to be contributing to facultative growth habit[9][10]. In order to pinpoint the responsible genes and gain more detailed information on how these genes are contributing to flowering time, the RIL population should be phenotyped for heading date after various vernalization and photoperiod regimes. Altering the photoperiod as well as vernalization time should increase the sensitivity of the GWAS analysis and aid in determining when these genes are activated and how much they are contributing to facultative growth habits.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a new technology that is being tested in many crop breeding programs and shows great promise, but has yet to be used in any released variety[11]. Due to the high expectations of its future contributions to plant breeding, the feasibility of its use in a wheat breeding program was tested. The first obstacle that was encountered was target sgRNA for a specific gene being functional *in vitro* but not *in vivo*, possibly due to an inaccessible chromatin state at the desired target site. To address this issue, a protoplast isolation and transformation protocol was developed for testing CRISPR gene targets *in vivo* before beginning full transformations. The protocol represents a simplified and more easily scalable version of the currently published isolation and transformation protocols. It consistently yields high numbers of protoplasts, does not require any specialized equipment, and returns clear positive and negative transformation results for sgRNA targets that show positive results *in vitro* but are inaccessible *in vivo*.

There is currently only one published protocol for transforming wheat with a completely non-transgenic CRISPR system[12][13]. Therefore, this protocol was tested for its feasibility of use in a wheat breeding program. After regenerating biolistically transformed plants using tissue culture, the average rate of transformation was 6.7 percent, which was comparable to that of the published protocol. Upon further testing, however, it became clear that the lack of a selective medium during tissue culture was resulting in chimeric plants. These chimeras lead to adult plantlets with no detectable transformation, as the ratio of transformed to untransformed cells had decreased

substantially during growth. The only way to obtain fully transformed plants would be to take every seed from the T_0 plants, grow them, then sample every plantlet in the T_1 generation to test for transformation. This would reduce the transformation efficiency down to less than one percent[14]. Therefore, the conclusion was reached that due to the vast amounts of time, resources, space, and capital needed to recover transformed plants using this method it is not considered a viable option for a breeding program. Three other options are proposed as potential methods to implement non-transgenic CRISPR transformation and should be tested further: first, electroporation of wheat pollen grains; second, electroporation of wheat microspores; and third, bombardment of shoot apical meristems of immature wheat embryos[15]-[18].

In summary, NGS and GWAS are extremely useful tools that can be employed to identify sources of important phenotypic variation and associated markers in adapted germplasm. The QTL and markers found in this study will aid in the identification of important genes and alleles in local varieties as well as immediately facilitate the incorporation of important disease resistance and facultative growth habit traits into new varieties. CRISPR transformation is expected to become an extremely important part of crop breeding, and wheat breeding would greatly benefit from the theoretical advancements it could supply. The only currently published study, however, is not a viable option. More testing is needed to find a method suitable for integration into wheat breeding programs that does not require tissue culture and the screening of thousands of plantlets. If this can be achieved, CRISPR will undoubtedly contribute a great deal to the rapid development of future wheat varieties.

7.1 References

- [1] Z. Peleg, T. Fahima, A. B. Korol, S. Abbo, and Y. Saranga, “Genetic analysis of wheat domestication and evolution under domestication,” *J Exp Bot*, vol. 62, no. 14, pp. 5051–5061, Oct. 2011.
- [2] P. Gustafson, O. Raskina, X. Ma, and E. Nevo, “Wheat evolution, domestication, and improvement,” in *Wheat Science and Trade*, Wiley-Blackwell, 2009, pp. 3–30.
- [3] Rosenzweig, “Climate Change and Extreme Weather Events; Implications for Food Production, Plant Diseases, and Pests,” *GLOBAL CHANGE*, vol. 2, no. 2, p. 15, 2001.
- [4] S. Mondal *et al.*, “Harnessing Diversity in Wheat to Enhance Grain Yield, Climate Resilience, Disease and Insect Pest Resistance and Nutrition Through Conventional and Modern Breeding Approaches,” *Front Plant Sci*, vol. 7, Jul. 2016.
- [5] W. Wang *et al.*, “Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat,” *The CRISPR Journal*, vol. 1, no. 1, pp. 65–74, Feb. 2018.
- [6] USDA-ARS, Wheat, Yr genes and QTL. In *GrainGenes*_. Available: <https://wheat.pw.usda.gov/cgi-bin/GG3/browse.cgi?class=map>. [Accessed: 15-Apr-2019].
- [7] D. K. Santra, M. Santra, R. E. Allan, K. G. Campbell, and K. K. Kidwell, “Genetic and Molecular Characterization of Vernalization Genes *Vrn-A1*, *Vrn-B1*, and *Vrn-D1* in Spring Wheat Germplasm from the Pacific Northwest Region of the U.S.A.,” *Plant Breeding*, vol. 128, no. 6, pp. 576–584, Dec. 2009.
- [8] A. Loukoianov, L. Yan, A. Blechl, A. Sanchez, and J. Dubcovsky, “Regulation of VRN-1 Vernalization Genes in Normal and Transgenic Polyploid Wheat,” *Plant Physiology*, vol. 138, no. 4, pp. 2364–2373, Aug. 2005.
- [9] J. Halliwell *et al.*, “Systematic Investigation of FLOWERING LOCUS T-Like Poaceae Gene Families Identifies the Short-Day Expressed Flowering Pathway Gene, TaFT3 in Wheat (*Triticum aestivum* L.),” *Front. Plant Sci.*, vol. 7, pp. 1–15, Jun. 2016.
- [10] X. Y. Zhao, M. S. Liu, J. R. Li, C. M. Guan, and X. S. Zhang, “The wheat TaGI1, involved in photoperiodic flowering, encodes an Arabidopsis GI ortholog,” *Plant Molecular Biology*, vol. 58, no. 1, pp. 53–64, May 2005.
- [11] D. Jaganathan, K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman, “CRISPR for Crop Improvement: An Update Review,” *Front Plant Sci*, vol. 9, no. 985, Jul. 2018.
- [12] Z. Liang *et al.*, “Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes,” *Nat Commun*, vol. 8, no. 14261, pp. 1–5, Jan. 2017.
- [13] Z. Liang *et al.*, “Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 *in vitro* transcripts or ribonucleoproteins,” *Nature Protocols*, vol. 13, no. 3, pp. 413–430, Mar. 2018.
- [14] Z. Zhang *et al.*, “Development of an Agrobacterium-delivered CRISPR/Cas9 system for wheat genome editing,” *Plant Biotechnology Journal*, vol. doi: 10.1111/pbi.13088, Feb. 2019.

- [15] J. Liu, “Identifying the Parameters for Wheat Pollen Electroporation Transformation,” Oregon State University, Corvallis, Oregon, 2016.
- [16] M. Kumar, “Genetic Modification of Wheat (*Triticum aestivum* L.) Straw for Potential Use as a Biofuel,” University of Idaho, Moscow, Idaho, 2004.
- [17] P. Bhowmik *et al.*, “Targeted mutagenesis in wheat microspores using CRISPR/Cas9,” *Sci Rep*, vol. 8, pp. 1–10, Apr. 2018.
- [18] H. Hamada, Y. Liu, Y. Nagira, R. Miki, N. Taoka, and R. Imai, “Biolistic-delivery-based transient CRISPR/Cas9 expression enables in planta genome editing in wheat,” *Scientific Reports*, vol. 8, no. 1, p. 14422, Sep. 2018.

Appendices

Appendix A

QTL Analysis of Adult Plant Resistance to Stripe Rust in a Winter Wheat Recombinant Inbred Population

Table A.1 ANOVA of disease severity as a function of Treatment, Taxa, and Treatment by Taxa Interaction. The p-values indicate that all three have a significant effect on the Disease Severity.

ANOVA					
	DF	Sum Sq	Mean Sq	F Value	p-value
Treatment	4	201799	50450	865.790	<2e-16***
Taxa	197	977808	4963	85.181	<2e-16***
Treatment*Taxa	788	318178	404	6.929	<2e-16***
Residuals	198	11538	58		

Table A.2 Tukey's HSD (Honest Significant Difference) Test of all Treatments compared to one another. The 'p adjusted' value shows that all treatments are significantly different from one another, except for Corvallis 2018 (Cor18) and Mount Vernon 2018 (MV18). Therefore, disease severity scores for Corvallis 2018 will be combined with scores from Mount Vernon 2018 for GWAS analysis. The others will be kept separate.

Tukey's HSD				
Treatment	Diff	Lower	Upper	p adjusted
MV18 – MV17	-4.6616162	-6.773765	-2.549468	0.0000001***
Pull17 – MV17	-18.8383838	-20.950532	-16.726235	0.0000000***
Pull18 – MV17	-38.0050505	-40.117199	-35.892902	0.0000000***
Cor18 – MV17	-4.9873737	-6.816548	-3.158200	0.0000000***
Pull17 – MV18	-14.1767677	-16.288916	-12.064619	0.0000000***
Pull18 – MV18	-33.3434343	-35.455583	-31.231286	0.0000000***
Cor18 – MV18	-0.3257576	-2.154932	1.503417	0.9881925
Pull18 – Pull17	-19.1666667	-21.278815	-17.054518	0.0000000***
Cor18 – Pull17	13.8510101	12.021836	15.680184	0.0000000***
Cor18 – Pull18	33.0176768	31.188503	34.846851	0.0000000***

Table A.3 Linkage Disequilibrium (LD) analysis of all significant SNPs. An R^2 of 1.0 indicates the two SNPs are completely correlated with one another in this population. An R^2 of 0.0 indicates that the two SNPs are completely unlinked from one another in this population.

Locus1	Position1	States1	Locus2	Position2	States2	R^2
3B	6309968	G:A	3B	6309966	C:G	1.00
3B	6309973	C:G	3B	6309966	C:G	1.00
3B	6309973	C:G	3B	6309968	G:A	1.00
3D	909572	G:T	3B	5601689	C:A	1.00
3D	4068757	C:G	3B	6309966	C:G	1.00
3D	4068757	C:G	3B	6309968	G:A	1.00
3D	4068757	C:G	3B	6309973	C:G	1.00
3D	4068759	G:A	3B	6309966	C:G	1.00

(Continued)

3D	4068759	G:A	3B	6309968	G:A	1.00
3D	4068759	G:A	3B	6309973	C:G	1.00
3D	4068759	G:A	3D	4068757	C:G	1.00
3D	4068764	C:G	3B	6309966	C:G	1.00
3D	4068764	C:G	3B	6309968	G:A	1.00
3D	4068764	C:G	3B	6309973	C:G	1.00
3D	4068764	C:G	3D	4068757	C:G	1.00
3D	4068764	C:G	3D	4068759	G:A	1.00
UN	34103779	G:A	3B	5601689	C:A	1.00
UN	34103779	G:A	3D	909572	G:T	1.00
UN	234960006	C:T	3B	5601689	C:A	1.00
UN	234960006	C:T	3D	909572	G:T	1.00
UN	234960006	C:T	UN	34103779	G:A	1.00
UN	242439365	G:C	3B	6309966	C:G	1.00
UN	242439365	G:C	3B	6309968	G:A	1.00
UN	242439365	G:C	3B	6309973	C:G	1.00
UN	242439365	G:C	3D	4068757	C:G	1.00
UN	242439365	G:C	3D	4068759	G:A	1.00
UN	242439365	G:C	3D	4068764	C:G	1.00
UN	242439370	C:T	3B	6309966	C:G	1.00
UN	242439370	C:T	3B	6309968	G:A	1.00
UN	242439370	C:T	3B	6309973	C:G	1.00
UN	242439370	C:T	3D	4068757	C:G	1.00
UN	242439370	C:T	3D	4068759	G:A	1.00
UN	242439370	C:T	3D	4068764	C:G	1.00
UN	242439370	C:T	UN	242439365	G:C	1.00
UN	242439372	G:C	3B	6309966	C:G	1.00
UN	242439372	G:C	3B	6309968	G:A	1.00
UN	242439372	G:C	3B	6309973	C:G	1.00
UN	242439372	G:C	3D	4068757	C:G	1.00
UN	242439372	G:C	3D	4068759	G:A	1.00
UN	242439372	G:C	3D	4068764	C:G	1.00
UN	242439372	G:C	UN	242439365	G:C	1.00
UN	242439372	G:C	UN	242439370	C:T	1.00
UN	242452400	G:C	3B	6309966	C:G	1.00
UN	242452400	G:C	3B	6309968	G:A	1.00
UN	242452400	G:C	3B	6309973	C:G	1.00
UN	242452400	G:C	3D	4068757	C:G	1.00
UN	242452400	G:C	3D	4068759	G:A	1.00
UN	242452400	G:C	3D	4068764	C:G	1.00
UN	242452400	G:C	UN	242439365	G:C	1.00
UN	242452400	G:C	UN	242439370	C:T	1.00
UN	242452400	G:C	UN	242439372	G:C	1.00
UN	242452405	C:T	3B	6309966	C:G	1.00
UN	242452405	C:T	3B	6309968	G:A	1.00
UN	242452405	C:T	3B	6309973	C:G	1.00
UN	242452405	C:T	3D	4068757	C:G	1.00
UN	242452405	C:T	3D	4068759	G:A	1.00
UN	242452405	C:T	3D	4068764	C:G	1.00
UN	242452405	C:T	UN	242439365	G:C	1.00
UN	242452405	C:T	UN	242439370	C:T	1.00

(Continued)

UN	242452405	C:T	UN	242439372	G:C	1.00
UN	242452405	C:T	UN	242452400	G:C	1.00
UN	242452407	G:C	3B	6309966	C:G	1.00
UN	242452407	G:C	3B	6309968	G:A	1.00
UN	242452407	G:C	3B	6309973	C:G	1.00
UN	242452407	G:C	3D	4068757	C:G	1.00
UN	242452407	G:C	3D	4068759	G:A	1.00
UN	242452407	G:C	3D	4068764	C:G	1.00
UN	242452407	G:C	UN	242439365	G:C	1.00
UN	242452407	G:C	UN	242439370	C:T	1.00
UN	242452407	G:C	UN	242439372	G:C	1.00
UN	242452407	G:C	UN	242452400	G:C	1.00
UN	242452407	G:C	UN	242452405	C:T	1.00
UN	36153637	C:T	3B	6309968	G:A	0.96
UN	242439372	G:C	UN	36153637	C:T	0.96
UN	242452400	G:C	UN	36153637	C:T	0.96
UN	242452407	G:C	UN	36153637	C:T	0.96
UN	242439365	G:C	UN	36153637	C:T	0.96
UN	242439370	C:T	UN	36153637	C:T	0.96
UN	242452405	C:T	UN	36153637	C:T	0.96
UN	36153637	C:T	3B	6309966	C:G	0.96
UN	36153637	C:T	3D	4068759	G:A	0.96
UN	36153637	C:T	3B	6309973	C:G	0.96
UN	36153637	C:T	3D	4068764	C:G	0.95
UN	36153637	C:T	3D	4068757	C:G	0.95
3B	6309968	G:A	3B	5601689	C:A	0.91
UN	242439372	G:C	3B	5601689	C:A	0.91
UN	242452400	G:C	3B	5601689	C:A	0.91
UN	242452407	G:C	3B	5601689	C:A	0.91
UN	242439370	C:T	3B	5601689	C:A	0.91
UN	242452405	C:T	3B	5601689	C:A	0.91
UN	242439365	G:C	3B	5601689	C:A	0.91
3B	6309966	C:G	3B	5601689	C:A	0.91
3D	4068759	G:A	3B	5601689	C:A	0.91
3B	6309973	C:G	3B	5601689	C:A	0.91
3D	909572	G:T	3B	6309968	G:A	0.91
UN	242439372	G:C	3D	909572	G:T	0.91
UN	242452400	G:C	3D	909572	G:T	0.91
UN	242452407	G:C	3D	909572	G:T	0.91
UN	242439370	C:T	3D	909572	G:T	0.91
UN	242452405	C:T	3D	909572	G:T	0.91
3D	4068764	C:G	3B	5601689	C:A	0.91
UN	242439365	G:C	3D	909572	G:T	0.91
3D	4068757	C:G	3B	5601689	C:A	0.91
3D	909572	G:T	3B	6309966	C:G	0.91
3D	4068759	G:A	3D	909572	G:T	0.91
3D	909572	G:T	3B	6309973	C:G	0.91
3D	4068757	C:G	3D	909572	G:T	0.91
3D	4068764	C:G	3D	909572	G:T	0.91
UN	234960006	C:T	3B	6309968	G:A	0.89
UN	242439372	G:C	UN	234960006	C:T	0.89

(Continued)

UN	242452400	G:C	UN	234960006	C:T	0.89
UN	242452407	G:C	UN	234960006	C:T	0.89
UN	242439370	C:T	UN	234960006	C:T	0.89
UN	242452405	C:T	UN	234960006	C:T	0.89
UN	242439365	G:C	UN	234960006	C:T	0.89
UN	234960006	C:T	3B	6309966	C:G	0.89
UN	234960006	C:T	3D	4068759	G:A	0.89
UN	234960006	C:T	3B	6309973	C:G	0.89
UN	36153637	C:T	3B	5601689	C:A	0.89
UN	234960006	C:T	3D	4068757	C:G	0.89
UN	234960006	C:T	3D	4068764	C:G	0.89
UN	36153637	C:T	3D	909572	G:T	0.89
UN	234960006	C:T	UN	36153637	C:T	0.87
UN	34103779	G:A	3B	6309968	G:A	0.87
UN	242439372	G:C	UN	34103779	G:A	0.87
UN	242452400	G:C	UN	34103779	G:A	0.87
UN	242452407	G:C	UN	34103779	G:A	0.87
UN	36153637	C:T	UN	34103779	G:A	0.87
UN	242439365	G:C	UN	34103779	G:A	0.87
UN	242439370	C:T	UN	34103779	G:A	0.87
UN	242452405	C:T	UN	34103779	G:A	0.87
UN	34103779	G:A	3B	6309966	C:G	0.87
UN	34103779	G:A	3D	4068759	G:A	0.87
UN	34103779	G:A	3B	6309973	C:G	0.87
UN	34103779	G:A	3D	4068757	C:G	0.87
UN	34103779	G:A	3D	4068764	C:G	0.87
UN	36153637	C:T	3B	10644041	T:A	0.77
3D	4068764	C:G	3B	10644041	T:A	0.76
3D	4068757	C:G	3B	10644041	T:A	0.76
3B	10644041	T:A	3B	6309968	G:A	0.75
UN	242439372	G:C	3B	10644041	T:A	0.75
UN	242452400	G:C	3B	10644041	T:A	0.75
UN	242452407	G:C	3B	10644041	T:A	0.75
UN	242439365	G:C	3B	10644041	T:A	0.75
UN	242439370	C:T	3B	10644041	T:A	0.75
UN	242452405	C:T	3B	10644041	T:A	0.75
3B	10644041	T:A	3B	6309966	C:G	0.74
3D	4068759	G:A	3B	10644041	T:A	0.74
3B	10644041	T:A	3B	6309973	C:G	0.74
UN	234960006	C:T	3B	10644041	T:A	0.71
UN	34103779	G:A	3B	10644041	T:A	0.70
3B	10644041	T:A	3B	5601689	C:A	0.70
3D	909572	G:T	3B	10644041	T:A	0.70

Appendix B

Genome-Wide Analysis of Vernalization Requirements in a Winter Wheat Recombinant Inbred Population

Table B.1 ANOVA of Heading Date as a function of Treatment, Taxa, and Treatment by Taxa Interaction. The p-values indicate that all three have a significant effect on the Heading Date.

ANOVA					
	DF	Sum Sq	Mean Sq	F Value	p-value
Treatment	3	1431690	477230	1332.938	<2e-16***
Taxa	197	1901755	9654	26.963	<2e-16***
Treatment*Taxa	591	914867	1548	4.324	<2e-16***
Residuals	976	349436	358		

Table B.2 Tukey's HSD (Honest Significant Difference) Test of all treatments compared to one another. 'The p adjusted' value shows that all treatments are significantly different from one another and therefore should be analyzed separately for the GWAS.

Tukey's HSD				
Treatment	Diff	Lower	Upper	p adjusted
Two – Zero	5.114488	1.943701	8.285275	0.0002105***
Four – Zero	-50.407846	-53.566411	-47.249280	0.0000000***
Six – Zero	-59.398259	-62.576507	-56.220010	0.0000000***
Four – Two	-55.522334	-59.002779	-52.041889	0.0000000***
Six – Two	-64.512747	-68.011064	-61.014430	0.0000000***
Six – Four	-8.990413	-12.477657	-5.503169	0.0000000***

Table B.3 Linkage Disequilibrium (LD) analysis of all significant SNPs. An R^2 of 1.0 indicates the two SNPs are completely correlated with one another in this population. An R^2 of 0.0 indicates that the two SNPs are completely unlinked from one another in this population.

Locus	Position1	States1	Locus	Position2	States2	R^2
5B	565507635	G:A	5B	565507633	G:A	1.00
5B	569243058	C:G	5B	569141895	G:A	1.00
5D	477960042	T:C	5B	585398903	T:C	1.00
5B	569243058	C:G	5B	565507635	G:A	0.96
5B	569243058	C:G	5B	565507633	G:A	0.96
5B	588093888	T:C	5B	585398903	T:C	0.94
5B	569083742	C:T	1D	439252922	G:A	0.94
5B	569243058	C:G	5B	569083742	C:T	0.93
5D	477960042	T:C	5B	588093888	T:C	0.91
5B	569141895	G:A	5B	565507635	G:A	0.90
5B	569141895	G:A	5B	565507633	G:A	0.90
5B	569141895	G:A	1D	439252922	G:A	0.90

(Continued)

5B	569141895	G:A	5B	569083742	C:T	0.90
5B	581028391	G:A	5B	569141895	G:A	0.88
5B	569243058	C:G	1D	439252922	G:A	0.87
5B	585398903	T:C	5B	581028391	G:A	0.83
5D	477960042	T:C	5B	581028391	G:A	0.83
5B	569083742	C:T	5B	565507633	G:A	0.83
5B	569083742	C:T	5B	565507635	G:A	0.81
5B	565507633	G:A	1D	439252922	G:A	0.81
5B	581028391	G:A	5B	569243058	C:G	0.80
5B	590557005	G:A	5B	588093888	T:C	0.80
5B	565507635	G:A	1D	439252922	G:A	0.79
5D	477960042	T:C	5B	590557005	G:A	0.75
5B	590557005	G:A	5B	585398903	T:C	0.73
5B	588093888	T:C	5B	581028391	G:A	0.72
5B	581028391	G:A	5B	565507635	G:A	0.67
5B	581028391	G:A	5B	565507633	G:A	0.67
5D	477960042	T:C	5B	569141895	G:A	0.66
5B	585398903	T:C	5B	569141895	G:A	0.65
5D	477960042	T:C	5B	569243058	C:G	0.64
5B	585398903	T:C	5B	569243058	C:G	0.64
5B	581028391	G:A	5B	569083742	C:T	0.63
5B	581028391	G:A	1D	439252922	G:A	0.61
5B	590557005	G:A	5B	581028391	G:A	0.59
5B	588093888	T:C	5B	569243058	C:G	0.56
5B	565507633	G:A	5B	563716653	A:C	0.54
5B	565507635	G:A	5B	563716653	A:C	0.53
5D	477960042	T:C	5B	565507635	G:A	0.52
5B	585398903	T:C	5B	565507635	G:A	0.52
5D	477960042	T:C	5B	565507633	G:A	0.52
5B	585398903	T:C	5B	565507633	G:A	0.52
5D	477960042	T:C	5B	569083742	C:T	0.51
5B	588093888	T:C	5B	569141895	G:A	0.50
5D	477960042	T:C	1D	439252922	G:A	0.50
5B	585398903	T:C	5B	569083742	C:T	0.50
5B	585398903	T:C	1D	439252922	G:A	0.48
5B	588093888	T:C	5B	565507635	G:A	0.45
5B	588093888	T:C	5B	565507633	G:A	0.45
5B	569141895	G:A	5B	563716653	A:C	0.45
5B	588093888	T:C	5B	569083742	C:T	0.44
5B	569243058	C:G	5B	563716653	A:C	0.41
5B	588093888	T:C	1D	439252922	G:A	0.40
5B	590557005	G:A	5B	569141895	G:A	0.39

(Continued)

5B	569083742	C:T	5B	563716653	A:C	0.39
5B	581028391	G:A	5B	563716653	A:C	0.37
5B	563716653	A:C	1D	439252922	G:A	0.37
5B	590557005	G:A	5B	565507635	G:A	0.36
5B	590557005	G:A	5B	569243058	C:G	0.36
5B	590557005	G:A	5B	565507633	G:A	0.35
5B	565507633	G:A	3B	184436222	A:G	0.33
5B	569243058	C:G	3B	184436222	A:G	0.32
5B	590557005	G:A	5B	569083742	C:T	0.32
5B	565507635	G:A	3B	184436222	A:G	0.31
5B	569083742	C:T	3B	184436222	A:G	0.31
5B	590557005	G:A	1D	439252922	G:A	0.30
3B	184436222	A:G	1D	439252922	G:A	0.30
5B	581028391	G:A	3B	184436222	A:G	0.27
5B	585398903	T:C	5B	563716653	A:C	0.27
5B	588093888	T:C	5B	563716653	A:C	0.27
5D	477960042	T:C	5B	563716653	A:C	0.26
5B	569141895	G:A	3B	184436222	A:G	0.25
5B	585398903	T:C	3B	184436222	A:G	0.19
5D	477960042	T:C	3B	184436222	A:G	0.19
5B	563716653	A:C	3B	184436222	A:G	0.18
5B	590557005	G:A	5B	563716653	A:C	0.17
5B	588093888	T:C	3B	184436222	A:G	0.16
5B	590557005	G:A	3B	184436222	A:G	0.14