Leveraging High Throughput Sequencing for Fine Fescue (*Festuca* spp.) Breeding and Genetics

A dissertation submitted to the faculty of the

University of Minnesota

By

Yinjie Qiu

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Dr. Eric Watkins, Advisor

University of Minnesota, Twin Cities

March 2020

© Yinjie Qiu 2020

Acknowledgements

First, and most of all, I would like to thank Dr. Eric Watkins, for his guidance and patience throughout the four and half years of my Ph.D. dissertation research at the University of Minnesota. I would also like to thank my Ph.D. committee members, Dr. Adrian Hegeman, Dr. Cory Hirsch, and Dr. Ya Yang, for their support, discussion, suggestions, and good conversations. Thank you to Kris Moncada, Dr. Dominic Petrella, and Dr. Florence Sessoms for all the support and helpful suggestions on my research projects.

I would like to extend my thanks to all the people working on metabolomics in Department of Horticultural Science, for valuable assistance during the later stages of my Ph.D. research.

I would also like to thank my friends who I shared an office with, and everyone else who helped me throughout my graduate study; to my best friends Sofia (twincita) and Matthew, thanks for keeping me company; to Judy and Del Johnson, who are functioning as my American parents in South Dakota, thanks for being there for me for the past 8 years! Finally, to my parents, thank you for all your support and letting me pursue my passions and dreams in a foreign country. I love you! 亲爱的爸爸妈妈，谢谢你们对我从小到大的培养和无限的付出，谢谢你们对我无穷无尽的关爱和支持。没有你们就没有我今天的成绩。我爱你们!

Abstract

Fine fescues (*Festuca* L., Poaceae) are turfgrass species that perform well in low-input environments. Improvement of these grass species through breeding and genetics have been limited due to the difficulty of species identification and lack of genomic resources. The objectives of this dissertation were to develop an improved method for fine fescue species identification, generate the first reference transcriptome for hard fescue, and use the reference transcriptome for transcriptome studies. In my first project, I used flow cytometry, chloroplast genome sequencing, and molecular marker development to provide new fine fescue identification methodology. Next, I used flow cytometry to characterize ploidy level in the USDA *F. ovina* collection. My third project used PacBio Isoform sequencing to develop the reference transcriptome using four tissue types for hard fescue; by using a phylotranscriptomic approach, the reference transcriptome provided information of the allopolyploid origin of the hexaploid species. Finally, the reference transcriptome was used to study how hard fescue responded to propiconazole fungicide application; in addition, untargeted metabolomics was used to study changes in metabolites caused by fungicide application. This dissertation developed methods for fine fescue by a combination of flow cytometry and molecular markers. Methods and genomics resources developed in this dissertation will benefit fine fescue breeding and genetics programs.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Turfgrass has been used as ground cover and maintained by humans for more than 10 centuries. It has several functional aspects such as to soil erosion protection, groundwater recharge, biodegradation of organic pollutants, and soil restoration (Clark and Kenna, 2010). Approximately 83% of the 84 million households in the United States are in urban areas, and 79% of these households have private lawns (Whitmore et al., 1993). Turfgrass is also found in a number of other situations including parks, school grounds, roadsides, and professionally managed sports surfaces. For instance, as of 2009, there were more than 16,000 golf courses in the United States that covers around 1.2 million acres of irrigated turfgrass (Throssell et al., 2009).

To maintain healthy turfgrass for home and garden use, the U.S. Environmental Protection Agency estimated that around 12 million pounds of insecticides and 4.5 million pounds of fungicide were applied in 1996 (Joyce, 1998). A consumer survey conducted in 2016 in the United States and Canada suggested a need for turfgrass cultivars requiring less use of water, fertilizers, and reduced mowing frequency (Yue et al., 2017).

Fine fescues are a group of turfgrass that perform well under low-input environments (Ruemmele et al., 1995; Watkins et al., 2010). Fine fescues include hard fescue (*Festuca brevipila* Tracey, 2n=6x=42), sheep fescue (*Festuca ovina* L., 2n=4x=28), strong creeping red fescue (*Festuca rubra* ssp. *rubra* [L.] 2n=8x=56), slender creeping red

fescue (*Festuca rubra* ssp. *littoralis* [G. Mey.] Auquier 2n=6x=42), and Chewings fescue (*Festuca rubra* ssp. *fallax* [Thuill.] Nyman 2n=6x=42; Ruemmele et al., 1995). Based on morphological similarity, fine fescues are divided into the *Festuca ovina* complex which includes hard and sheep fescue; and the *Festuca rubra* complex which includes strong creeping red, slender creeping red, and Chewings fescue (Huff and Palazzo, 1998). The species identification of fine fescue is challenging due to the indistinguishable morphological characters within each complex and naming differences between different continents (Hubbard, 1968; Beard, 1972).

Breeding efforts for fine fescue have been limited due to the difficulties of species identification and shortage of genetics information about these species. Due to ploidy differences, DNA content-based determination provided the first method of non-morphological taxa identification (Huff and Palazzo, 1998). Nevertheless, this is problematic when taxa have the same ploidy level. To solve this problem, developing molecular markers based on plastid DNA sequence polymorphisms could provide a more accurate species identification result (Demesure et al., 1995). Besides species identification, the lack of a reference genome for any turfgrass makes transcriptome studies difficult. With the advancement of sequencing technologies, the cost to sequence plastid genomes and transcriptome is possible for non-model species like fine fescues.

In this dissertation, the chloroplast genomes of five fine fescue were sequenced, annotated, and used for molecular markers development. Additionally, a chloroplast genome-based phylogenetic tree was reconstructed for fine fescues and other species in the *Festuca – Lolium* complex. To evaluate the ploidy of germplasm from the USDA PI

accession from the USDA-ARS National Plant Germplasm System, flow cytometry was performed for 127 USDA PIs to estimate DNA content and ploidy level.

To improve the quality of transcriptome studies in hard fescue, PacBio Isoform was used to generate a reference transcriptome using four tissue types (root, crown, leaf, and seedhead). The PacBio Isoseq transcriptome showed improved sequence lengths and completeness over the de novo assembly from short reads sequencing. In addition, it provided the power to perform phylotranscriptomic analysis on determining the hard fescue polyploid genome. Finally, the interaction between the commonly used fungicide propiconazole and hard fescue was studied using both transcriptome and metabolomics approaches. The long-term effect of using propiconazole on plants was studied by inoculating pink snow mold (*Microdochium nivale*).

Information generated from this dissertation provides a genetic and genomic foundation to advance fine fescue breeding and genetics into molecular marker-assisted breeding and gene discovery.

Chapter 2

Towards Improved Molecular Identification Tools in Fine Fescue (*Festuca* L., Poaceae) Turfgrasses: Nuclear Genome Size, Ploidy, and Chloroplast Genome Sequencing

Yinjie Qiu[1], Cory D. Hirsch[2], Ya Yang[3] and Eric Watkins[1]

[1] Department of Horticultural Science, University of Minnesota, St. Paul, MN, 55108, USA;

[2] Department of Plant Pathology, University of Minnesota, St. Paul, MN, 55108, USA;

[3] Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, 55108, USA

This manuscript is published in Frontiers in Genetics (Qiu et al., 2019)Copyright © 2019 Qiu, Hirsch, Yang and Watkins. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Author contribution: YQ performed the experiments, analyzed the data, and wrote the manuscript. CDH helped analyze data and wrote custom computational scripts. YY helped with the phylogenetic analysis. EW secured funding for this project, supervised the research, and provided suggestions and comments. All authors contributed to the revision of the manuscript and approved the final version.

Summary

Fine fescues (*Festuca* L., Poaceae) are turfgrass species that perform well in low-input environments. Based on morphological characteristics, the most commonly utilized fine fescues are divided into five taxa: three are subspecies within *F. rubra* L. and the remaining two are treated as species within the *F. ovina* L. complex. Morphologically, these five taxa are very similar; both identification and classification of fine fescues remain challenging. In an effort to develop identification methods for fescues, we used flow cytometry to estimate genome size, ploidy level, and sequenced the chloroplast genome of all five taxa. Fine fescue chloroplast genome sizes ranged from 133,331 to 133,841 bp and contained 113 to 114 genes. Phylogenetic relationship reconstruction using whole chloroplast genome sequences agreed with previous work based on morphology. Comparative genomics suggested unique repeat signatures for each fine fescue taxon that could potentially be used for marker development for taxon identification.

Keywords: Fine fescue, chloroplast genome, phylogeny, comparative genomics

Introduction

With ca. 450 species, fescues (*Festuca* L., Poaceae) is a large and diverse genus of perennial grasses (Clayton and Renvoize, 1986). Fescue species are distributed mostly in temperate zones of both the northern and southern hemispheres, but most commonly found in the northern hemisphere (Jenkin, 1959). Several of the fescue species have been commonly used as turfgrass. Based on both leaf morphology and nuclear ITS sequences, fescue species can be divided into two groups: broad-leaved fescues and fine-leaved fescues (Torrecilla and Catalán, 2002). Broad-leaved fescues commonly used as turfgrass include tall fescue (*Festuca arundinacea* Schreb.) and meadow fescue (*Festuca pratensis* Huds.). Fine-leaved fescues are a group of cool-season grasses that include five commonly used taxa called fine fescues. Fine fescues include hard fescue (*Festuca brevipila* Tracey, 2n=6x=42), sheep fescue (*Festuca ovina* L., 2n=4x=28), strong creeping red fescue (*Festuca rubra* ssp. *rubra* 2n=8x=56), slender creeping red fescue (*Festuca rubra* ssp. *littoralis* [G. Mey.] Auquier 2n=6x=42), and Chewings fescue (*Festuca rubra* ssp. *fallax* [Thuill. Nyman] 2n=6x=42) (Ruemmele et al., 1995). All five taxa share very fine and narrow leaves and have been used for forage, turf, and ornamental purposes. They are highly tolerant to shade and drought, prefer low pH (5.5-6.5), and low fertility soils (Beard, 1972). Additionally, fine fescues grow well in the shade or sun, have reduced mowing requirements, and do not need additional fertilizer or supplemental irrigation (Ruemmele et al., 1995).

Based on morphological and cytological features, fine fescues are currently divided into two groups referred to as the *F. rubra* complex (includes *F. rubra* ssp. *littoralis*, *F. rubra* ssp. *rubra*, *F. rubra* ssp. *fallax*) and the non-rhizomatous *F. ovina* complex (includes

F. brevipila and *F. ovina*; Ruemmele et al., 1995). While it is relatively easy to separate fine fescue taxa into their proper complex based on the presence and absence of rhizome, it is challenging to identify taxon within the same complex. In the *F. rubra* complex, both ssp. *littoralis* and ssp. *rubra* are rhizomatous while ssp. *fallax* is non-rhizomatous. However, the separation of ssp. *littoralis* from ssp. *rubra* using rhizome length is challenging. Taxon identification within the *F. ovina* complex relies heavily on leaf characters; however, abundant morphological and ecotype diversity within *F. ovina* makes taxa identification difficult (Piper, 1906). This is further complicated by inconsistent identification methods between different continents. For example, in the United States, sheep fescue is described as having a bluish-gray leaf color and hard fescue leaf blade color is considered green (Beard, 1972), while in Europe, it is the opposite (Hubbard, 1968). Because the ploidy level of the five taxa varies from tetraploid to octoploid, beyond morphological classifications, laser flow cytometry has been used to determine ploidy level of fine fescues and some other fescue species (Huff and Palazzo, 1998). A wide range of DNA content within each complex suggests that the evolutionary history of each named species is complicated, and interspecific hybridization might interfere with species determination using this approach. Plant breeders have been working to improve fine fescues for turf use for several decades, with germplasm improvement efforts focused on disease resistance, traffic tolerance, and ability to perform well under heat stress (Casler, 2003). Turfgrass breeders have utilized germplasm collections from old turf areas as a source of germplasm (Bonos and Huff, 2013); however, confirming the taxon identity in these collections has been challenging. A combination of molecular markers and flow

cytometry could be a valuable tool for breeders to identify fine fescue germplasm (Hebert et al., 2003).

Due to the complex polyploidy history of fine fescues, sequencing plastid genomes provides a more cost-effective tool for taxon identification than the nuclear genome because it is often maternally inherited, lacks heterozygosity, is present in high copies, and usable even in partially degraded material (Bryan et al., 1999; Provan et al., 2001). Previous studies have developed universal polymerase chain reaction (PCR) primers to amplify non-coding polymorphic regions for DNA barcoding in plants for species identification (Baldwin et al., 1995; Demesure et al., 1995). However, the polymorphisms discovered from these regions are often single nucleotide polymorphisms that are difficult to apply using PCR screening methods. For these reasons, it would be helpful to assemble chloroplast genomes and identify simple sequence repeat (SSR) and tandem repeats polymorphisms. Chloroplast genome sequencing has been simplified due to improved sequencing technology. In turfgrass species, high throughput sequencing has been used to assemble the chloroplast genomes of perennial ryegrass (*Lolium perenne* cv. Cashel), (Diekmann et al., 2009); tall fescue (*Lolium arundinacea* cv. Schreb), (Cahoon et al., 2010); diploid *Festuca ovina*, *Festuca pratensis*, *Festuca altissima* (Hand et al., 2013); and bermudagrass (*Cynodon dactylon*), (Huang et al., 2017). To date, there is limited molecular biology information on fine fescue taxon identification and their phylogenetic position among other turfgrass species (Hand et al., 2013; Cheng et al., 2016). In this study, we used flow cytometry to confirm the ploidy level of five fine fescue cultivars, each representing one of the five commonly utilized fine fescue taxon. We then reported the complete chloroplast genome sequences of these five taxa, carried out comparative genomics and

phylogenetic inference. Based on the genome sequence we identified unique genome features among fine fescue taxa and predicted taxon specific SSR and tandem repeat loci for molecular marker development.

Materials and Methods

Plant Material

Seeds from fine fescue cultivars were obtained from the 2014 National Turfgrass Evaluation Program (www.ntep.org, USA) and planted in the Plant Growth Facility at the University of Minnesota, St. Paul campus under 16 hours daylight (25 ℃) and 8 hours dark (16 ℃) with weekly fertilization. Single genotypes of *F. brevipila* cv. Beacon, *F. rubra* ssp. *littoralis* cv. Shoreline, *F. rubra* ssp. *rubra* cv. Navigator II, *F. rubra* ssp. *fallax* cv. Treazure II, and *F. ovina* cv. Quatro were selected and used for chloroplast genome sequencing.

Flow Cytometry

To determine the ploidy level of the cultivars used for sequencing and compare them to previous work (2n=4x=28: *F. ovina*; 2n=6x=42: *F. rubra* ssp. *littoralis*, *F. rubra* ssp. *fallax*, and *F. brevipila*; 2n=8x=56: *F. rubra* ssp. *rubra*), flow cytometry was carried out using *Lolium perenne* cv. Artic Green (2n=2x=14) as the reference. Samples were prepared using CyStain PI Absolute P (Sysmex, product number 05-5022). Briefly, to prepare the staining solution for each sample, 12 µL propidium iodide (PI) was added to 12 mL of Cystain UV Precise P staining buffer with 6 µL RNase A. To prepare plant tissue, a total size of 0.5 cm x 0.5 cm leaf sample of the selected fine fescue was excised into small pieces using a razor blade in 500 µL CyStain UV Precise P extraction buffer and passed through a 50 µm size filter (Sysmex, product number 04-004-2327). The staining solution was added to the flow-through to stain nuclei in each sample. Samples were stored on ice before loading the flow cytometer. Flow cytometry was carried out using the BD LSRII H4760 (LSRII) instrument with PI laser detector using 480V with 2,000 events at the University

of Minnesota Flow Cytometry Resource (UCRF). Data was visualized and analyzed on BD FACSDiva 8.0.1 software. To estimate the genome size, *L. perenne* DNA (5.66 pg/2C) was used as standard (Arumuganathan et al., 1999), USDA PI 230246 (2n=2x=14) was used as diploid fine fescue relative (unpublished data). To infer fine fescues ploidy, estimation was done using equations (1) and (2) (Doležel et al., 2007).

$$(1) \quad \text{Sample 2C DNA Content} = \text{Standard 2C DNA Content (pg DNA)} \times \frac{(\text{Sample G1 Peak Mean})}{(\text{Standard G1 Peak Mean})}$$

$$(2) \quad \text{Sample Ploidy} = \frac{(2n \times \text{Sample pg/Nucleus})}{(\text{Diploid Relative pg/Nucleus})}$$

DNA Extraction and Sequencing

To extract DNA for chloroplast genome sequencing, 1 g of fresh leaves were collected from each genotype and DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega, USA) following manufacturer instructions. DNA quality was examined on 0.8% agarose gel and quantified via PicoGreen (Thermo Fisher, Catalog number: P11496). Sequencing libraries were constructed by NovoGene, Inc. (Davis, CA) using Nextera XT DNA Library Preparation Kit (Illumina) and sequenced in 150 bp paired-end mode, using the HiSeq X platform (Illumina Inc., San Diego, CA, USA) with an average of 10 million reads per sample. All reads used in this study were deposited in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA512126.

Chloroplast Genome Assembly and Annotation

Raw reads were trimmed of Illumina adaptor sequences using Trimmomatic (v. 0.32) (Bolger et al., 2014). Chloroplast genomes were *de novo* assembled using NovoPlasty v. 2.0 (Dierckxsens et al., 2016). Briefly, *rbcL* gene sequence from diploid *F. ovina* (NCBI accession number: JX871940) was extracted and used as the seed to initiate the assembly.

NovoPlasty assembler configuration was set as follows: *k-mer* size = 39; insert size = auto; insert range = 1.8; and insert range strict 1.3. Reads with quality score above 25 were used to complete the guided assembly using *F. ovina* (NCBI accession number: JX871940) as the reference. Assembled plastid genomes for each taxon were manually corrected by inspecting the alignments of reads used in the assembly. The assembled chloroplast genomes were deposited under BioProject PRJNA512126, GenBank accession numbers MN309822-MN309826.

The assembled chloroplast genomes were annotated using the GeSeq pipeline (Tillich et al., 2017) and corrected using DOGMA online interface (https://dogma.ccbb.utexas.edu; Wyman et al., 2004). BLAT [a BLAST-like alignment tool (Kent, 2002)] protein, tRNA, rRNA, and DNA search identity threshold was set at 80% in the GeSeq pipeline using the default reference database with the generate codon-based alignments option turned on. tRNAs were also predicted via tRNAscan-SE v2.0 and ARAGORN v 1.2.38 using the bacterial/plant chloroplast genetic code (Lowe and Eddy, 1997; Laslett and Canback, 2004). The final annotation was manually inspected and corrected using results from both pipelines. The circular chloroplast map was drawn by the OrganellarGenomeDRAW tool (OGDRAW; Lohse et al., 2007).

Nucleotide Polymorphism of Fine Fescue Species

To identify genes with the most single nucleotide polymorphism, quality trimmed sequencing reads of the five fine fescues were mapped to the diploid *Festuca ovina* chloroplast genome (NCBI accession number: JX871940) using BWA v.0.7.17 (Li and Durbin, 2009). SNPs and short indels were identified using bcftools v.1.9 with the setting "mpileup -Ou" and called via bcftools using the -mv function (Quinlan and Hall, 2010).

Raw SNPs were filtered using bcftools filter -s option to filter out SNPs with low quality (Phred score cutoff 20, coverage cutoff 20). The subsequent number of SNPs per gene and InDel number per gene was calculated using a custom perl script SNP_vcf_from_gene_gff.pl (https://github.com/qiuxx221/fine-fescue-).

To identify simple sequence repeat (SSR) markers for plant identification, MIcroSAtellite identification tool (MISA v 1.0) was used with a threshold of 10, 5, 4, 3, 3, 3 repeat units for mono-, di-, tri-, tetra-, penta-, and hexanucleotide SSRs, respectively (Thiel et al., 2003). The identification of repetitive sequences and structure of whole chloroplast genome was done via REPuter program online server (https://bibiserv.cebitec.uni-bielefeld.de/reputer; Kurtz et al., 2001). Program configuration was set with minimal repeat size set as 20 bp and with sequence identify above 90%. Data was visualized using ggplot2 in R (v 3.5.3). Finally, the sliding window analysis was performed using DnaSP (v 5) with a window size of 600 bp, step size 200 bp to detected highly variable regions in the fine fescue chloroplast genome (Librado and Rozas, 2009).

Comparative Chloroplast Genomics Analysis

To compare fine fescue species chloroplast genome sequence variations, the five complete chloroplast genomes were aligned and visualized using mVISTA, an online suite of computation tools with LAGAN mode (Brudno et al., 2003;Frazer et al., 2004). The diploid *Festuca ovina* (NCBI accession number: JX871940) chloroplast genome and annotation were used as the template for the alignment.

Phylogenetic Analysis of Fine Fescues and Related *Festuca* Species

To construct the phylogenetic tree of the fine fescues using the whole chloroplast genome sequence, chloroplast genomes of eight species were downloaded from GenBank. Of the eight downloaded genomes, perennial ryegrass (*Lolium perenne*, AM777385), Italian ryegrass (*Lolium multiflorum*, JX871942), diploid *Festuca ovina* (JX871940), tall fescue (*Festuca arundiancea*, FJ466687), meadow fescue (*Festuca pratensis*, JX871941), and wood fescue (*Festuca altissima*, JX871939) were within the *Festuca-Lolium* complex. Turfgrass species outside of *Festuca-Lolium* complex including creeping bentgrass (*Agrostis stolonifera* L., EF115543) and *Cynodon dactylon* (KY024482.1) were used as an outgroup. All chloroplast genomes were aligned using the MAFFT program (v 7; Katoh and Standley, 2013); alignments were inspected and manually adjusted. Maximum likelihood (ML) analyses was performed using the RAxML program (v 8.2.12) under GTR+G model with 1,000 bootstrap (Stamatakis, 2006). The phylogenetic tree was visualized using FigTree (v 1.4.3, https://github.com/rambaut/figtree), (Rambaut, 2012).

Results

Species Ploidy Level Confirmation

We used flow cytometry to estimate the ploidy levels of five fine fescue taxa by measuring the DNA content in each nucleus. DNA content was reflected by the flow cytometry mean PI-A value. Overall, fine fescue taxa had mean PI-A values roughly from 110 to 180 (**Figure 2.1** and **Figure S2.1**). *Festuca rubra* ssp. *rubra* cv. Navigator II (2n=8x=56) had the highest mean PI-A value (181.434, %rCV 4.4). *Festuca rubra* ssp. *littoralis* cv. Shoreline (2n=6x=42) and *F. rubra* ssp. *fallax* cv. Treazure II (2n=6x=42) had similar mean PI-A values of 137.852, %rCV 3.7 and 145.864, %rCV 3.5, respectively. *Festuca brevipila* cv. Beacon (2n=6x=42) had a mean PI-A of 165.25, %rCV 1.9, while *F. ovina* cv. Quatro (2n=4x=28) had a mean PI-A of 108.43, %rCV 2.9. Standard reference *L. perenne* cv. Artic Green (2n=2x=14) had a G1 phase mean PI-A of 63.91, %rCV 3.0. USDA *F. ovina* PI 230246 (2n=2x=14) had a G1 mean PI-A of 52.73 (histogram not shown). The estimated genome size of USDA PI 230246 was 4.67 pg/2C. Estimated ploidy level of *F. brevipila* cv. Beacon was 6.3, *F. ovina* cv. Quatro was 4.11, *F. rubra* ssp. *rubra* cv. Navigator II was 6.9*, F. rubra* ssp. *littoralis* cv. Shoreline was 5.2, and *F. rubra* ssp. *fallax* cv. Treazure II was 5.5 (**Table 2.1**). All newly estimated ploidy levels roughly correspond to previously reported ploidy levels based on chromosome counts.

Plastid Genome Assembly and Annotation of Five Fescue Taxa

A total of 47,843,878 reads were produced from the five fine fescue taxa. After Illumina adaptor removal, we obtained 47,837,438 reads. The assembled chloroplast genomes ranged from 133,331 to 133,841 bp. The large single copy (LSC) and small single copy (SSC) regions were similar in size between the sequenced fine fescue accessions (78

kb and 12 kb, respectively). *Festuca ovina* and *F. brevipila* in the *F. ovina* complex had exactly the same size inverted repeat (IR) region (42,476 bp). In the *F. rubra* complex, *F. rubra* ssp. *rubra* and *F. rubra* ssp. *littoralis* had the same IR size (21,235 bp). Species in the *F. rubra* complex had a larger chloroplast genome size compared to species in the *F. ovina* complex. All chloroplast genomes shared similar GC content (38.4%) (**Figure 2.2** and **Table 2.2**). The fine fescue chloroplast genomes encoded for 113-114 genes, including 37 transfer RNAs (tRNA), 4 ribosomal RNAs (rRNA), and 72 protein-coding genes (**Table 2.2**). Genome structures were similar among all five fine fescue taxa sequenced, except that the pseudogene *accD* was annotated in all three subspecies of *F. rubra,* but not in the *F. ovina* complex (**Table S2.1**).

Chloroplast Genome IR Expansion and Contraction

Contraction and expansion of the IR regions resulted in the size variation of chloroplast genomes. We examined the four junctions in the chloroplast genomes, LSC/IRa, LSC/IRb, SSC/IRa, and SSC/IRb, of the fine fescue and the model turfgrass species *L. perenne*. Although the chloroplast genome of fine fescue taxa were highly similar, some structural variations were still found in the IR/LSC and IR/SSC boundary. Similar to *L. perenne*, fine fescue taxa chloroplast genes *rpl22-rps19, rps19-psbA* were located in the junction of IR and LSC; *rps15-ndhF* and *ndhH-rps15* were located in the junction of IR/SSC. In the *F. ovina* complex, the *rps19* gene was 37 bp into the LSC/IRb boundary while in the *F. rubra* complex and *L. perenne*, the *rps19* gene was 36 bp into the LSC/IRb boundary (**Figure 2.3**). The *rsp15* gene was 308 bp from the IRb/SSC boundary in *F. ovina* complex, 307 bp in *F. rubra* complex, and 302 bp in *L. perenne*. Both the *ndhH* and the pseudogene fragment of the *ndhH* (*ψndhH)* genes spanned the junction of the

IR/SSC. The *ψndhH* gene crossed the IRb/SSC boundary with 32 bp into SSC in *F. brevipila* and *F. ovina*, 9 bp in *F. rubra* ssp. *rubra* and *F. rubra* ssp. *littoralis*, 10 bp in *F. rubra* ssp. *fallax,* and 7 bp in *L. perenne*. The *ndhF* gene was 88 bp from the IRb/SSC boundary in *F. brevipila* and *F. ovina*, 91 bp in *F. rubra* ssp. *rubra*, 84 bp in *F. rubra* ssp. *littoralis*, 77 bp in *F. rubra* ssp. *fallax*, and 102 bp in *L. perenne*. Finally, the *psbA* gene was 87 bp apart from the IRa/LSC boundary into the LSC in *L. perenne* and *F. ovina* complex taxa but 83 bp in the *F*. *rubra* complex taxa.

Whole Chloroplast Genome Comparison and Repetitive Element Identification

Genome-wide comparison among five fine fescue taxa showed high sequence similarity with most variations located in intergenic regions (**Figure 2.4**). To develop markers for species screening, we predicted a total of 217 SSR markers for fine fescue taxa sequenced (*F. brevipila* 39; *F. ovina* 45; *F. rubra* ssp. *rubra* 45; *F. rubra* ssp. *littoralis* 46; *F. rubra* ssp. *fallax* 42) that included 17 different repeat types for the fine fescue species (**Figure 2.5a** and **Table S2.2**). The most frequent repeat type was A/T repeats, followed by AT/AT. The pentamer AAATT/AATTT repeat was only presented in the rhizomatous *F. rubra* ssp. *littoralis* and *F. rubra* ssp. *rubra*, while ACCAT/ATGGT was only found in *F. ovina* complex species *F. brevipila* and *F. ovina*. Similar to SSR loci prediction, we also predicted long repeats for the fine fescue species and identified a total of 171 repeated elements ranging in size from 20 to 51 bp (**Figure 2.5b**, **Table S2.3**). Complementary (C) matches were only identified in *F. brevipila* and *F. ovina*; *F. rubra* species had more palindromic (P) and reverse (R) matches. Number of forward (F) matches were similar among five taxa. Selected polymorphic regions were validated by PCR and gel electrophoresis assay (**Figure S2.2**).

SNP and InDel Distribution in the Coding Sequence of Five Fine Fescue Species

To identify single nucleotide polymorphisms (SNPs, non-reference allele in this content), we used the diploid *F. ovina* chloroplast genome (JX871940.1) as the reference for the mapping and used the genome annotation file to identify genic and non-genic SNPs. The total genic and non-genic sequence of (JX871940.1) were 60,582 and 72,583 bp, respectively. We found SNP polymorphisms were over-present within intergenic regions in the *F. ovina* complex (~0.3 SNP/Kbp more), while were under-present in the *F. rubra* complex (~0.5 SNP/Kbp less). Most InDels were located in intergenic regions of the fine fescue species (**Table 2.3**). Between *F. ovina* and the *F. rubra* complex, *the ropC2* gene had the most SNPs (4 vs 31). The *rbcL* gene also had a high level of variation (1 vs 14.3). In addition, *rpoB*, *ccsA*, NADH dehydrogenase subunit genes (*ndhH*, *ndhF*, *ndhA*), and ATPase subunit genes (*atpA*, *atpB*, *aptF*) also showed variation between *F. ovina* and *F. rubra* complexes. Less SNP and InDel variation were found within each complex (**Table 2.3**, **Table S2.4** and **S2.5**).

Nucleotide Diversity Calculation

A sliding window analysis successfully detected highly variable regions in the fine fescue chloroplast genomes (**Figure 2.6** and **Table S2.6**). The average nucleotide diversity (Pi) among fine fescue taxa was relatively low (0.00318). The IR region showed lower variability than the LSC and SSC region. There were several divergent loci having a Pi value greater than 0.01 (*psbK-psbI*, *trnfM-trnE*, *trnC-rpoB*, *psbH-petB*, *trnL-trnF*, *trnS-rps4*, *aptB-rbcL-psaI*, and *rpl32-trnL*), but mostly within intergenic regions. The *rbcL-psaI* region contained a highly variable *accD-like* region in some fine fescue taxa, so we looked

at the structural variation of 10 taxa in the *Festuca-Lolium* complex. We found taxa in broad-leaved fescue and *F. rubra* complex had similar structure, while *F. ovina* (2x, 4x) and *F. brevipila* had a 276 bp deletion in the *rbcL-psaI* intergenic region (**Figure 2.7**).

Phylogenetic Reconstruction of Fine Fescue Species

We reconstructed the phylogenetic relationships of taxa within the *Festuca-Lolium* complex using the chloroplast genomes sequenced in our study and eight publicly available complete chloroplast genomes including six taxa within the *Festuca-Lolium* complex (**Figure 2.8**). The dataset included 125,824 aligned characters, of which 3,923 were parsimony-informative and 91.11% characters were constant. The five fine fescue taxa were split into two clades ([ML]BS=100). In the *F. ovina* complex, two *F. ovina* accessions included in the phylogenetic analysis, a diploid one from GenBank, and a tetraploid one newly sequenced in this study are paraphyletic to *F. brevipila* ([ML]BS=100). All three subspecies of *F. rubra* formed a strongly supported clade ([ML]BS=100). Together they are sisters to the *F. ovina* complex ([ML]BS=100).

Discussion

In this study, we used flow cytometry to determine the ploidy level of five fine fescue cultivars, assembled the chloroplast genomes for each, and identified structural variation and mutation hotspots. We also identified candidate loci for marker development to facilitate fine fescue species identification. Additionally, we reconstructed the phylogenetic relationships of the *Festuca-Lolium* complex using plastid genome information generated in this study along with other publicly available plastid genomes.

While most crop plants are highly distinctive from their close relatives, *Festuca* is a species-rich genus that contains species with highly similar morphology and different ploidy level. Consequently, it is difficult for researchers to interpret species identity. In our case, flow cytometry was able to successfully separate fine fescue taxa *F. brevipila* cv. Beacon, *F. ovina* cv. Quatro and *F. rubra* ssp. *rubra* cv. Navigator II based on the estimated ploidy levels. However, it is difficult to distinguish between *F. rubra* ssp. *littoralis* cv. Shoreline and *F. rubra* ssp. *fallax* cv. Treazure II as they had similar PI-A values based on flow cytometry.

We noticed that the average mean PI-A of the diploid *L. perenne* (63.91) was higher than the mean PI-A of diploid *F. ovina* (52.73), suggesting that *F. ovina* taxa have smaller genome size than *L. perenne*. The ploidy estimation in the *F. ovina* complex are fairly consistent while the estimations of genome sizes in the *F. rubra* complex are smaller than we expected, even though these two complexes are closely related. Indeed, a similar finding was reported by Huff et al., (1998) who reported that *F. brevipila* has a larger genome size than *F. rubra* ssp. *littoralis* and *F. rubra* ssp. *fallax*, both of which have the same ploidy level as *F. brevipila*. The range of variation in DNA content within each

complex suggests a complicated evolutionary history in addition to polyploidization (Huff and Palazzo, 1998).

When we cannot identify taxon based on the ploidy level, we need different approaches to identify them. The presence and absence of rhizome formation could be taken into consideration; for example, *F. rubra* ssp. *fallax* cv. Treazure II is a bunch-type turfgrass, while *F. rubra* ssp. *littoralis* cv. Shoreline forms short and slender rhizomes (Meyer and Funk, 1989). This method may not be effective because rhizome formation can be greatly affected by environmental conditions (Yang et al., 2015; Ma and Huang, 2016).

To further develop molecular tools to facilitate species identification, we carried out chloroplast genome sequencing. We assembled the complete chloroplast genomes of five low-input turfgrass fine fescues using Illumina sequencing. Overall, the chloroplast genomes had high sequence and structure similarity among all five fine fescue taxa sequenced, especially within each complex. All five chloroplast genomes share similar gene content except for the three species in the *F. rubra* complex that have a pseudogene Acetyl-coenzyme A carboxylase carboxyl transferase subunit (*accD*). The *accD* pseudogene is either partially or completely absent in some monocots. Instead, a nuclear-encoded ACC enzyme has been found to replace the plastic *accD* gene function in some angiosperm lineages (Rousseau-Gueutin et al., 2013). Indeed, even though the *accD* pseudogene is missing in the *F. brevipila* chloroplast genome, the gene transcript was identified in a transcriptome sequencing dataset (unpublished data), suggesting that this gene has been translocated to nucleus genome. Previous studies have shown that broad-leaf fescues, *L. perenne*, *O. sativa*, and *H. vulgare* all had the pseudogene *accD* gene, while it was absent in diploid *F. ovina*, *Z. mays*, *S. bicolor*, *T. aestivum,* and *B. distachyon* (Hand

et al., 2013). Broad-leaf and fine-leaf fescues diverged around 9 Mya ago (Fjellheim et al., 2006), which raises an interesting question about the mechanisms of the relocation of *accD* among closely related taxa in the *Festuca-Lolium* complex and even within fine fescue species.

In plants, chloroplast genomes are generally considered "single copy" and lack recombination due to maternal inheritance (Ebert and Peakall, 2009). For this reason, chloroplast genomes are convenient for developing genetic markers for distinguishing species and subspecies. We have identified a number of repeat signatures that are unique to a single species or species complex in fine fescue. For example, complement match is only identified in *F. ovina* complex, and *F. rubra* complex has more reversed matches. We also identified two SSR repeats unique to each of the two complexes. The first one consists of AAATT/AATTT repeat units is unique to *F. rubra* ssp. *littoralis* and *F. rubra* ssp. *rubra*, and the second one consists of ACCAT/ATGGT repeat units is unique to *F. brevipila* and *F. ovina*. In cases like the identification of hexaploid *F. brevipila*, *F. rubra* ssp. *fallax*, and *Festuca rubra* ssp. *littoralis*, it is critical to have these diagnostic repeats given all three taxa share similar PI-A values from flow cytometry. Taxon-specific tandem repeats could also aid the SSR repeats for species identification. We used chloroplast sequence developed candidate primer sets to solve the problem. The first primer set provided a clear separation of *F. rubra* ssp. *littoralis* cv. Shoreline and *F. rubra* ssp. *fallax* cv. Treazure II when flow cytometry was not able to separate them. The second primer set provided clear separation of *F. brevipila* cv. Beacon and *F. ovina* cv. Quatro, which provided an alternative method for *F. ovina* complex taxa identification. By combining both flow cytometry and candidate

primer sets developed in this study, researchers will be able to identify fine fescue taxa within and between two complexes.

Nucleotide diversity analysis suggested that several variable genome regions exist among the five fine fescue taxa sequenced in this study. These variable regions included previously known highly variable chloroplast regions such as *trnL-trnF* and *rpl32-trnL* (Demesure et al., 1995; Dong et al., 2012). These regions have undergone rapid nucleotide substitution and are potentially informative molecular markers for characterization of fine fescue species.

Phylogeny inferred from DNA sequence is valuable for understanding the evolutionary context of a species. The phylogenetic relationship of fine fescue using whole plastid genome sequences agrees with previous classification based on genome size estimation and morphology (Huff and Palazzo, 1998; Cheng et al., 2016). The *F. ovina* complex includes *F. ovina* and *F. brevipila* and the *F. rubra* complex includes *F. rubra* ssp. *rubra*, *F. rubra* ssp. *littoralis* and *F. rubra* ssp. *fallax*, with the two rhizomatous subspecies (ssp. *rubra* and ssp. *literalis*) being sisters to each other. Within the *Festuca-Lolium* complex, fine fescues are monophyletic and together sisters to a clade consisting of broad-leaved fescues and *Lolium*. In our analysis, *F. brevipila* (6x) is nested within *F. ovina* and sister to the diploid *F. ovina*. It is likely that *F. brevipila* arose from the hybridization between *F. ovina* (2x) and *F. ovina* (4x). Considering the complex evolutionary history of this genus, further research using nuclear loci sequences is needed to provide a more accurate phylogeny tree and validate this hypothesis.

The diversity of fine fescue provides valuable genetic diversity for breeding and cultivar development. Breeding fine fescue cultivars for better disease resistance, heat

tolerance, and traffic tolerance could be achieved through screening wild accessions and by introgressing desired alleles into elite cultivars. Some work has been done using *Festuca* accessions in the USDA Germplasm Resources Information Network (GRIN) (https://www.ars-grin.gov) to breed for improved forage production in fescue species (Robbins et al., 2016). To date, there are 229 *F. ovina* and 486 *F. rubra* accessions in the USDA GRIN. To integrate this germplasm into breeding programs, plant breeders and other researchers need to confirm the ploidy level using flow cytometry and further identify them using molecular markers. Resources developed in this study could provide the tools to screen the germplasm accessions and refine the species identification so breeders can efficiently use these materials for breeding and genetics improvement of fine fescue species.

Figures and Tables

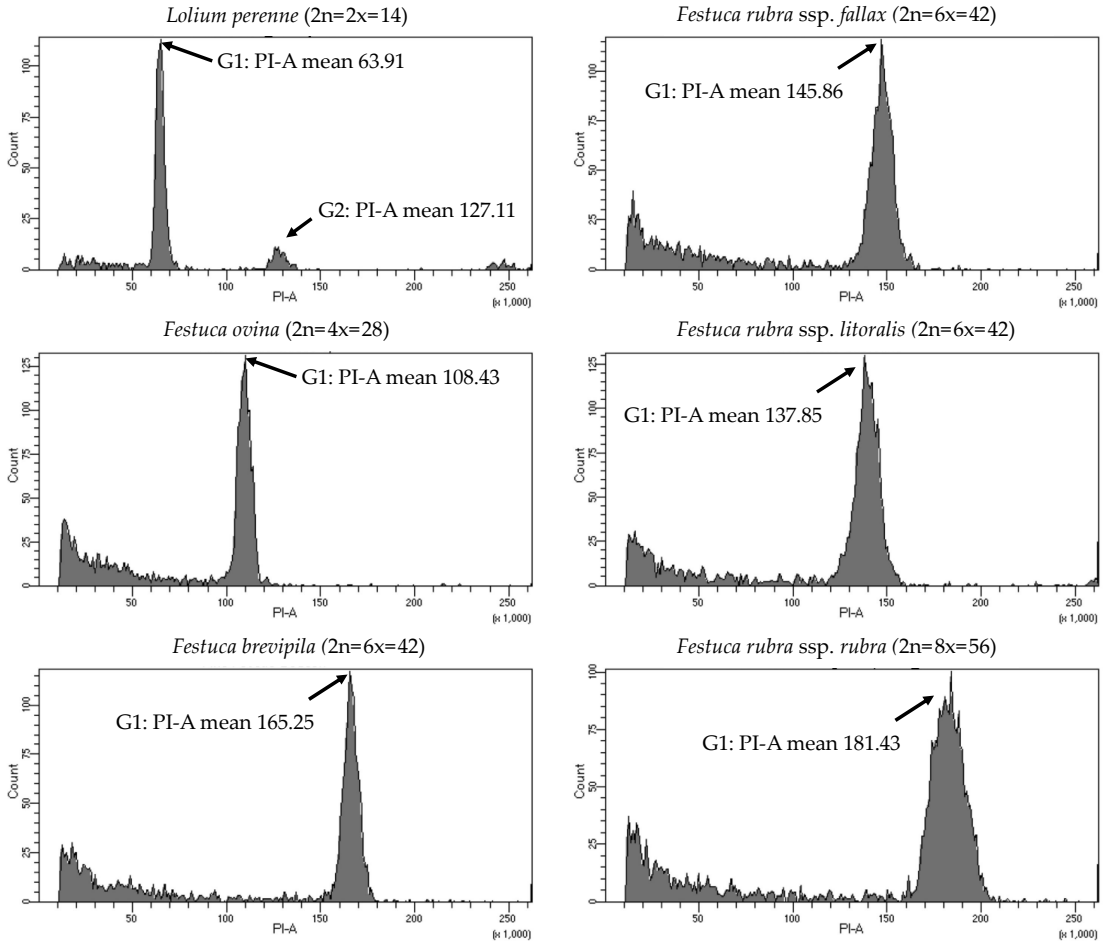

Figure 2.1. Flow cytometry results for the five fine fescue taxa. *Lolium perenne* (2n=2x=14) was used as the reference. Flow cytometry was able to separate *F. rubra* ssp. *rubra* from the other two subspecies in the *F. rubra* complex. The mean PI-A values of *F. rubra* ssp. *fallax* and *F. rubra* ssp. *littoralis* were similar (145.86 to 137.85).

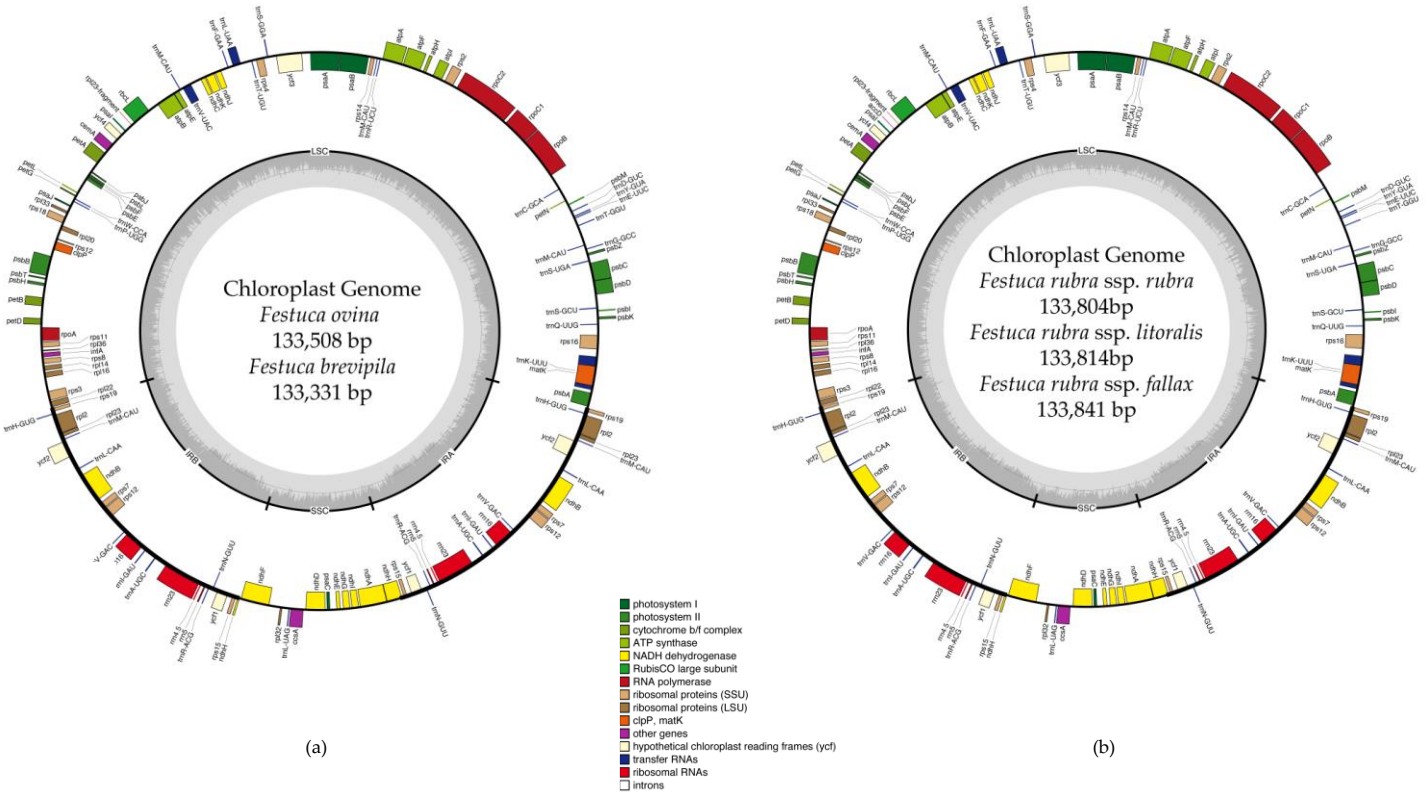

Figure 2.2. Whole chloroplast genome structure of *F. ovina* complex (a) and *F. rubra* complex (b). Genes inside the circle are transcribed clockwise, genes outside are transcribed counterclockwise. Genes belonging to different functional groups are color coded. GC content is represented by the dark gray inner circle, the light gray corresponds to the AT content. IRA(B), inverted repeat A(B); LSC, large single copy region; SSC, small single copy region.

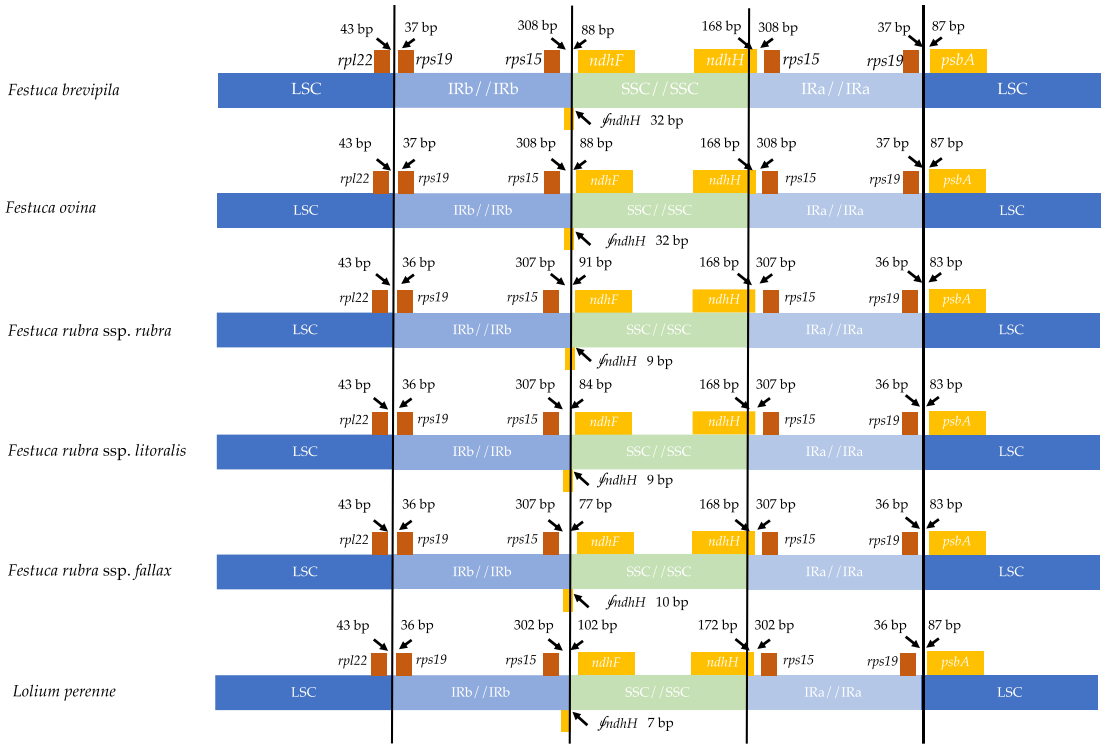

Figure 2.3. Comparison for border positions of LSC, SSC and IR regions among five fine fescues and *L. perenne.* Genes are denoted by boxes, and the gap between the genes and the boundaries are indicated by the number of bases unless the gene coincides with the boundary. Extensions of genes are also indicated above the boxes.

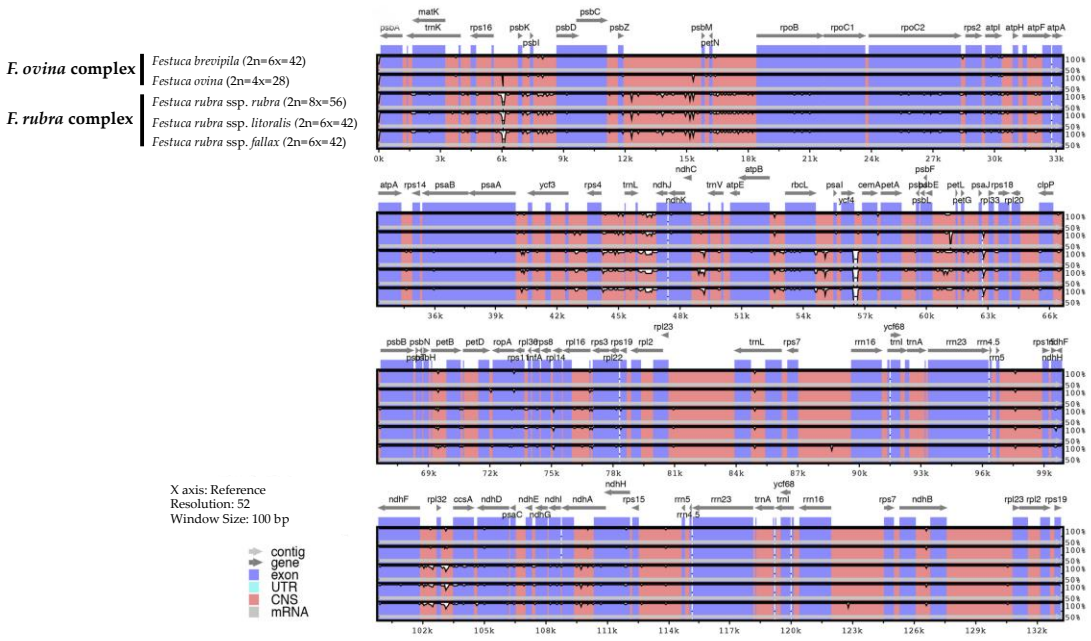

Figure 2.4. Sequence identity plot of fine fescues chloroplast genome sequences with *F. ovina* (2x) as the reference using mVISTA. A cut-off of 70% identify was used for the plots, and the percent of identity varies from 50% to 100% as noted on the y-axis. Most of the sequence variations between fine fescues were in intergenic regions. Taxa in the *F. ovina* complex, *F. brevipila* and *F. ovina* showed high sequence similarity. Similarly, subspecies within *F. rubra* complex also showed high sequence similarity.

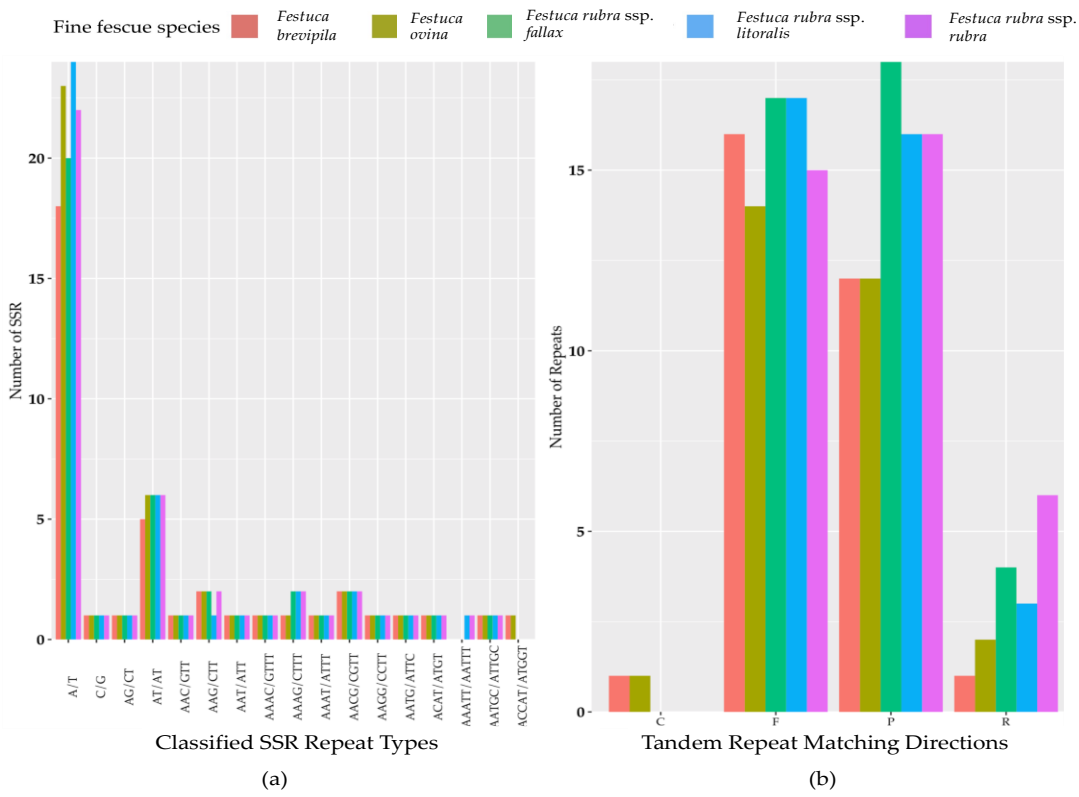

Figure 2.5. (a) SSR repeat type and numbers in the five fine fescue taxa sequenced. Single nucleotide repeat type has the highest frequency. No hexanucleotide repeats were identified in the fine fescue chloroplast genomes sequenced. One penta-nucleotide repeat type (AAATT/AATTT) is unique to *F. rubra* ssp. *rubra* and *F. rubra* ssp. *littoralis*; one penta-nucleotide repeat type (ACCAT/ATGGT) is unique to *F. brevipila* and *F. ovina*. (b) Long repeats sequences in fine fescue chloroplast genomes. Complement (C) match was only identified in the F. ovina complex; Reverse (R) match has the most number variation in fine fescues.

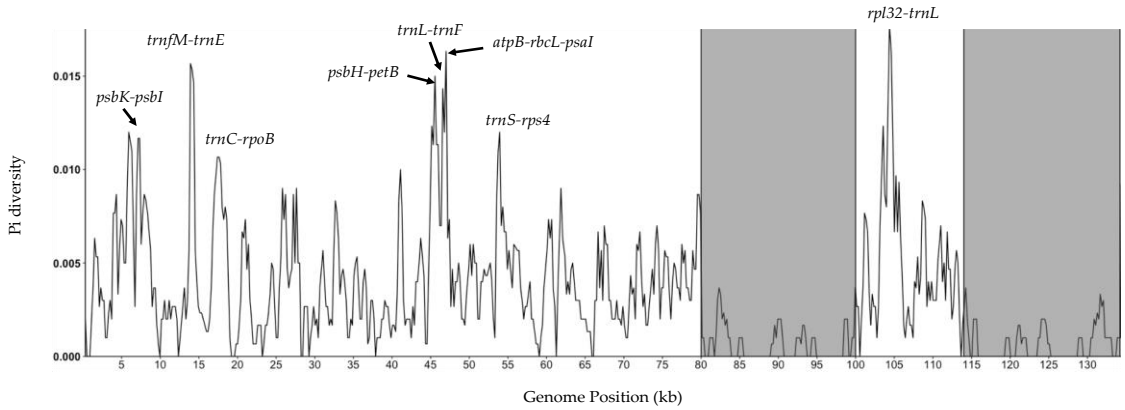

Figure 2.6. Sliding window analysis of fine fescue whole chloroplast genomes. Window size: 600 bp, step size: 200 bp. X-axis: the position of the midpoint of a window (kb). Y-axis: nucleotide diversity of each window. Inverted repeat regions are highlighted in gray. *rpl32-trnL* region has the most nucleotide diversity followed by *psbH-petB-trnL-trnF-trnS-rps4* region.

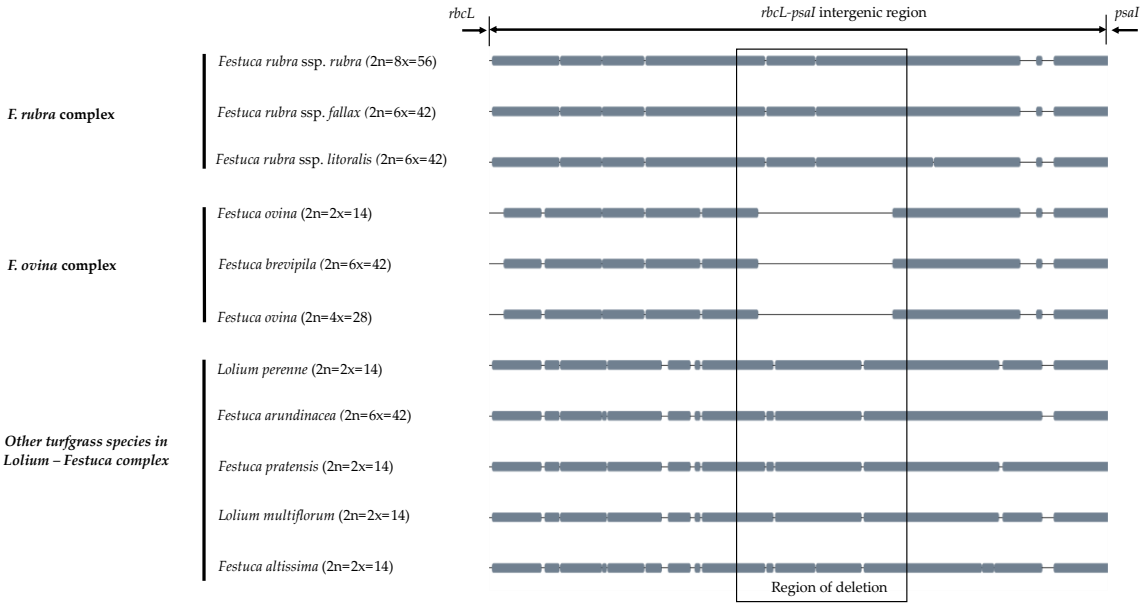

Figure 2.7. The alignment of *rbcL-psaI* intergenic sequence shows that the pseudogene *accD* is missing in both *F. ovina* (2x, 4x) and *F. brevipila* but present in the *F. rubra* complex and other species examined in this study. Species were ordered by complexes.

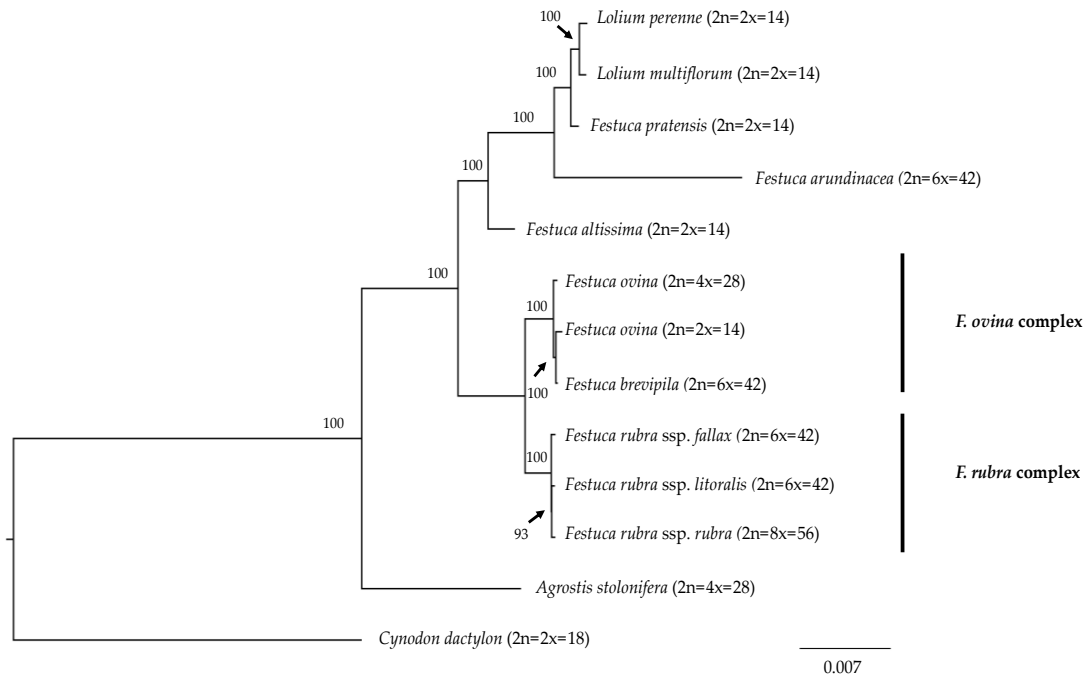

Figure 2.8. Maximum likelihood (ML) phylogram of the *Festuca-Lolium* complex with 1,000 bootstrap replicates. Fine fescues were grouped into previous named complexes (*F. ovina* and *F. rubra*), sister to broad leaved fescues in the *Festuca-Lolium* complex.

Table 2.1. Summary of flow cytometry statistics, genome size estimation, and ploidy level estimation of fine fescue species. *Lolium perenne* 2C DNA content was used to calculate fine fescue and USDA *F. ovina* PI 230246 genome size, calculated PI 230246 DNA content was used as reference to infer fine fescue ploidy level.

Species name	Chromosome count	Cultivar name	Mean PI-A	%rCV *	Estimated Genome Size (pg/Nuclei)	Estimated Ploidy Level
F. brevipila	2n=6x=42	Beacon	165.3	1.9	14.6	6.3
F. ovina	2n=4x=28	Quatro	108.4	2.9	9.6	4.1
F. ovina PI 230246	2n=2x=14	NA	52.7	3.1	4.7	1.7
F. rubra ssp. *rubra*	2n=8x=56	Navigator II	181.4	4.4	16.1	6.9
F. rubra ssp. *littoralis*	2n=6x=42	Shoreline	137.9	3.7	12.2	5.2
F. rubra ssp. *fallax*	2n=6x=42	Treazure II	145.9	3.5	12.9	5.5
L. perenne	2n=2x=14	Artic Green	63.9	3.0	5.7	2.0

* %rCV: Quality of laser alignment. Low %rCV suggests high resolution sensitivity.

Table 2.2. Characteristics of fine fescue chloroplast genomes.

	F. brevipila cv. Beacon	*F. ovina* cv. Quatro	*F. rubra* ssp. *rubra* cv. Navigator II	*F. rubra* ssp. *littoralis* cv. Shoreline	*F. rubra* ssp. *fallax* cv. Treazure II
NCBI GenBank ID	MN309822	MN309824	MN309825	MN309823	MN309826
Total Genome Size (bp)	133,331	133,508	133,804	133,814	133,841
Large Single Copy (bp)	78,462	78,632	78,888	78,909	78,882
Small Single Copy (bp)	12,393	12,400	12,446	12,435	12,451
Inverted Repeat (bp)	42,476	42,476	42,470	42,470	42,508
Ratio of LSC (%)	58.85	58.9	58.96	58.97	58.94
Ratio of SSC (%)	9.29	9.29	9.3	9.29	9.3
Ratio of IRs (%)	31.86	31.82	31.74	31.74	31.76
GC content (%)	38.4	38.4	38.4	38.4	38.4

Table 2.3. Distribution of SNPs and InDels for the five fine fescue taxa sequenced in this study.

Species	*F. brevipila*	*F. ovina*	*F. rubra* ssp. *rubra*	*F. rubra* ssp. *littoralis*	*F. rubra* ssp. *fallax*
Total number of SNPs	98	134	638	615	624
SNPs in the coding region	35	52	306	301	300
SNPs in intergenic region	63	82	332	314	324
SNPs per Kbp in genic region	0.5777	0.8583	5.0510	4.9685	4.9520
SNPs per Kbp in non-genic region	0.8680	1.1297	4.5741	4.3261	4.4639
Total number of InDels	112	102	149	156	149
InDels in the coding region	22	17	27	26	27
InDels in intergenic region	90	85	122	130	122
Percentage of InDels in the intergenic region	80.36	83.33	81.88	83.33	81.88
Average sequencing depth	171.61	86.81	101.58	77.04	50.94

Chapter 3

DNA Content and Ploidy Estimation of Turfgrass *Festuca ovina* Germplasm Accessions by Flow Cytometry

Yinjie Qiu, Sierra Hamernick, Joan Barreto Ortiz, and Eric Watkins

Author Affiliations:

Department of Horticultural Science, University of Minnesota, St. Paul, MN, 55108, USA

This manuscript has been released on bioRxiv (Qiu et al., 2020a).

Author Contributions: YQ and SH performed the flow cytometry experiments and analyzed the data. JO performed the image analysis. YQ wrote the manuscript. EW secured the funding, supervised this research.

Summary

Festuca ovina is a fine fescue that is used as a low-input turfgrass. The ploidy levels of *F. ovina* accessions held by the USDA National Plant Germplasm System (NPGS) are unknown, limiting the use of the germplasm in breeding programs. The objective of this study was to determine DNA content and estimate ploidy of these 127 accessions. Among the accessions, we identified a wide range of ploidy levels from diploid to octoploid. We also found the accessions with higher ploidy levels usually had larger seed size. These results will be informative to plant breeders and researchers using germplasm from the *F. ovina* collection and point to challenges in maintaining polyploid, outcrossing germplasm seed stocks in common nurseries.

Introduction

Fine fescues (*Festuca* spp. L.) are a diverse group of grasses characterized by fine leaf texture. Fine fescues are native to Eurasia but have been introduced and naturalized to many temperate regions in the world (Vasey, 1883; Beard, 1972; Barkworth et al., 2007). These grasses are used for forage, ornamental purposes, and particularly as low-input turfgrasses. The group comprises genetically diverse taxa, including hard fescue (*Festuca brevipila* Tracey, 2n=6x=42), sheep fescue (*F. ovina* L., 2n=4x=28), strong creeping red fescue (*Festuca rubra* ssp. *rubra* 2n=8x=56), slender creeping red fescue [*Festuca rubra* ssp. *littoralis* (G. Mey.) Auquier 2n=6x=42], and Chewings fescue [*Festuca rubra* ssp. *fallax* (Thuill.) Nyman 2n=6x=42] (Ruemmele et al., 1995).

Based on their morphology, cytology, and chloroplast-genome-based phylogenetic relationship, fine fescues are divided in two complexes: the *F. ovina* complex, which includes *F. brevipila* and *F. ovina*, and the *F. rubra* complex, which includes *F. rubra* ssp. *littoralis*, *F. rubra* ssp. *rubra*, and *F. rubra* ssp. *fallax* (Wilkinson and Stace, 1991; Huff and Palazzo, 1998; Qiu et al., 2019). Species in *F. ovina* are bunch-type and non-rhizomatous while subspecies in the *F. rubra* complex can be rhizomatous (ssp. *littoralis* and ssp. *rubra*) and bunch-type (ssp. *fallax*). Taxon identification within the *F. ovina* complex is difficult because of morphological and ecotype diversity (Piper, 1906; Schmit et al., 1974). Previously, leaf color has been used for species identification; however, in the United States, sheep fescue is described as having a bluish-gray leaf color and hard fescue leaf blade color is considered green (Beard, 1972), while in Europe, it is the opposite (Hubbard, 1968). The *F. ovina* complex also includes some other taxa beyond those commonly used as turfgrasses that have

ploidy levels varying from diploid (2n=2x=14) such as *Festuca ovina* ssp. *ovina*, tetraploid *Festuca armoricana*, and hexaploid *Festuca huonii* (Seal, 1983; Wilkinson and Stace, 1991; Stace, 2010). All of these factors make the identification of *F. ovina* challenging.

Turfgrass breeding and genetics objectives are focused on aesthetic beauty, disease resistance, drought tolerance, and traits associated with reduced inputs (Casler, 2003; Bonos et al., 2006; Clarke et al., 2006; Bonos and Huff, 2013). Plant breeders seek desirable alleles in exotic accessions and attempt to introgress them into the existing cultivars. One of the most valuable resources for breeders is the USDA-ARS National Plant Germplasm System (NPGS), which holds more than 500,000 accessions that represent more than 10,000 plant species. These accessions have been widely used in plant breeding programs for abiotic and biotic stress improvement (Nelson et al., 1987; Dilday et al., 1994; Christensen et al., 2007; Leng et al., 2016; Chang and Hartman, 2017). To improve turfgrass cultivars by utilizing the germplasm accessions, it is important to know the ploidy level of the accessions used to avoid hybridizing plants with different ploidy levels that results in nonviable offspring. Additionally, without knowing the ploidy level, genetics and phenotyping of these germplasm could lead to incorrect interpretation of results.

Traditional plant breeding methods that emphasize hybridizing elite germplasm usually result in the loss of genetic diversity and heterosis which could result in greater susceptibility to important stresses (Melchinger, 1999; Christiansen et al., 2002; Reif et al., 2005). The use of exotic germplasm has been a common practice in maintaining plant genetic diversity in the breeding process to reduce these problems and avoid breeding

bottlenecks (Van Esbroeck and Bowman, 1998; Goodman, 1999; Mikel et al., 2010; Prasanna, 2012). In both cool-season and warm-season turfgrasses, genetic diversity of available germplasm has been studied using either molecular markers or sequencing arrays (Budak et al., 2004; Chen et al., 2009; Baird et al., 2012).

For NPGS accessions from the *F. ovina* complex, it is necessary to determine ploidy level before conducting germplasm selection and performing hybridization; this has traditionally been done by counting chromosomes (Vargas et al., 1999; Maluszynska, 2003), a reliable but time-consuming method which is made all the more challenging when dealing with the numerous and small (even under magnification) chromosomes of the fine fescue species. Flow cytometry is a powerful tool for DNA content measurement and allows researchers to estimate the ploidy level by comparing to known standards. This method is less time consuming, cheaper, and proven to work in grasses where it has been used to calculate the DNA content and estimate ploidy levels in Texas bluegrass (*Poa arachnifera* Torr.), buffalograss [*Bouteloua dactyloides* (Nutt.) Engelm.], perennial ryegrass (*Lolium perenne* L.) and fine fescues (Arumuganathan and Earle, 1991; Huff and Palazzo, 1998; Johnson et al., 1998; Johnson et al., 2001; Goldman, 2015; Qiu et al., 2019).

We used flow cytometry to determine the DNA content and ploidy level of 127 USDA *F. ovina* PI collections. In addition, we used image analysis to measure and compare the seed size on selected PI accessions among ploidy levels.

Material and Methods

Plant Material

A total of 127 accessions labeled as *Festuca ovina* from 20 countries were obtained from the USDA Germplasm Resources Information Network (GRIN) in 2016 (**Table S3.1**).

Seeds of each accession were sown into greenhouse pots (four-inch size) filled with BRK Promix soil (Premier Tech, USA) at the Plant Growth Facility at the University of Minnesota in St. Paul. After reaching four to five leaf stage, five seedlings per accession were randomly selected and transplanted into individual one-inch size cone container. Plants were grown with 16 hr day and 8 hr night with bi-daily watering and weekly fertilization using Peat-lite 20-10-20 fertilizer (J.R. Peters Inc.) with supplemental ammonium sulfate and Sprint 330 (BASF, USA). The five genotypes of each accession were vegetatively cloned into six replications of each genotype and transplanted to a field nursery in at the Minnesota Agricultural Experiment Station in St. Paul, MN. *F. ovina* cv. Quatro and *F. brevipila* cv. Beacon were also planted in the field as standards.

Flow Cytometry Procedure

To determine the nucleus DNA content of accessions, flow cytometry was carried out using the method described by Arumuganathan and Earle (1991). Because seeds used in this study resulted from open pollination, we evaluated three genotypes for each accession to obtain a more accurate ploidy representation of the population. When the three selected genotypes were not at the same ploidy level, a fourth genotype was evaluated. The ploidy level of the accession was determined by the ploidy of the majority genotypes (75%).

Fresh mature *F. ovina* leaf samples in the nursery field were harvested between 9-11 a.m. and trimmed to 1-2 cm and used for flow cytometry. Perennial ryegrass leaf tissue was harvested at the same time in the greenhouse. For each sample the flow cytometry staining solution contained 4.29 µL propidium iodide, 0.71 mL of CyStain UV Precise P staining buffer (Sysmex), and 2.14 µL RNAseA. To prepare plant tissue, a 0.5 cm x 0.5 cm leaf sample was excised into small pieces using a razor blade in 500 µL CyStain UV Precise P extraction buffer and passed through a 50-µm size filter (Sysmex). The staining solution was added to the flow-through to stain the nuclei in each sample. Samples were stored on ice before loading the flow cytometer. Flow cytometry was carried out using the BD LSRII H4760 (LSRII) instrument (BD Biosciences, USA) with PI laser detector using 480V with a minimum of 1,000 events at the University of Minnesota Flow Cytometry Resource (UCRF). Data were visualized and analyzed on BD FACSDiva 8.0.1 software.

DNA Content Estimation and Ploidy Level Determination

Species DNA content was estimated following the method described by Dolezel and Bartoš (Doležel and Bartoš, 2005). The perennial ryegrass (*Lolium perenne*) 2C DNA content (2C = 5.66 pg/2C) served as the diploid DNA content standard (Arumuganathan et al., 1999). Sample 2C DNA content was calculated using Equation 1. To estimate the ploidy level in *F. ovina* complex, diploid *F. ovina* PI 230246 measured in previous study was used (2C = 4.7 pg/2C); species ploidy level was estimated using Equation 2.

Equation 1 $$\text{Sample 2C DNA content} = \frac{\text{(sample G1 peak mean)}}{\text{(standard G1 peak mean)}} \times \text{standard 2C DNA content (pg DNA)}$$

Equation 2 $$\text{Sample ploidy} = \frac{\text{2n x sample pg/nucleus}}{\text{PI 230246 pg/nucleus}}$$

Seed Size Measurement and Comparison

For seed size measurements, we randomly picked 10 seeds from each of three accessions for each calculated ploidy level (a total of 12 accessions). Tetraploid *F. ovina* cv. Quatro and hexaploid *F. brevipila* cv. Beacon were also included as references.

Seeds were spread on a digital scanner (Epson Perfection v6 flatbed scanner, Nagano, Japan) and scanned at 1200 dots per inch (dpi) with the size of 2097 x 1624 pixels. The images were processed with a custom MATLAB script (https://github.com/qiuxx221/Fine-fescue-/) that transformed the original images to measure the seed length and width respectively as the length (in pixels) of the major and minor axes of a fitted ellipse, whereas the area was calculated as the number of pixels in the seed. The seed area was calculated for each of 10 seeds for each of the 14 entries. The seed area was used to analyze the correlation between ploidy level and seed size and visualized using the ggplot2 package in R (Kahle and Wickham, 2013).

Results

DNA Content Measurement and Ploidy Estimation

Festuca ovina accessions had a high level of DNA content variation with the smallest DNA content of 3.77 pg (2C) from PI 115358, and the largest genome 19.66 pg (2C) from PI 302899 (**Table S3.2**). Flow cytometry revealed the 127 accessions represented a range of ploidy levels from diploid to octoploid (**Figure 3.1**).

The largest standard deviation within accession was 1.25 pg (PI 274619) while the lowest was 0.01 pg (W6 23622). The average of standard deviation of the 126 accessions was 0.45 pg. The majority of accessions had consistent DNA content (judged by standard deviation) within the three genotypes examined. There were cases in which the sampled genotypes had more than one ploidy level. For example, for PI 330706, three out of the four plants where tetraploid and one was triploid (**Figure 3.2**). Three out of the five genotypes examined in PI 235072 accession had three different ploidies (data not shown) and were therefore excluded from further analysis. The DNA content estimation of the 126 accessions is summarized in **Figure 3.3**.

Ploidy Estimation

To estimate ploidy, estimated DNA content was divided by the DNA content of the known diploid and round up to an integer. Ploidy estimates for the 126 accessions are summarized in **Table 3.1**. The majority of the accessions (82%) were di-, tetra-, hexa-, and octoploids while 18% of accessions (classified as “others”) were potential tri-, penta-, and septaploids. The seven accessions estimated as triploid had the estimated ploidy level between 3.23-3.47; the 14 pentaploid accessions had estimated ploidy between 4.70-5.47;

and the 2 septaploids had estimated ploidy level of 6.55 and 6.56. Ploidy distribution by country of origin is shown in **Figure 3.4**.

Besides DNA content variation, we observed leaf color variation at different ploidy levels. For all ploidy levels, there were the presence of blue-greenish color and green color (**Figure 3.5**).

Seed Size and Ploidy

We found statistical differences in seed size among ploidy levels. In general, higher ploidy level was associated with larger seed size. Diploid accessions had an average seed size between 1.5 and 2.5 mm^2, tetraploid accessions seed size varied between 2 and 4 mm^2, and hexaploid and octoploid had a similar seed size between 4-6.5 mm^2 (**Figure 3.6**).

Analysis of the main effects model suggested that the ploidy level explains over 70% of the variation in seed area and length, and more than 55% of the variation in seed length. We rejected the hypothesis that ploidy level has no significant effect on the seed size of fine fescues (*p-value*: < 2.2e-16) (**Table S3.3**, **S3.4**).

Discussion

Accessions from the USDA NPGS can serve as a valuable source of genetic diversity for crop improvement by providing an important gene pool for the improvement of both abiotic and biotic stress tolerance (Rubenstein et al., 2006). Numerous studies have been conducted to characterize PI germplasm with topics varying from soybean maturity groups (Nelson et al., 1987) to the identification of allelopathic accessions in rice (Dilday et al., 1994). These accessions have also been used to select for reproductive characteristics and seed production in garlic (Jenderek and Hannan, 2004), screening for disease resistance in dry beans (Pastor-Corrales, 2003), and the evaluation of drought tolerance in watermelon at seedling stages (Zhang et al., 2011). In forage crops and turfgrasses, tall wheatgrass was evaluated for forage yield and quality (Vogel and Moore, 1998) and Kentucky bluegrass was studied for the reproductive mode (Wieners et al., 2006). It is clear that USDA PI accessions provide a diverse gene pool for cultivar abiotic and biotic stress tolerance improvement (Rubenstein et al., 2006). The USDA accessions are particularly important for turfgrass breeding and genetics because commercial turfgrass cultivars have undergone heavy selection and lost genetic diversity.

Different from major crops, most turfgrass species have numerous ploidy levels and some are morphologically indistinguishable; therefore, germplasm characterization is very important. Ploidy determination of 200 perennial ryegrass (*Lolium perenne*) suggested that six accessions were tetraploid and 194 were diploid (Wang et al., 2009). Screening of buffelgrass [*Cenchrus ciliaris* L. syn. *Pennisetum* ciliare (L.) Link] germplasm suggested multiple ploidy levels existed in the PI collection (Burson et al.,

2012). A similar result was found in buffalograss, where multiple ploidy levels were found in the PI collections (Johnson et al., 1998).

The *F. ovina* complex includes seven species and three additional subspecies that are highly outcrossing and visually indistinguishable (Watson, 1958; Wilkinson and Stace, 1991; Stace, 2010). The 126 *F. ovina* USDA PI accessions used in our study have ploidy levels ranging from diploid to octoploid. The DNA content predicted by flow cytometry showed low average standard deviation, suggesting the methods for DNA content measurement is consistent. We measured DNA content variation between genotypes at same ploidy level, suggesting complex genome composition variation, which could be explained by post-polyploid diploidization events with differential gene loss (Mandáková and Lysak, 2018). This variation could also be the result of the change of transposon elements (Kidwell, 2002), and hybridization in the field (Leitch and Leitch, 2008; Soltis et al., 2015) or different chromosome size in different populations (Ceccarelli et al., 1992).

Although the DNA content of diploid samples was clearly separated from higher ploidies, there is no clear separation between accessions with higher ploidy levels. We also noticed ploidy level variation within some accessions. For example, of the four genotypes within the accession PI 330706 examined in this study, three were tetraploid while one was triploid. The triploid plant is likely the result of hybridization between the tetraploid plant with the pollen from some diploid relative. Another example is PI 235072, where we found three ploidy levels represented in the five genotypes we examined. It is known that *F. ovina* species can easily hybridize with relatives even at different ploidy levels (Jenkin and Jenkin, 1955). Under open pollination conditions, it is

not surprising that the seed purity is low. Our results suggest that resources should be allocated such that the relevant USDA NPGS center managing open-pollinated species can properly isolate collections during seed increase.

While most accessions we surveyed can be assigned to discrete ploidy bins, 18% suggested DNA content that fell between ploidy levels. These accessions may have originated through hybridization between different diploid parents. It is also possible that these accessions are either aneuploid or dysploid, both of which have experienced chromosome gain/loss, or rearrangements. Further evaluation using a number of different approaches, ranging from morphological classification to genotyping, will be needed to fully classify these accessions.

In our study, all 16 accessions from China were found to be diploids and a majority of plants from Iran were hexaploid. It is known that environmental and geographical factors played a role in *Festuca* ssp. genome size evolution (Ceccarelli et al., 1992; Šmarda et al., 2008). It would be interesting to see if the geographic location would be a factor to explain the distribution of ploidy levels. However, there is a lack of information on the specific area each wild accession was collected. Geographic information added to each accession in the NPGS collection would be useful; Rubenstein et al. (2006) found that NPGS users were more likely to utilize accessions when additional and accurate information was given about accessions. Beyond plant breeding, this new knowledge about the *F. ovina* collection might inspire other avenues of exploration such as investigating how geographic origin plays a role in taxon adaptation.

Seed size comparison suggests that accessions with higher ploidy levels tend to have a bigger seed size in *F. ovina* complex. Seed size often has an important impact on

germination and plant development. Bretagnolle et al. (1995) found that larger seed size is correlated with higher ploidy level in *Dactylis glomerata* L. and the larger seed size had a positive influence on robust seedling growth (Bretagnolle et al., 1995). Larger seeds contain more carbohydrates that provide the seedling vigor to help increase the competitive advantage in the natural environment (Te Beest et al., 2011). Besides observing DNA content variation, we also observed phenotypic variation on leaf color within the *F. ovina* collection. Green and bluish leaf color was observed in all ploidy levels, suggesting that leaf-color-based fine fescue identification is not reliable (Hubbard, 1968; Beard, 1972).

Conclusion

We evaluated 127 USDA PI accession and provided their DNA content and ploidy level estimation for 126 accessions. A total of 102 accessions were assigned to discrete ploidy levels, with the remaining had DNA content between discrete ploidy levels. Because of the cross-pollinating nature of *Festuca ovina* complex, better pollen control during the germplasm maintenance period could potentially reduce the chance of contaminations. Meanwhile, researchers should examine the PI collections to determine their ploidy level prior to adapting the accessions in their breeding program. This research builds the ground work for turfgrass researchers for using the *F. ovina* in their breeding program.

Acknowledgment

The authors would like to thank Dr. Ya Yang at the University of Minnesota for discussion about polyploid genome evolution. The authors would also like to thank Drs. Adrian Hegeman and Cory Hirsch at the University of Minnesota for reviewing this manuscript and providing comments and feedback. This research is funded by the National Institute of Food and Agriculture, U.S. Department of Agriculture, Specialty Crop Research Initiative under award number 2017-51181-27222.

Figures and Tables

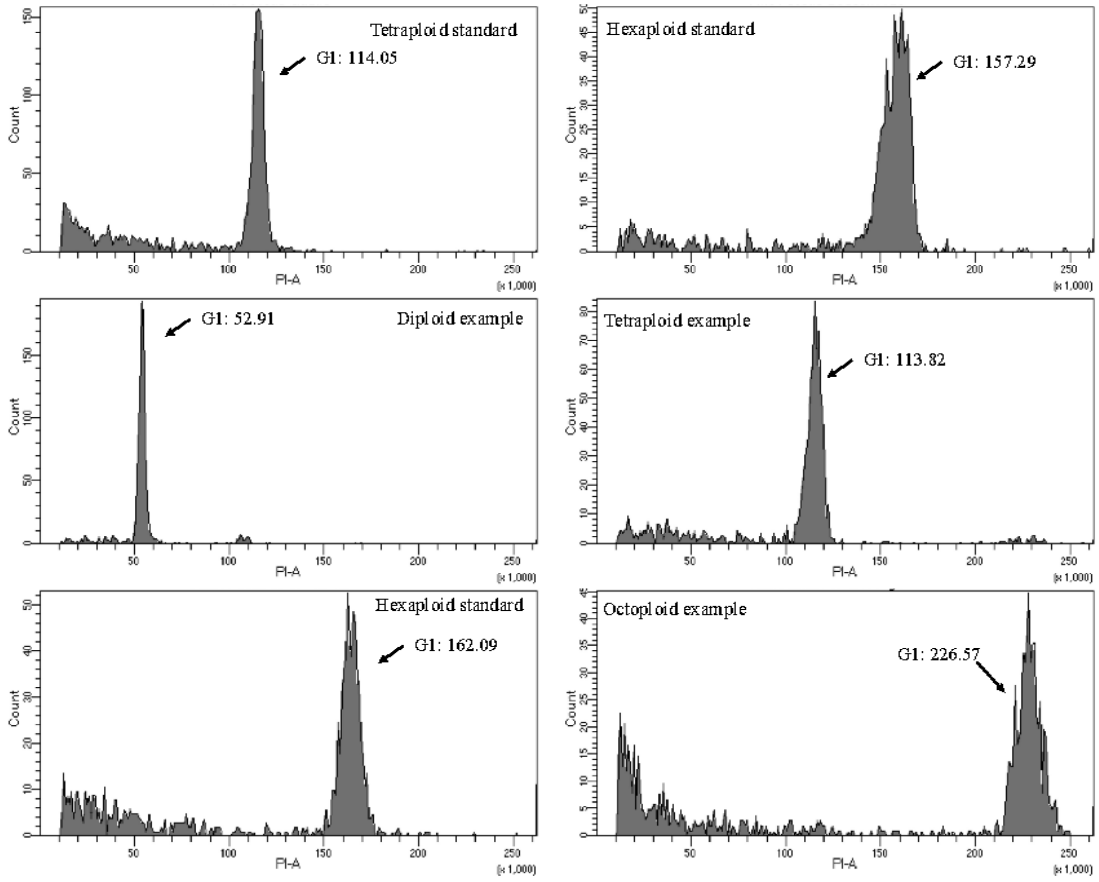

Figure 3.1. Flow cytometry data example of the PI accessions. Tetraploid sheep fescue cultivar 'Quatro' and hexaploid hard fescue cultivar 'Beacon' were included as standards. PI accessions included taxa that cover at least four ploidy levels.

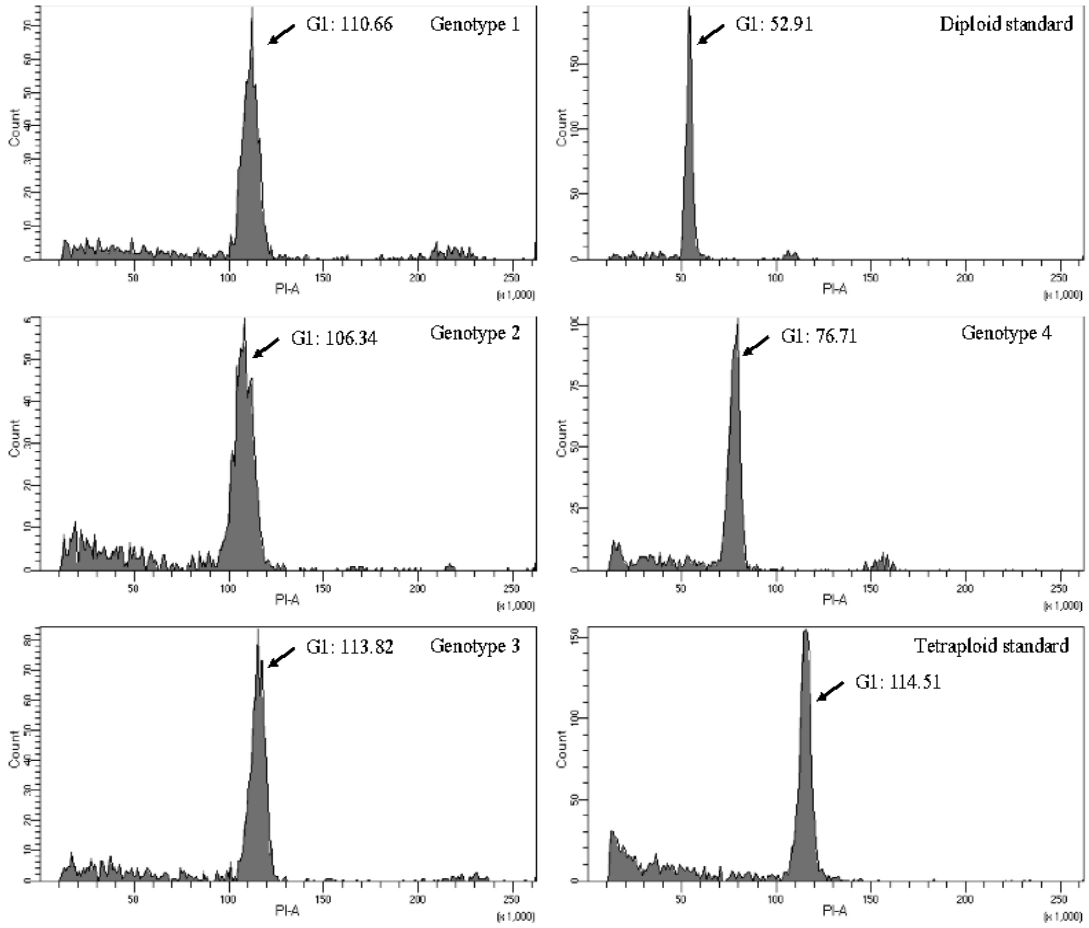

Figure 3.2. Flow cytometry histogram of PI 330706. Three genotypes from this accession had similar DNA content compared to the tetraploid standard cultivar 'Quatro'. One genotype from this accession had a DNA content between diploid and tetraploid standard and was estimated to be triploid based on the DNA content estimation.

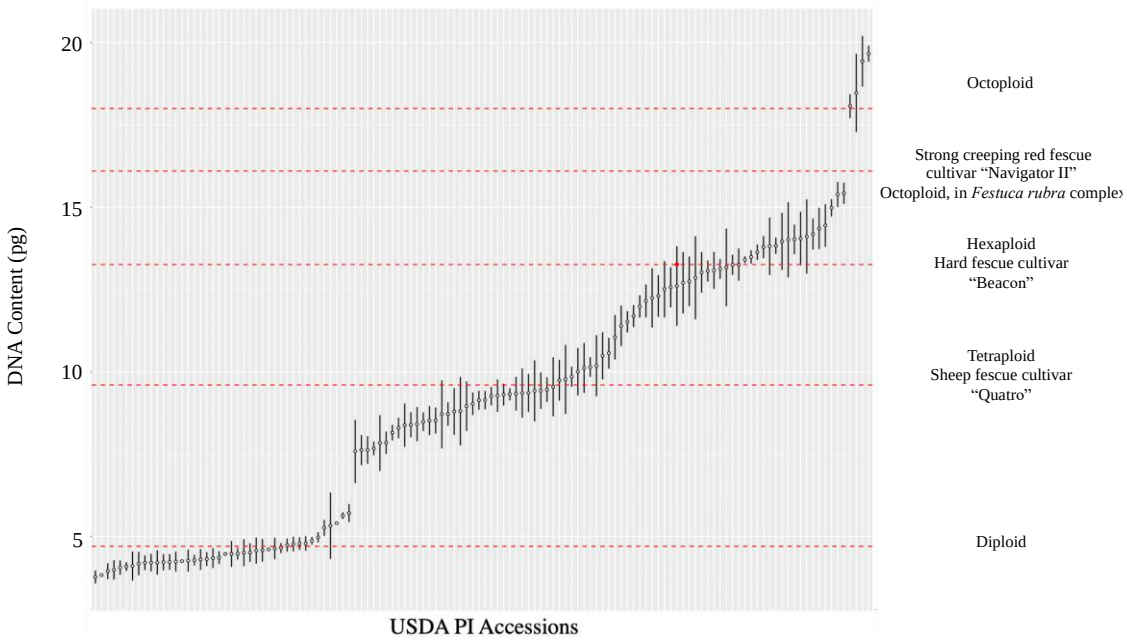

Figure 3.3. PI accessions sorted by the estimated DNA content. Dashed red lines represent the ploidy level estimation of cultivars as listed to the right of the figure.

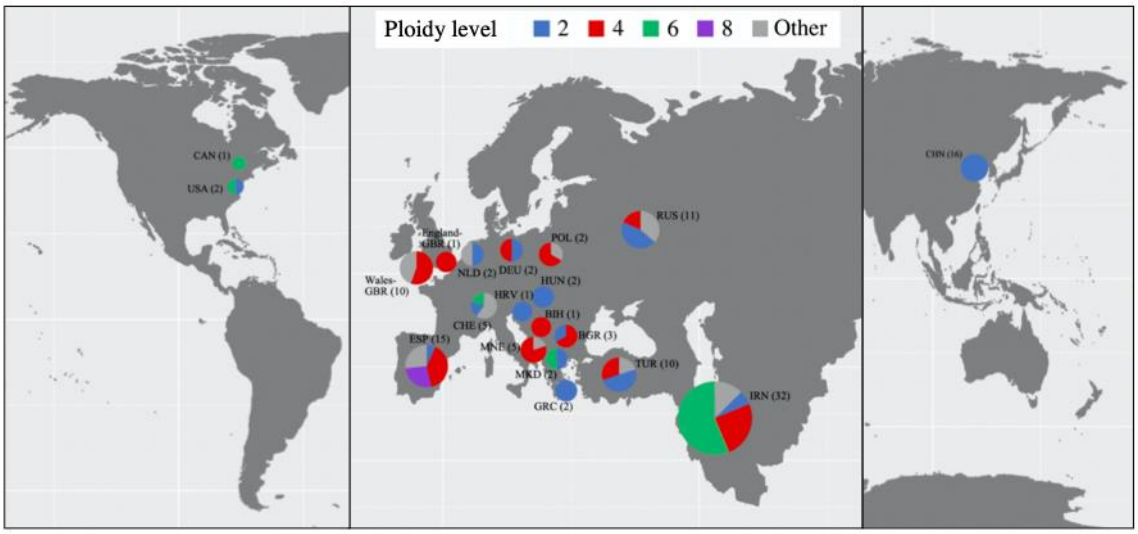

Figure 3.4. Ploidy estimation by DNA content for the 126 USDA PI accessions.

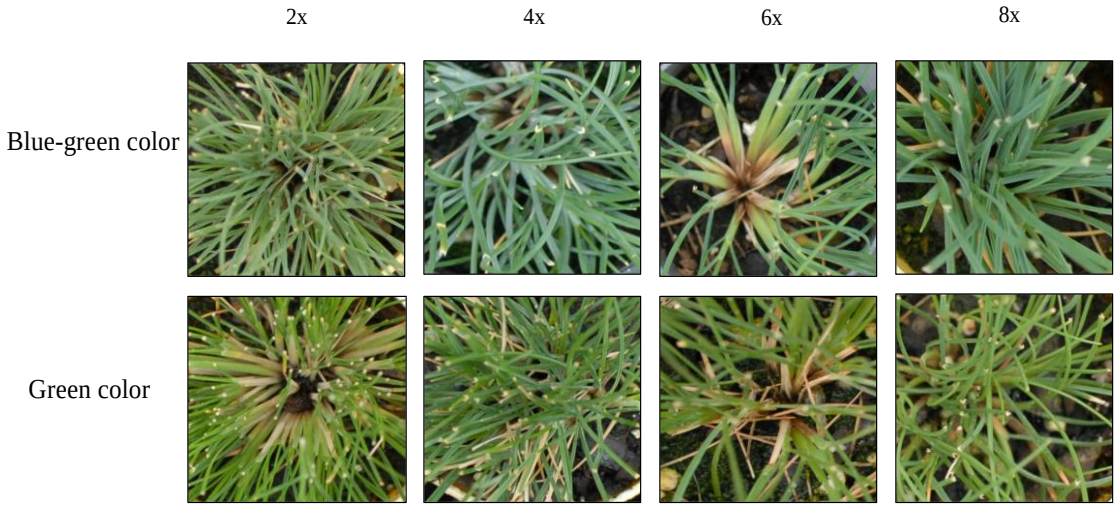

Figure 3.5. Leaf color variation. Both green and blue-greenish leaf color were observed at four ploidy levels.

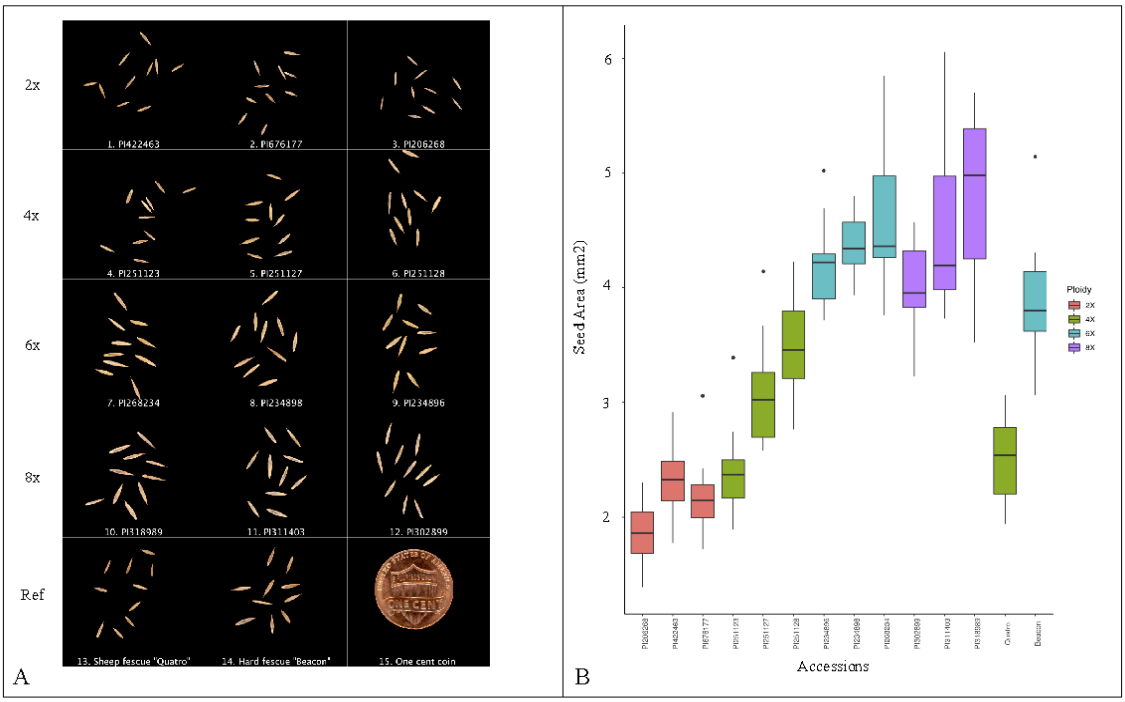

Figure 3.6. Comparison of seed area and ploidy level. The seed morphology of the examined 12 accessions and two fine fescues cultivars were shown in (A) with the ploidy of each accession labeled on the left. Box plot of seed size comparison is shown in (B); larger seed size was associated with higher ploidy level.

Table 3.1. The ploidy level estimation of the 126 USDA PI accessions.

Ploidy level	Diploid	Tetraploid	Hexaploid	Octoploid	Other Ploidy
Accession Count	42	34	22	4	24
DNA content range (pg)	3.76 ~ 5.71	8.29 ~ 10.48	13.02~ 14.98	18.07 ~ 19.66	NA

Chapter 4

Building a Reference Transcriptome for the Hexaploid Hard Fescue Turfgrass (*Festuca brevipila*) using a Combination of Iso-Seq and Illumina Sequencing

Authors: Yinjie Qiu[1], Ya Yang[2], Cory D. Hirsch[3], Eric Watkins[1]

Author Affiliations:

[1] Department of Horticultural Science, University of Minnesota, St. Paul, MN, 55108, USA,

[2] Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, 55108, USA,

[3] Department of Plant Pathology, University of Minnesota, St. Paul, MN, 55108, USA

Author Contributions: Y. Q. performed the experiments, analyzed the data, and wrote the manuscript; Y. Y. helped with phylogenetic analysis; C. D. H. helped with the genomics analysis; E. W. secured funding for this project, supervised this research, provided suggestions, and comments. All authors contributed to the revision of the manuscript and approved the final version.

This manuscript has been released on bioRxiv (Qiu et al., 2020a).

Summary

Hard fescue (*Festuca brevipila* Tracey, 2n=6x=42) is a cool season turfgrass with a fine leaf texture that performs well under low-input management. Breeding and genetics studies of *F. brevipila* have been limited due to the complexity of its hexaploid genome. To advance our knowledge of *F. brevipila* genomics, we used PacBio isoform sequencing to develop a reference transcriptome for this species. Here, we report the *F. brevipila* reference transcriptome generated from root, crown, leaf, and seed head tissues. We obtained 59,510 full-length transcripts, of which 38,595 were non-redundant full-length transcripts. The longest and shortest transcripts were 11,487 and 58 bp, respectively. To test the polyploid origin of *F. brevipila,* we sequenced three additional transcriptomes using closely related species on an Illumina platform. The results of our phylotranscriptomic analysis supported the allopolyploid origin of *F. brevipila*. Overall, the *F. brevipila* Pacbio Isoseq reference transcriptome provides the foundation for transcriptome studies and allows breeders for gene discovery in this important turfgrass species.

Keywords: hard fescue, Pacbio Isoseq, RNA sequencing, phylogenetics, turfgrass

Introduction

Festuca (*Festuca* L., Poaceae) is a large and diverse genus that includes more than 450 species (Clayton and Renvoize, 1986). Several species in this genus have been used as ground cover turfgrass: (1) broader-leaf fescues that includes tall fescue (*Festuca arundinacea*) and meadow fescue (*Festuca pratensis*) and (2) fine-leaf fescues (Wilkinson and Stace, 1991). Five fine-leaved fescue taxa are of more interest to turfgrass breeders because of their performance in shade and drought tolerance, and adaption to low fertility and acidic soils (Carroll, 1943; Hanson and Juska, 1969; Reiter et al., 2017). These five fine fescue taxa include hard fescue (*Festuca brevipila* Tracey, 2n=6x=42), sheep fescue (*Festuca ovina*, 2n=4x=28), both of which belong to *Festuca ovina* complex; and *Festuca rubra* complex taxa: strong creeping red fescue (*Festuca rubra* ssp. *rubra* 2n=8x=56), slender creeping red fescue (*Festuca rubra* ssp. *littoralis* 2n=6x=42), and Chewings fescue (*Festuca rubra* ssp. *fallax* 2n=6x=42) (Huff and Palazzo, 1998). A previous phylogenetic study using plastid genomes have resolved the relationships among these five taxa (Qiu et al., 2019). Of the five fescue taxa, *F. brevipila* is a particularly attractive target for breeding and genetics improvement due to its excellent performance in low input environments (Ruemmele et al., 1995).

Modern breeding programs in model crop species rely heavily on the use of genomic and molecular biology tools, such as marker-assisted breeding and CRISPR-Cas 9 based genome editing to improve consumer and field production traits (Varshney et al., 2005; Gupta et al., 2010; Shan et al., 2013; Phing Lau et al., 2016). To apply these tools in turfgrass breeding, there is a need to build genomic and transcriptome resources. However, adapting molecular breeding techniques and genomic studies into *F. brevipila*

has been challenging because of the lack of genetic information. As of September 2019, there were 174 nucleotide sequences for *F. brevipila* on National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov), while perennial ryegrass (*Lolium perenne*), the turfgrass species for which the most genomics work has been done, has a reference genome available, 45,149 Sequence Read Archive (SRA) accessions, and more than 600,000 nucleotide sequences on GenBank.

RNA sequencing is a powerful tool to generate genomic data for marker development and enable researchers to study the dynamics of gene expression of plants under stress. As the cost of Illumina sequencing continues to drop, more and more researchers are using this approach for gene discovery and gene identification. Without a reference genome, researchers use *de novo* assembled short reads to generate a reference to conduct gene expression study (Grabherr et al., 2011). However, this approach is error-prone in polyploids with highly similar subgenomes (Chen et al., 2019), making it difficult to obtain full-length transcripts or identify alternative splicing (Alkan et al., 2011; Ozsolak and Milos, 2011).

PacBio sequencing, also known as long-read sequencing, has average read length over 10 kb (Rhoads and Au, 2015) and is particularly suitable for generating reference transcriptomes in polyploid species. With the capability to obtain full-length RNA sequences (Isoseq), PacBio Isoseq has been used for the discovery of new genes, isoforms, and alternative splicing events; this technology is improving genome annotation in crop species such as maize (*Zea mays*), rice (*Oryza sativa*), sugarcane (*Saccharum officinarum*), and sorghum (*Sorghum bicolor*) (Wang et al., 2016; Hoang et al., 2017; McCormick et al., 2018; Zhang et al., 2019). In turfgrass and ground cover

genetics research, PacBio Isoseq has been used to study the transcriptome and phylogeny of bermudagrass (*Cynodon dactylon*) and red clover (*Trifolium pratense*) (Chao et al., 2018; Zhang et al., 2018). Due to the highly outcrossing and polyploid nature of *F. brevipila*, PacBio isoform sequencing is essential for generating a reference transcriptome useful for plant breeding and genetics research.

For plant breeding and germplasm improvement, USDA PI accession have proven to be a valuable resources (Pastor-Corrales, 2003; Chang and Hartman, 2017). Our previous work showed out of 127 USDA accessions of the closely related *Festuca ovina*, only 17.3% were to be hexaploid, while 33.1% were diploid and 26.8% were tetraploid (Qiu et al., 2020a). For this reason, it is important to understand the subgenomes so breeders could potentially create synthetic hexaploids using species with lower ploidy levels for traits introgression. Previous GC content, average chromosome size, and genome size evidence suggested the presence of allo/auto-polyploid in *Festuca* ssp. (Šmarda et al., 2008). However, so far there lacks nuclear sequence-based evidence to test the polyploid origin of *F. brevipila*. Previous study using the amplified fragment length polymorphism (AFLP) markers followed by neighbor-joining analysis suggested that *F. valesiaca* (2n=2x=14) is closely related to *F. brevipila* (Ma et al., 2014). Through sequencing the chloroplast genomes our previous work suggested *F. ovina* ssp. *ovina* (2n=2x=14) and *F. ovina* (2n=4x=28) were both closely related to *F. brevipila* (Qiu et al., 2019). Transcriptome sequencing followed by phylogenetic analyses has been previously used to resolve allopolyploid origin of strawberry (Edger et al., 2019). Therefore here we adopt a similar phylotranscriptomic approach to test the allopolyploid origin of *F. brevipila* and investigate the relationship among its close relatives.

In this study, we used the Pacbio Sequel II platform to conduct isoform sequencing to develop a reference transcriptome for the hexaploid *F. brevipila*. We then evaluated and annotated the reference transcriptome using publicly available protein databases. We also used Illumina HiSeq 4000 RNA sequencing on leaf samples of *F. brevipila*, and closely related taxa *F. ovina* ssp. *ovina* (two accessions, one diploid and one tetraploid) and *F. valesiaca* to test the allopolyploid origin of *F. brevipila*.

Material and Methods

Plant Materials

Festuca brevipila breeding material SPHD15-3 from the University of Minnesota breeding program was used for this study (plant materials available upon request). A clonal population was vegetatively propagated and grown using the BRK Promix soil (Premier Tech, USA) in the Plant Growth Facility at the University of Minnesota, St. Paul campus under 14 hours daylight and 8 hours darkness with bi-daily irrigation and weekly fertilization (906 grams of Ammonium sulfate, 950 grams of Peat-lite 20-10-20 and 38 grams of Sprint 330 mixed into 5 gallons of water with a Dositron fertilizer injector set at a 1:100 ratio).

To capture a broad representation of the *F. brevipila* transcriptome using PacBio Isoform sequencing, we collected root, crown, leaf, and inflorescence tissues. Plants were vernalized in a 4℃ cold room for 3 months with 8 hours daylight and 16 hours darkness. The plants were then moved back to greenhouse conditions as described above with day time temperature of 25℃ to recover for one month and allow inflorescence development. The root, crown, leaf, and inflorescence tissues were harvested and flash frozen for RNA extraction (**Figure 4.1**).

The plant materials used in this study were identified following guidelines described by (Wilkinson and Stace, 1991) using vein number, leaf morphology, and flow cytometry (Qiu et al., 2019). Plant materials were in the process of inducing flowering and we will deposit a voucher, tentative voucher number Qiu 1 (MIN), into the University of Minnesota herbarium. Dr. Ya Yang, co-author and Curator of the University of Minnesota Herbarium, will facilitate the process.

To perform phylogenetic study on *F. brevipila* and closely related taxa, leaf samples of USDA PI 676177, identified as *F. ovina* ssp. *ovina*; PI 422463, identified as *F. valesiaca*; *F. ovina* cv. Quatro (obtained from National Turfgrass Evaluation Program, http://www.ntep.org) and *F. brevipila* were grown in the greenhouse with the same conditions described above and leaf samples were harvested and flash frozen for RNA extraction.

RNA Extraction and Sequencing

To maintain RNA integrity, RNA extractions for Isoseq were done using a Quick-RNA Miniprep Kit (ZYMO Research, Catalog number R1055) in a cold room following the manufacturer's instructions. RNA extraction for Illumina sequencing was performed at room temperature using the same method. A Qubit Fluorometer (ThermoFisher Scientific) was used for initial RNA quantification, before an Agilent 2100 Bioanalyzer (Agilent) was used to assess the RNA integrity. Only samples with a RIN >8.0 were used for sequencing library construction.

For PacBio Isoform sequencing, four sequencing libraries (one for each tissue type) were constructed by NovoGene (China), using the SMARTer PCR cDNA Synthesis Kit (ClonTech, Takara Bio Inc., Shiga, Japan) by tissue type with no size selection Sequencing was performed on a PacBio Sequel II (PacBio, CA) instrument. Illumina sequencing libraries were constructed using a TruSeq® Stranded mRNA Library Prep kit (Illumina) and sequenced on an Illumina HiSeq 4000 instrument in 150 bp paired-end sequencing mode. All sequencing data generated from this study have been deposited in NCBI SRA under bioProject PRJNA598357.

Illumina Sequencing Data Processing and *de novo* Assembly

Illumina adaptor sequences were trimmed using Trimmomatic with the default settings (v 0.32) (Bolger et al., 2014). Quality trimming was performed using the seqtk tool with -q 0.01 (https://github.com/lh3/seqtk). Trinity (v 2.4.0) was used to assemble transcriptomes of *F. ovina* subsp *ovina*, *F. valeciaca*, *F. ovina* cv. Quatro, and *F. brevipila* (Grabherr et al., 2011) with max memory 62G, 24 CPUs, bflyCalculateCPU, and minimum contig size of 200 bp.

PacBio isoform sequencing (Isoseq) Data Processing

The software SMRTlink (v 2.3.0, PacBio) was used to filter and process original sequencing files, with minLength 0, minReadScore 0.8 to produce the subreads file. To identify full-length transcripts, the Isoseq 3 (v 3.1.0, PacificBiosciences) pipeline was installed locally using conda (Anaconda-2.3.0) under bioconda following instruction by PacBio (https://github.com/PacificBiosciences/Isoseq_SA3nUP). Circular consensus sequences (ccs) bam files were generated using ccs command with --noPolish, --maxPoaCoverage 10, and 1 minimum passes options. To classify full-length (FL) reads, we identified and trimmed the 5' and 3' cDNA primers (primer sequences can be found in **Table S4. 1**) in the ccs using lima with --dump-clips, --no-pbi, --peak-guess option. Poly(A) tails in FL reads that were at least 20 bp long were removed with --require-polya. Bam files produced from the previous steps for four tissues types were merged into one dataset; the source files were also merged into a subreadset.xml file. Clustered reads were polished using the Isoseq3 polish function to produce high-quality FL transcripts (expected accuracy ≥ 99% or QV ≥ 30).

To correct the PacBio Isoseq FL transcripts, high quality Illumina paired-end reads from *F. brevipila* leaf tissue were used in the LoRDEC -correct function in LoRDEC program (v 0.6) with configuration *k-mer* size = 19 and 21, solid-*k-mer* = 3, error rate 0.4, and maximum branch time 200 (Salmela and Rivals, 2014).

Removing Sequence Contamination and Collapsing Redundant Sequences

To remove microbial sequence contamination in both Isoseq and Illumina *de novo* transcriptomes, bacterial and virus genomes were downloaded from NCBI ftp://ftp.ncbi.nih.gov/genomes/archive/old_refseq/Bacteria/all.fna.tar.gz, ftp://ftp.ncbi.nih.gov/genomes/Viruses/all.fna.tar.gz, a mapping index was built using bwa64 (v 0.5.9-r16) included in the deconseq program (v 0.4.3) (Schmieder and Edwards, 2011). Contaminant transcripts were removed using the deconseq.pl script in the deconseq program before performing downstream analysis. To merge and remove redundant transcripts in each transcriptome 'cd-hit-est' from the CD-HIT (v 4.8.1) package was used with the following parameters: -c 0.99 -G 0 -aL 0.00 -aS 0.99 -AS 30 -M 0 -d 0 -p 1 (Isoseq transcriptome); -c 0.98 -p 1 -T 10 (*de novo* assembly) (Li and Godzik, 2006).

Transcriptome Completeness Analysis

To evaluate the completeness of the assembled *F. brevipila* transcriptomes, we used 1335 core embryophyte genes (embryophyta_odb10) from Benchmarking Universal Single-Copy Orthologs (BUSCO v3) (Simão et al., 2015). Besides individual evaluation of Isoseq and *de novo* transcriptomes, we also combined non-redundant full length (NR-FL) transcripts and non-redundant (NR) *de novo* transcriptomes and used cd-hit-est to cluster redundant sequences with 95% sequence identity (Li and Godzik, 2006). The

completeness of the combined assembly was also evaluated using the same BUSCO method described above. To evaluate how combined assembly could improve short read mapping, we mapped RNA sequencing reads used for *de novo* assembly to the two individual and the combined assembly using bowtie2 program (Langmead and Salzberg, 2012). Mapping statistics was summarized using the samstat tool (v 1.5.1) (Lassmann et al., 2010).

Coding Genome Reconstruction

To reconstruct the coding genome of *F. brevipila*, non-redundant full length transcripts were processed using the COding GENome reconstruction Tool (Cogent, v 3.3, https://github.com/Magdoll/Cogent) with minimap (v 2.17-r941, https://github.com/lh3/minimap2), Mash (v 1.0.2, https://github.com/marbl/Mash/releases), and BioPython (v. 1.71). Briefly, *k-mer* similarity was used to partition the non-redundant Pacbio Isoseq transcriptome into gene families. Undirected weight graphs were constructed based on the percentage of matching *k-mer* sketches. Python scripts run_preCluster.py, generate_batch_cmd_for_Cogent_family_finding.py (https://github.com/Magdoll/Cogent) with *k-mer* size 30, and reconstruct_contig.py were used to reconstruct the coding genome.

Transcriptome Functional Annotation

The NR-FL transcriptome was searched against NCBI NR protein and SwisProt, UniProt protein database using diamond BLASTx (v 0.9.13) with e-value < 1e-5. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using the KEGG Automatic Annotation Server (KASS,

https://www.genome.jp/kegg/kaas/). *Oryza sativa*, *Brassica napus*, *Zea mays*, and *Arabidopsis thaliana* were set as references with a single-directional best hit model. Gene Ontology (GO) terms were produced by the interproscan program (Quevillon et al., 2005).

Identification of lncRNAs and miRNA from PacBio Sequences

The prediction of long non-coding RNAs (lncRNAs) was performed using an improved *k-mer* scheme (PLEK) tool (v 1.2) with a minimum sequence length of 200 bp (Li et al., 2014). Putative miRNA precursors were identified by BLASTn of NR-FL sequences against the plant miRNAs database (http://www.mirbase.org) that includes both hairpin and mature miRNA with e-value ≤ 10^{-5} (Kozomara and Griffiths-Jones, 2013). For each sequence the BLASTn hit with the highest bit score was kept.

Phylotranscriptomic Analyses of *Festuca-Lolium* Taxa

To test the allopolyploid origin of *F. brevipila*, we also generated the transcriptome for three additional taxa (*F. ovina* ssp. *ovina*, *F. valeciaca*, *F. ovina*) in this study using *de novo* assembly method described above. In addition, we included four publicly available transcriptomes from closely related species selected according to previous phylogenetic analyses based on chloroplast sequences (**Table 4.1**) (Hand et al., 2013; Qiu et al., 2019). Protein and coding sequences of all *Festuca-Lolium* taxa were predicted using the TransDecoder program (https://github.com/TransDecoder) (Wu and Watanabe, 2005). *Brachypodium distachyon* was included as the outgroup.

Orthology inference was carried out using protein sequences of all nine taxa in OrthoFinder 2 (Emms and Kelly, 2019). Single-copy protein sequences were aligned using the MUSCLE program (v 3.8.31) with the parameters -maxiters 16 -diags -sv

(Edgar, 2004). To improve the phylogenetic resolution among closely related taxa, we used the PAL2NAL program (v 14) to align coding sequences according to the protein alignment (Suyama et al., 2006). The resulting coding sequence alignments were trimmed using TrimAl with the parameters -gt 0.9 -cons 60, which removes all positions in the alignment with gaps in 10% or more of the sequences and keeps at least 60% of the sequence alignment (Capella-Gutiérrez et al., 2009).

To trace the maternal relationships of the focal taxa, chloroplast gene sequences for *F. ovina* ssp. *ovina*, *F. valeciaca*, and *F. brevipila* were extracted from *de novo* assembled transcriptomes by aligning assembled contigs to the *F. ovina* reference chloroplast genome (MN309824) using Burrows-Wheeler Aligner (Li and Durbin, 2009). Sequences for the other five taxa were obtained from NCBI by BLASTn using *F. ovina* reference chloroplast genome sequences as the query. *Brachypodium* chloroplast sequence was used as the outgroup (**Table 4.1**). Three fragments of chloroplast genome sequence that covered *rps2-atpA*, *rpl32 -ccsA*, and *clpP – psbH* region had the highest completeness and taxon representation and were aligned using MAFFT (v 7) and concatenated (Katoh and Standley, 2013).

Phylogenetic trees were constructed using the RAxML program (v 8.2.12) for both the concatenated nuclear and chloroplast alignments under the GTR+GAMMA model with 1,000 bootstrap (Stamatakis, 2006). The phylogenetic trees were visualized using FigTree v 1.4.3 (Rambaut, 2012). Because the chloroplast sequences and transcriptomes for *L. perenne, L. multiflorum, F. pratensis, F. arundinacea,* and *B. distachyon* were downloaded from NCBI/EBI, they might not come from the same

genotypes; for these taxa, we only present the nuclear and chloroplast phylogenies without carrying out conflict analyses.

Next, we further explored phylogenetic conflict between chloroplast and nuclear gene trees conflicts among *F. ovina* ssp. *ovina*, *F. ovina,* and *F. valeciaca* using transcriptomes generated in this study and the reference transcriptome for *F. brevipila*. *Lolium perenne* and *B. distachyon* were included as outgroups. We carried out a new OrthoFinder run with a reduced set of taxa and we used single-copy genes predicted from the new run. The coding sequence alignment for each single-copy gene was generated using the methods described above. Gene trees were estimated using the RAxML program (v 8.2.12) under GTR+GAMMA model with 1000 bootstrap. A species tree was estimated using ASTRAL v 5.6.3 using gene trees from previous step (Mirarab et al., 2014). The concordance and conflicts among gene trees and were visualized using Phypart (Smith et al., 2015) by mapping single-copy gene trees rooted by *Brachypodium distachyon* to the rooted species tree from ASTRAL. Tree concordance and conflicts were plotted using the python script phypartspiecharts.py available at https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts (last accessed on Feb 2, 2020).

Results

Festuca brevipila Transcriptome Assembly

Using Pacbio Isoseq, a total of 76.3 GB of sequence was generated from four SMRT cells, 73.91 Gb (48,788,662) of which were subreads. The Isoseq 3 pipeline identified 60,719 putative high-quality full length (FL) transcripts and 487 low-quality FL transcripts. After removing microbial contamination and plastid genome sequences, 59,510 FL transcripts were retained (98.01%) with a N50 of 2,668 bp, a total length of 130,657,908 bp. For Illumina sequencing on *F. brevipila* leaf tissue, a total of 37,100,522 paired-end reads were generated (SRR10995913). The reads were used for PacBio reads error correction. Finally, we generated the *F. brevipila* transcriptome which included 59,510 FL transcripts with a N50 of 2,667 bp and a total length of 130,631,252 bp. After removing redundant sequences, we obtained 38,556 (64.78%) unique transcripts with a total length of 82,782,352 bp and a N50 of 2,584 bp. The longest and shortest transcripts were 11,487 and 58 bp, respectively. The GC% content of the transcriptome was 51.5%, with the majority of transcripts size above 1,000 bp (83.3%) (**Figure 4.2**).

De novo assembly of llumina reads recovered 325,781 assembled contigs. After contamination and redundancy removal, 302,471 contigs were retained with a N50 of 1,157 bp. Because of the complexity of the hexaploid genome, a short and long read combined assembly was generated that included 248,318 transcripts and an N50 of 1,405 bp. A total of 20,063 transcripts were included from the Isoseq transcriptome by cd-hit-est (**Table 4.2**).

When evaluating the completeness of the assemblies, the Illumina *de novo* assembly showed better coverage (94.6%, fragmented BUSCO included) than the PacBio

Isoseq transcriptome (71.85%, fragmented BUSCO included). The combined assembly outperformed the individual assemblies for better coverage (96.07%, fragmented BUSCO included) and completeness (89.24%) (**Figure 4.3**).

Around 60% of the fragmented and 35% of the missing BUSCO genes from the Illumina-only assembly were complete in the hybrid assembly (**Figure 4.4A**). Similarly, 54% of the fragmented and 78% of the missing BUSCO genes from the PacBio assembly were marked as completed in the hybrid assembly (**Figure 4.4B**).

When mapping Illumina reads from the *F. brevipila* leaf tissue back to the three assemblies, the percent of unmapped reads for PacBio Isoseq, Illumina, and combined assembly was 34.5%, 38.2%, and 28.5%, respectively.

Coding Genome Reconstruction

In the absence of a reference genome, we reconstructed the coding genome for detecting potential splicing events in future studies. After *k-mer* clustering and minimizing edges of graphs assuming each gene family had the similar k-mer pattern, the Cogent program reconstructed 6,043 gene families (**Figure 4.5**). Of these, 4,384 gene families had one path, 1,285 had 2 paths, and one had the most paths of 19. A total of 153 transcripts were absent from the final coding genome.

Transcript Functional Annotation

We carried out functional annotation of the PacBio Isoseq transcriptome using multiple databases including NCBI NR protein, UniRef90, SwisProt, KEGG, and Pfam (GO terms) (**Figure 4.6**). Around 25% transcripts had annotation in all five databases and 36,067 (93.54%) transcripts had annotation from more than one database.

Identification of lncRNAs and miRNA from PacBio Sequences

In the NR PacBio Isoseq transcriptome, we identified 7,868 transcripts that were long non-coding RNA. A total of 145 transcripts were identified as hairpin miRNA, and non-plant hits were removed from further analysis. A total of 39.04% of the miRNA was annotated in *Hordeum vulgare*, 19.18% was annotated in *O. sativa*, and 17.8% was annotated in *F. arundinacea* (**Table S4.2**). The most copies found was hvu-MIR6179 with 16 copies, followed by far-MIR1119 and osa-MIR2927, both of which had 11 copies (**Figure 4.7**).

Phylotranscriptomic Analyses of the *Festuca-Lolium* Complex

A total of 86,445,109 paired end reads were generated from the Illumina sequencing for *F. ovina* ssp. *ovina*, *F. valeciaca*, and *F. ovina*. After removing microbes contaminates, the assembly statistics was summarized in **Table 4.3**.

OrthoFinder identified 29 single copy nuclear genes across the 9 taxa to reconstruct the nuclear gene trees and three chloroplast genome regions that included 9,680 bp sequence which covered the *rps2-atpA*, *rpl32 -ccsA*, and *clpP – psbH* regions. They were likely to be DNA contamination in the RNA sequencing library, because these contigs spanned gene and intergenic regions. The topology between concatenated nuclear and plastome datasets were mostly congruent, except the placement of *F. brevipila* and *F. valesiaca* (**Figure 4.8**).

To explore the conflict among gene trees within the *F. ovina* complex, we assembled a second phylotranscriptomic dataset with all four taxa in the complex and two outgroups. OrthoFinder recovered 286 single copy genes from this six-taxon dataset. The resulting species tree topology by ASTRAL was congruent with the concatenated tree

recovered from the nine-taxon nuclear dataset (**Figure 4.9**). However, significant proportion of informative (bootstrap >50) gene trees supported alternative placements of *F. brevipila* (red and green in **Figure 4.9**) being either sister to *F. valesiaca*, *F. ovina* 2x, or *F. ovina* ssp. *ovina* 4x.

Discussion

In this study, we used the PacBio isoform sequencing platform to generate a reference transcriptome of *F. brevipila* using root, leaf, crown, and seed head tissue types. We used multiple protein and pathway databases to annotate the transcriptome and reconstruct the coding genome in the absence of a reference genome. We used Illumina sequencing data from *F. brevipila* leaf tissue to generate a *de novo* assembly and a combined assembly to improve the coverage and of the transcriptome. In addition, we carried out phylotranscriptomic analyses to test the allopolyploid origin of *F. brevipila*.

We took advantage of both short and long reads sequencing platform to develop an improved *F. brevipila* reference transcriptome. Comparing our PacBio Isoseq, Illumina *de novo*, and the combined assembly, the PacBio Isoseq transcriptome had the best N50 of 2,584 bp while *de novo* assembly only had N50 of 1,157 bp. The combined contigs had the highest BUSCO completes (89.2%), reduced fragmented BUSCO (6.8%), and the lowest missing BUSCO (4%). This suggested the combined contigs improved the overall completeness of the assembly. In addition, when mapping Illumina sequencing reads to the three assemblies, the combined assembly has the least unmapped reads, followed by Pacbio Isoseq, and *de novo* assembly. The PacBio assembly has the best mapping quality (42.9% MAPQ >=30), followed by the *de novo,* and then the hybrid assembly. The combined assembly had the highest number of reads with MAPQ <3 (Mapping to multiple region of the transcriptome), which suggested that during the process to combine multiple assemblies using cd-hit-est, the sequence identity threshold played a crucial role. In our case, we used a 95% sequencing identity cut off. By using a lower cut off, we would likely increase the unique mapping (MAPQ>=30); however, we

would potentially lose homolog or paralog information due to the hexaploidy nature of the taxon. Similarly, if we used a more stringent cutoff, we would be likely to maintain more sequence information but lose more unique reads. This is an inherent problem in polyploid transcriptomics and future whole genome sequencing will help researchers to overcome these limitations.

Although PacBio Isoseq has proven to improve the transcriptome quality, it has several limitations. First of all, it has a higher cost and lower sequencing coverage compared to the Illumina dataset. In our case, based on the BUSCO evaluation, 83 out of the 988 present BUSCO genes are fragmented. When comparing the BUSCO assessment result, the PacBio Isoseq assembly had 28.8% BUSCO genes missing while the short read *de novo* assembly had only 5.4% missing even though the data came from single tissue. In addition, PacBio sequencing has a higher error rate. However, all these limitation could be reduced by increasing the number of SMRT cell for sequencing, improvement of sequencing chemistry, enzyme (Hi-Fi), and a combination of using Illumina short reads sequencing and long reads (Rhoads and Au, 2015; Wenger et al., 2019).

When a genome is large and complex, obtaining a well-annotated genome is challenging. Pacbio Isoseq allowed us to reconstruct the transcribed regions of the genome (coding genome) using isoform transcripts to simplify the process. When constructing the *F. brevipila* coding genome using k-mer based gene family clustering approach, 153 transcripts failed to cluster into gene families. This is likely due to the undetected cycles in the graph. Using different k-mer lengths could potentially resolve the problem, although we also tried k-mers of 40 and 50 and still failed to reconstruction

the transcripts. In our dataset, the majority of genes families had one (72.54%) or two (21.26%) paths. Having multiple paths is likely due to a single path, which cannot resolve the variation from paralogs, exon skipping, and sequencing artifacts. The coding genome generated in this study will be used in future *F. brevipila* genomics research to predict and visualize RNA alternative splicing events.

In addition to the reconstruction of coding genome, which enables researchers to look for RNA alternative splicing events, we also identified the potential microRNA precursors (miRNA). When searching miRNA precursor transcripts in our dataset, we found MIR6179, MIR1119, and MIR2927 had the highest presence in the four tissue types sampled. Previous studies have shown that MIR1119 is essential in regulating plant drought tolerance in wheat and dehydration stress in barley (Kantar et al., 2010; Shi et al., 2018). MIR2927 has shown to regulate abiotic (salt) and biotic (virus) stress in rice (Sanan-Mishra et al., 2009). The function of MIR6179 was found to associate with gametocidal action in wheat (Wang et al., 2018). It is possible that the high abundancy of those miRNAs reflect the importance of the pathways being targeted in *F. brevipila*.

Finally, to test the allopolyploid origin of *F. brevipila*, we used the transcriptome data to construct the phylogenetic relationship among closely related taxa in the *Festuca-Lolium* complex. When comparing the phylogeny from maternal (chloroplast) vs. bi-parentally inherited (nuclear) genes, we noticed a conflict at the placement of *F. brevipila* (6x) and *F. valeciaca* (2x; **Figure 4.8**). Among nuclear gene trees (**Figure 4.9**), 40.25% and 40.85% of informative gene trees supported the nuclear species tree topology, with no single dominant alternative topology. Given the short internal branch length among *F. brevipila* (6x), *F. valeciaca* and *F. ovina* (2x and 4x), hybridization, incomplete lineage

sorting, and assembly artifact may contribute to the conflict among the nuclear genes. However, when we plotted the synonymous substitutions per site (*Ks*), we did not observe a clear peak in any of our ingroup species (**Figure S4.1**), indicating that the polyploidy event happened relatively recently. Together our data had no clear support for or against an allo- vs. autopolyploidy origin of *F. brevipila.* Previous studies suggest that *Festuca* taxa can hybridize with relatives easily (Jenkin and Jenkin, 1955). It is possible that *F. brevipila*, *F. ovina*, *F. ovina* ssp. *ovina* share some subgenomes. Ideally, we could use ortholog group information to further investigate this event, however, due to the lack of a polyploid reference genome, it is challenging to tease apart homeologs in our current dataset. Future studies could be focused on whole genome sequencing and screening additional taxa that are closely related to *F. brevipila* to identify the potential subgenome of this important hexaploid turfgrass taxon.

Conclusion

In this study, we developed the reference transcriptome of *F. brevipila* using PacBio isoform sequencing and Illumina sequencing. We also carried out phylotranscriptomic analyses to test the allopolyploid origin of *F. brevipila*. However, the results were unable to distinguish auto- vs. allopolyploid due to the recent polyploidy events in *F. brevipila*. This dataset provided insight about a complicated hexaploid transcriptome and laid the foundation for future genomics research on this important turfgrass taxon.

Funding: This research is funded by the National Institute of Food and Agriculture, U.S. Department of Agriculture, Specialty Crop Research Initiative under award number 2017-51181-27222

Acknowledgments: The authors would like to thank Minnesota Supercomputing Institute for the high-performance computing clusters.

Conflicts of Interest: The authors declare no conflict of interest.

Figures and Tables

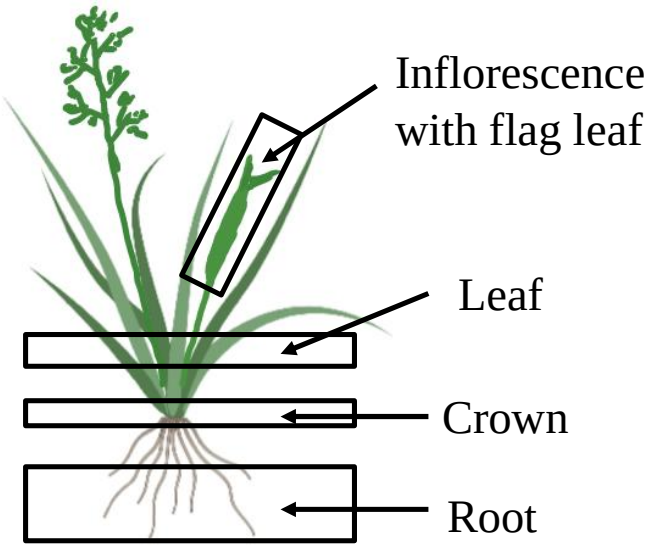

Figure 4.1. Plant tissues used for PacBio Isoform and RNA sequencing.

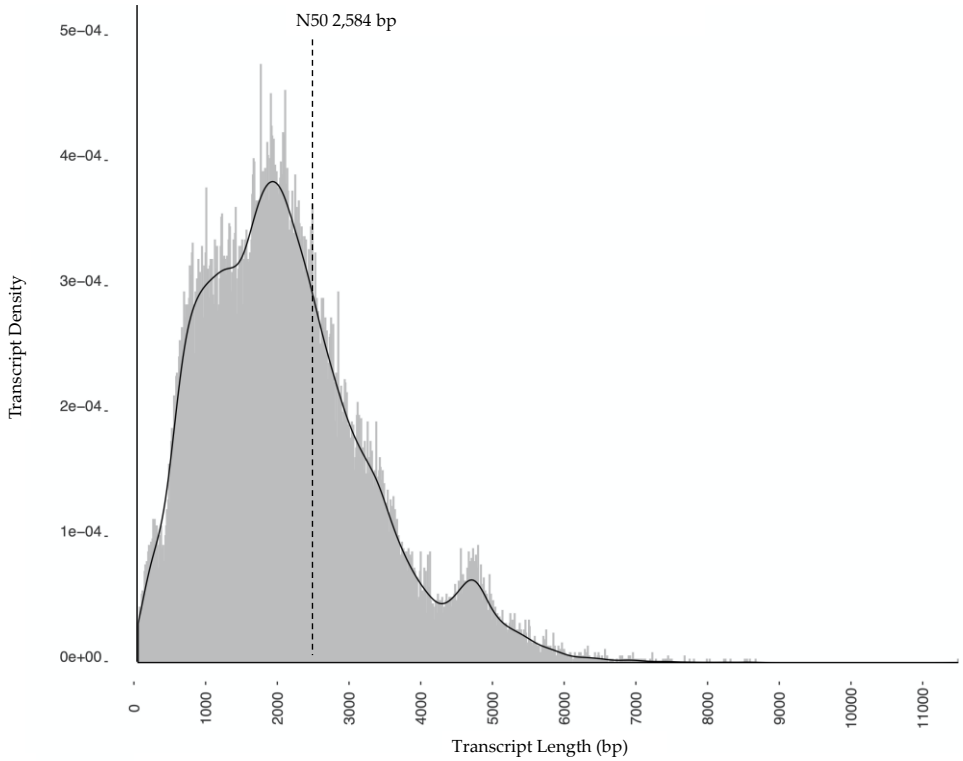

Figure 4.2. Transcript length distribution after redundant sequence removal.

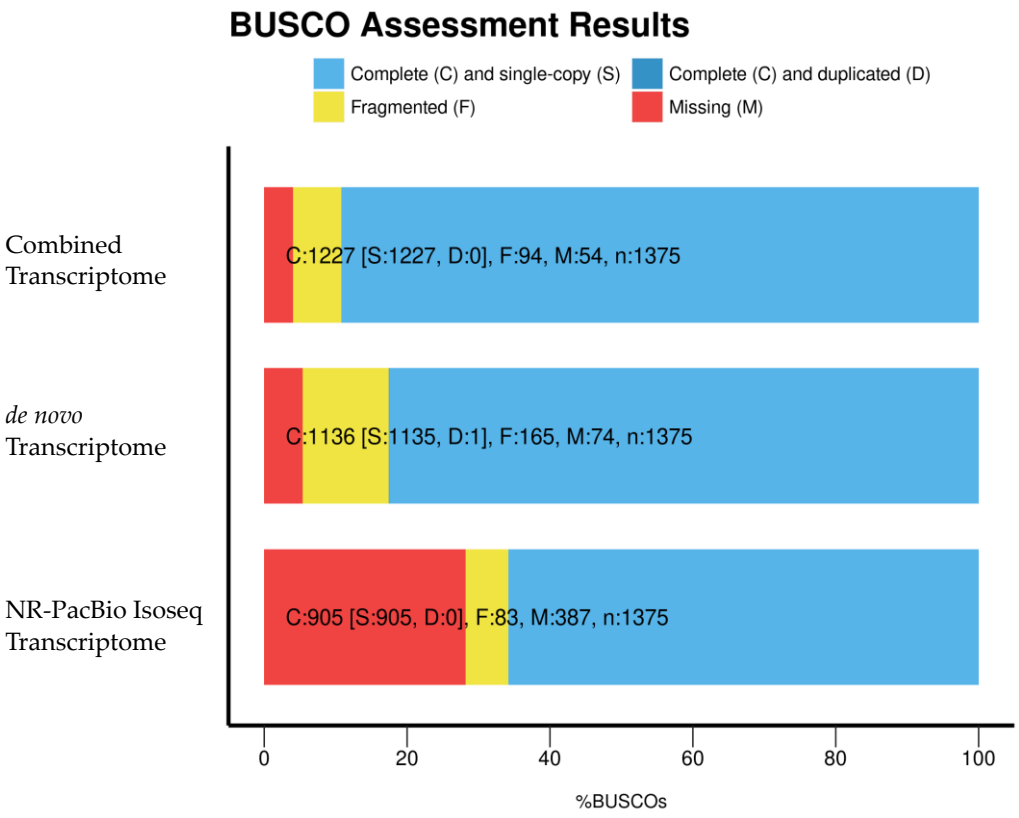

Figure 4.3. BUSCO assessment of the three assemblies. The combined assembly had the most complete BUSCO genes and the least missing genes.

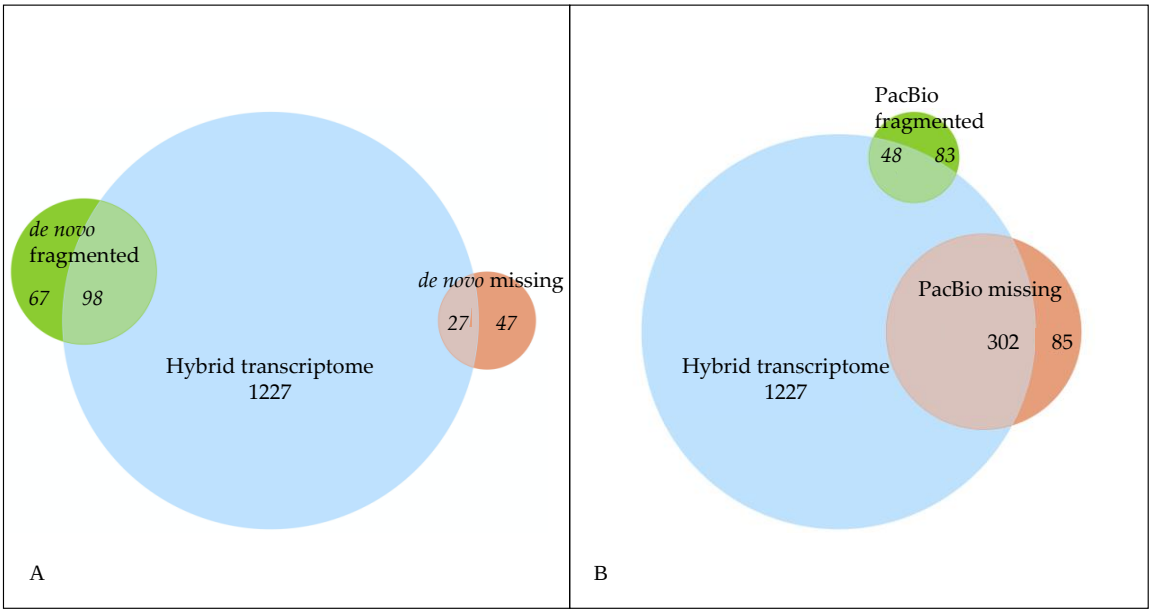

Figure 4.4. Comparison of BUSCO completeness. Fragmented or missing BUSCO in the Illumina assembly (A) or the PacBio assembly (B) but were complete (1,227) in the hybrid assembly.

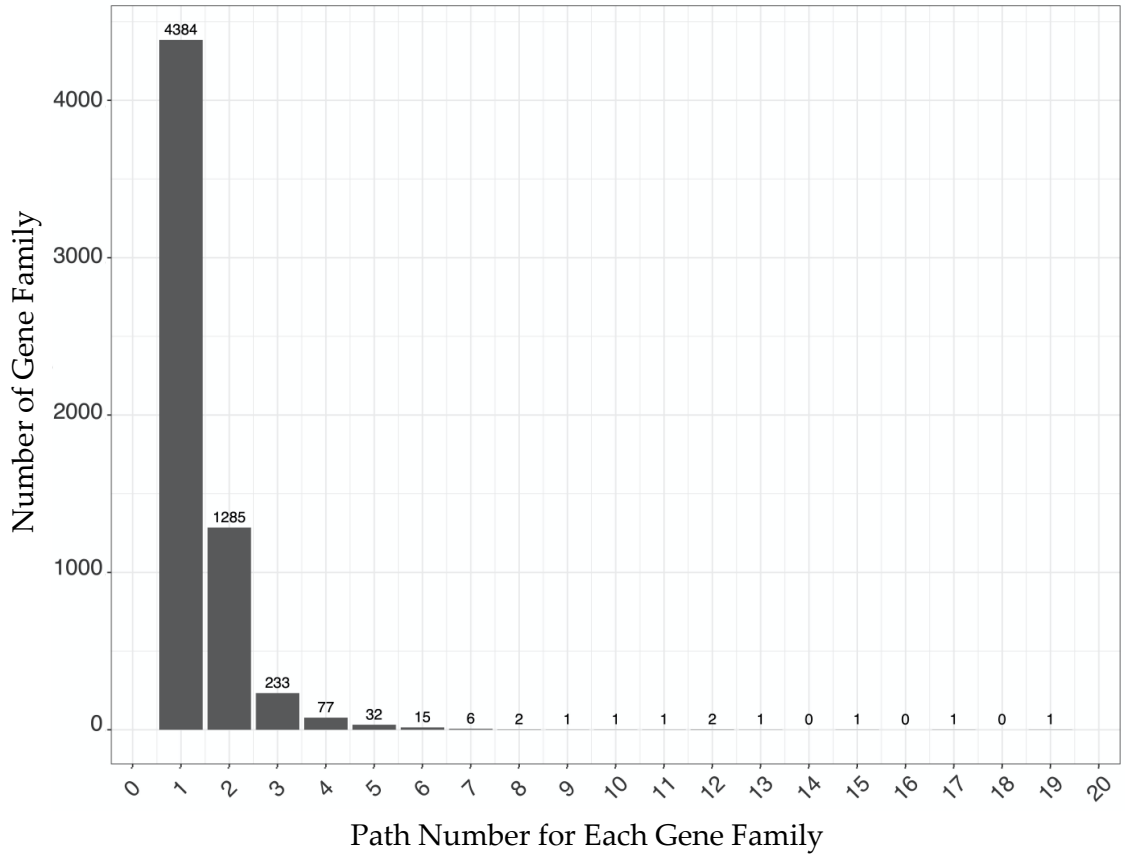

Figure 4.5. Distribution of number of paths for constructed gene families from coding genome reconstruction. The majority of gene families had one path. A total of eight gene families had more than ten paths in the reconstructed coding genome.

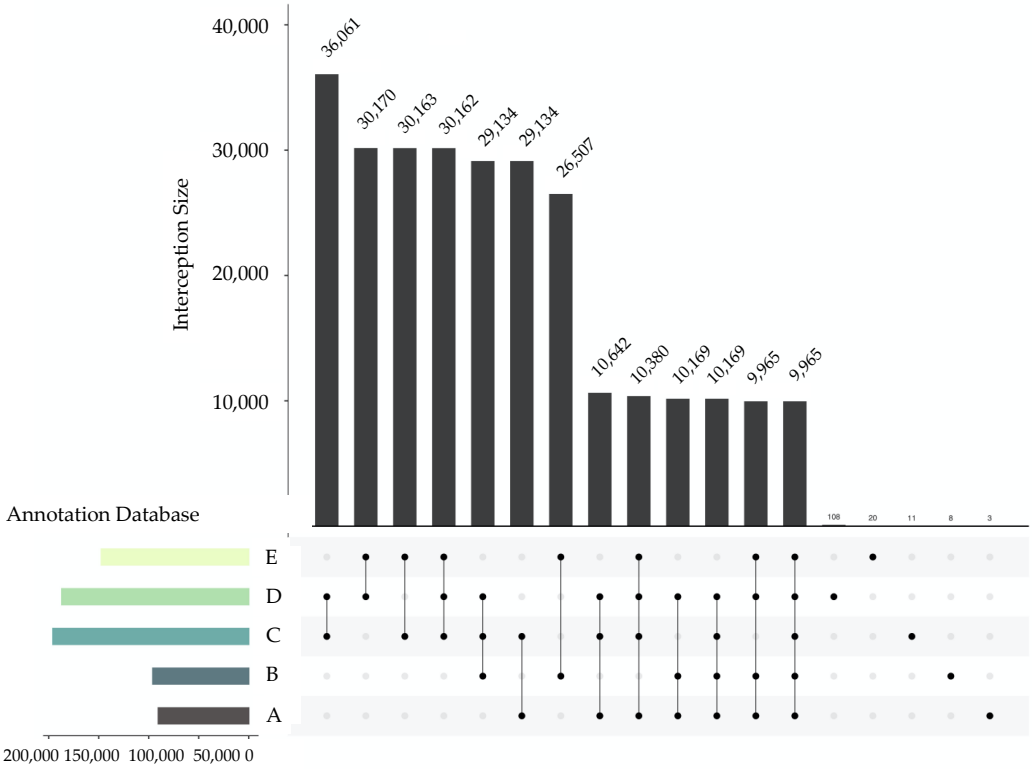

Figure 4.6. Comparison of *Festuca brevipila* PacBio Isoseq transcriptome annotation using different databases. A: Kyoto Encyclopedia of Genes and Genomes; B: Gene Ontology; C: NCBI NR Protein; D: UniRef90; E: SwisProt.

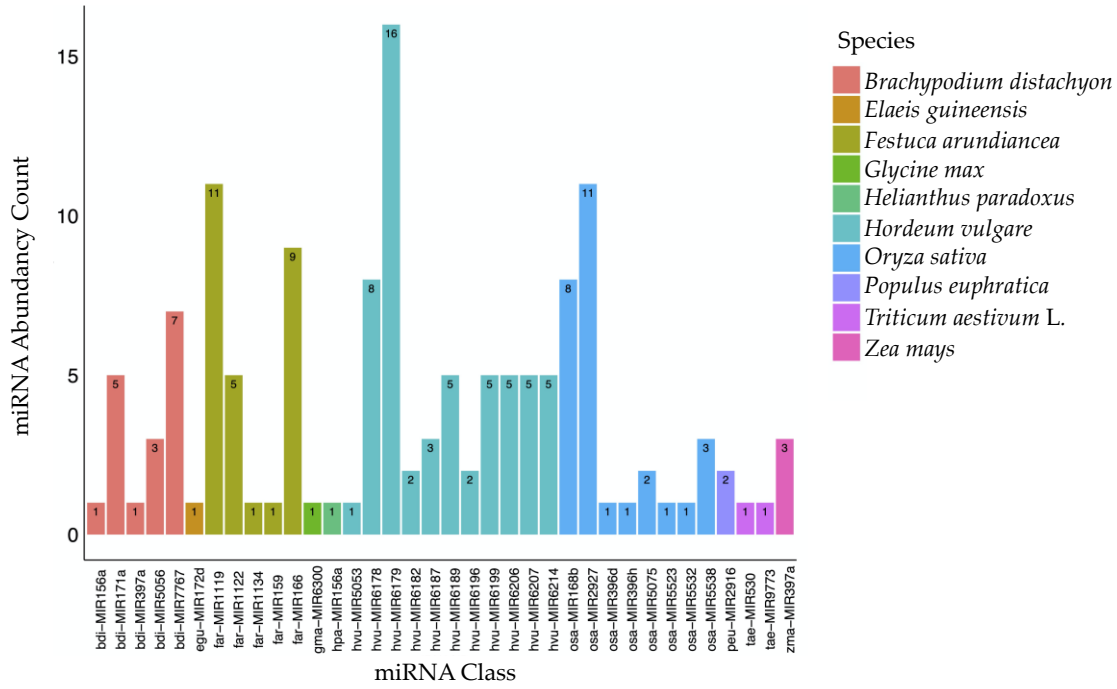

Figure 4.7. miRNA identified in the PacBio Isoseq reference transcriptome. Most of the homologous miRNAs were identified in barley (*Hordeum vulgare*) and tall fescue (*Festuca arundiancea*). With MIR6179, MIR1119, and MIR2927 had the most abundance.

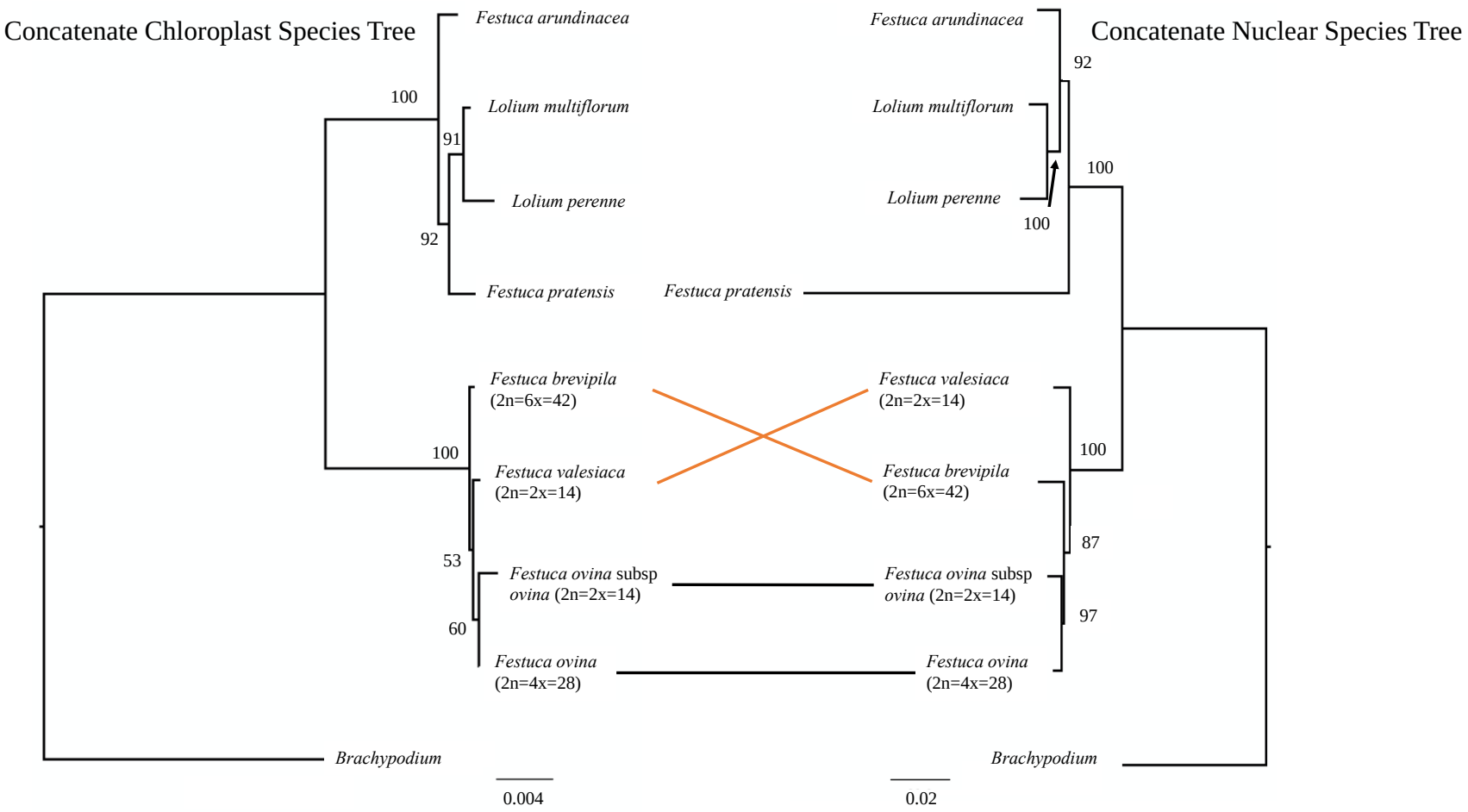

Figure 4.8. Nuclear and chloroplast species tree of the 9 taxa used in this study. In *F. ovina* complex, concordance relationship was linked in black while conflicts were linked in orange.

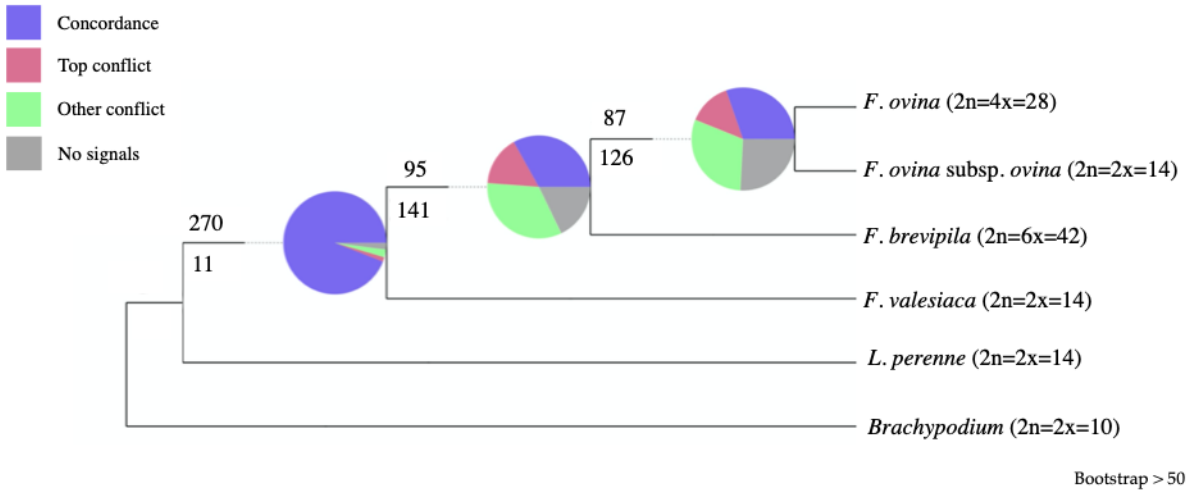

Figure 4.9. Species tree of *F. ovina* complex reconstructed by ASTRAL using 286 single-copy genes. Pie charts at nodes indicate the number of gene trees that were concordant (blue) or conflict (top conflict in red and all other conflicts in green) with the species tree topology. Total number of concordance gene trees were above the branch and the total number of conflicting gene trees were below the branch.

Table 4.1 Sources of transcriptome and chloroplast genome sequences used in this study. Sequences newly generated in this study were in bold.

Taxon	Transcriptome Sources	Chloroplast Genome Sequence
Festuca ovina ssp. *ovina* (2n=2x=14)	Illumina (SRR10997225)	*de novo contigs*
Festuca ovina (2n=4x=28)	Illumina (SRR10997223)	MN309824.1 (same genotype for used for Illumina sequencing)
Festuca valesiaca (2n=2x=14)	Illumina (SRR10997224)	*de novo contigs*
Festuca brevipila (2n=6x=42)	Pacbio Isoseq with Illumine short reads correction (http://hdl.handle.net/11299/211413)	*de novo contigs*
Lolium perenne	Perennial Ryegrass Genome Sequencing Project	AM777385.2
Lolium multiflorum	ftp.ebi.ac.uk GBXX01	JX871942.1
Festuca pratensis	ftp.ebi.ac.uk GBXZ01	JX871941.1
Festuca arundinacea	ftp.ebi.ac.uk GCTP01	KM974751.1
Brachypodium distachyon	https://plants.ensembl.org (v 3.0)	LT558603.1

Table 4.2. Statistics from *Festuca brevipila* PacBio Isoseq, *de novo* transcriptome, and combined assemblies. No decontamination was carried out for the combined assembly.

Assembly	Total Contigs	Contigs After Contaminants Removal	Contigs After Redundancy Removal	N50 (bp)
de novo	325,781	322,110	302,471 (92.84%)	1,157
PacBio Isoseq	60,719	59,510	38,556 (64.79%)	2,584
Combined Transcritpome (cd-hit-est)	341,027	NA	248,314 (72.81%)	1,405

Table 4.3. Summary statistics of Illumina sequencing and *de novo* transcriptome assembly. No decontamination was carried out for the three assemblies.

Taxon	Number of raw reads	Number of Contigs	N50 (bp)
F. ovina (4x)	26,640,768	242,794	1,230
Festuca ovina ssp. *ovina* (2x)	24,719,943	179,393	1,664
F. valeciaca (2x)	35,084,398	157,926	1,770

Chapter 5

Evaluating Effects of Propiconazole fungicide on Hard Fescue Turfgrass (*Festuca brevipila*) on Gene Expression and Metabolites

Summary

Propiconazole is often used to remove fungal endophytes from turfgrass to study the effects of *Epichloë* endophytes. However, besides a fungicidal effect, propiconazole can bind to the genes in the cytochrome 90 family and affect the biosynthesis of brassinosteroids. For this reason, the long-term effect of propiconazole on plants needs to be evaluated. In this study, we used a combination of RNA sequencing and liquid chromatography–mass spectrometry (LC-MS) to study how plants responded to the high dose of fungicide use. Plants were sprayed with a high dose of the fungicide. An individual tiller was taken from each plant at a month post the last fungicide treatment, and the same procedure was done at 4 months post the first transplanting. After 30 days (a total of 6 months days after fungicide application), plants were either inoculated with *Microdochium nivale* (causal agent of pink snow mold) or not inoculated. Propiconazole-treated plants had strong pink snow mold resistance, likely through the early expression of cold responsive genes, disease resistance genes; conversely, non-inoculated plants showed high levels of pathogen growth. This study suggested that the high dose use of fungicide resulted in important physiological changes in the plant. For this reason, researchers should be aware of the interference caused by fungicide effect when interpreting the function of fungal endophyte. Differentially expressed genes that were identified from the snow mold experiment provided breeders some gene targets for breeding for pink snow mold resistance in hard fescue.

Keywords: hard fescue, propiconazole, metabolomics, brassinosteroids

Introduction

Turfgrass diseases are one of the greatest challenges to turfgrass managers. Reduction of turfgrass diseases can be achieved through management methods such as increased air flow, reducing soil compaction, and the use of fungicides (Vargas, 2018). Of these, fungicide applications appear to be the most effective way for common turfgrass disease control. For example, most golf course superintendents apply fungicides every 14 to 21 days during summer months to control dollar spot (*Sclerotinia homoeocarpa* F.T. Bennett) on creeping bentgrass putting greens (Vincelli, 2004). Vincelli et al. (2003) observed the best dollar spot management control when spraying triadimefon (0.38 kg a.i./ha) and chlorothalonil (8.0 kg a.i./ha) in 407 liter of water. For pink snow mold control, the use of five fungicides showed the reduction of disease severity (Jung et al., 2007). The heavy use of fungicide application can control diseases, but creates issues such as fungicide runoff (Watschke et al., 2000) and fungicide resistance (Jo et al., 2008). For these reasons, significant turfgrass breeding and genetics research has focused on disease resistance (Bonos and Huff, 2013). Turfgrass breeders can develop more resistance in germplasms through either traditional breeding or transgenic approaches (Belanger et al., 2004; Bonos et al., 2006; Dong et al., 2007).

In turfgrasses, turfgrass breeders have exploited fungal endophytes as a strategy to enhance biotic stress tolerance (Bacon et al., 1997). Endophytes have been found to improve the resistance to sod webworms (*Crambus* spp.) in perennial ryegrass (*Lolium perenne*) (Funk et al., 1983), and increased dollar spot and red thread resistance in fine fescues (*Festuca* spp.) (Bonos et al., 2005; Clarke et al., 2006). Fungal endophytes were also found to improve abiotic stress tolerance to heat (Tian et al., 2015) and drought

(White et al., 1992) in turfgrass species. However, recent research has found no effect for both an abiotic stress (low temperature tolerance; Heineck et al. 2018) and a biotic stress (crown and stem rust; Heineck et al., 2020).

When studying how endophytes improve plants stress tolerance, most studies carry out a systemic fungicide application with high rates and multiple applications to eliminate endophytes in the turfgrasses to create an endophyte-free population (Hesse et al., 2003; Kane, 2011; Heineck et al., 2018). Propiconazole (cis–trans-1-[2-(2,4-dichlorophenyl)-4-propyl-1,3-dioxolan-2-ylmethyl]-1H-1,2,4-triazole) shows the best results for fungal endophyte removal (Harvey et al., 1982; Latch and Christensen, 1982). Propiconazole is a systematic triazole fungicide that targets fungal 14-alpha demethylase enzyme (14-alpha demethylase inhibitor) from demethylating a precursor to ergosterol (Kwok and Loeffler, 1993).

Besides fungicidal effects, propiconazole targets the brassinosteroid (BR) biosynthesis pathway in plants through binding to CYP90B1, CYP90C1, and CYP90D1 to inhibit the side chain hydroxylation between campesterol and teasterone (Hallahan et al., 1988; Asami et al., 2001; Ohnishi et al., 2006; Hartwig et al., 2012). The brassinosteroids are a group of important phytohormones that function in multiple aspects of plants such as growth and development, along with stress tolerance (Mandava, 1988; Nakashita et al., 2003; Bajguz and Hayat, 2009). Plants with deficient BR usually show a dwarf phenotype (Noguchi et al., 1999). The BR and gibberellin acid pathway work together to regulate plant growth and development (Li et al., 2012). For these reasons, the high rate application of propiconazole could confound any effects observed from the loss

of fungal endophyte. Therefore, it is important to study both short- and long-term effects of high propiconazole application rates for the endophyte removal process.

Because previous fungal endophyte studies focused on their anti-fungal effect, we decided to test how high rates of propiconazole affected plant responses to an important turfgrass pathogen. Pink snow mold (*Microdochium nivale*) is a winter turfgrass disease in northern regions (Tronsmo et al., 2001). *M. nivale* does not require snow cover to develop symptoms. Studies have shown that *M. nivale* is becoming more abundant in recent years and starting to spread to less temperate climates like Hawaii and California (Tronsmo, 2013; Wong, 2006). Disease resistance to *M. nivale* has been found to be associated with the winter hardiness of the plant which includes but is not limited to cell wall modification, hormone regulation, pathogenesis-related protein and the accumulation of anti-freezing proteins (Pociecha et al., 2008; Bertrand et al., 2011; Kovi et al., 2016).

We previously reported a reference transcriptome of this species developed using the PacBio Isoform sequencing [see Chapter 4 of this thesis and Qiu et al. (2020b)]. In this study, we are using the same genotype of *F. brevipila* to study the short- and long-term effects of high rate use of propiconazole via next generation sequencing and LC-MS approaches.

Material and Methods

Plant Material and Sampling

A single genotype of *F. brevipila* (SPHD15-3), from the University of Minnesota turfgrass breeding program, which was used to generated the *F. brevipila* reference transcriptome (Chapter 4 of this thesis), was vegetatively propagated into individual 5-tiller plants in soilless media (BRK Promix soil, Premier Tech, USA). This genotype was previously confirmed as endophyte free using a commercial Phytoscreen Immunoblot Kit [ENDO797-3, Agrinostics, http://www.agrinostics.com]. Because a previous study showed the effect of propiconazole on creeping bentgrass (*Agrostis stolonifera* L.), we used a single genotype 'Penncross' creeping bentgrass as a positive control and used the same propagation method described above to generate the plant population. All plant material was grown in the Minnesota Agricultural Experiment Station Plant Growth Facilities in St. Paul, MN under a 16-hour photoperiod with every-other-day watering and weekly fertilization using a liquid fertilizer solution (200 ppm N, 22 ppm P, 83 ppm K, 114 ppm S, 2.5 ppm Fe, 750 ppb Mg, 100 PPb B, 50 ppb Cu, 280 ppb Mn, 500 ppb Mo, and 81 ppb Zn). Plants at approximate eight tillers were used for fungicide spray.

Fungicide Application

Fungicide application was carried out using Kestrel Mex fungicide which has propiconazole as the active ingredient (Phoenix Environmental Care, USA). The fungicide was sprayed at the rate of 1.31 ml m^{-2} between 9 and 10 am at one-week intervals (a total of five applications) in the greenhouse condition starting at the end of February. A total of five fungicide applications were made.

Tissue Sampling

One week after the final fungicide application, leaf samples were harvested for transcriptome sequencing and metabolomics profiling (workflow illustrated in **Figure 5.1**). Briefly, 1-2 cm of leaf tissue that was 2 cm from the leaf tip was harvested and flash frozen into 1.5 mL microcentrifuge tubes for RNA extraction. The plant tissues that were 1-2 cm below the cutting site for RNA sequencing were used to for LC-MS (**Figure 5.1**). All samples were stored at -80˚C prior to analysis. Prior to tissue sampling, data was collected for turfgrass quality (using a 1-9 scale with 9 = best turfgrass quality as described by Kran et al., 2007) and phenotypic changes including plant leaf color, plant size change, were recorded using a Nikon D500 camera.

RNA Extraction and Sequencing

RNA extraction was done using Quick-RNA Miniprep Kit (ZYMO Research, Catalog number R1055) following the manufacturer's instructions. Briefly, samples were ground in liquid nitrogen and RNA lysis buffer was added. After filtering through the column, DNase I was added to remove genomic DNA contamination. Finally, RNA prep and wash buffer were added to precipitate and clean the RNA. The RNA was diluted in DNase/RNase-Free water and stored at -80°C. A Qubit Fluorometer (ThermoFisher Scientific) was used to accurately quantify the RNA concentration. Agilent 2100 Bioanalyzer (Agilent) was used to precisely detect the integrity of RNA. Samples with RNA Integrity Number (RIN) >7.5 were used for sequencing library construction.

Illumina sequencing libraries were constructed using TruSeq® Stranded mRNA Library Prep kit (Illumina) and sequenced on Illumina HiSeq 4000 instrument with 150 bp paired-end sequencing mode. All sequencing data generated from this study was deposited at NCBI under bioProject PRJNA606332.

Illumina Sequencing Data Processing and *de novo* Assembly

To process Illumina raw reads, adaptor sequences were trimmed using the Trimmomatic program (v 0.32) with the default setting (Bolger et al., 2014). Quality trimming was performed using seqtk tool with -q 0.05 (https://github.com/lh3/seqtk). Trinity (v 2.4.0) was used to assemble transcriptomes using sequencing reads generated from both fungicide and snow mold inoculation experiments (Grabherr et al., 2011). Trinity program parameters were set as max memory 200G, 24 CPUs, and bflyCalculateCPU, with minimum contig size of 200 bp. To improve assembly quality, a previously published *F. brevipila* Pacbio Isoform sequencing transcriptome was used and combined with the *de novo* assembly generated in this study (see Chapter 4). To merge and remove redundant transcripts in the combined transcriptome, 'cd-hit-est' from the CD-HIT (v 4.8.1) package was used to cluster redundant sequences with 95% sequence identity (Li and Godzik, 2006). To evaluate the completeness of the *F. brevipila de novo* and hybrid transcriptomes, we used 1335 core embryophyte genes (embryophyta_odb10) from Benchmarking Universal Single-Copy Orthologs (BUSCO v3) (Simão et al., 2015).

Transcriptome Functional Annotation

The combined transcriptome was blast searched against NCBI non-redundant (NR) protein and SwisProt, UniProt protein database using diamond blastx (v 0.9.13) with e-value < 1e-5. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using the KEGG Automatic Annotation Server (KASS, https://www.genome.jp/kegg/kaas/). *Oryza sativa*, *Brassica napus*, *Zea mays*, *Arabidopsis thaliana*, and *Aegilops tauschii* were used as references with a single-directional best hit model. The TransDecoder pogram was used to generate transcriptome

protein sequences to perform Gene Ontology (GO) search using interproscan program (Quevillon et al., 2005).

Gene Expression Analysis

To investigate the gene expression change between propiconazole- and water-treated groups, expected number of fragments per kilobase of transcript sequence per millions mapped reads (FPKM) was calculated using RSEM program (Trapnell et al., 2010; Li and Dewey, 2011) and further processed using edgeR programs implemented in the Trinity pipeline (Robinson et al., 2010; Trapnell et al., 2010). By using DE_analysis.pl script in Trinity pipeline, differentially expressed genes (DEGs) were determined and further filtered with a False Discovery Rate (FDR) of 0.05 and minimum fold change of 2.0. The DEGs annotation was done using annotation results generated above. The workflow to identify differential expressed genes is shown in **Figure 5.2**.

Differentially Expressed Cytochrome P450 genes

Because of high sequence feature similarity within the P450 gene family, a two-step approach was taken to annotate CYP450 genes identified in this study. The first step used a phylogenetics method. Briefly, *A. thaliana* cytochrome P450 protein sequences were downloaded from "The Cytochrome P450 homepage" reported by D. R. Nelson (http://drnelson.uthsc.edu/CytochromeP450.html) (Nelson et al., 2004). Pseudogene sequences were removed and the remaining named *Arabidopsis* P450 sequences were used as a reference. Because the CYP92 gene family was lost in *Arabidopsis* lineage (Nelson et al., 2004), protein sequences of eight genes within CYP92 family were downloaded from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu) (**Table S5.1**). CYP protein sequences of the differentially expressed CYP450 genes were

added to the P450 references prior to performing phylogenetic relationship reconstruction. Protein sequences were aligned using MAFFT with –auto option (Katoh and Standley, 2013). The alignment was inspected using Jalview program (v 1.0). Sequences before the starting codon were trimmed using TrimAl (Capella-Gutiérrez et al., 2009). To construct the maximum likelihood (ML) tree, the best-fit model was predicted using iqtree (v 1.6.11) with -st AA -m TESTONLY option (Kalyaanamoorthy et al., 2017). The model with the best Bayesian Information Criterion (BIC) was selected to reconstruct phylogenetic tree via iqtree (v 1.6.11) with 1000 bootstrap replications (Nguyen et al., 2014). The phylogenetic tree was visualized using FigTree (v 1.4.3) (https://github.com/rambaut/figtree) and color coded by CYP Clan (Rambaut, 2012). The second step was done by blasting protein sequences of candidate genes via NCBI blastp. Top hit (Max score) annotation that was consistent with the phylogenetic results was considered as the final annotation for the transcripts.

Untargeted Metabolomics Analysis

Propiconazole-treated plants were also analyzed for metabolomic changes compared to water-treated plants. To improve the accuracy of the LC-MS run, 4-5 subsamples were analyzed for each biological replicate for water and propiconazole treatments. Samples were extracted in 90% aqueous methanol (Sigma-Aldrich HPLC Plus grade methanol and Fisher Scientific Optima™ water) using 10 µL solvent per mg of fresh-frozen tissue. After suspending samples in solvent, microcentrifuge tubes were placed back into a -80°C freezer for 48 hr. One 5-mm tungsten bead was added to each sample tube, and samples were ground in a bead mill (Genoginder SPEX SamplePrep) for 12 mins (1,400 RPM for 2 mins, and 700 RPM for 10 mins) in -20°C grinding blocks.

Samples were vortexed and centrifuged at 14,000 g for 10 mins (4℃), and the supernatant was transferred into new tubes and re-centrifuged. To remove chlorophyll from the supernatant, all samples were diluted 1:2 in Optima™ water and were placed at -80℃ overnight. Diluted extracts were then centrifuged at 16,000 g (4℃) for 30 mins. Final extracts were aliquoted into glass autosampler vial inserts, which were then inserted into autosampler vials for LC-MS analysis. A quality control pooled sample was also made using aliquots of extracts from all samples.

Untargeted metabolite analysis was performed using both reverse phase and normal phase methods. Both pool and blank samples were sampled four times throughout the run as quality controls. The reverse phase was performed using Waters Acquity UPLC HSS T3 1.8 Microm, 2.1 x 100 column (C18), and normal phase was performed with a Millipore SeQuant ZIC-cHILIC 3um, 100A, 100 x 2.1 mm column (cHILIC). Analyses were performed using an UltiMate 3000 High Performance Liquid Chromatography (ThermoFisher), and Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (ThermoFisher).

For the C18 column run, the Thermo UltiMate 3000 UPLC pump, auto sampler, column compartment, and diode array detector (DAD) were set with the following parameters: flow rate = 0.4 mL per min; 10% B (0.1% Formic acid in acetonitrile) 1-2 min, 10-95% B; 2-25 min 95% B; 25-28 min 95-10% B 28-30 min; 210, 254, 340, and 520 nm were analyzed using the DAD along with full scan UV-vis ; column temperature=40℃. The mass spectrometer parameters were set as: probe height = C; Full MS Scan range = 200 - 1500 m/z; Positive and negative ionization were monitored simultaneously using polarity switching; Sheath gas flow rate = 50; Aux gas flow rate =

20; Sweep gas flow rate = 1; Spray voltage = 3.8 kV; Capillary temp = 350 ℃; S-lens RF level = 50; Aux gas heater temp = 300 ℃.

For the cHILIC run, the Thermo UltiMate 3000 UPLC pump, auto sampler, column compartment, and diode array detector (DAD) were set with the following parameters: flow rate = 0.4 mL per min; 98% B hold 0-2 min; 98-55% B 2-32 min; 55-95% B 32-34 min; 95% B hold 34-37 min; and the UV-vis = NA; column temperature=40 ℃. The mass spectrometer parameters were set as: probe height = C; Full MS Scan range = 50 - 750 m/z; positive and negative ionization were monitored simultaneously using polarity switching; sheath gas flow rate = 5 Aux; Gas flow rate = 0; sweep gas flow rate = 0; spray voltage = 3.3 kV; capillary temp = 320 ℃; S-lens RF level = 55 Aux; gas heater temp = 0.

Liquid Chromatography-Mass Spectrometry Data Processing

LC-MS data processing was done using the Work4Metabolomics Galaxy platform (galaxy.workflow4metabolomics.org, version 3.3) for untargeted data analysis. Based on polarity, raw data files were split into positive and negative mzML files using the msconvert program (ProteoWizard release: 3.0.11252). The workflow for the data processing was done following instructions by Galaxy training for mass spectrometry: LC-MS analysis (https://galaxyproject.github.io/) with adjusted bandwidth for peak identification parameter. Detail preprocessing and postprocessing workflow is shown in **Figure S5.2**, and **Figure S5.3**. LC-MS features were used for unsupervised multivariant analysis and scatter plots were produced with PC1 and PC2 data to evaluate differences in water and propiconazole treatments. The identity of the top 10 metabolite features that were driving principal component (PC) 1 for positive and negative phase of both columns

were predicted by searching accurate mass m/z data via KEGG compound and the human Metabolome Database (HMDB, hmdb.ca/spectra/ms/search), and Metlin.

Long-Term Effect of Propiconazole on *F. brevipila*

To study the long-term effects of high rate propiconazole on *F. brevipila*, we inoculated both propiconazole-treated and untreated plants with pink snow mold isolate PSM to evaluate the disease severity (isolate was obtained from Dr. Paul Koch, University of Wisconsin, Madison). To avoid the effect from residual fungicide propiconazole, we carried out two rounds of vegetative propagation. Briefly, one month after the fifth fungicide treatment, single tillers were split from the original plant and were transplanted into fresh soilless media. The newly transplanted tillers were grown in the greenhouse in the condition described above for four months. Tillers of these transplants were split and transplanted in fresh soilless media again and were grown for one month to reach approximately five tillers before performing snow mold inoculation. Therefore, the plant had experienced the final propiconazole applications six months prior to inoculation. The scheme of the plant material preparation was shown in **Figure 5.3**.

Inoculum Preparation

Microdochium nivale inoculum was prepared following the protocol described by Tronsmo (1993), Hofgaard (2006), and Kovi (2016) with modification. Briefly, the isolate was first cultured on potato dextrose agar (PDA) and incubated at 12 °C in the dark. Agar plugs (5 mm in diameter) were exercised from the PDA plate, transferred in a flask that contained 50 mL potato dextrose broth (PDB) and were incubated at 12 °C. After 15 days of growth, the mycelium was harvested by filtering through cheesecloth

and homogenized in distilled water containing 0.01% TWEEN 20 (Sigma) using PRO Scientific Bio-Gen PRO200 Homogenizer. The inoculum was diluted to an optical destiny of 0.5 at 430 nm prior inoculation.

Pink Snow Mold Inoculation

Plants were transferred into a dark walk-in cooler set at 10 °C one hour prior to inoculation. Inoculation was performed by dropping 1 mL inoculum to the crown region for the treatment group, and for mock-treated plants water drops were used as the control. Moist paper towels were used to cover plants to mimic snow cover and maintain moisture. One day post inoculation, plant tissues (1-3 cm above ground, **Figure S5.4**) were harvested for RNA sequencing for gene expression analysis (Propiconazole-pretreatment plant with/without snow mold, water-pretreatment plant with/without snow mold, three biological repeats each). Sequencing and DEG identification were carried out using the method described earlier. Fourteen days post inoculation, the maximum mycelium spread length on the leaf (between crown region) was measured with a ruler to compare the different disease severity.

Results

Plant Phenotype Observation after Propiconazole Treatment

One week following the fifth propiconazole application, the positive control creeping bentgrass plants showed wider and thicker leaf tissue along with stunted plant growth as expected (**Figure 5.4 A**, **B**). *Festuca brevipila* had a similar response with slower growth and darker leaf color with propiconazole treatment (**Figure 5.4 C**, **D**) (photos were taken from the same height). One month after the last fungicide application, propiconazole-treated *F. brevipila* had quality differences (tiller, plant size, and turf color) compared with the control groups (size rating score of 7.67±0.58 for treated plants vs 9.0 ± 0.00 for non-treated).

Transcriptome Assembly and Annotation

A total of 392,578,625 raw reads were generated to study plant-fungicide interaction; after adaptor removal and quality trimming, a total of 392,556,876 reads were retained. To study the different responses to pink snow mold inoculation in fungicide- and water-pretreated plants, a total of 970,893,276 reads were generated from Illumina sequencing. After adaptor removal and quality trimming 919,321,188 reads were retained. All sequencing reads were used to build the *de novo* assembly that included 982,775 contigs with N50 of 1,047 bp. After combining *de novo* and reference transcriptome assemblies and removal of sequence with high similarity, we obtained a combined assembly with 734,009 contigs and N50 of 1,042 bp.

The BUSCO assessment showed 422 of the BUSCO genes were fragmented in the *de novo* assembly and the number was reduced to 185 in the combined assembly. Additionally, only 43 BUSCO transcripts were showing as missing in the hybrid

assembly, while the *de novo* assembly potentially missed 113 genes (**Figure 5.5 A**). The annotation of combined assembly statistics was shown in **Figure 5.5 B** with 25,239 transcripts had annotation from all five databases. NCBI NR and UniRef had the most annotation for this combined assembly.

Differentially Expressed Genes Analysis

All RNA sequencing reads were mapped to the transcriptome reference using the combined assembly, isoseq only, and *de novo* assembly. The RNA sequencing reads mapping statistics are summarized in **Table 5.1**. The combined assembly had the least unmapped reads, followed by Isoseq reference. The *de novo* assembly had the most unmapped reads.

The RSEM mapping result using the combined assembly as the reference was used for differential gene expression analysis. At FDR 0.05 with a minimum fold change of 2.0, a total of 204 differential expressed genes (DEGs) were identified with 117 genes upregulated and 87 downregulated in propiconazole-treated plants (**Figure 5.6**). Based on the conserved domain, two of the 117 upregulated genes were identified as CYP450 genes.

CYP450 Related Differentially Expressed Genes

Transcript_34799 (logFC: -2.37) and Trinity_dn78694_c0_g4_i1(LogFC: -9.57) were identified as CYP450 genes. To provide a phylogenetic relationship between these two transcripts with known CYP450 genes, a total of 274 Arabidopsis and rice CYP genes were used as the reference. After sequence alignment and trimming, iqtree model testing suggested the JTT+F+I+G4 as the best model which was then used to reconstruct the phylogenetic tree. Transcript_34799 was a sister to rice CYP92A9 and

Trinity_dn78694_c0_g4_i1 was close to the *Arabidopsis* CYP71A gene family. Both transcripts were nested the CYP71 Clan (**Figure 5.7**).

The NCBI blastP results suggested Transcript_34799 had the best hit to CYP92A44-3 gene (GenBank: AER39773.1), with a max score of 772, 99.21% protein sequence similarity. The CYP92A44-3 was identified and cloned from a close relative of *F. brevipila*: *Festuca rubra* ssp. *fallax*. The second hit of this transcript was CYP71A genes in *Aegilops tauschii* ssp. *tauschii* (697, 90.50%), and followed by a flavonoid 3'-monooxygenase in *Hordeum vulgare*. Since the phylogenetic data suggested Transcript_34799 was a sister to the rice CYP92 gene family, the final annotation of this transcript in this study would be considered as CYP92A family gene.

The Trinity_dn78694_c0_g4_i1 transcript had highest sequence similarity to CYP450 71A1-like protein (XP_020192280.1) in *Aegilops tauschii* ssp. *tauschii*, and a CYP450 71A1 protein (GenBank: EMS50742.1) in *Triticum urartu*, which was within expectations of the phylogenetic data. The final annotation of this transcript in this study would be considered as a CYP71A family gene.

Non-CYP450 Differentially Expressed Genes

The non-CYP450 DEGs were selected based on gene annotation using five databases, and further grouped by metabolic activities and transcriptional factor/signaling functions (**Figure 5.8**). For example, we observed upregulation of gibberellin 2-oxidase, cinnamoyl-CoA reductase, CBF protein, and multiple WRKY (5, 30, 41, 53) transcriptional factors induced in the propiconazole-treated plants while the expression of NADH ubiquinone oxidoreductase and cinnamyl alcohol dehydrogenase were downregulated. All the transcript ID and fold changes are summarized in **Table S5.2**.

In addition to the transcriptional factors and metabolite-associated enzymes, such as photosystem II activity, several expansin-related genes activities were also down regulated in propiconazole-treated plants (**Table S5.2**).

Metabolites Profile Change

The metabolite profile from the HILIC column run was analyzed with retention time set to 0.5-25 minutes. After quality control and CV-based filtering, the unsupervised metabolite clustering for positive (720 features) and negative (692 features) modes were shown in **Figure 5.9**, where principle components 1 (PC1) explained 37.8% and 16.4% variation in positive and negative mode, respectively. The initial quality control showed in one of the four subsamples from one of the water-treated plant was clustered with propiconazole treated plants in the C18 dataset. This was likely the result of sampling contamination. Since there were four subsamples from the same plant, this subsample was removed from the downstream analysis. The first pool sample was also removed in the HILIC column data analysis because it deviated from the center of the PCA.

The metabolite profile changes from the C18 column was analyzed using the same method except the retention time was subset to 1-15 minutes. The positive mode included 2,933 metabolite features, and the negative mode included 4,090 metabolite features. The unsupervised clustering for the metabolite profiles is shown in **Figure 5.10** with PC1 explaining 44.5% variation in positive mode, and 48.2% in negative mode.

The top 10 metabolites for each mode and column (a total of 40) predicted by accurate m/z mass are shown in **Table 5.2**. Because numerous metabolites could match with accurate mass data, more than one database hit was included for each feature.

In addition to the top 10 metabolites for each mode and column, we also identified a total of 93 candidate metabolites (C18: 92; cHILIC 1) that were unique only in propiconazole treated plants (**Table S5.3**)

Propiconazole Pretreatment Increased the *M. nivale* Resistance of *F. brevipila*

To evaluate the long-term effect of propiconazole on *F. brevipila*, plants were inoculated with *M. nivale*. Fourteen days post inoculation, mycelium diameter growth was measured for the FS and WS groups. WS plants had significant mycelium growth compared to FS plants (**Figure 5.11**).

To compare gene expression difference between FS and WS one day post snow mold inoculation, RNA sequencing identified a total of 4,118 differentially expressed genes (DEGs) at FDR 0.05 with the minimum LogFC of 2.0. Filtered by NCBI NR protein database annotation, the 4,118 DEGs were split into *Microdochium* fungi-specific (1,730 DEGs) and plant-specific groups (1,871 DEGs). DEGs without NCBI NR annotation were excluded from further analysis. Heatmaps suggested a higher level of *Microdochium* activity in the WS group (**Figure 5.12 A**), and some *Microdochium* activity in the FS group; there was also a clear expression pattern separation in the plant-specific genes expression (**Figure 5.12 B**).

A total of 143 upregulated genes in the FS group had KEGG annotation and 821 upregulated genes in the WS group had KEGG annotation. KEGG mapper reconstruction suggested that specific pathway modules such as fatty acid and α-linolenic acid metabolism were upregulated in the WS group comparing to the FS group.

After inspecting NR-based annotation for each group using candidate genes information from previous literature, we found the ice recrystallization inhibition proteins

(IRIPs), cold responsive proteins, and disease resistance proteins were upregulated in the FS group at significant levels, while chitinase and pathogenesis-related proteins had significantly higher expression in the WS group (**Figure 5.13**, **Table 5.3**). The conserved domain search of the six IRIPs showed four out of the six IRIPs contained at least one leucine-rich repeat receptor-like (LRR) protein kinase domain, and three contained ice binding domains (**Figure S5.5**). One identified IRIPs contained CAP (cysteine rich secretory proteins, antigen 5 and pathogenesis-related 1 protein domain).

Discussion

Propiconazole is a triazole fungicide known as a demethylation inhibitor (DMI), that targets the C14-alpha demethylase enzyme from removing the C-14α-methyl group from lanosterol to stop ergosterol production in fungi (Peyton et al., 2015). PCZ is a known regulator of plant growth (Fletcher et al., 2000) and has been shown to enhance root development of redroot pigweed (*Amaranthus retroflexus*) (Hanson et al., 2003). PCZ treatment has also been shown to reduce salt stress in Kentucky bluegrass (*Poa pratensis* L.) (Nabati et al., 1994), speed up the recovery of Kentucky bluegrass sod from heat injury (Zhang et al., 2003), and provide a greater turf quality in creeping bentgrass (Ervin et al., 2004) by increasing superoxide dismutase (SOD) activity. In addition, PCZ was found to be an inhibitor for brassinosteroid biosynthesis in *Arabidopsis* and maize (Sekimata et al., 2002; Hartwig et al., 2012; Oh et al., 2016) and also showed an inhibition effect on obtusifoliol 14R-demethylase activity (Burden et al., 1989; Gilley and Fletcher, 1997).

Major brassinosteroid biosynthetic genes include but are not limited to, cytochromes P450s 85A, 90B, 90C, 90D and 92A (Nomura and Bishop, 2006). Studies have shown the BR inhibitor such as brassinazole, and triadimefon, binds to CYP90B, C and D to suppress BR biosynthesis (Asami et al., 2003). In our study, after five propiconazole applications, the positive control creeping bentgrass samples used in this study showed darker leaf color, wider leaves, and stunted growth as expected. Because the leaves of hard fescue are narrow, we did not measure leaf size change. We observed the dwarf phenotype with darker green leaf colors. Through transcriptome analysis, we identified two CYP 450 (CYP92A and CYP71A) that were upregulated in PCZ-treated

plants. CYP92A was found to convert typhasterol to castasterone both early and late in the C6 brassinosteroid biosynthesis pathway and was believed to function as BR C-2 hydroxylase (Nomura and Bishop, 2006). The application of P450 inhibitors [Ancymidol and 1-amin-obenzotriazole (ABT)] on Chewings fescue (*Festuca rubra* ssp. *fallax*), a close relative of *F. brevipila*, also showed the upregulation of CYP92A gene (Huang et al., 2012).

Our finding agrees with previous studies that in BR-deficient plants, genes committed to BR biosynthesis were upregulated at the transcription level (Noguchi et al., 2000). In our case, we did not observe differential expression of CYP90 and CYP85 genes. One possible reason is that CYP90A1 and CYP85A2 have a circadian rhythm of gene expression and are not affected by the feedback regulation (Bancos et al., 2006), or simplify because the variation induced by leaf differs with age.

The CYP71A gene family is thought to have highly specific functions across and within individual species (Hamberger and Bak, 2013). For example, the CYP71A1 gene was suggested to play a role in oxidation of monoterpenoids in avocado fruit (Bozak et al., 1990), while CYP71A13 was found to be involved in the camalexin biosynthesis (Nafisi et al., 2007). In our study, the protein identified as CYP71A1-like did not cluster close to a specific CYP71A gene in *Arabidopsis*; thus, its function remains unknown.

Sekimata et al. (2012) suggested that PCZ affects the BR biosynthetic pathway through a BR independent pathway, while Asami et al.,(2003) suggested triadimefon affects both the GA and BR pathways. In our study, we observed the up-regulation of giberellin 2-oxidiase 6 in PCZ-treated plants. The activiation of giberellin 2-oxidiase 6 was found to deactivate active giberellin in rice and created a semi-dwarf phenotype

(Huang et al., 2010). Interestingly, GA and BR interact via DELLA proteins to regulate plant growth and development (Li et al., 2012), and the dwarf phenotype we observed in this study could be the combination of BR and GA pathway effects. Future study is needed to address this question.

Through transcriptome analysis, we also observed that multiple plant signaling and defense pathways were upregulated by PCZ treatment. The upregulation of WRKY 30 has been associated with plant salt resistance by regulating reactive oxygen species-scavenging in grape (Zhu et al., 2019), and increased endogenous jasmonic acid accumulation and PR gene expression in rice (Peng et al., 2012). WRKY5 was found to be related to SA signaling and induced by elicitors (Cormack et al., 2002; Schilirò et al., 2012). The CBF family is usually regulated by cold temperatures (Gilmour et al., 1998; Medina et al., 1999). Interestingly, we also observed the upregulation of the AP2 domain CBF protein. These results suggest the fungicide potentially functions similar to a stress stimulus, such as cold stress.

We observed some gene expression variations within biological repeats that might be due to the leaf age variations during the sampling. Metabolomics feature (C18 and HILIC) scatter plots using unsupervised clustering showed separation between propiconazole and water-treated plants, suggesting a treatment effect on *F. brevipila*. Of the forty features driving the feature clustering of C18 and cHILIC metabolite profiles, metabolites such as methyl dihydrojasmonate, gibberellin acid isomers, secoisolariciresinol (lignan, a type of phenylpropanoid compound), and 4-O-beta-D-glucosyl-sinapate (a lignin monomer with a hexose) are particularly interesting because transcriptome data suggested potential change in GA signaling pathway (GA 2-oxidase),

jasmonic acid accumulation (WRKY30), and lignin biosynthesis [cinnamoyl-CoA reductase, cinnamyl alcohol dehydrogenase (Kawasaki et al., 2006; Ma, 2010)]. However, the accurate mass search returned with multiple putative metabolites. It is difficult to confirm the metabolite features based on information available. Future research will be done to using standard compounds, tandem MS, or even nuclear magnetic resonance (NMR) to characterize and confirm the predicted metabolites from this dataset.

We found that propiconazole-pretreated plants showed better resistance to snow mold than water-pretreated plants six months after treatment. Pink snow mold resistance is associated with the winter hardiness of the plant. In winter rye, chitinase, thaumatin-like protein (GLP), and endo-β-1,3-glucanase (GLPs) were found to confer snow mold disease resistance (Hiilovaara-Teijo et al., 1999; Kuwabara et al., 2002). Multiple reports suggested the cold hardiness-related traits such as cold-induced PR, GLP proteins have antifungal activities (Bertrand et al., 2011; Kovi et al., 2016). Cytological analysis on wheat suggested that cold hardened wheat delayed the infection of the snow mold penetration and showed better disease resistance compared to the non-harden plants (Dubas et al., 2011). The modification of cell wall components was also found to be associated with the snow mold resistance in perennial ryegrass (Kovi et al., 2016). In our study, when comparing the fungicide- and water-pretreated plants, transcriptome data showed more pink snow mold gene activity in plants that were pretreated with water. This suggests that the pathogen had a faster penetration or growth on water-pretreated plants in the first 24 hours. When comparing the PR gene expression between WS and WC groups, we observed the significant upregulation of PR proteins in WS, suggesting

the expression of PR proteins were likely triggered by the pathogen and not the cold temperature. We observed multiple chitinases, pathogenesis-related (PR) protein, and GLP protein that were upregulated upon infection in WS plants, suggesting the activation of a pathway by the plant similar to a snow mold pathogen defense pathway.

Interestingly, we identified a set of highly up-regulated ice recrystallization inhibition proteins (IRIP), cold responsive proteins, and some disease resistance proteins expression in FS plants compared to the WS group. The primary function of the IRIPs is to inhibit ice recrystallization and minimize the physical damage caused by larger ice crystals, thereby enhancing plant freezing tolerance (Griffith et al., 1997; Tremblay et al., 2005; John et al., 2009; Zhang et al., 2010). Additionly, the IRIPs have the LRR domain to the N-terminal protion and studies have suggested the pathogenesis-related function from the IRIP (Hon et al., 1995; Yeh et al., 2000; Tremblay et al., 2005). The expression of IRIPs and disease resistance proteins are likely the result of a combination of pink snow mold and cool temperature stress. It is still unknown if the IRIPs produced in the propiconazole-pretreated plants have antifungal activity and the function of the disease resistance protein; future research such as *in vitro* expression of these genes in *E. coli* will provide evidence for their potential function beyond ice recryslization inhibtion function. The disease resistance protein RGA2 was found to restrict the pathogen growth (Song et al., 2003). One possible hypothesis based on these results is that the plants gained pink snow mold resistance as a result of the fast expression of cold responsivie protein, disease resistance protein, and IRIPs expression.

Plants used in the snow mold study were treated with large amounts of fungicide six months prior to snow mold inoculation. The half-lives of propiconazole in organic-

rich soil is 20 days (Thorstensen and Lode, 2001). To avoid the fungicide residual in the soil or plant tissue, we used newly developed tillers and transplanted to fresh soil twice to conduct the experiment six months post fungicide treatment. Considering the additional frequent watering, we do not expect that our observations are a result of fungicide residual interference. Both FS and WS plants have an identical genetic background; thus, the observed difference to cold and snow mold infection likely resulted from the rapid use of propiconazole. During the fungicide spray, the TRINITY_DN72600_C2_G3_I3 transcript, which contained ice binding and multiple LRR domain, had higher expression but not at a significant level due to one outlier in the control group (LogFC 0.61, p-value 0.36; average transcript count: 214 to 133). It has been shown non-low-temperature stresses such as drought can also trigger antifreeze activity-related gene expression (Yu and Griffith, 2001). It is possible that the propiconazole treatment induced a similar effect so when the plants are exposed to a similar condition, the plants are able to respond faster. Since we observed the potential propiconazole effect on lignin biosynthesis one week post fungicide application, future cytogenetics work will be done to see if there is a change in cell wall structure.

Conclusion

Fungicide applications are widely used in turfgrass research to create fungal endophyte-free populations to study the effect of fungal endophytes. We found that fungicide applications at rates and frequencies typically used in endophyte research affected the brassinosteroid and gibberellins pathways of plants. The long-term effect of the use of fungicide induced stronger disease resistance to *Microdochium nivale*, likely through the expression of cold responsive genes, disease resistance proteins, and anti-freezing proteins. For this reason, researchers should think carefully about fungicide use in fungal endophyte – grass host research experiments; furthermore, results from these studies should be taken with caution. The cold-responsive, disease-resistance, and anti-freezing proteins identified in this study potentially provide the breeders the new methods for developing turfgrass cultivars for better pink snow resistance.

Figures and Tables

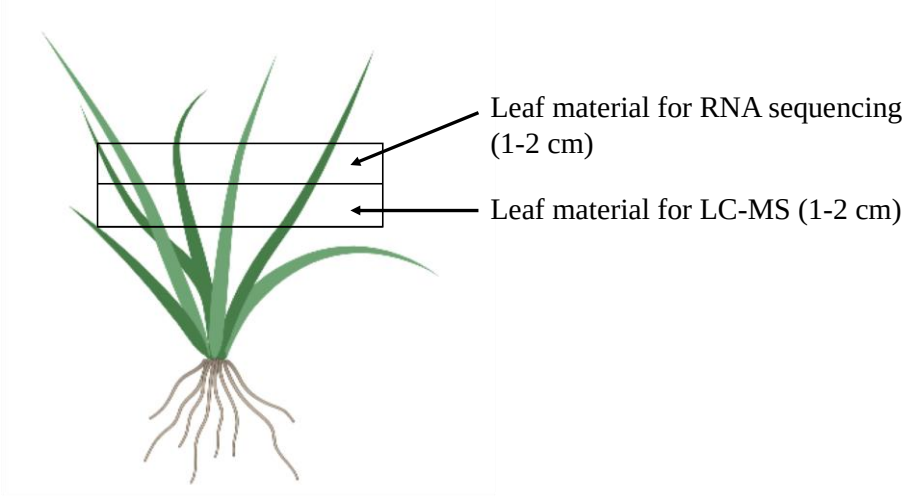

Figure 5.1. Destructive sampling of *F. brevipila* a week after the fifth fungicide application for transcriptome and metabolomics studies. For transcriptome sequencing, leaf samples were taken 1-2 cm from the leaf tip.

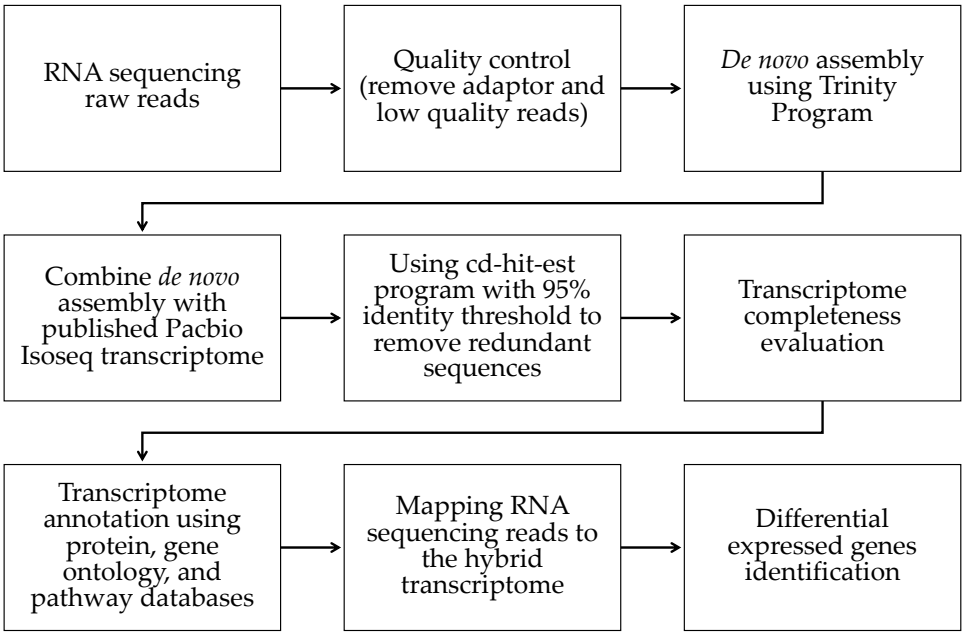

Figure 5.2. The workflow of RNA sequencing study for *F. brevipila* and propiconazole interaction.

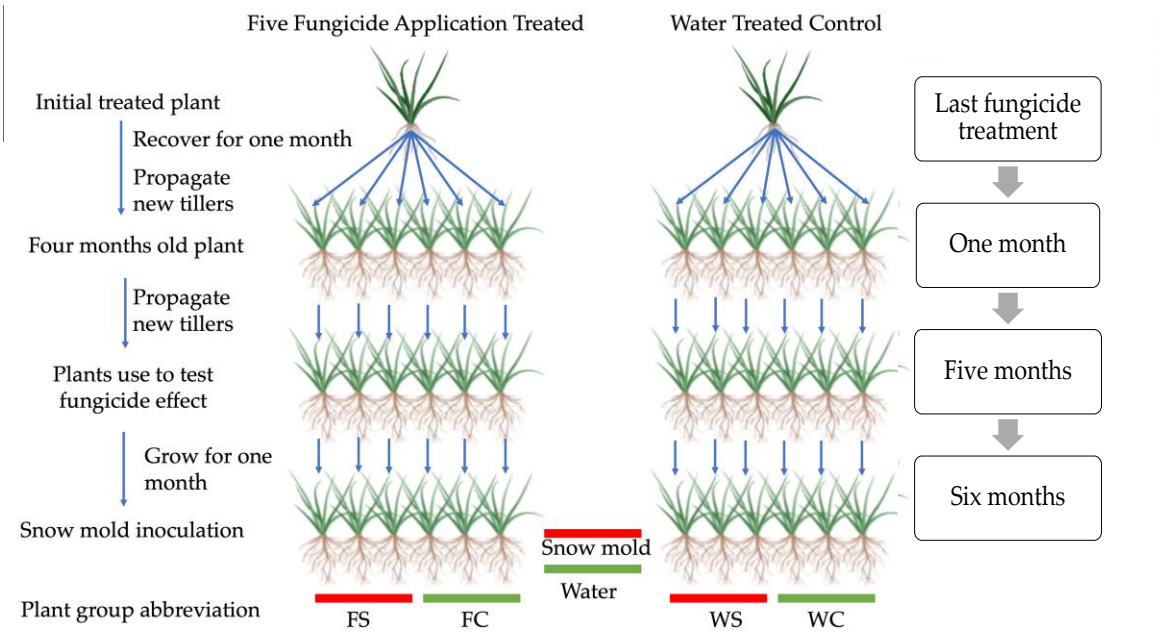

Figure 5.3. Workflow in preparation of plants for snow mold inoculation. Throughout the process, plants were transplanted twice into new soil. The snow mold inoculation was carried out six months after the last fungicide treatment.

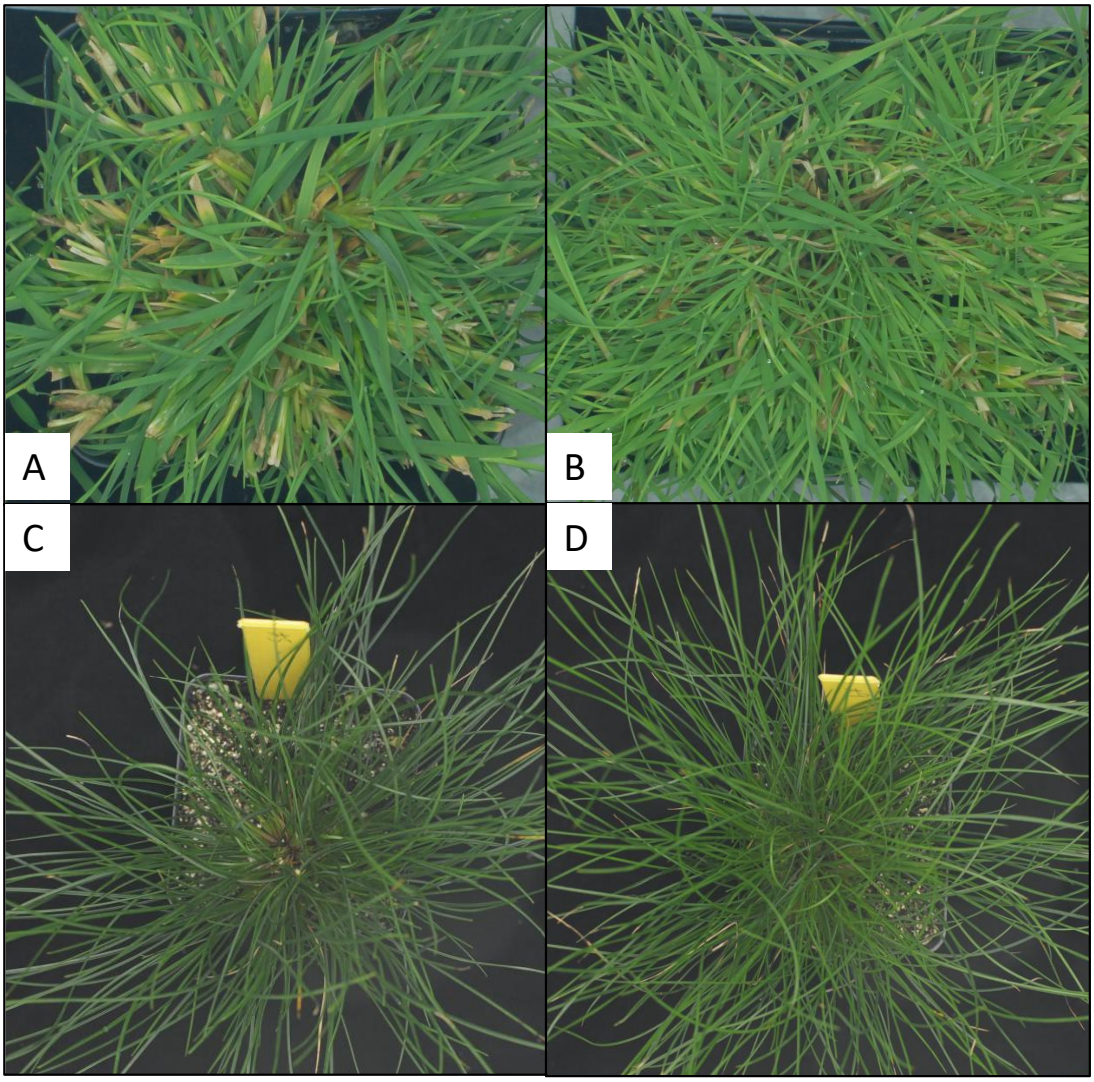

Figure 5.4. *Agrostis stolonifera* and *F. brevipila* phenotypes a week after five fungicide applications. The plant treated with propiconazole fungicide showed darker green leaf color and slower growth (A, C). *Agrostis stolonifera* also showed wider leaves (A), while the water-treated plants showed normal growth and color (B, D).

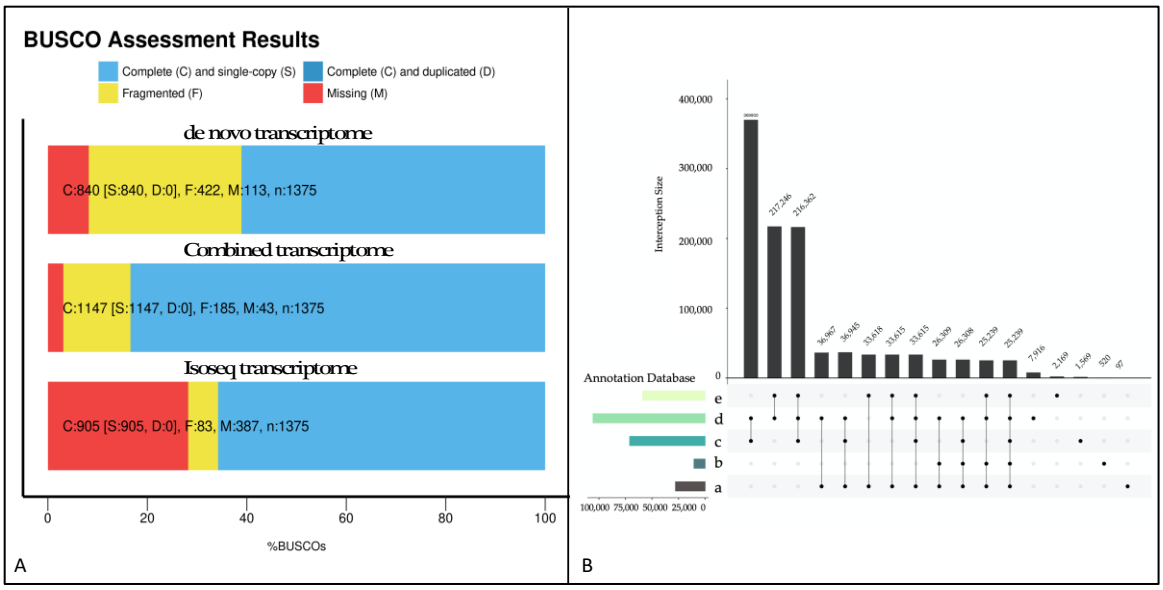

Figure 5.5. The transcriptome completeness evaluation (A) and transcriptome annotation using five protein databases (B). Protein databases in panel (B): a: Kyoto Encyclopedia of Genes and Genomes; b: Gene Ontology; c: NCBI NR Protein; d: UniRef90; e: SwisProt.

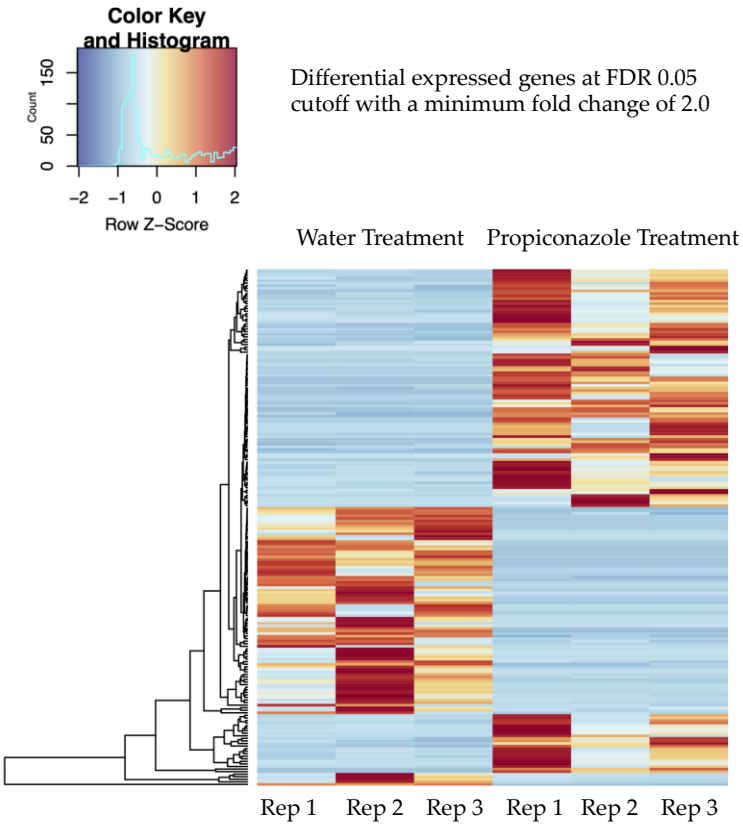

Figure 5.6. Differentially expressed genes between propiconazole-treated and water-treated *control* plants at FDR 0.05, minimum Log fold change of 2.0.

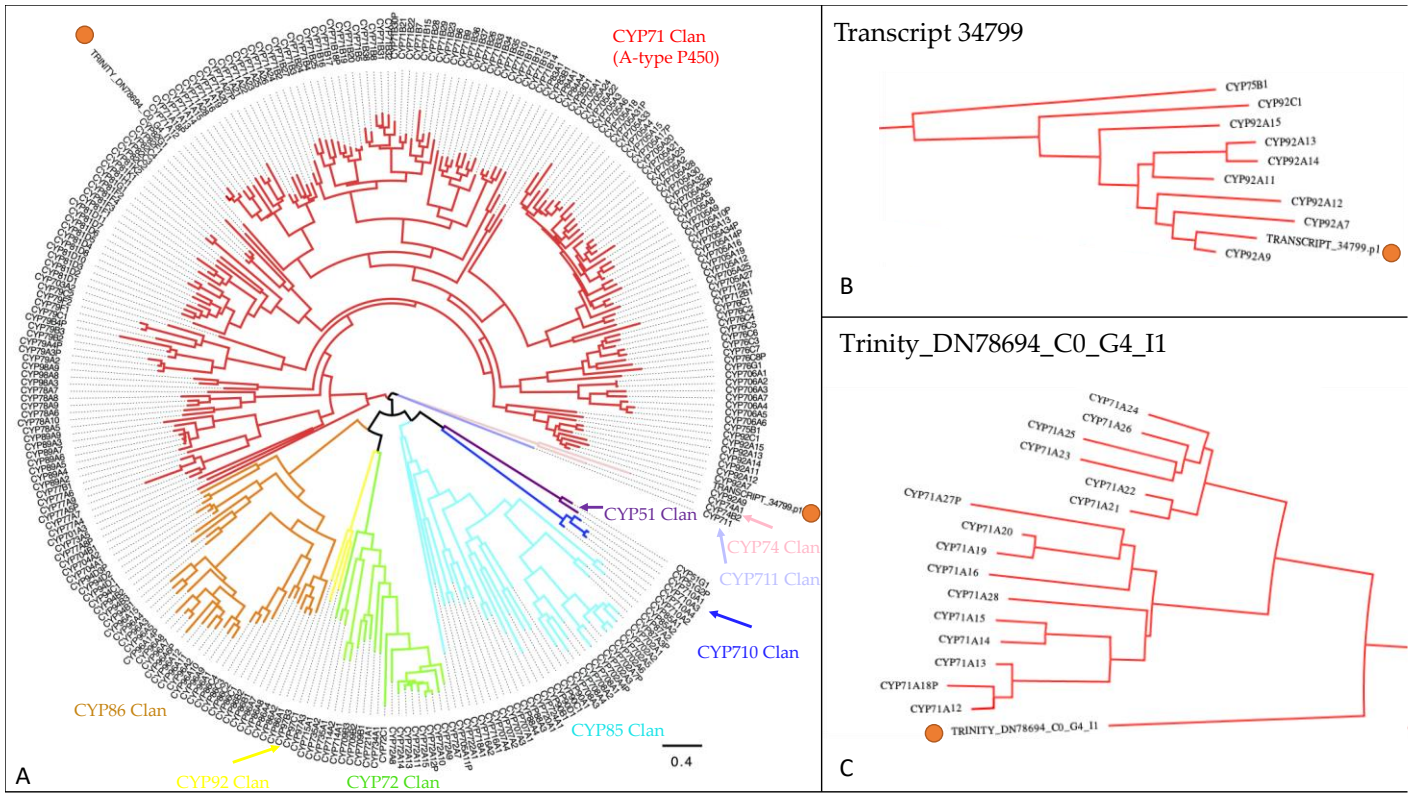

Figure 5.7. The phylogenetic relationship of the two P450 transcripts that were upregulated under the propiconazole treatment with Arabidopsis and rice P450 genes. Nine CYP clans were color coded, and the two transcripts from *F. brevipila* were highlighted in orange dots (A). The zoomed regions where the two transcripts were located (B, C). The transcript_34799 was a sister to rice CYP92A9 and Trinity_dn78694_c0_g4_i1 was close to the *Arabidopsis* CYP71A gene family.

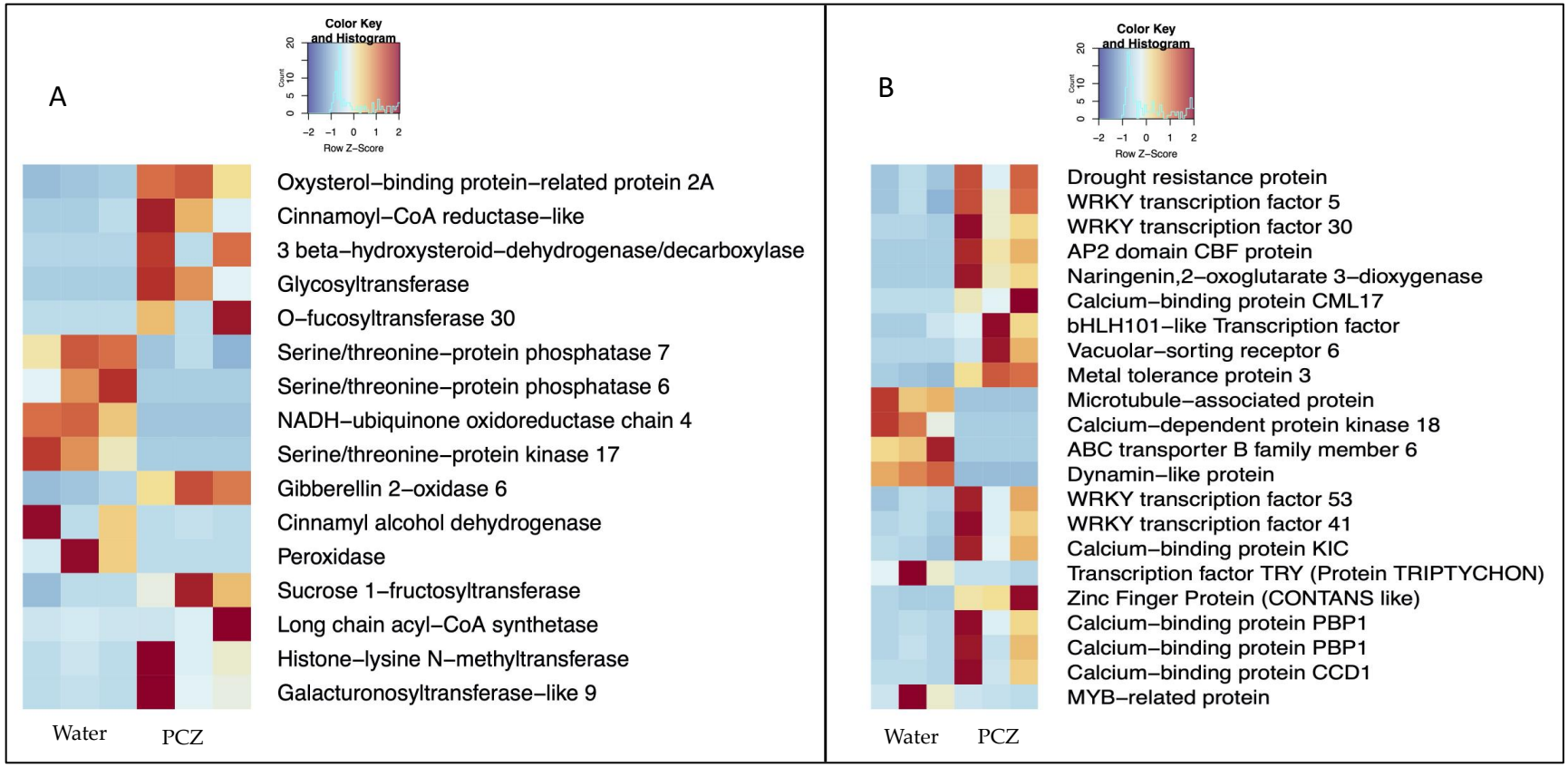

Figure 5.8. Subsets of enzymes (A) and transcriptional factors (B) affected by the propiconazole (PCZ) fungicide application in *F. brevipila*.

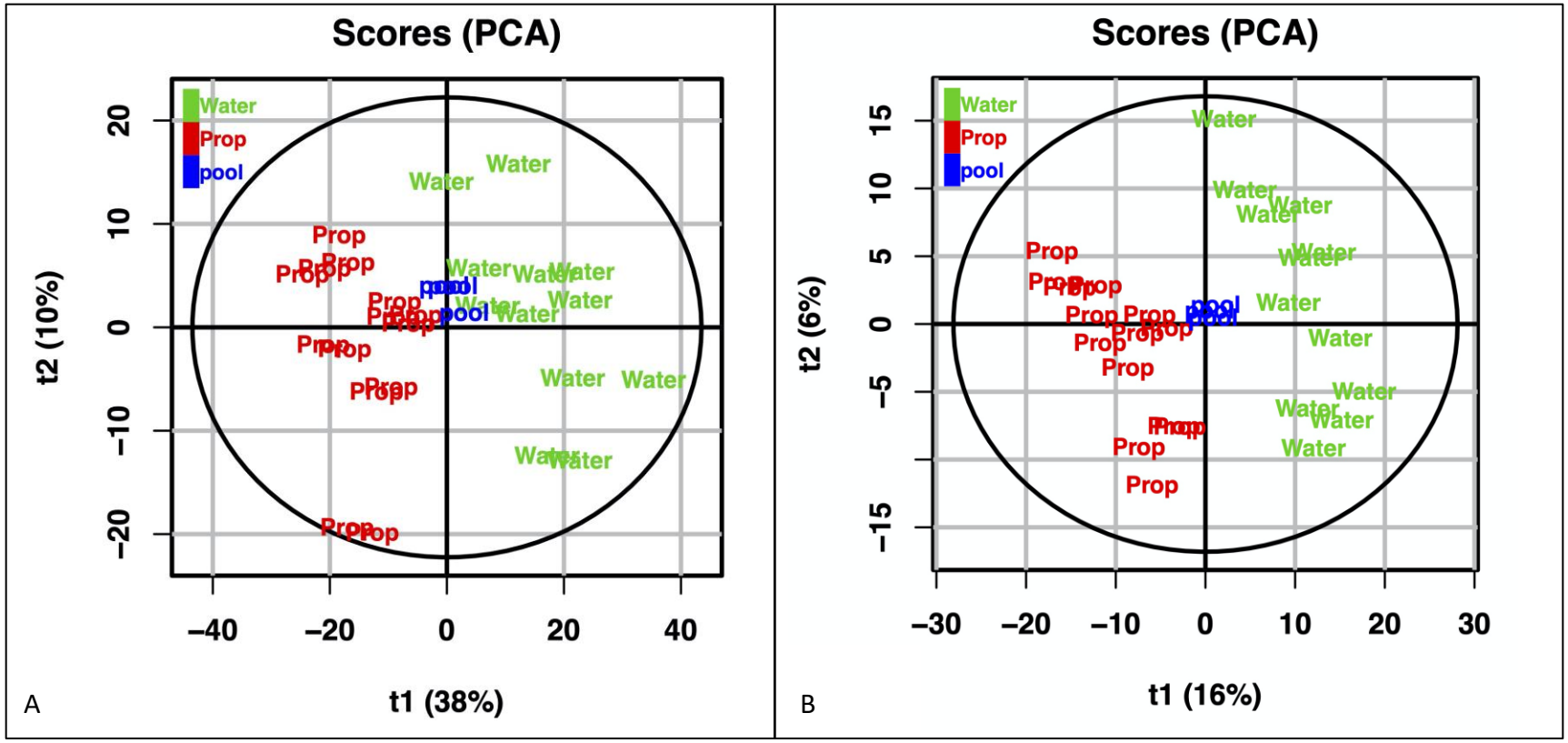

Figure 5.9. The multivariate analysis of metabolite features generated from HILIC column. The unsupervised metabolite clustering for positive (A) and negative (B) modes showed treatment effects on the primary metabolites profile.

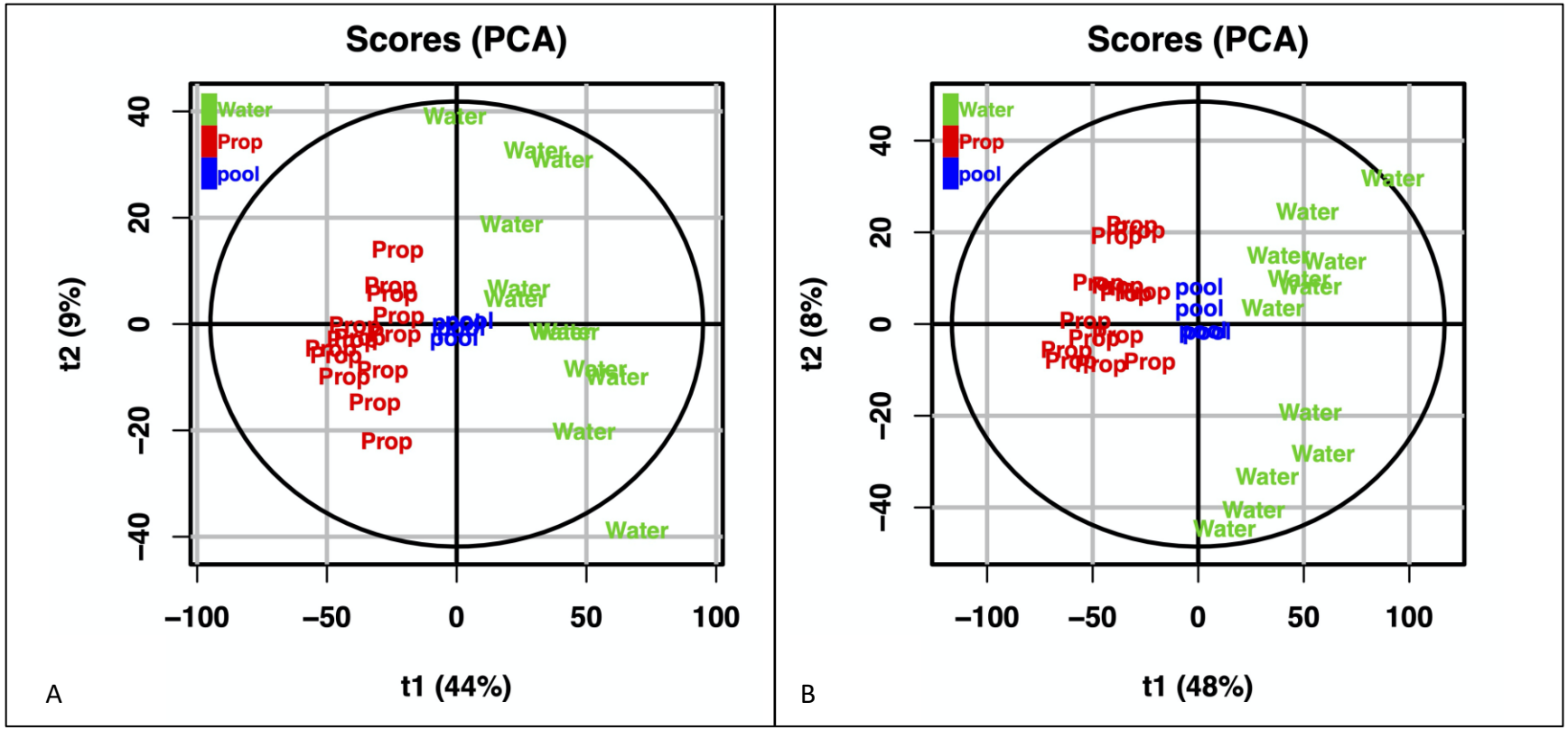

Figure 5.10. The multivariate analysis of metabolite features generated from C18 column. The unsupervised metabolite clustering for C18 positive (A) and C18 negative (B) modes showed treatment effect on the primary metabolites profile.

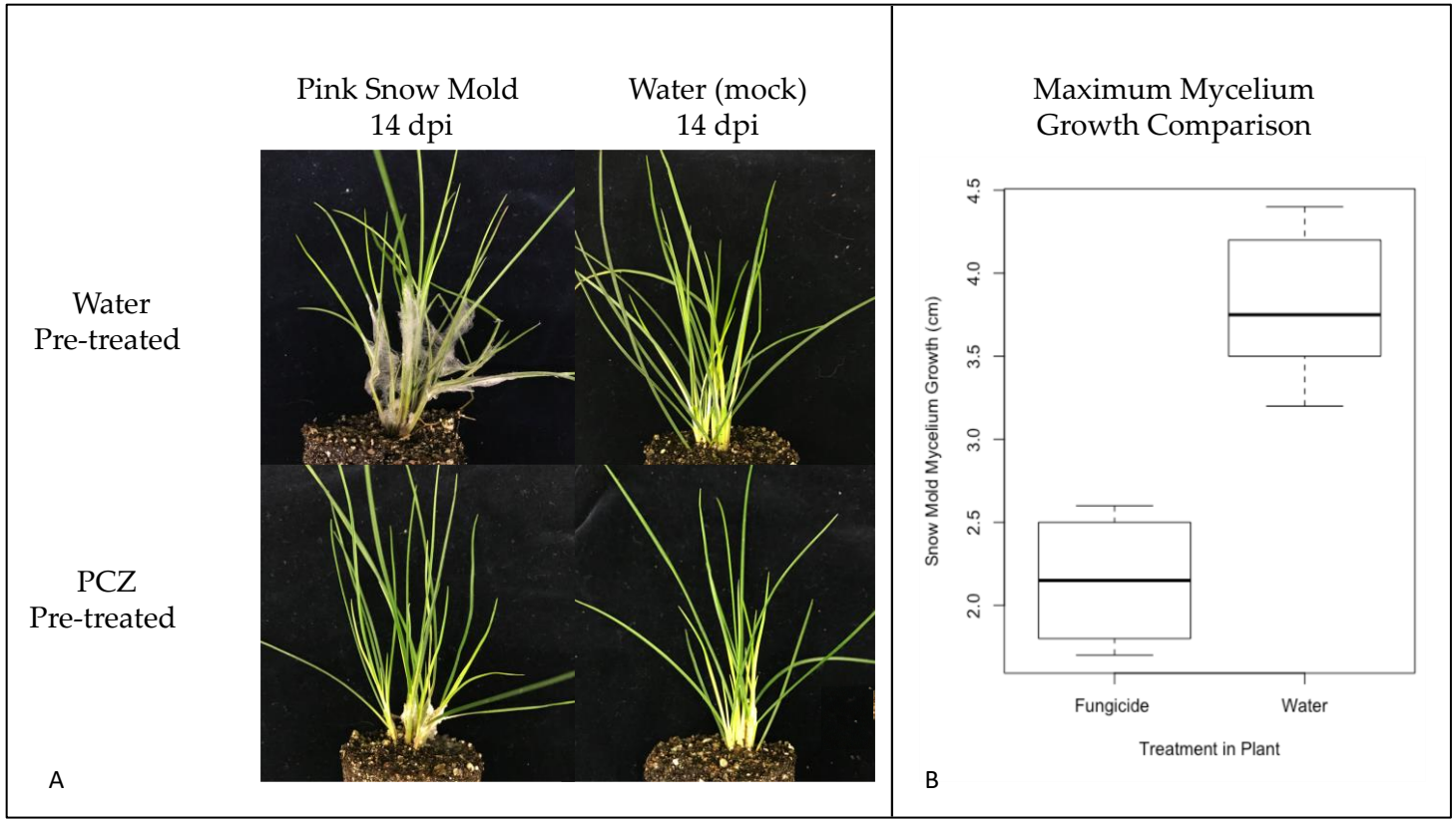

Figure 5.11. Pink snow mold inoculation showed WS plants had significantly more snow mold spread compared to other treatments. FS plants had some level of snow mold infection. PCZ: propiconazole; WS: water pretreated plant with snow mold; FS: fungicide pretreated plant with snow mold.

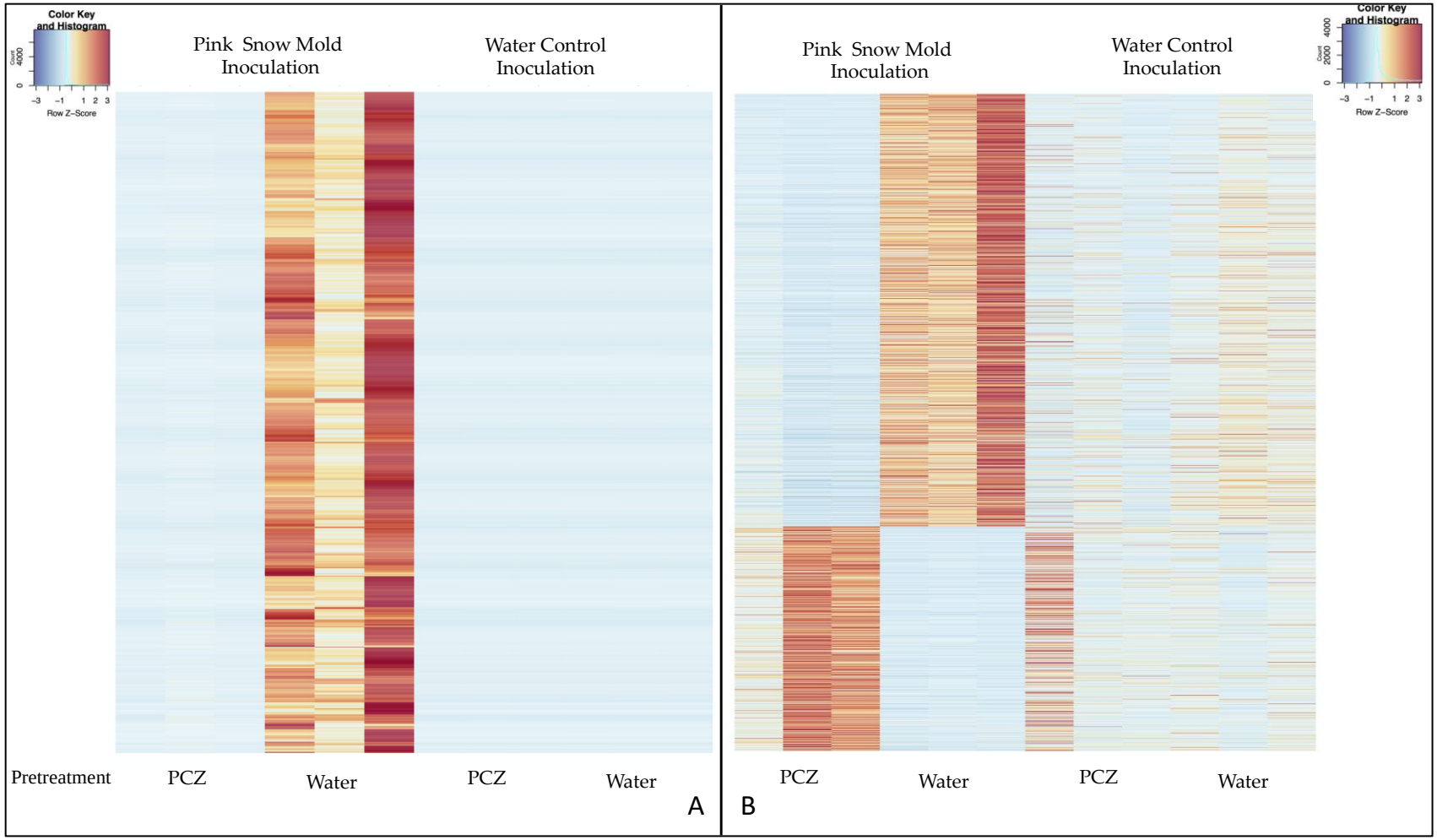

Figure 5.12. Differentially expressed *Microdochium*-related genes (A), and plant-related genes (B). Propiconazole (PCZ)

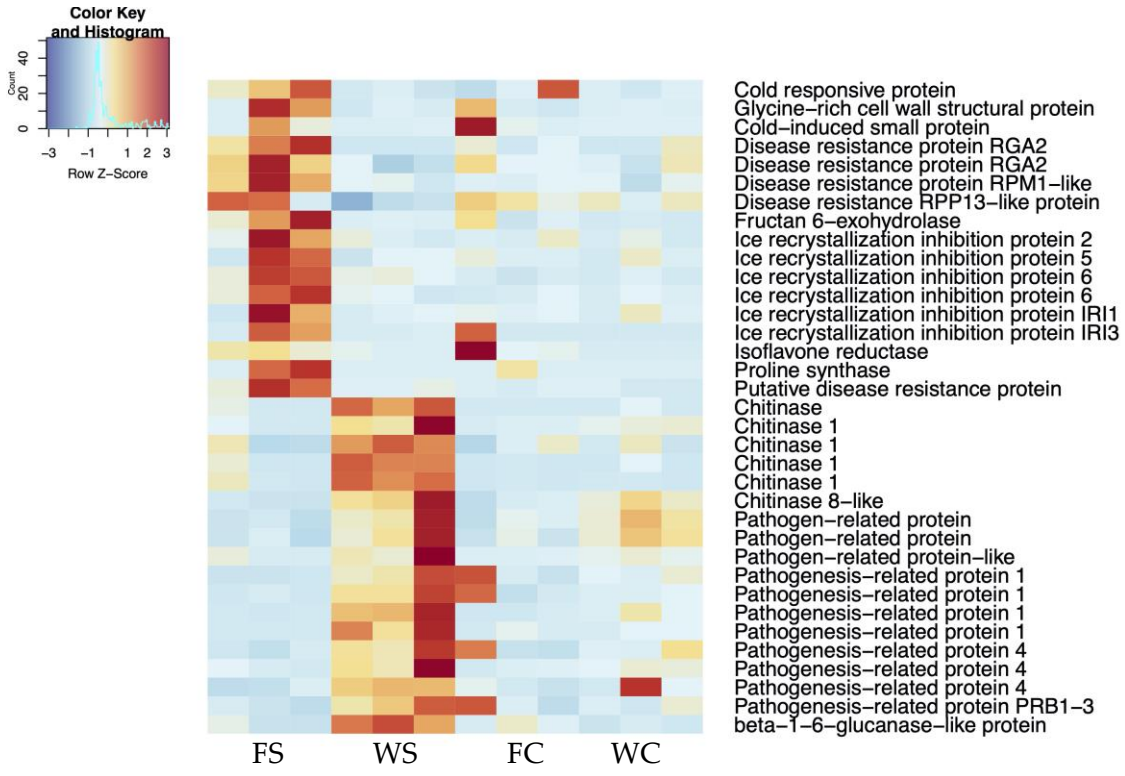

Figure 5.13. Cold and pathogen infection related differentially expressed genes between the fungicide pretreated plants with and without snow mold inoculation vs water pretreated plants with and without snow mold inoculation.

Table 5.1. RNA-Seq yield and reads mapping for the propiconazole study using Isoseq, *de novo*, and combine transcriptome, respectively. (PCZ: Propiconazole)

Treatment	Replicates	Total reads	High quality reads	Reference	Mapped reads (%)	Uniquely mapped reads (%) (MAPQ>3)
PCZ	R1	62,266,322	62,262,686	Isoseq	64	57.1
				de novo	48.6	44.7
				Combined	72.4	66.8
	R2	72,353,963	72,349,738	Isoseq	62.7	57.1
				de novo	48.1	44.2
				Combined	71.9	66.2
	R3	65,990,410	65,986,554	Isoseq	65.5	59
				de novo	45.9	42.9
				Combine	72	66.9
Water	R1	64,102,358	64,098,632	Isoseq	63.5	57.2
				de novo	47.7	44
				Combined	71.6	66.3
	R2	59,400,130	59,397,440	Isoseq	64.1	57.2
				de novo	48.4	44.6
				Combined	71.8	65.9
	R3	68,465,442	68,461,826	Isoseq	63.7	56.9
				de novo	48.2	44.5
				Combined	71.9	66.3

Table 5.2. Accurate mass-based metabolite feature prediction using the human metabolome database and KEGG compound database.

Column	ID	Acc. mass mz	rt (min)	Hit(s)	Adduct	Monoisotipic mass	PPM error
C18 Negative Mode	1	431.114346	3.7	Luteolinidin 3-O-glucoside	M-H	433.113472	36
	1	431.114346	3.7	1-O-Sinapoylglucose	M+FA-H	386.121297	13
	2	431.114477	3.7	8-Cinnamoyl-3,4-dihydro-5,7-dihydroxy-4-phenylcoumarin	M+FA-H	386.115424	1
	3	430.118426	3.7	Stylisterol A/C			
	4	581.330936	7.53	hydroxyisovaleroyl carnitine	2M+Hac-H	261.157623	3
	4	581.330936	7.53	Cyanidin 3-O-(beta-D-xylosyl-(1->2)-beta-D-galactoside)			
	5	431.118602	3.7	1-O-Sinapoylglucose	M+FA-H	386.121297	2
	5	431.118602	3.7	4-O-beta-D-Glucosyl-sinapate	M+FA-H	386.121297	2
	5	431.118602	3.7	trans (cis)-Zeatin riboside monophosphate			
	6	621.092739	6.18	Quercetin 3-(2''-caffeylglucuronide)	M-H2O-H	640.106435	7
	6	621.092739	6.18	Quercetin 3-(2-caffeoylglucuronoside)	M-H2O-H	640.106435	7
	7	621.114766	6.18	Apigenin 7,4'-diglucuronide	M-H	622.116999	8
	8	259.120209	7.98	UNKNOWN			
	9	621.100141	6.18	UNKNOWN			
	10	621.107399	6.18	Apigenin 7,4'-diglucuronide	M-H	622.116999	3

C18 Positive Mode	11	541.285259	9.33	Glycinoeclepin B	M+H-H2O	558.282883	0
	12	541.279291	9.33	UNKNOWN			
	13	541.291234	9.33	Medicagenic acid (terpenoid)	M+K	502.329439	5
	14	559.303626	7.55	Auriculine (alkloid in orchid)			
	15	542.28991	9.33	4-Coumaroyl-2-hydroxyputrescine	2M+ACN+H	250.131743	14
	16	563.227831	5.45	UNKNOWN			
	17	385.160119	5.45	Secoisolariciresinol	M+Na	362.172939	5
	17	385.160119	5.45	Gibberellin A38	M+Na	362.172939	
	17	385.160119	5.45	Gibberellin A19, 65, 101,115, 124	M+Na	362.172939	5
	18	319.153044	2.73	UNKNOWN			
	19	563.221492	5.45	Geranylfarnesyl diphosphate	M+2Na-H	518.256227	8
	20	563.215151	5.45	(+)-7-epi-Syringaresinol 4'-glucoside	M+H-H2O	580.215591	4
cHILIC Negative Mode	21	596.266442	12.34	UNKNOWN			
	22	449.20306	9.46	Hydrocinnamic acid	3M-H	150.06808	13
	23	432.196476	6.97	UNKNOWN			
	24	591.15752	16.41	UNKNOWN			
	25	442.17208	5.31	UNKNOWN			
	26	250.995649	16.28	Xylulose 5-phosphate (sugar phosphate)	M+Na-2H	230.019154	7
	27	189.076748	7.73	Ethyl beta-D-glucopyranoside	M-H2O-H	208.094688	2
	27	189.078558	7.73	Di-4-coumaroylputrescine	M-2H	380.173607	5
	27	189.078558	7.73	3-Hydroxysuberic acid	M-H	190.084124	9

	28	227.034079	16.26	Quinic acid	M+Cl	192.063388	6
	28	227.034079	16.26	3-Hydroxysuberic acid	M+K-2H	190.084124	6
	29	278.099268	16.02	UNKNOWN			
	30	630.819578	19.26	N1, N5,N10-Tricaffeoyl spermidine			
cHILIC Positive Mode	31	431.118632	13.9	UNKNOWN			
	32	448.145051	13.89	UNKNOWN			
	33	432.121856	13.89	3-O-Caffeoyl-4-O-methylquinic acid	M+ACN+Na	368.110732	11
	34	432.120673	13.89	5-Feruloylquinic acid	M+ACN+Na	368.110732	14
	35	209.153758	6.12	3-Hydroxy-beta-ionone	M+H	208.14633	0
	35	209.153787	12.34	Methyl dihydrojasmonate	M+H-H2O	226.156895	2
	36	204.086871	19.82	beta-N-Acetylglucosamine	M+H-H2O	221.089937	2
	36	204.086871	19.82	Avenic acid B	M+H-H2O	221.089937	2
	37	689.211234	22.39	Glycogen (hexose polymer)	M+Na	666.221858	0
	38	228.099577	19.77	Eugenol	M+ACN+Na	164.08373	0
	38	228.099577	19.77	Benzenebutanoic acid	M+ACN+Na	164.08373	0
	39	100.988462	19.2	UNKNOWN			
	40	261.037152	19.82	Hexose phosphate (eg. Glucose 6-phosphate)	M+H	260.029719	1

Table 5.3. Cold and pathogen infection related DEGs between the FS and WS treatments.

Transcript ID	NCBI NR Based Transcript Annotation	LogFC (FS/WS)	*p*-value
TRINITY_DN72353_C1_G2_I2	Cold responsive protein	2.60	2.0E-06
TRINITY_DN69307_C0_G1_I7	Glycine-rich cell wall structural protein	3.34	2.0E-03
TRINITY_DN88735_C0_G1_I2	Cold-induced small protein	3.79	1.5E-03
TRINITY_DN74175_C0_G1_I10	Disease resistance protein RGA2	7.77	1.9E-05
TRINITY_DN74175_C0_G1_I3	Disease resistance protein RGA2	2.66	1.3E-04
TRANSCRIPT_12036	Disease resistance protein RPM1-like	2.18	1.9E-04
TRINITY_DN81777_C0_G1_I12	Disease resistance RPP13-like protein	2.28	8.6E-04
TRINITY_DN100146_C0_G1_I9	Fructan 6-exohydrolase	2.50	1.5E-04
TRINITY_DN97131_C2_G5_I3	Ice recrystallization inhibition protein 2	2.70	2.7E-03
TRANSCRIPT_46772	Ice recrystallization inhibition protein 5	2.11	1.5E-03
TRANSCRIPT_48329	Ice recrystallization inhibition protein 6	2.05	3.3E-04
TRINITY_DN86392_C0_G9_I1	Ice recrystallization inhibition protein 6	2.81	2.7E-04
TRINITY_DN77946_C1_G1_I3	Ice recrystallization inhibition protein IRI1	2.26	2.2E-03
TRINITY_DN72600_C2_G3_I3	Ice recrystallization inhibition protein IRI3	4.50	4.9E-04
TRINITY_DN100575_C2_G2_I8	Isoflavone reductase	2.13	3.7E-04
TRINITY_DN105886_C6_G1_I5	Proline synthase	8.02	4.0E-04
TRANSCRIPT_3235	Putative disease resistance protein	2.16	6.5E-04
TRANSCRIPT_53735	Chitinase	-3.91	2.4E-03
TRINITY_DN87228_C0_G3_I17	Chitinase 1	-3.82	1.0E-05
TRINITY_DN80789_C0_G3_I4	Chitinase 1	-2.47	2.4E-03
TRINITY_DN80789_C0_G3_I2	Chitinase 1	-3.53	1.6E-03
TRANSCRIPT_54666	Chitinase 1	-3.13	1.4E-03

TRANSCRIPT_50472	Chitinase 8-like	-2.07	2.0E-05
TRINITY_DN98644_C1_G1_I2	Pathogen-related protein	-2.55	5.5E-05
TRINITY_DN73403_C1_G1_I1	Pathogen-related protein	-2.67	1.7E-03
TRANSCRIPT_52665	Pathogen-related protein-like	-3.08	3.9E-04
TRINITY_DN91197_C0_G3_I1	Pathogenesis-related protein 1	-2.23	2.3E-04
TRINITY_DN91197_C0_G2_I1	Pathogenesis-related protein 1	-2.23	1.0E-03
TRANSCRIPT_49936	Pathogenesis-related protein 1	-2.80	2.8E-09
TRANSCRIPT_47468	Pathogenesis-related protein 1	-6.88	9.6E-24
TRINITY_DN91197_C0_G1_I4	Pathogenesis-related protein 4	-2.69	8.5E-05
TRINITY_DN87228_C0_G3_I5	Pathogenesis-related protein 4	-3.47	8.5E-07
TRINITY_DN87228_C0_G3_I4	Pathogenesis-related protein 4	-2.05	2.8E-04
TRINITY_DN102652_C0_G1_I23	Pathogenesis-related protein PRB1-3	-2.99	3.0E-06

Chapter 6

General Discussion and Perspectives

The work presented in this dissertation covered diverse topics regarding fine fescue breeding and genetics. In this work, I first proposed a molecular phylogeny of fine fescues using chloroplast genome sequences of five representative cultivars. Based on these chloroplast genome sequences, I developed chloroplast molecular markers for rapid fine fescue germplasm screening. For fine fescue identification, by using flow cytometry, we can identify *F. ovina* and *F. rubra* ssp. *rubra*. Because *F. brevipila*, *F. rubra* ssp. *fallax* and *F. rubra* ssp. *littoralis* have the same ploidy level, molecular marker assays can be carried out using chloroplast genome-based primers developed in this dissertation. These primers allow for separation of *F. ovina* from *F. brevipila*; *F. rubra* ssp. *fallax* from *F. rubra* ssp. *littoralis* and *F. brevipila*. Rhizome development provides additional separation between *F. rubra* ssp. *littoralis* and *F. brevipila*. By combining these approaches, we are now able to provide a more accurate fine fescue identification. Following the fine fescue identification topics, flow cytometry was performed to estimate the DNA content and ploidy levels of USDA *F. ovina* collections. The result showed species with multiple ploidy levels were presented in the collections, suggesting that plant breeders should inspect and evaluate the germplasm requested from USDA prior to adapting them in the breeding programs.

To develop genomics resources for *Festuca brevipila* transcriptomics study, the PacBio Isoform sequencing was carried out using four tissues (root, crown, leaf, and

seedhead) to present diverse transcript pools. The non-redundant PacBio Isoform transcriptome has proven to improve the short reads mapping quality and also helped to resolve the question on the allohexaploid nature of this species. Finally, an experiment was carried out to evaluate the propiconazole-fungicide plant interaction, and study if the rapid fungicide application would create a long-term effect in the stress tolerance of plants. Transcriptome and metabolomics data suggested the fungicide affected the brassinosteroid biosynthesis pathway, and potentially affect the lignin biosynthesis. The long-term effect of propiconazole fungicide on *F. brevipila* showed an increase of the pink snow mold disease resistance, likely through a fast response to produce antifreezing proteins to enhance the plant's performance. Future study will be to carry out to study if these antifreezing proteins also have antifungal effect. The result also suggests the potential confounding result for using fungicide to remove endophytes to study endophyte effect.

The genomics work produced in this dissertation lay the foundation for a more accurate gene expression quantification in hard fescues. Future whole genome sequencing project will create more genomics resources for this turfgrass species. Methods developed for fine fescue identification and genomics resources generated in this thesis provide breeders and turfgrass researchers tools to advance the fine fescue breeding in a molecular breeding era.

Bibliography

Alkan, C., Sajjadian, S., and Eichler, E.E. (2011). Limitations of next-generation genome sequence assembly. *Nature Methods* 8, 61.

Arumuganathan, K., and Earle, E. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* 9, 208-218.

Arumuganathan, K., Tallury, S., Fraser, M., Bruneau, A., and Qu, R. (1999). Nuclear DNA content of thirteen turfgrass species by flow cytometry. *Crop Science* 39, 1518-1521.

Asami, T., Mizutani, M., Fujioka, S., Goda, H., Min, Y., Shimada, Y., Nakano, T., Takatsuto, S., Matsuyama, T., and Nagata, N. (2001). Selective interaction of triazole derivatives with DWF4, a P450 monooxygenase of the brassinosteroid pathway, correlates with brassinosteroid deficiency in planta. *J Biol Chem* 276, 25687-25691.

Asami, T., Mizutani, M., Shimada, Y., Goda, H., Kitahata, N., Sekimata, K., Han, S.-Y., Fujioka, S., Takatsuto, S., and Sakata, K. (2003). Triadimefon, a fungicidal triazole-type P450 inhibitor, induces brassinosteroid deficiency-like phenotypes in plants and binds to DWF4 protein in the brassinosteroid biosynthesis pathway. *Biochemical Journal* 369, 71-76.

Bacon, C.W., Richardson, M., and White, J.F. (1997). Modification and uses of endophyte-enhanced turfgrasses: a role for molecular technology. *Crop Science* 37, 1415-1425.

Baird, J.H., Kopecký, D., Lukaszewski, A.J., Green, R.L., Bartoš, J., and Doležel, J. (2012). Genetic diversity of turf-type tall fescue using diversity arrays technology. *Crop Science* 52, 408-412.

Bajguz, A., and Hayat, S. (2009). Effects of brassinosteroids on the plant responses to environmental stresses. *Plant Physiology and Biochemistry* 47, 1-8.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., and Donoghue, M.J. (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, 247-277.

Bancos, S., Szatmári, A.-M., Castle, J., Kozma-Bognár, L., Shibata, K., Yokota, T., Bishop, G.J., Nagy, F., and Szekeres, M. (2006). Diurnal regulation of the brassinosteroid-biosynthetic CPD gene in Arabidopsis. *Plant Physiology* 141, 299-309.

Barkworth, M., Capels, K., Long, S., Anderton, L., and Piep, M. (2007). "Magnoliophyta Commelinidae (in part: Poaceae, part 1). Flora of North America North of Mexico. Vol. 24". Oxford University Press: New York).

Beard, J.B. (1972). Turfgrass: Science and culture. N.J., Prentice-Hall

Belanger, F., Bonos, S., and Meyer, W. (2004). Dollar spot resistant hybrids between creeping bentgrass and colonial bentgrass. *Crop Science* 44, 581-586.

Bertrand, A., Castonguay, Y., Azaiez, A., Hsiang, T., and Dionne, J. (2011). Cold-induced responses in annual bluegrass genotypes with differential resistance to pink snow mold (*Microdochium nivale*). *Plant science* 180, 111-119.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.
Bonos, S.A., Clarke, B.B., and Meyer, W.A. (2006). Breeding for disease resistance in the major cool-season turfgrasses. *Annu. Rev. Phytopathol.* 44, 213-234.
Bonos, S.A., and Huff, D.R. (2013). Cool-season grasses: Biology and breeding. *Turfgrass: Biology, Use, and Management*, 591-660.
Bonos, S.A., Wilson, M.M., Meyer, W.A., and Reed Funk, C. (2005). Suppression of red thread in fine fescues through endophyte-mediated resistance. *Applied Turfgrass Science* 2, 0-0.
Bozak, K.R., Yu, H., Sirevåg, R., and Christoffersen, R.E. (1990). Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit. *Proceedings of the National Academy of Sciences* 87, 3904-3908.
Bretagnolle, F., Thompson, J., and Lumaret, R. (1995). The influence of seed size variation on seed germination and seedling vigour in diploid and tetraploid *Dactylis glomerata* L. *Annals of Botany* 76, 607-615.
Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., Batzoglou, S., and Program, N.C.S. (2003). LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Research* 13, 721-731.
Bryan, G., Mcnicoll, J., Ramsay, G., Meyer, R., and De Jong, W. (1999). Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. *Theoretical and Applied Genetics* 99, 859-867.
Budak, H., Shearman, R., Gaussoin, R.E., and Dweikat, I. (2004). Application of sequence-related amplified polymorphism markers for characterization of turfgrass species. *HortScience* 39, 955-958.
Burden, R.S., Cooke, D.T., and Carter, G.A. (1989). Inhibitors of sterol biosynthesis and growth in plants and fungi. *Phytochemistry* 28, 1791-1804.
Burson, B., Actkinson, J., Hussey, M., and Jessup, R. (2012). Ploidy determination of buffel grass accessions in the USDA National Plant Germplasm System collection by flow cytometry. *South African Journal of Botany* 79, 91-95.
Cahoon, A.B., Sharpe, R.M., Mysayphonh, C., Thompson, E.J., Ward, A.D., and Lin, A. (2010). The complete chloroplast genome of tall fescue (*Lolium arundinaceum*; Poaceae) and comparison of whole plastomes from the family Poaceae. *American Journal of botany* 97, 49-58.
Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972-1973.
Carroll, J. (1943). Effects of drought, temperature and nitrogen on turf grasses. *Plant Physiology* 18, 19.
Casler, M.D. (2003). *Turfgrass biology, Genetics, and Breeding.* John Wiley & Sons.
Ceccarelli, M., Falistocco, E., and Cionini, P. (1992). Variation of genome size and organization within hexaploid *Festuca arundinacea*. *Theoretical and Applied Genetics* 83, 273-278.

Chang, H.-X., and Hartman, G.L. (2017). Characterization of insect resistance loci in the USDA soybean germplasm collection using genome-wide association studies. *Frontiers in Plant Science* 8, 670.
Chao, Y., Yuan, J., Li, S., Jia, S., Han, L., and Xu, L. (2018). Analysis of transcripts and splice isoforms in red clover (*Trifolium pratense* L.) by single-molecule long-read sequencing. *BMC Plant Biology* 18, 300.
Chen, L.-Y., Morales-Briones, D.F., Passow, C.N., and Yang, Y. (2019). Performance of gene expression analyses using de novo assembled transcripts in polyploid species. *Bioinformatics* 35, 4314-4320.
Chen, Z., Wang, M.L., Waltz, C., and Raymer, P. (2009). Genetic diversity of warm-season turfgrass: seashore paspalum, bermudagrass, and zoysiagrass revealed by AFLPs. *Floriculture Ornamental Biotechnol* 3, 20-24.
Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., and Zhou, S. (2016). Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources* 16, 138-149.
Christensen, S., Pratt, D.B., Pratt, C., Nelson, P., Stevens, M., Jellen, E.N., Coleman, C.E., Fairbanks, D.J., Bonifacio, A., and Maughan, P.J. (2007). Assessment of genetic diversity in the USDA and CIP-FAO international nursery collections of quinoa (*Chenopodium quinoa* Willd.) using microsatellite markers. *Plant Genetic Resources* 5, 82-95.
Christiansen, M., Andersen, S.B., and Ortiz, R. (2002). Diversity changes in an intensively bred wheat germplasm during the 20 th century. *Molecular Breeding* 9, 1-11.
Clark, J.M., and Kenna, M.P. (2010). "Lawn and turf: Management and environmental issues of turfgrass pesticides," in *Hayes' Handbook of Pesticide Toxicology*. Elsevier), 1047-1076.
Clarke, B.B., White Jr, J.F., Hurley, R.H., Torres, M.S., Sun, S., and Huff, D.R. (2006). Endophyte-mediated suppression of dollar spot disease in fine fescues. *Plant Disease* 90, 994-998.
Clay, K. (1989). Clavicipitaceous endophytes of grasses: their potential as biocontrol agents. *Mycological Research* 92, 1-12.
Clayton, W.D., and Renvoize, S.A. (1986). Genera graminum. Grasses of the world. *Genera graminum. Grasses of the World.* 13.
Cormack, R.S., Eulgem, T., Rushton, P.J., Köchner, P., Hahlbrock, K., and Somssich, I.E. (2002). Leucine zipper-containing WRKY proteins widen the spectrum of immediate early elicitor-induced WRKY transcription factors in parsley. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* 1576, 92-100.
Demesure, B., Sodzi, N., and Petit, R. (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4, 129-134.
Diekmann, K., Hodkinson, T.R., Wolfe, K.H., Van Den Bekerom, R., Dix, P.J., and Barth, S. (2009). Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium perenne* L.). *DNA research* 16, 165-176.

Dierckxsens, N., Mardulyn, P., and Smits, G. (2016). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45, e18-e18.
Dilday, R., Lin, J., and Yan, W. (1994). Identification of allelopathy in the USDA-ARS rice germplasm collection. *Australian Journal of Experimental Agriculture* 34, 907-910.
Doležel, J., and Bartoš, J. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of botany* 95, 99-110.
Doležel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* 2, 2233.
Dong, S., Tredway, L.P., Shew, H.D., Wang, G.-L., Sivamani, E., and Qu, R. (2007). Resistance of transgenic tall fescue to two major fungal diseases. *Plant science* 173, 501-509.
Dong, W., Liu, J., Yu, J., Wang, L., and Zhou, S. (2012). Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PloS one* 7, e35071.
Dubas, E., Golebiowska, G., Zur, I., and Wedzony, M. (2011). *Microdochium nivale* (Fr., Samuels & Hallett): cytological analysis of the infection process in triticale (× *Triticosecale* Wittm.). *Acta physiologiae plantarum* 33, 529-537.
Ebert, D., and Peakall, R. (2009). Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Molecular Ecology Resources* 9, 673-690.
Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792-1797.
Edger, P.P., Poorten, T.J., Vanburen, R., Hardigan, M.A., Colle, M., Mckain, M.R., Smith, R.D., Teresi, S.J., Nelson, A.D., and Wai, C.M. (2019). Origin and evolution of the octoploid strawberry genome. *Nature Genetics* 51, 541-547.
Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *BioRxiv*, 466201.
Ervin, E., Zhang, X., Goatley, J., and Askew, S. (2004). Trinexapac-ethyl, propiconazole, iron, and biostimulant effects on shaded creeping bentgrass. *HortTechnology* 14, 500-506.
Fjellheim, S., Rognli, O.A., Fosnes, K., and Brochmann, C. (2006). Phylogeographical history of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. *Journal of Biogeography* 33, 1470-1478.
Fletcher, R.A., Gilley, A., Sankhla, N., and Davis, T.D. (2000). Triazoles as plant growth regulators and stress protectants. *Horticultural Reviews* 24, 55-138.
Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Research* 32, W273-W279.
Funk, C., Halisky, P., Johnson, M., Siegel, M., Stewart, A., Ahmad, S., Hurley, R., and Harvey, I. (1983). An endophytic fungus and resistance to sod webworms: Association in *Lolium perenne* L. *Bio/technology* 1, 189.
Gao, F., Dai, C., and Liu, X. (2010). Mechanisms of fungal endophytes in plant protection against pathogens. *Afr J Microbiol Res* 4, 1346-1351.

Gilley, A., and Fletcher, R. (1997). Relative efficacy of paclobutrazol, propiconazole and tetraconazole as stress protectants in wheat seedlings. *Plant Growth Regulation* 21, 169-175.

Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M., and Thomashow, M.F. (1998). Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *The Plant Journal* 16, 433-442.

Goldman, J.J. (2015). DNA contents in Texas bluegrass (*Poa arachnifera*) selected in Texas and Oklahoma determined by flow cytometry. *Genetic Resources and Crop Evolution* 62, 643-647.

Goodman, M.M. (1999). Broadening the genetic diversity in maize breeding by use of exotic germplasm. *The Genetics and Exploitation of Heterosis in Crops*, 139-148.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., and Zeng, Q. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29, 644.

Griffith, M., Antikainen, M., Hon, W.C., Pihakaski-Maunsbach, K., Yu, X.M., Chun, J.U., and Yang, D.S. (1997). Antifreeze proteins in winter rye. *Physiologia Plantarum* 100, 327-332.

Gupta, P., Langridge, P., and Mir, R. (2010). Marker-assisted wheat breeding: present status and future possibilities. *Molecular Breeding* 26, 145-161.

Hallahan, D.L., Heasman, A.P., Grossel, M.C., Quigley, R., Hedden, P., and Bowyer, J.R. (1988). Synthesis and biological activity of an azido derivative of paclobutrazol, an inhibitor of gibberellin biosynthesis. *Plant Physiology* 88, 1425-1429.

Hamberger, B., and Bak, S. (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20120426.

Hand, M.L., Spangenberg, G.C., Forster, J.W., and Cogan, N.O. (2013). Plastome sequence determination and comparative analysis for members of the *Lolium-Festuca* grass species complex. *G3: Genes, Genomes, Genetics*, g3. 112.005264.

Hanson, A.A., and Juska, F.V. (1969). Turfgrass science. *Turfgrass science.*

Hanson, B.D., Mallory-Smith, C.A., Brewster, B.D., Wendling, L.A., and Thill, D.C. (2003). Growth regulator effects of propiconazole on redroot pigweed (Amaranthus retroflexus). *Weed Technology* 17, 777-781.

Hartwig, T., Corvalan, C., Best, N.B., Budka, J.S., Zhu, J.-Y., Choe, S., and Schulz, B. (2012). Propiconazole is a specific and accessible brassinosteroid (BR) biosynthesis inhibitor for Arabidopsis and maize. *PloS one* 7.

Harvey, I., Fletcher, L., and Emms, L. (1982). Effects of several fungicides on the Lolium endophyte in ryegrass plants, seeds, and in culture. *New Zealand Journal of Agricultural Research* 25, 601-606.

Hebert, P.D., Cywinska, A., Ball, S.L., and Dewaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270, 313-321.

Heineck, GC, Qiu, Y, Ehlke, NJ, Watkins, E. The fungal endophyte *Epichloë festucae* var. *lolii* plays a limited role in mediating crown rust severity in perennial ryegrass. (2020). *Crop Science*. 1– 15. https://doi.org/10.1002/csc2.20095

Heineck, G.C., Watkins, E., and Ehlke, N.J. (2018). The Fungal Endophyte *Epichloë festucae* var. *lolii* Does Not Improve the Freezing Tolerance of Perennial Ryegrass. *Crop Science* 58, 1788-1800.

Hesse, U., Schöberlein, W., Wittenmayer, L., Förster, K., Warnstorff, K., Diepenbrock, W., and Merbach, W. (2003). Effects of Neotyphodium endophytes on growth, reproduction and drought-stress tolerance of three *Lolium perenne* L. genotypes. *Grass and Forage Science* 58, 407-415.

Hiatt, E., Hill, N., Bouton, J., and Mims, C. (1997). Monoclonal antibodies for detection of Neotyphodium coenophialum. *Crop Science* 37, 1265-1269.

Hiilovaara-Teijo, M., Hannukkala, A., Griffith, M., Yu, X.-M., and Pihakaski-Maunsbach, K. (1999). Snow-mold-induced apoplastic proteins in winter rye leaves lack antifreeze activity. *Plant Physiology* 121, 665-674.

Hoang, N.V., Furtado, A., Mason, P.J., Marquardt, A., Kasirajan, L., Thirugnanasambandam, P.P., Botha, F.C., and Henry, R.J. (2017). A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genomics* 18, 395.

Hon, W.-C., Griffith, M., Mlynarz, A., Kwok, Y.C., and Yang, D.S. (1995). Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiology* 109, 879-889.

Huang, J., Tang, D., Shen, Y., Qin, B., Hong, L., You, A., Li, M., Wang, X., Yu, H., and Gu, M. (2010). Activation of gibberellin 2-oxidase 6 decreases active gibberellin levels and creates a dominant semi-dwarf phenotype in rice (Oryza sativa L.). *Journal of Genetics and Genomics* 37, 23-36.

Huang, T., Rehak, L., and Jander, G. (2012). meta-Tyrosine in Festuca rubra ssp. commutata (Chewings fescue) is synthesized by hydroxylation of phenylalanine. *Phytochemistry* 75, 60-66.

Huang, Y.-Y., Cho, S.-T., Haryono, M., and Kuo, C.-H. (2017). Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. *PloS one* 12, e0179055.

Hubbard, C.E. (1968). Grasses. A guide to their structure, identification, uses, and distribution in the British Isles. *Grasses. A guide to their structure, identification, uses, and distribution in the British Isles.*

Huff, D.R., and Palazzo, A.J. (1998). Fine fescue species determination by laser flow cytometry. *Crop Science* 38, 445-450.

Jenderek, M.M., and Hannan, R.M. (2004). Variation in reproductive characteristics and seed production in the USDA garlic germplasm collection. *HortScience* 39, 485-488.

Jenkin, T., and Jenkin, T. (1955). Interspecific and intergeneric hybrids in herbage grasses-XIV. The breeding affinities of *Festuca ovina*. *Journal of Genetics* 53.

Jenkin, T.J. (1959). Fescue Species (*Festuca* L.). *In: Roemer, T. & W. Rudorf. Handbuch der Pflanzenzüchtung*, 418–434.

Jo, Y.-K., Won Chang, S., Boehm, M., and Jung, G. (2008). Rapid development of fungicide resistance by *Sclerotinia homoeocarpa* on turfgrass. *Phytopathology* 98, 1297-1304.

John, U.P., Polotnianka, R.M., Sivakumaran, K.A., Chew, O., Mackin, L., Kuiper, M.J., Talbot, J.P., Nugent, G.D., Mautord, J., and Schrauf, G.E. (2009). Ice recrystallization inhibition proteins (IRIPs) and freeze tolerance in the cryophilic Antarctic hair grass *Deschampsia antarctica* E. Desv. *Plant, Cell & Environment* 32, 336-348.

Johnson, P.G., Kenworthy, K.E., Auld, D.L., and Riordan, T.P. (2001). Distribution of buffalograss polyploid variation in the southern Great Plains. *Crop Science* 41, 909-913.

Johnson, P.G., Riordan, T., and Arumuganathan, K. (1998). Ploidy level determinations in buffalograss clones and populations. *Crop Science* 38, 478-482.

Joyce, S. (1998). Why the grass isn't always greener. *Environmental Health Perspectives* 106, A378-A385.

Jung, G., Chang, S. W., & Jo, Y. K. (2007). A fresh look at fungicides for snow mold control. Golf Course Management, 7, 91-94.

Kahle, D., and Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal* 5, 144-161.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14, 587.

Kane, K.H. (2011). Effects of endophyte infection on drought stress tolerance of *Lolium perenne* accessions from the Mediterranean region. *Environmental and Experimental Botany* 71, 337-344.

Kantar, M., Unver, T., and Budak, H. (2010). Regulation of barley miRNAs upon dehydration stress correlated with target gene expression. *Functional & Integrative Genomics* 10, 493-507.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30, 772-780.

Kawasaki, T., Koita, H., Nakatsubo, T., Hasegawa, K., Wakabayashi, K., Takahashi, H., Umemura, K., Umezawa, T., and Shimamoto, K. (2006). Cinnamoyl-CoA reductase, a key enzyme in lignin biosynthesis, is an effector of small GTPase Rac in defense signaling in rice. *Proceedings of the National Academy of Sciences* 103, 230-235.

Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research* 12, 656-664.

Kidwell, M.G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115, 49-63.

Kovi, M.R., Abdelhalim, M., Kunapareddy, A., Ergon, Å., Tronsmo, A.M., Brurberg, M.B., Hofgaard, I.S., Asp, T., and Rognli, O.A. (2016). Global transcriptome

changes in perennial ryegrass during early infection by pink snow mould. *Scientific Reports* 6, 28702.
Kozomara, A., and Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* 42, D68-D73.
Krans, J.V., and Morris, K. (2007). Determining a profile of protocols and standards used in the visual field assessment of turfgrasses: A survey of national turfgrass evaluation program-sponsored university scientists. *Applied Turfgrass Science* 4, 0-0.
Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29, 4633-4642.
Kuwabara, C., Takezawa, D., Shimada, T., Hamada, T., Fujikawa, S., and Arakawa, K. (2002). Abscisic acid-and cold-induced thaumatin-like protein in winter wheat has an antifungal activity against snow mould, *Microdochium nivale*. *Physiologia Plantarum* 115, 101-110.
Kwok, I.M.Y., and Loeffler, R.T. (1993). The biochemical mode of action of some newer azole fungicides. *Pesticide Science* 39, 1-11.
Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357.
Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research* 32, 11-16.
Lassmann, T., Hayashizaki, Y., and Daub, C.O. (2010). SAMStat: monitoring biases in next generation sequencing data. *Bioinformatics* 27, 130-131.
Latch, G., and Christensen, M. (1982). Ryegrass endophyte, incidence, and control. *New Zealand Journal of Agricultural Research* 25, 443-448.
Leitch, A., and Leitch, I. (2008). Genomic plasticity and the diversity of polyploid plants. *Science* 320, 481-483.
Leng, Y., Wang, R., Ali, S., Zhao, M., and Zhong, S. (2016). Sources and genetics of spot blotch resistance to a new pathotype of *Cochliobolus sativus* in the USDA National small grains collection. *Plant Disease* 100, 1988-1993.
Li, A., Zhang, J., and Zhou, Z. (2014). PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics* 15, 311.
Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754-1760.
Li, Q.-F., Wang, C., Jiang, L., Li, S., Sun, S.S., and He, J.-X. (2012). An interaction between BZR1 and DELLAs mediates direct signaling crosstalk between brassinosteroids and gibberellins in Arabidopsis. *Sci. Signal.* 5, ra72-ra72.
Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658-1659.
Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451-1452.

Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics* 52, 267-274.
Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25, 955-964.
Ma, Q.-H. (2010). Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. *Journal of Experimental Botany* 61, 2735-2744.
Ma, X., and Huang, B. (2016). Gibberellin-stimulation of rhizome elongation and differential GA-responsive proteomic changes in two grass species. *Frontiers in Plant Science* 7, 905.
Ma, Y., Staub, J.E., Robbins, M.D., Johnson, P.G., and Larson, S.R. (2014). Phenotypic and genetic characterization of Kyrgyz fine-leaved *Festuca valesiaca* germplasm for use in semi-arid, low-maintenance turf applications. *Genetic Resources and Crop Evolution* 61, 185-197.
Maluszynska, J. (2003). "Cytogenetic tests for ploidy level analyses—chromosome counting," in *Doubled haploid production in crop plants*. Springer, 391-395.
Mandáková, T., and Lysak, M.A. (2018). Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology* 42, 55-65.
Mandava, N.B. (1988). Plant growth-promoting brassinosteroids. *Annual Review of Plant Physiology and Plant Molecular Biology* 39, 23-52.
Mccormick, R.F., Truong, S.K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., Kennedy, M., Amirebrahimi, M., Weers, B.D., and Mckinley, B. (2018). The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *The Plant Journal* 93, 338-354.
Medina, J., Bargues, M., Terol, J., Pérez-Alonso, M., and Salinas, J. (1999). The Arabidopsis CBF gene family is composed of three genes encoding AP2 domain-containing proteins whose expression is regulated by low temperature but not by abscisic acid or dehydration. *Plant Physiology* 119, 463-470.
Mejía, L.C., Rojas, E.I., Maynard, Z., Van Bael, S., Arnold, A.E., Hebbar, P., Samuels, G.J., Robbins, N., and Herre, E.A. (2008). Endophytic fungi as biocontrol agents of Theobroma cacao pathogens. *Biological Control* 46, 4-14.
Melchinger, A. (1999). "Genetic diversity and heterosis". chapter.
Meyer, W.A., and Funk, C.R. (1989). Progress and Benefits to Humanity from Breeding Cool-Season Grasses for Turf 1. *Contributions from Breeding Forage and Turf Grasses*, 31-48.
Mikel, M.A., Diers, B.W., Nelson, R.L., and Smith, H.H. (2010). Genetic diversity and agronomic improvement of North American soybean germplasm. *Crop Science* 50, 1219-1229.
Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, i541-i548.
Nabati, D., Schmidt, R., and Parrish, D. (1994). Alleviation of salinity stress in Kentucky bluegrass by plant growth regulators and iron. *Crop Science* 34, 198-202.

Nafisi, M., Goregaoker, S., Botanga, C.J., Glawischnig, E., Olsen, C.E., Halkier, B.A., and Glazebrook, J. (2007). Arabidopsis cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. *The Plant Cell* 19, 2039-2052.

Nakashita, H., Yasuda, M., Nitta, T., Asami, T., Fujioka, S., Arai, Y., Sekimata, K., Takatsuto, S., Yamaguchi, I., and Yoshida, S. (2003). Brassinosteroid functions in a broad range of disease resistance in tobacco and rice. *The Plant Journal* 33, 887-898.

Nelson, D.R., Schuler, M.A., Paquette, S.M., Werck-Reichhart, D., and Bak, S. (2004). Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiology* 135, 756-772.

Nelson, R.L., Amdor, P.J., and Orf, J.H. (1987). "Evaluation of the USDA soybean germplasm collection: Maturity groups 000 to IV (PI 273.483 to PI 427.107)".).

Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32, 268-274.

Noguchi, T., Fujioka, S., Choe, S., Takatsuto, S., Tax, F.E., Yoshida, S., and Feldmann, K.A. (2000). Biosynthetic pathways of brassinolide in Arabidopsis. *Plant Physiology* 124, 201-210.

Noguchi, T., Fujioka, S., Choe, S., Takatsuto, S., Yoshida, S., Yuan, H., Feldmann, K.A., and Tax, F.E. (1999). Brassinosteroid-insensitive dwarf mutants of Arabidopsis accumulate brassinosteroids. *Plant Physiology* 121, 743-752.

Nomura, T., and Bishop, G.J. (2006). Cytochrome P450s in plant steroid hormone synthesis and metabolism. *Phytochemistry Reviews* 5, 421-432.

Oh, K., Matsumoto, T., Hoshi, T., and Yoshizawa, Y. (2016). In vitro and in vivo evidence for the inhibition of brassinosteroid synthesis by propiconazole through interference with side chain hydroxylation. *Plant Signaling & Behavior* 11, e1158372.

Ohnishi, T., Szatmari, A.-M., Watanabe, B., Fujita, S., Bancos, S., Koncz, C., Lafos, M., Shibata, K., Yokota, T., and Sakata, K. (2006). C-23 hydroxylation by Arabidopsis CYP90C1 and CYP90D1 reveals a novel shortcut in brassinosteroid biosynthesis. *The Plant Cell* 18, 3275-3288.

Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics* 12, 87-98.

Pastor-Corrales, M. (2003). Sources, genes for resistance and pedigrees of 52 rust and mosaic resistant dry bean germplasm lines released by the USDA Beltsville bean project in collaboration. *Science* 29, 834-835.

Peng, X., Hu, Y., Tang, X., Zhou, P., Deng, X., Wang, H., and Guo, Z. (2012). Constitutive expression of rice WRKY30 gene increases the endogenous jasmonic acid accumulation, PR gene expression and resistance to fungal pathogens in rice. *Planta* 236, 1485-1498.

Peyton, L., Gallagher, S., and Hashemzadeh, M. (2015). Triazole antifungals: a review. *Drugs Today (Barc)* 51, 705-718.

Phing Lau, W.C., Latif, M.A., Y. Rafii, M., Ismail, M.R., and Puteh, A. (2016). Advances to improve the eating and cooking qualities of rice by marker-assisted breeding. *Critical Reviews in Biotechnology* 36, 87-98.

Piper, C.V. (1906). *North American species of Festuca.* US Government Printing Office.

Pociecha, E., Płażek, A., Janowiak, F., and Zwierzykowski, Z. (2008). ABA level, proline and phenolic concentration, and PAL activity induced during cold acclimation in androgenic Festulolium forms with contrasting resistance to frost and pink snow mould (*Microdochium nivale*). *Physiological and Molecular Plant Pathology* 73, 126-132.

Prasanna, B. (2012). Diversity in global maize germplasm: characterization and utilization. *Journal of Biosciences* 37, 843-855.

Provan, J., Powell, W., and Hollingsworth, P.M. (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* 16, 142-147.

Qiu, Y., Hamernick, S., Ortiz, J.B., and Watkins, E. (2020a). DNA Content and Ploidy Estimation of *Festuca ovina* Accessions by Flow Cytometry. *bioRxiv*, 2020.2002.2006.938100.

Qiu, Y., Hirsch, C.D., Yang, Y., and Watkins, E. (2019). Towards Improved Molecular Identification Tools in Fine Fescue (*Festuca* L., Poaceae) Turfgrasses: Nuclear Genome Size, Ploidy, and Chloroplast Genome Sequencing. *Frontiers in Genetics* 10.

Qiu, Y., Yang, Y., Hirsch, C.D., and Watkins, E. (2020b). Building a Reference Transcriptome for the Hexaploid Hard Fescue Turfgrass (*Festuca brevipila*) Using a Combination of PacBio IsoSeq and Illumina Sequencing. *bioRxiv*, 2020.2002.2026.966952.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Research* 33, W116-W120.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.

Rambaut, A. (2012). "FigTree v1. 4".

Reif, J.C., Zhang, P., Dreisigacker, S., Warburton, M.L., Van Ginkel, M., Hoisington, D., Bohn, M., and Melchinger, A.E. (2005). Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics* 110, 859-864.

Reiter, M., Friell, J., Horgan, B., Soldat, D., and Watkins, E. (2017). Drought Response of Fine Fescue Mixtures Maintained as a Golf Course Fairway. *International Turfgrass Society Research Journal* 13, 65-74.

Rhoads, A., and Au, K.F. (2015). PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics* 13, 278-289.

Robbins, M.D., Staub, J.E., and Bushman, B.S. (2016). Development of fine-leaved Festuca grass populations identifies genetic resources having improved forage production with potential for wildfire control in the western United States. *Euphytica* 209, 377-393.

Robinson, M.D., Mccarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
Rousseau-Gueutin, M., Huang, X., Higginson, E., Ayliffe, M., Day, A., and Timmis, J.N. (2013). Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (accD) gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiology* 161, 1918-1929.
Rubenstein, K.D., Smale, M., and Widrlechner, M.P. (2006). Demand for genetic resources and the US National Plant Germplasm System. *Crop Science* 46, 1021-1031.
Ruemmele, B., Brilman, L., and Huff, D. (1995). Fine fescue germplasm diversity and vulnerability. *Crop Science* 35, 313-316.
Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506-3514.
Sanan-Mishra, N., Kumar, V., Sopory, S.K., and Mukherjee, S.K. (2009). Cloning and validation of novel miRNA from basmati rice indicates cross talk between abiotic and biotic stresses. *Molecular Genetics and Genomics* 282, 463.
Schilirò, E., Ferrara, M., Nigro, F., and Mercado-Blanco, J. (2012). Genetic responses induced in olive roots upon colonization by the biocontrol endophytic bacterium Pseudomonas fluorescens PICF7. *PLoS One* 7.
Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PloS one* 6, e17288.
Schmit, R., Duell, R., and Funk, C. (Year). "Isolation barriers and self-compatibility in selected fine fescues 1", in: *Proceedings of the Second International Turfgrass Research Conference*: American Society of Agronomy, Crop Science Society of America), 9-17.
Seal, A. (1983). DNA variation in Festuca. *Heredity* 50, 225.
Sekimata, K., Han, S.-Y., Yoneyama, K., Takeuchi, Y., Yoshida, S., and Asami, T. (2002). A specific and potent inhibitor of brassinosteroid biosynthesis possessing a dioxolane ring. *Journal of Agricultural and Food Chemistry* 50, 3486-3490.
Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., and Qiu, J.-L. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology* 31, 686.
Shi, G.-Q., Fu, J.-Y., Rong, L.-J., Zhang, P.-Y., Guo, C.-J., and Kai, X. (2018). TaMIR1119, a miRNA family member of wheat (Triticum aestivum), is essential in the regulation of plant drought tolerance. *Journal of Integrative Agriculture* 17, 2369-2378.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210-3212.
Šmarda, P., Bureš, P., Horová, L., Foggi, B., and Rossi, G. (2008). Genome size and GC content evolution of *Festuca*: ancestral expansion and subsequent reduction. *Annals of Botany* 101, 421-433.

Smith, S.A., Moore, M.J., Brown, J.W., and Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15, 150.

Soltis, P.S., Marchant, D.B., Van De Peer, Y., and Soltis, D.E. (2015). Polyploidy and genome evolution in plants. *Current Opinion in Genetics & Development* 35, 119-125.

Song, J., Bradeen, J.M., Naess, S.K., Raasch, J.A., Wielgus, S.M., Haberlach, G.T., Liu, J., Kuang, H., Austin-Phillips, S., and Buell, C.R. (2003). Gene RB cloned from Solanum bulbocastanum confers broad spectrum resistance to potato late blight. *Proceedings of The National Academy of Sciences* 100, 9128-9133.

Stace, C. (2010). *New flora of the British Isles.* Cambridge University Press.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688-2690.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34, W609-W612.

Te Beest, M., Le Roux, J.J., Richardson, D.M., Brysting, A.K., Suda, J., Kubešová, M., and Pyšek, P. (2011). The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* 109, 19-45.

Thiel, T., Michalek, W., Varshney, R., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106, 411-422.

Thorstensen, C.W., and Lode, O. (2001). Laboratory degradation studies of bentazone, dichlorprop, MCPA, and propiconazole in Norwegian soils. *Journal of Environmental Quality* 30, 947-953.

Throssell, C.S., Lyman, G.T., Johnson, M.E., Stacey, G.A., and Brown, C.D. (2009). Golf course environmental profile measures water use, source, cost, quality, management and conservation strategies. *Applied Turfgrass Science* 6, 0-0.

Tian, Z., Huang, B., and Belanger, F.C. (2015). Effects of *Epichloë festucae* fungal endophyte infection on drought and heat stress responses of strong creeping red fescue. *Journal of the American Society for Horticultural Science* 140, 257-264.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45, W6-W11.

Torrecilla, P., and Catalán, P. (2002). Phylogeny of broad-leaved and fine-leaved *Festuca* lineages (Poaceae) based on nuclear ITS sequences. *Systematic Botany* 27, 241-252.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28, 511.

Tremblay, K., Ouellet, F., Fournier, J., Danyluk, J., and Sarhan, F. (2005). Molecular characterization and origin of novel bipartite cold-regulated ice recrystallization inhibition proteins from cereals. *Plant and Cell Physiology* 46, 884-891.

Tronsmo, A.M., Hsiang, T., Okuyama, H., and Nakajima, T. (2001). Low temperature diseases caused by *Microdochium nivale*. *Low Temperature Plant Microbe Interactions Under Snow*, 75-86.
Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research* 40, e115-e115.
Van Esbroeck, G., and Bowman, D.T. (1998). Cotton germplasm diversity and its importance to cultivar development. *Journal of Cotton Science*.
Vargas, J.M. (2018). *Management of turfgrass diseases.* Routledge.
Vargas, P., Mcallister, H.A., Morton, C., Jury, S.L., and Wilkinson, M.J. (1999). Polyploid speciation inHedera (Araliaceae): Phylogenetic and biogeographic insights based on chromosome counts and ITS sequences. *Plant Systematics and Evolution* 219, 165-179.
Varshney, R.K., Graner, A., and Sorrells, M.E. (2005). Genomics-assisted breeding for crop improvement. *Trends in Plant Science* 10, 621-630.
Vasey, G. (1883). *Grasses of the United States.* US Government Printing Office.
Vincelli, P. (2004). Simulations of fungicide runoff following applications for turfgrass disease control. *Plant Disease* 88, 391-396.
Vincelli, P., Dixon, E., Williams, D., and Burrus, P. (2003). Efficacy of fungicides for control of dollar spot of creeping bentgrass managed as a fairway. *Fungicide and Nematicide Tests* 59, T007.
Vogel, K.P., and Moore, K.J. (1998). Forage yield and quality of tall wheatgrass accessions in the USDA germplasm collection. *Crop Science* 38, 509-512.
Wang, B., Tseng, E., Regulski, M., Clark, T.A., Hon, T., Jiao, Y., Lu, Z., Olson, A., Stein, J.C., and Ware, D. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nature Communications* 7, 11708.
Wang, D., Ling, L., Zhang, W., Bai, Y., Shu, Y., and Guo, C. (2018). Uncovering key small RNAs associated with gametocidal action in wheat. *Journal of Experimental Botany* 69, 4739-4756.
Wang, Y., Bigelow, C.A., and Jiang, Y. (2009). Ploidy level and DNA content of perennial ryegrass germplasm as determined by flow cytometry. *HortScience* 44, 2049-2052.
Watkins, E., Hollman, A.B., and Horgan, B.P. (2010). Evaluation of alternative turfgrass species for low-input golf course fairways. *HortScience* 45, 113-118.
Watschke, T., Mumma, R., Linde, D., Borger, J., and Harrison, S. (2000). "Surface runoff of selected pesticides applied to turfgrasses." ACS Publications.
Watson, P.J. (1958). The distribution in Britain of diploid and tetraploid races within the Festuca ovina group. *The New Phytologist* 57, 11-18.
Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., and Olson, N.D. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* 37, 1155-1162.

White, R.H., Engelke, M.C., Morton, S.J., Johnson-Cicalese, J.M., and Ruemmele, B.A. (1992). Acremonium endophyte effects on tall fescue drought tolerance. *Crop Science* 32, 1392-1396.
Whitmore, R., Kelly, J., Reading, P., Brandt, E., and Harris, T. (1993). "National home and garden pesticide use survey." ACS Publications.
Wieners, R.R., Fei, S.-Z., and Johnson, R.C. (2006). Characterization of a USDA Kentucky bluegrass (Poa pratensis L.) core collection for reproductive mode and DNA content by flow cytometry. *Genetic Resources and Crop Evolution* 53, 1531-1541.
Wilkinson, M.J., and Stace, C.A. (1991). A new taxonomic treatment of the Festuca ovina L. aggregate (Poaceae) in the British Isles. *Botanical Journal of the Linnean Society* 106, 347-397.
Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859-1875.
Wyman, S.K., Jansen, R.K., and Boore, J.L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252-3255.
Yang, M., Zhu, L., Pan, C., Xu, L., Liu, Y., Ke, W., and Yang, P. (2015). Transcriptomic analysis of the regulation of rhizome formation in temperate and tropical lotus (*Nelumbo nucifera*). *Scientific Reports* 5, 13059.
Yeh, S., Moffatt, B.A., Griffith, M., Xiong, F., Yang, D.S., Wiseman, S.B., Sarhan, F., Danyluk, J., Xue, Y.Q., and Hew, C.L. (2000). Chitinase genes responsive to cold encode antifreeze proteins in winter cereals. *Plant Physiology* 124, 1251-1264.
Yu, X.M., and Griffith, M. (2001). Winter rye antifreeze activity increases in response to cold and drought, but not abscisic acid. *Physiologia Plantarum* 112, 78-86.
Yue, C., Wang, J., Watkins, E., Bonos, S.A., Nelson, K.C., Murphy, J.A., Meyer, W.A., and Horgan, B.P. (2017). Heterogeneous consumer preferences for turfgrass attributes in the United States and Canada. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 65, 347-383.
Zhang, B., Liu, J., Wang, X., and Wei, Z. (2018). Full-length RNA sequencing reveals unique transcriptome composition in bermudagrass. *Plant Physiology and Biochemistry* 132, 95-103.
Zhang, C., Fei, S.-Z., Arora, R., and Hannapel, D.J. (2010). Ice recrystallization inhibition proteins of perennial ryegrass enhance freezing tolerance. *Planta* 232, 155-164.
Zhang, G., Sun, M., Wang, J., Lei, M., Li, C., Zhao, D., Huang, J., Li, W., Li, S., and Li, J. (2019). PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice. *The Plant Journal* 97, 296-305.
Zhang, H., Gong, G., Guo, S., Ren, Y., Xu, Y., and Ling, K.-S. (2011). Screening the USDA watermelon germplasm collection for drought tolerance at the seedling stage. *HortScience* 46, 1245-1248.
Zhang, X., Ervin, E., and Schmidt, R. (2003). Plant growth regulators can enhance the recovery of Kentucky bluegrass sod from heat injury. *Crop Science* 43, 952-956.

Zhu, D., Hou, L., Xiao, P., Guo, Y., Deyholos, M.K., and Liu, X. (2019). VvWRKY30, a grape WRKY transcription factor, plays a positive regulatory role under salinity stress. *Plant Science* 280, 132-142.

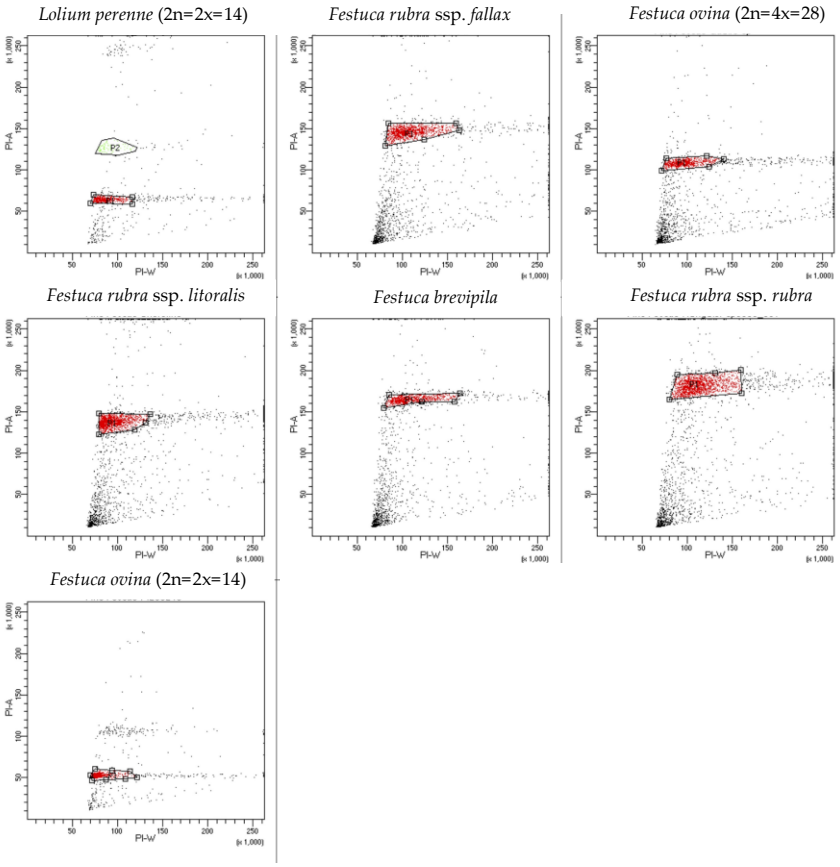

Figure S2.1. Flow cytometry nuclei population distribution of *L. perenne*, fine fescues, and the diploid USDA PI accession. G1 population for each sample is gated in red, G2 population is gated only in *L. perenne* with green color.

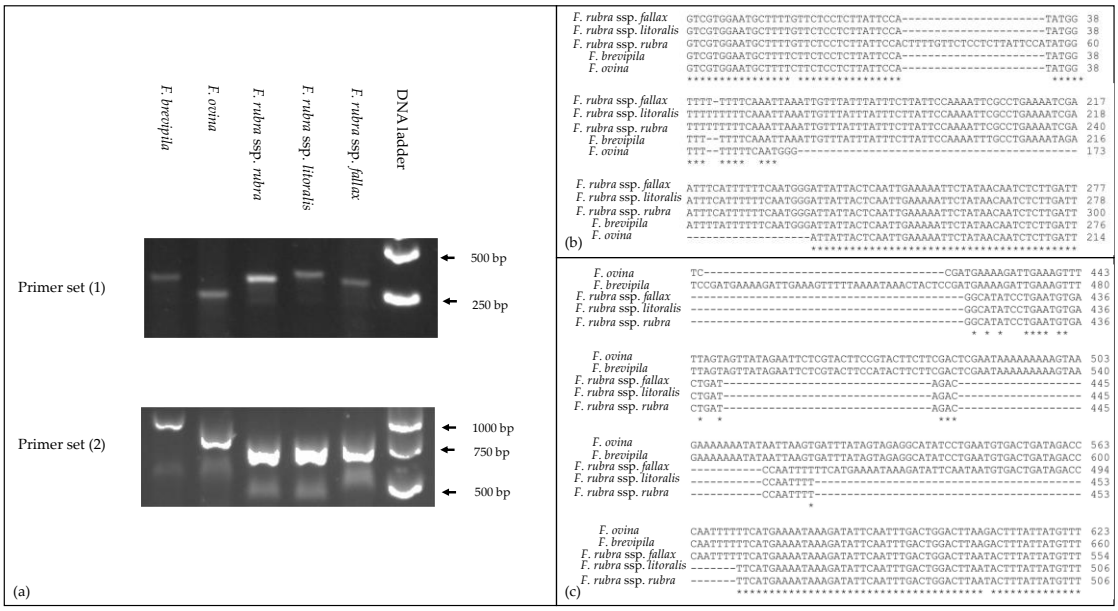

Figure S2.2. Examples of PCR validation of predicted repeat regions based on fine fescue chloroplast genomes. PCR primers were developed using Primer3 module (Untergasser et al., 2012). Primers used for the PCR assays (1) Forward primer 5'-GTCGTGGAATGCTTTTGTTCTC-3'; Reverse primer 5'-AGTGGATTCATCAGATGATACA-3'; (2) Forward primer 5'-TTCCTCTTTTCATTG-CAAAGTGGT AT-3'; Reverse primer 5'-TACTCGGAGGTTCGAATCCTTCC-3'. PCR products were examined on 1% agarose gel and gel images showed fragment size separation between different taxa (a). Figure (b) and (c) showed partial sequence alignment of regions amplified by primer sets (1 and 2).

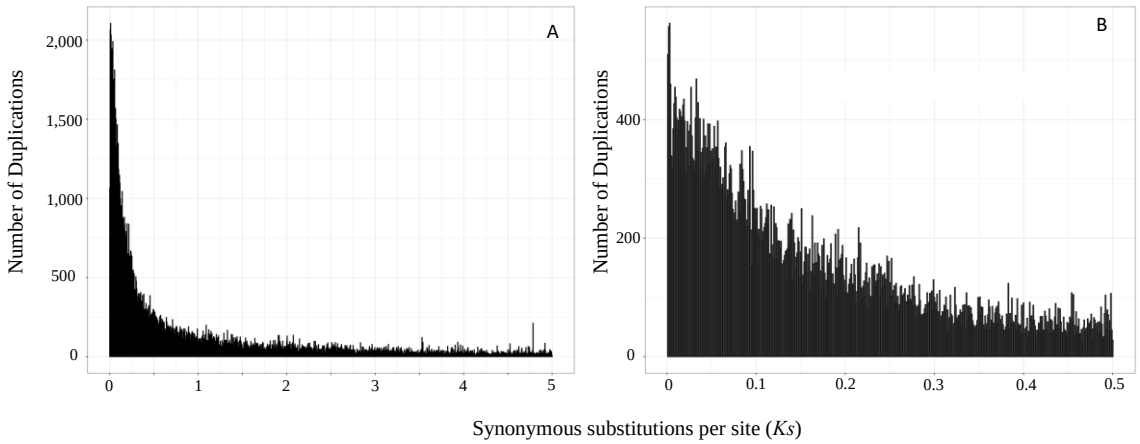

Figure S4.1. The Synonymous substitutions per site (Ks) analysis of *F. brevipila* using *de novo* contigs that were mapped to PacBio Isoseq trnascriptome. Analysis was done using FASTKs program (https://github.com/mrmckain/FASTKs). *Ks* was subset to 5.0 in (A), and 0.5 in (B). No peaks were identified based on this analysis

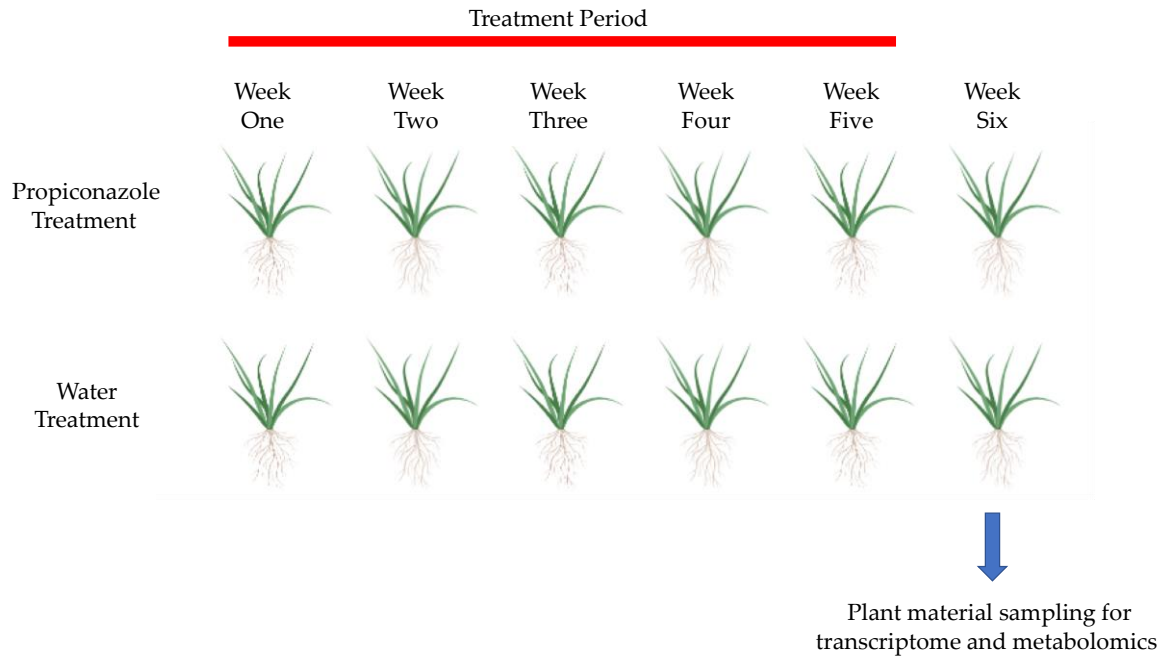

Figure S5.1. The workflow of fungicide spray and plant tissue sampling.

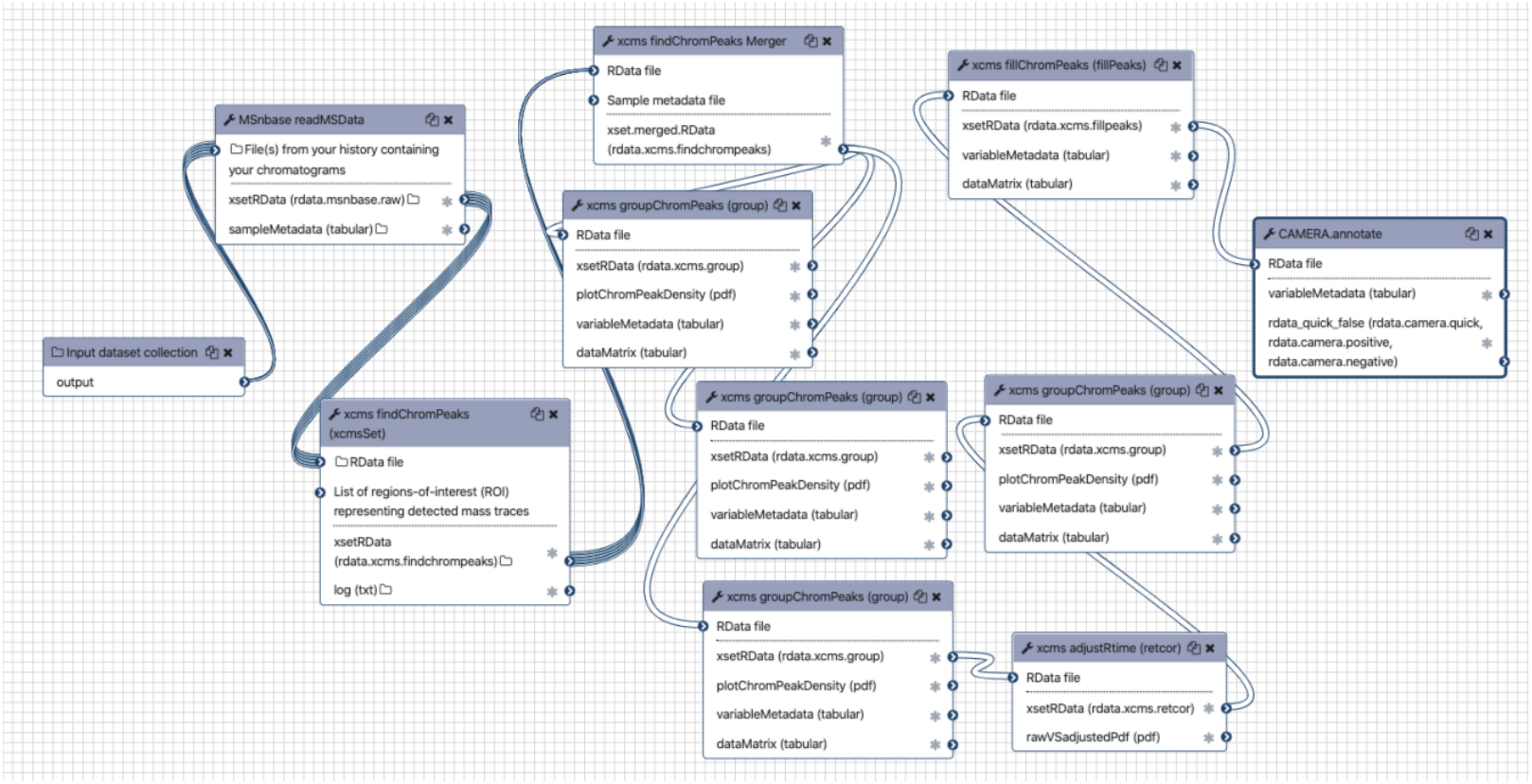

Figure S5.2. Galaxy metabolites data preprocessing workflow.

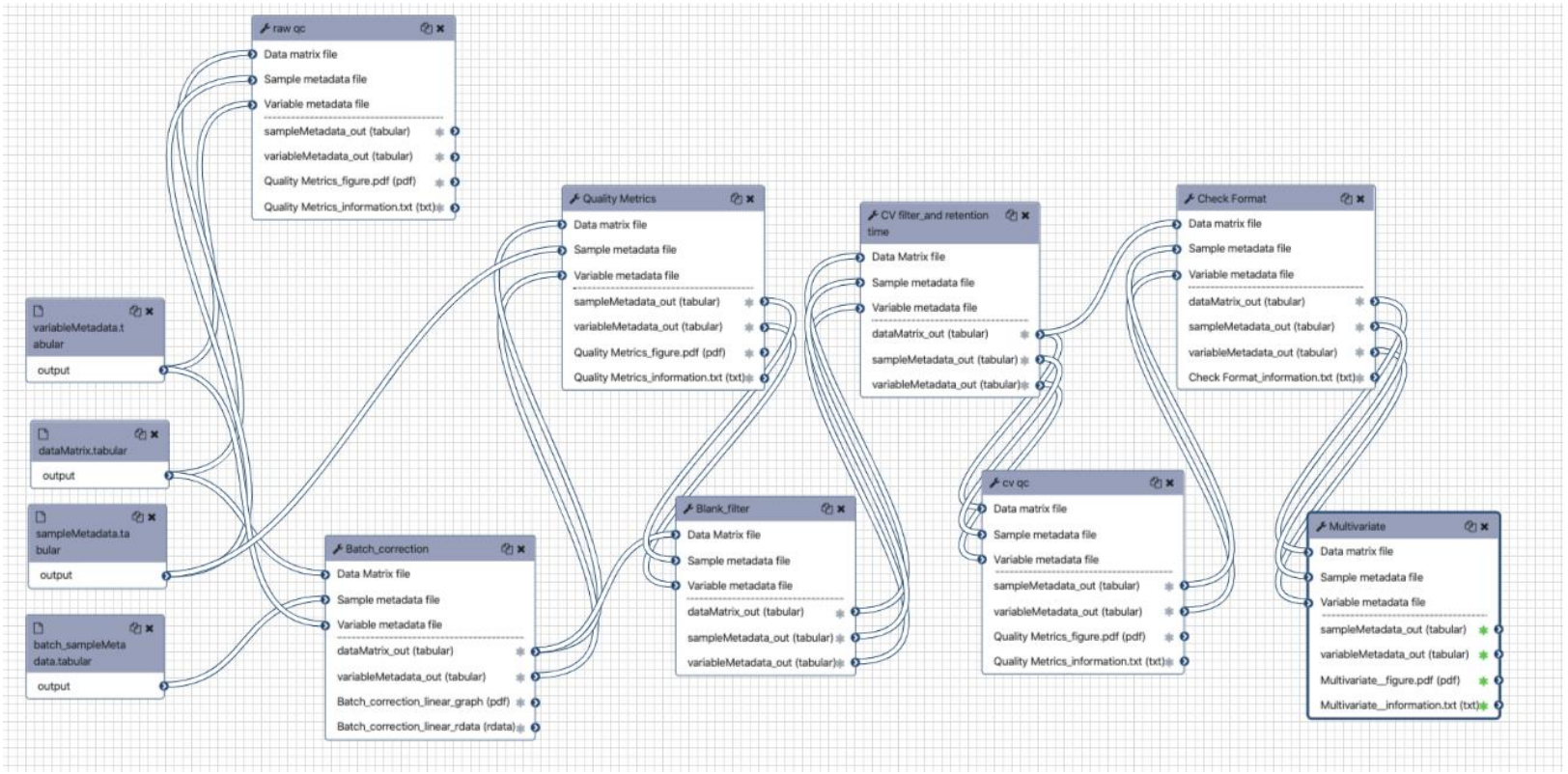

Figure S5.3. Galaxy metabolites data postprocessing workflow.

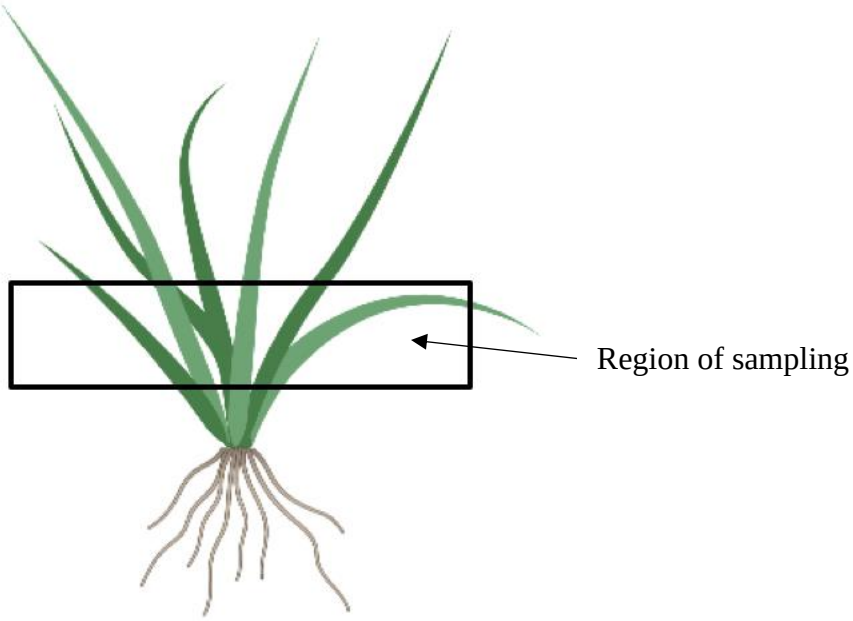

Figure S5.4. Figure illustration of plant tissues sampled for RNA sequencing in the long-term effect study.

TRANSCRIPT_46772
(Ice recrystallization inhibition protein 5)

TRINITY_DN97131_C2_G5_I3
(Ice recrystallization inhibition protein 2)

TRANSCRIPT_48329
(Ice recrystallization inhibition protein 6)

TRINITY_DN72600_C2_G3_I3
(Ice recrystallization inhibition protein IRI3)

TRINITY_DN86392_C0_G9_I1
(Ice recrystallization inhibition protein 6)

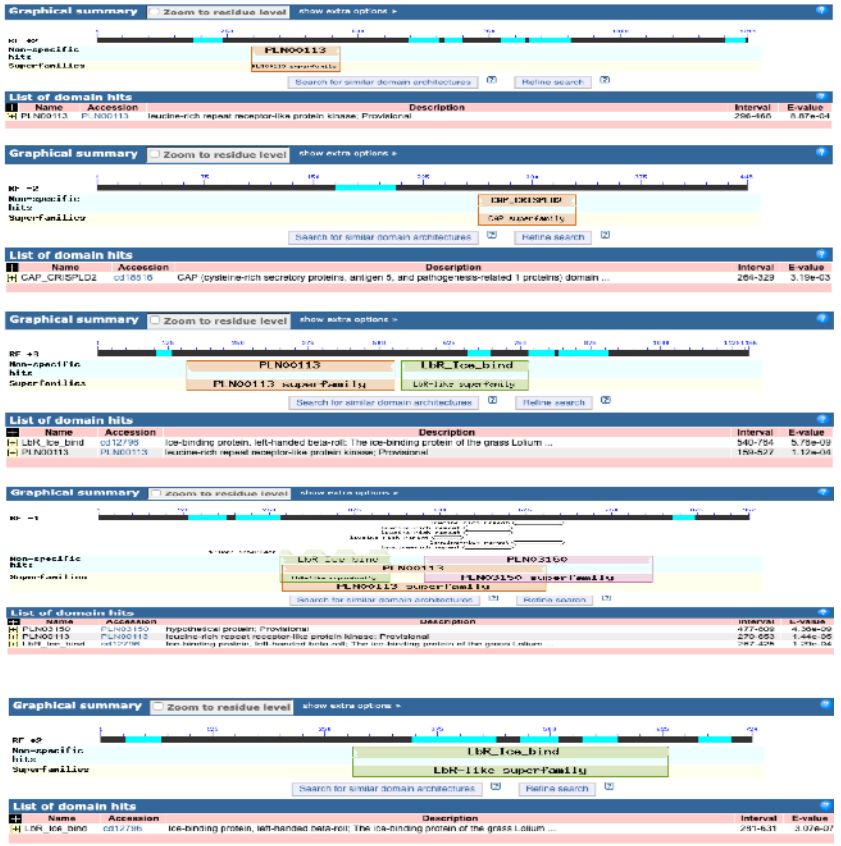

Figure S5.5. Conserved domain search in upregulated IRIPs via NCBI blastP.

Table S2.1. Fine fescue chloroplast genomes gene content by gene category.

	Group of Gene[&]	Name of gene			
Self-replication (58/77)	Ribosomal RNA genes (4/8)	*rrn4.5* [a]	*rrn5* [a]	*rrn16* [a]	*rrn23* [a]
	Transfer RNA genes (27/38)	*trnA-UGC* *[a]	*trnC-GCA*	*trnD-GUC*	*trnE-UUC*
		trnF-GAA	*trnG-GCC*	*trnH-GUG* [a]	*trnI-GAU* *[a]
		*trnK-UUU**	*trnL-CAA* [a]	*trnL-UAA**	*trnL-UAG*
		trnM-CAU [c]	*trnN-GUU* [a]	*trnP-UGG*	*trnQ-UUG*
		trnR-ACG [a]	*trnR-UCU*	*trnS-GCU*	*trnS-GGA*
		trnS-UGA	*trnT-GGU*	*trnT-UGU*	*trnV-GAC*[a]
		*trnV-UAC**	*trnW-CCA*	*trnY-GUA*	
	Small subunit of ribosome (12/16)	*rps2*	*rps3*	*rps4*	*rps7* [a]
		rps8	*rps11*	*rps12* *[ab]	*rps14*
		rps15 [a]	*rps16**	*rps18*	*rps19* [a]
	Large subunit of ribosome (9/11)	*rpl2**[a]	*rpl14*	*rpl16*	*rpl20*
		rpl22	*rpl23*[a]	*rpl32*	*rpl33*
		rpl36			
	RNA polymerase subunits (4)	*rpoA*	*rpoB*	*rpoC1*	*rpoC2*
Photosynthesis (45/46)	Subunits of Photosystem I (6)	*psaA*	*psaB*	*psaC*	*psaI*
		psaJ	*ycf3***		
	Subunits of Photosystem II (15)	*psbA*	*psbB*	*psbC*	*psbD*
		psbE	*psbF*	*psbH*	*psbI*
		psbJ	*psbK*	*psbL*	*psbM*
		psbN	*psbT*	*psbZ*	
	Subunits of cytochrome (6)	*petA*	*petB**	*petD**	*petG*
		petL	*petN*		
		atpA	*atpB*	*atpE*	*atpF**

	Subunits of ATP synthase (6)	*atpH*	*atpI*		
	Large subunit of Rubisco (1)	*rbcL*			
	Subunits of NADH Dehydrogenase (11/12)	*ndhA**	*ndhB*[a]*	*ndhC*	*ndhD*
		ndhE	*ndhF*	*ndhG*	*ndhH*
		ndhI	*ndhJ*	*ndhK*	
	Translational initiation factor (1)	*infA*			
	Maturase (1)	*matK*			
	Envelope membrane protein (1)	*cemA*			
Other genes (5)	C-type cytochrome (1)	*cssA*			
	Protease (1)	*clp*			
	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	*accD*[$]			
Unknown function (5)	Conserved open reading frames (3/5)	*ycf1*[a]	*ycf2*[a]	*ycf4*	

[&]Group of Genes were presented by gene family, followed by the number of unique gene count and total number of genes count (including genes with two or more copies) in the bracket. [a] Two gene copies in IRs; [b] Gene divided into two independent transcription units; [c] Gene that has five copies; [*] One intron-containing genes; [**] Two intron-containing genes. [$] Gene annotated in *F. rubra* spp. only. Fine fescue taxa chloroplast genomes share high structure similarity and gene content. Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (*accD*) pseudogene is annotated in *F. rubra* ssp.

Table S2.2. SSR loci types and number distributions of fine fescue species predicted using MISA program.

Repeats	SSR	Cultivar name	Taxon
A/T	24	Shoreline	*F. rubra* ssp. *littoralis*
C/G	1	Shoreline	*F. rubra* ssp. *littoralis*
AG/CT	1	Shoreline	*F. rubra* ssp. *littoralis*
AT/AT	6	Shoreline	*F. rubra* ssp. *littoralis*
AAC/GTT	1	Shoreline	*F. rubra* ssp. *littoralis*
AAG/CTT	1	Shoreline	*F. rubra* ssp. *littoralis*
AAT/ATT	1	Shoreline	*F. rubra* ssp. *littoralis*
AAAC/GTTT	1	Shoreline	*F. rubra* ssp. *littoralis*
AAAG/CTTT	2	Shoreline	*F. rubra* ssp. *littoralis*
AAAT/ATTT	1	Shoreline	*F. rubra* ssp. *littoralis*
AACG/CGTT	2	Shoreline	*F. rubra* ssp. *littoralis*
AAGG/CCTT	1	Shoreline	*F. rubra* ssp. *littoralis*
AATG/ATTC	1	Shoreline	*F. rubra* ssp. *littoralis*
ACAT/ATGT	1	Shoreline	*F. rubra* ssp. *littoralis*
AAATT/AATTT	1	Shoreline	*F. rubra* ssp. *littoralis*
AATGC/ATTGC	1	Shoreline	*F. rubra* ssp. *littoralis*
A/T	23	Quatro	*F. ovina*
C/G	1	Quatro	*F. ovina*
AG/CT	1	Quatro	*F. ovina*
AT/AT	6	Quatro	*F. ovina*
AAC/GTT	1	Quatro	*F. ovina*
AAG/CTT	2	Quatro	*F. ovina*
AAT/ATT	1	Quatro	*F. ovina*
AAAC/GTTT	1	Quatro	*F. ovina*
AAAG/CTTT	1	Quatro	*F. ovina*
AAAT/ATTT	1	Quatro	*F. ovina*
AACG/CGTT	2	Quatro	*F. ovina*
AAGG/CCTT	1	Quatro	*F. ovina*
AATG/ATTC	1	Quatro	*F. ovina*
ACAT/ATGT	1	Quatro	*F. ovina*
AATGC/ATTGC	1	Quatro	*F. ovina*
ACCAT/ATGGT	1	Quatro	*F. ovina*
A/T	18	Beacon	*F. brevipila*
C/G	1	Beacon	*F. brevipila*

AG/CT	1	Beacon	*F. brevipila*
AT/AT	5	Beacon	*F. brevipila*
AAC/GTT	1	Beacon	*F. brevipila*
AAG/CTT	2	Beacon	*F. brevipila*
AAT/ATT	1	Beacon	*F. brevipila*
AAAC/GTTT	1	Beacon	*F. brevipila*
AAAG/CTTT	1	Beacon	*F. brevipila*
AAAT/ATTT	1	Beacon	*F. brevipila*
AACG/CGTT	2	Beacon	*F. brevipila*
AAGG/CCTT	1	Beacon	*F. brevipila*
AATG/ATTC	1	Beacon	*F. brevipila*
ACAT/ATGT	1	Beacon	*F. brevipila*
AATGC/ATTGC	1	Beacon	*F. brevipila*
ACCAT/ATGGT	1	Beacon	*F. brevipila*
A/T	22	Navigator II	*F. rubra* ssp. *rubra*
C/G	1	Navigator II	*F. rubra* ssp. *rubra*
AG/CT	1	Navigator II	*F. rubra* ssp. *rubra*
AT/AT	6	Navigator II	*F. rubra* ssp. *rubra*
AAC/GTT	1	Navigator II	*F. rubra* ssp. *rubra*
AAG/CTT	2	Navigator II	*F. rubra* ssp. *rubra*
AAT/ATT	1	Navigator II	*F. rubra* ssp. *rubra*
AAAC/GTTT	1	Navigator II	*F. rubra* ssp. *rubra*
AAAG/CTTT	2	Navigator II	*F. rubra* ssp. *rubra*
AAAT/ATTT	1	Navigator II	*F. rubra* ssp. *rubra*
AACG/CGTT	2	Navigator II	*F. rubra* ssp. *rubra*
AAGG/CCTT	1	Navigator II	*F. rubra* ssp. *rubra*
AATG/ATTC	1	Navigator II	*F. rubra* ssp. *rubra*
ACAT/ATGT	1	Navigator II	*F. rubra* ssp. *rubra*
AAATT/AATTT	1	Navigator II	*F. rubra* ssp. *rubra*
AATGC/ATTGC	1	Navigator II	*F. rubra* ssp. *rubra*
A/T	20	Treazure II	*F. rubra* ssp. *fallax*
C/G	1	Treazure II	*F. rubra* ssp. *fallax*
AG/CT	1	Treazure II	*F. rubra* ssp. *fallax*
AT/AT	6	Treazure II	*F. rubra* ssp. *fallax*
AAC/GTT	1	Treazure II	*F. rubra* ssp. *fallax*
AAG/CTT	2	Treazure II	*F. rubra* ssp. *fallax*
AAT/ATT	1	Treazure II	*F. rubra* ssp. *fallax*
AAAC/GTTT	1	Treazure II	*F. rubra* ssp. *fallax*

AAAG/CTTT	2	Treazure II	*F. rubra* ssp. *fallax*
AAAT/ATTT	1	Treazure II	*F. rubra* ssp. *fallax*
AACG/CGTT	2	Treazure II	*F. rubra* ssp. *fallax*
AAGG/CCTT	1	Treazure II	*F. rubra* ssp. *fallax*
AATG/ATTC	1	Treazure II	*F. rubra* ssp. *fallax*
ACAT/ATGT	1	Treazure II	*F. rubra* ssp. *fallax*
AATGC/ATTGC	1	Treazure II	*F. rubra* ssp. *fallax*
AAATT/AATTT	0	Beacon	*F. brevipila*
AAATT/AATTT	0	Quatro	*F. ovina*
AAATT/AATTT	0	Treazure II	*F. rubra* ssp. *fallax*
ACCAT/ATGGT	0	Treazure II	*F. rubra* ssp. *fallax*
ACCAT/ATGGT	0	Shoreline	*F. rubra* ssp. *littoralis*
ACCAT/ATGGT	0	Navigator II	*F. rubra* ssp. *rubra*

Table S2.3. Tandem repeats for fine fescues taxa based on chloroplast genome sequence.

Repeat length	Repeat starting position	Repeat type	Repeat ending position	p-value	Species
48	102684	P	102684	6.31E-20	*Festuca brevipila*
33	54939	P	80775	6.78E-11	*Festuca brevipila*
33	54939	F	130985	6.78E-11	*Festuca brevipila*
30	64118	P	64118	4.34E-09	*Festuca brevipila*
30	85015	P	85015	4.34E-09	*Festuca brevipila*
30	85015	F	126748	4.34E-09	*Festuca brevipila*
30	126748	P	126748	4.34E-09	*Festuca brevipila*
29	7621	P	43236	1.73E-08	*Festuca brevipila*
26	6025	F	6062	1.11E-06	*Festuca brevipila*
26	74234	F	74252	1.11E-06	*Festuca brevipila*
24	63678	F	63720	1.78E-05	*Festuca brevipila*
23	74594	F	74617	7.11E-05	*Festuca brevipila*
22	15184	F	15205	2.84E-04	*Festuca brevipila*
21	7626	F	11271	1.14E-03	*Festuca brevipila*
21	11271	P	43239	1.14E-03	*Festuca brevipila*
21	12556	F	34811	1.14E-03	*Festuca brevipila*
21	14111	P	44447	1.14E-03	*Festuca brevipila*
21	25807	F	25951	1.14E-03	*Festuca brevipila*
21	31222	F	31619	1.14E-03	*Festuca brevipila*
21	88810	F	88829	1.14E-03	*Festuca brevipila*
21	88810	P	122943	1.14E-03	*Festuca brevipila*
21	88829	P	122962	1.14E-03	*Festuca brevipila*
21	122943	F	122962	1.14E-03	*Festuca brevipila*
20	3763	F	14345	4.55E-03	*Festuca brevipila*
20	16746	P	16746	4.55E-03	*Festuca brevipila*
20	23885	F	23906	4.55E-03	*Festuca brevipila*
20	31189	C	40311	4.55E-03	*Festuca brevipila*
20	36834	F	39058	4.55E-03	*Festuca brevipila*
20	48398	R	48398	4.55E-03	*Festuca brevipila*
20	100202	P	100202	4.55E-03	*Festuca brevipila*
48	102857	P	102857	6.33E-21	*Festuca ovina*
33	55106	P	80946	6.79E-11	*Festuca ovina*
33	55106	F	131162	6.79E-11	*Festuca ovina*
30	64294	P	64294	4.35E-09	*Festuca ovina*

30	85186	P	85186	4.35E-09	*Festuca ovina*
30	85186	F	126925	4.35E-09	*Festuca ovina*
30	126925	P	126925	4.35E-09	*Festuca ovina*
29	7588	P	43396	1.74E-08	*Festuca ovina*
26	74400	F	74418	1.11E-06	*Festuca ovina*
24	63854	F	63896	1.78E-05	*Festuca ovina*
23	49226	R	49226	7.12E-05	*Festuca ovina*
23	74760	F	74783	7.12E-05	*Festuca ovina*
22	15150	F	15171	2.85E-04	*Festuca ovina*
21	7593	F	11237	1.14E-03	*Festuca ovina*
21	11237	P	43399	1.14E-03	*Festuca ovina*
21	12522	F	34948	1.14E-03	*Festuca ovina*
21	14078	P	44607	1.14E-03	*Festuca ovina*
21	25944	F	26088	1.14E-03	*Festuca ovina*
21	31361	F	31758	1.14E-03	*Festuca ovina*
21	88981	F	89000	1.14E-03	*Festuca ovina*
21	88981	P	123120	1.14E-03	*Festuca ovina*
21	89000	P	123139	1.14E-03	*Festuca ovina*
21	123120	F	123139	1.14E-03	*Festuca ovina*
20	16883	P	16883	4.56E-03	*Festuca ovina*
20	24022	F	24043	4.56E-03	*Festuca ovina*
20	31328	C	40448	4.56E-03	*Festuca ovina*
20	36971	F	39195	4.56E-03	*Festuca ovina*
20	48560	R	48560	4.56E-03	*Festuca ovina*
20	100373	P	100373	4.56E-03	*Festuca ovina*
48	6034	F	6082	6.36E-20	*Festuca rubra* ssp. *fallax*
48	103127	P	103127	6.36E-20	*Festuca rubra* ssp. *fallax*
40	89228	F	89247	4.17E-15	*Festuca rubra* ssp. *fallax*
40	89228	P	123433	4.17E-15	*Festuca rubra* ssp. *fallax*
40	89247	P	123452	4.17E-15	*Festuca rubra* ssp. *fallax*
40	123433	F	123452	4.17E-15	*Festuca rubra* ssp. *fallax*
33	55247	P	81193	6.83E-11	*Festuca rubra* ssp. *fallax*
33	55247	F	131494	6.83E-11	*Festuca rubra* ssp. *fallax*
30	64559	P	64559	4.37E-09	*Festuca rubra* ssp. *fallax*
30	85433	P	85433	4.37E-09	*Festuca rubra* ssp. *fallax*
30	85433	F	127257	4.37E-09	*Festuca rubra* ssp. *fallax*
30	127257	P	127257	4.37E-09	*Festuca rubra* ssp. *fallax*
29	7510	P	43478	1.75E-08	*Festuca rubra* ssp. *fallax*

26	49467	F	49492	1.12E-06	*Festuca rubra* ssp. *fallax*
26	64117	F	64159	1.12E-06	*Festuca rubra* ssp. *fallax*
23	75032	F	75055	7.16E-05	*Festuca rubra* ssp. *fallax*
21	7515	F	11163	1.15E-03	*Festuca rubra* ssp. *fallax*
21	11163	P	43481	1.15E-03	*Festuca rubra* ssp. *fallax*
21	12608	F	35054	1.15E-03	*Festuca rubra* ssp. *fallax*
21	14157	P	44697	1.15E-03	*Festuca rubra* ssp. *fallax*
21	44539	R	44539	1.15E-03	*Festuca rubra* ssp. *fallax*
21	89228	F	89266	1.15E-03	*Festuca rubra* ssp. *fallax*
21	89228	P	123433	1.15E-03	*Festuca rubra* ssp. *fallax*
21	89266	P	123471	1.15E-03	*Festuca rubra* ssp. *fallax*
21	123433	F	123471	1.15E-03	*Festuca rubra* ssp. *fallax*
20	16979	P	16979	4.58E-03	*Festuca rubra* ssp. *fallax*
20	24120	F	24141	4.58E-03	*Festuca rubra* ssp. *fallax*
20	28683	R	28683	4.58E-03	*Festuca rubra* ssp. *fallax*
20	37077	F	39301	4.58E-03	*Festuca rubra* ssp. *fallax*
20	48666	R	48666	4.58E-03	*Festuca rubra* ssp. *fallax*
20	81284	P	81284	4.58E-03	*Festuca rubra* ssp. *fallax*
20	81284	F	131416	4.58E-03	*Festuca rubra* ssp. *fallax*
20	99359	P	99381	4.58E-03	*Festuca rubra* ssp. *fallax*
20	99359	F	113319	4.58E-03	*Festuca rubra* ssp. *fallax*
20	99381	F	113341	4.58E-03	*Festuca rubra* ssp. *fallax*
20	100640	P	100640	4.58E-03	*Festuca rubra* ssp. *fallax*
20	103475	R	103475	4.58E-03	*Festuca rubra* ssp. *fallax*
20	113319	P	113341	4.58E-03	*Festuca rubra* ssp. *fallax*
20	131416	P	131416	4.58E-03	*Festuca rubra* ssp. *fallax*
48	103126	P	103126	6.36E-20	*Festuca rubra* ssp. *littoralis*
37	26092	F	26113	2.67E-13	*Festuca rubra* ssp. *littoralis*
33	55233	P	81222	6.83E-11	*Festuca rubra* ssp. *littoralis*
33	55233	F	131469	6.83E-11	*Festuca rubra* ssp. *littoralis*
30	64581	P	64581	4.37E-09	*Festuca rubra* ssp. *littoralis*
30	85462	P	85462	4.37E-09	*Festuca rubra* ssp. *littoralis*
30	85462	F	127232	4.37E-09	*Festuca rubra* ssp. *littoralis*
30	127232	P	127232	4.37E-09	*Festuca rubra* ssp. *littoralis*
29	7463	P	43454	1.75E-08	*Festuca rubra* ssp. *littoralis*
26	49454	F	49479	1.12E-06	*Festuca rubra* ssp. *littoralis*
26	64139	F	64181	1.12E-06	*Festuca rubra* ssp. *littoralis*
25	49198	F	49223	4.47E-06	*Festuca rubra* ssp. *littoralis*

23	75056	F	75079	7.16E-05	*Festuca rubra* ssp. *littoralis*
22	15743	R	15743	2.86E-04	*Festuca rubra* ssp. *littoralis*
22	61461	F	61483	2.86E-04	*Festuca rubra* ssp. *littoralis*
21	7468	F	11116	1.15E-03	*Festuca rubra* ssp. *littoralis*
21	11116	P	43457	1.15E-03	*Festuca rubra* ssp. *littoralis*
21	12561	F	35023	1.15E-03	*Festuca rubra* ssp. *littoralis*
21	14111	P	44671	1.15E-03	*Festuca rubra* ssp. *littoralis*
21	89257	F	89276	1.15E-03	*Festuca rubra* ssp. *littoralis*
21	89257	P	123427	1.15E-03	*Festuca rubra* ssp. *littoralis*
21	89276	P	123446	1.15E-03	*Festuca rubra* ssp. *littoralis*
21	123427	F	123446	1.15E-03	*Festuca rubra* ssp. *littoralis*
20	16934	P	16934	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	24076	F	24097	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	37046	F	39270	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	48633	R	48633	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	81313	P	81313	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	81313	F	131391	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	99369	P	99391	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	99369	F	113313	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	99391	F	113335	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	100643	P	100643	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	103474	R	103474	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	113313	P	113335	4.58E-03	*Festuca rubra* ssp. *littoralis*
20	131391	P	131391	4.58E-03	*Festuca rubra* ssp. *littoralis*
51	49443	F	49468	9.93E-22	*Festuca rubra* ssp. *rubra*
48	103116	P	103116	6.36E-20	*Festuca rubra* ssp. *rubra*
33	55248	P	81201	6.82E-11	*Festuca rubra* ssp. *rubra*
33	55248	F	131459	6.82E-11	*Festuca rubra* ssp. *rubra*
30	64562	P	64562	4.37E-09	*Festuca rubra* ssp. *rubra*
30	85441	P	85441	4.37E-09	*Festuca rubra* ssp. *rubra*
30	85441	F	127222	4.37E-09	*Festuca rubra* ssp. *rubra*
30	127222	P	127222	4.37E-09	*Festuca rubra* ssp. *rubra*
28	7464	P	43443	6.99E-08	*Festuca rubra* ssp. *rubra*
26	49443	F	49493	1.12E-06	*Festuca rubra* ssp. *rubra*
26	64120	F	64162	1.12E-06	*Festuca rubra* ssp. *rubra*
23	75035	F	75058	7.16E-05	*Festuca rubra* ssp. *rubra*
21	7468	F	11116	1.14E-03	*Festuca rubra* ssp. *rubra*
21	11116	P	43446	1.14E-03	*Festuca rubra* ssp. *rubra*

21	12561	F	35011	1.14E-03	*Festuca rubra* ssp. *rubra*
21	14117	P	44662	1.14E-03	*Festuca rubra* ssp. *rubra*
21	32476	R	32476	1.14E-03	*Festuca rubra* ssp. *rubra*
21	44504	R	44504	1.14E-03	*Festuca rubra* ssp. *rubra*
21	49372	R	49372	1.14E-03	*Festuca rubra* ssp. *rubra*
21	89236	F	89255	1.14E-03	*Festuca rubra* ssp. *rubra*
21	89236	P	123417	1.14E-03	*Festuca rubra* ssp. *rubra*
21	89255	P	123436	1.14E-03	*Festuca rubra* ssp. *rubra*
21	123417	F	123436	1.14E-03	*Festuca rubra* ssp. *rubra*
20	16939	P	16939	4.58E-03	*Festuca rubra* ssp. *rubra*
20	24079	F	24100	4.58E-03	*Festuca rubra* ssp. *rubra*
20	28642	R	28642	4.58E-03	*Festuca rubra* ssp. *rubra*
20	37034	F	39258	4.58E-03	*Festuca rubra* ssp. *rubra*
20	48635	R	48635	4.58E-03	*Festuca rubra* ssp. *rubra*
20	81292	P	81292	4.58E-03	*Festuca rubra* ssp. *rubra*
20	81292	F	131381	4.58E-03	*Festuca rubra* ssp. *rubra*
20	99348	P	99370	4.58E-03	*Festuca rubra* ssp. *rubra*
20	99348	F	113303	4.58E-03	*Festuca rubra* ssp. *rubra*
20	99370	F	113325	4.58E-03	*Festuca rubra* ssp. *rubra*
20	100629	P	100629	4.58E-03	*Festuca rubra* ssp. *rubra*
20	103464	R	103464	4.58E-03	*Festuca rubra* ssp. *rubra*
20	113303	P	113325	4.58E-03	*Festuca rubra* ssp. *rubra*
20	131381	P	131381	4.58E-03	*Festuca rubra* ssp. *rubra*

Table S2.4. Numbers of SNPs per gene for the five fine fescue taxa sequenced in this study. rpoC2 gene has the most SNPs (31) in *F. rubra* complex comparing to *F. ovina* species.

	F. brevipila	*F. ovina*	*F. rubra* ssp. *rubra*	*F. rubra* ssp. *littoralis*	*F. rubra* ssp. *fallax*
atpA	-	-	7	5	6
atpB	1	-	11	10	10
atpE	1	1	3	3	3
atpF	-	-	11	9	9
atpI	-	-	2	2	2
ccsA	1	1	12	12	10
cemA	-	-	1	1	2
clpP	1	2	5	6	5
infA	-	-	4	4	4
ndhA	1	2	15	15	15
ndhC	-	-	1	1	1
ndhD	4	4	9	8	8
ndhE	-	1	2	2	2
ndhF	2	3	13	13	12
ndhG	1	-	2	2	2
ndhH	6	6	14	14	14
ndhI	-	-	2	2	2
ndhJ	-	-	3	4	3
ndhK	-	-	5	5	5
petA	-	1	2	2	2
petB	-	1	6	6	7
petD	-	1	4	5	4
psaA	1	2	7	6	6
psaB	1	2	9	9	9
psaC	-	-	3	3	3
psbB	-	1	6	6	6
psbC	-	-	4	4	4
psbD	-	-	3	4	2
psbH	-	-	1	1	1
psbJ	-	-	1	1	1

psbL	-	-	1	1	1
rbcL	1	1	15	14	14
rpl14	-	-	3	3	3
rpl16	1	-	8	8	7
rpl20	-	-	1	-	1
rpl22	-	-	3	3	3
rpl32	-	-	2	2	2
rpl33	1	1	1	1	1
rpoA	-	1	6	6	8
rpoB	1	3	13	14	13
rpoC1	-	1	4	2	2
rpoC2	3	5	31	31	31
rps11	-	-	1	1	1
rps14	-	-	1	1	1
rps16	3	4	9	11	11
rps18	-	-	2	2	2
rps2	-	-	1	1	1
rps3	-	-	3	3	3
ycf3	1	3	7	8	8
ycf4	-	-	2	2	3

Table S2.5. Number of InDels per gene for the five fine fescue taxa sequenced in this study. *ndhA* had the most InDels among the *F. rubra* complex. *atpI* had the most InDesl between the two two species in the *F. ovina* complex (*F. brevipila* and *F. ovina*).

	F. brevipila	*F. ovina*	*F. rubra* ssp. *rubra*	*F. rubra* ssp. *littoralis*	*F. rubra* ssp. *fallax*
atpF	2	1	2	2	2
atpI	7	3	3	2	2
clpP	-	-	1	1	1
ndhA	2	2	6	4	3
ndhF	-	-	1	2	1
ndhH	-	-	2	1	2
ndhK	-	1	-	-	-
petB	2	2	3	3	5
rbcL	-	-	-	1	-
rpl16	1	1	1	1	3
rpl22	4	-	-	1	-
rpoA	-	1	-	-	-
rps16	-	2	2	1	3
ycf3	1	2	3	4	2

Table S2.6. Sliding window analysis for average nucleotide diversity calculation of the fine fescue chloroplast genomes.

Midpoint	Pi	Position	Midpoint	Pi	Position
306	0.002	0.306	7335	0.01167	7.335
506	0	0.506	7536	0.006	7.536
706	0	0.706	7736	0.00767	7.736
906	0	0.906	7937	0.00867	7.937
1106	0.002	1.106	8150	0.00833	8.15
1306	0.00367	1.306	8356	0.00767	8.356
1506	0.00633	1.506	8557	0.00667	8.557
1713	0.00533	1.713	8757	0.00567	8.757
1913	0.00533	1.913	8957	0.00267	8.957
2113	0.00267	2.113	9157	0.00367	9.157
2313	0.00367	2.313	9357	0.00367	9.357
2513	0.003	2.513	9557	0.00167	9.557
2713	0.003	2.713	9757	0.00067	9.757
2913	0.001	2.913	9957	0	9.957
3113	0.001	3.113	10157	0.002	10.157
3313	0.00233	3.313	10357	0.002	10.357
3513	0.003	3.513	10557	0.003	10.557
3714	0.002	3.714	10757	0.002	10.757
3922	0.00767	3.922	10957	0.002	10.957
4122	0.00767	4.122	11157	0.003	11.157
4322	0.00867	4.322	11357	0.002	11.357
4526	0.00333	4.526	11557	0.00267	11.557
4726	0.00567	4.726	11757	0.00267	11.757
4927	0.00733	4.927	11957	0.00267	11.957
5127	0.007	5.127	12157	0.002	12.157
5327	0.005	5.327	12357	0	12.357
5527	0.005	5.527	12717	0.00167	12.717
5727	0.009	5.727	12922	0.00267	12.922
5927	0.012	5.927	13122	0.00367	13.122
6329	0.011	6.329	13322	0.002	13.322
6529	0.00667	6.529	13522	0.003	13.522
6729	0.00267	6.729	13722	0.00567	13.722
6929	0.00867	6.929	13922	0.01567	13.922
7129	0.01167	7.129	14129	0.01533	14.129

14331	0.01467	14.331	22215	0.00067	22.215
14531	0.00567	14.531	22415	0.00067	22.415
14734	0.004	14.734	22615	0.00167	22.615
14934	0.00267	14.934	22815	0.00167	22.815
15134	0.00233	15.134	23015	0.00167	23.015
15342	0.00233	15.342	23215	0	23.215
15563	0.002	15.563	23415	0.00067	23.415
15991	0.00133	15.991	23615	0.00167	23.615
16192	0.00133	16.192	23815	0.00167	23.815
16392	0.002	16.392	24015	0.00233	24.015
16598	0.00367	16.598	24215	0.00333	24.215
16798	0.00667	16.798	24415	0.005	24.415
16998	0.00867	16.998	24615	0.00467	24.615
17205	0.00967	17.205	24815	0.00267	24.815
17409	0.01067	17.409	25015	0.001	25.015
17609	0.01067	17.609	25215	0.001	25.215
17809	0.01033	17.809	25415	0.00367	25.415
18009	0.008	18.009	25615	0.00533	25.615
18214	0.00733	18.214	25815	0.009	25.815
18415	0.008	18.415	26015	0.00733	26.015
18615	0.00733	18.615	26215	0.00867	26.215
18815	0.004	18.815	26436	0.00467	26.436
19015	0.001	19.015	26636	0.00367	26.636
19215	0	19.215	26836	0.00433	26.836
19415	0	19.415	27036	0.00467	27.036
19615	0	19.615	27236	0.00867	27.236
19815	0.00067	19.815	27436	0.005	27.436
20015	0.00067	20.015	27636	0.009	27.636
20215	0.00167	20.215	27836	0.005	27.836
20415	0.003	20.415	28036	0.005	28.036
20615	0.00667	20.615	28236	0	28.236
20815	0.00633	20.815	28436	0	28.436
21015	0.00733	21.015	28636	0.00267	28.636
21215	0.00467	21.215	28836	0.00267	28.836
21415	0.006	21.415	29047	0.00267	29.047
21615	0.003	21.615	29251	0	29.251
21815	0.002	21.815	29451	0.001	29.451
22015	0.00067	22.015	29651	0.00167	29.651

29851	0.00267	29.851	37468	0.003	37.468
30057	0.00167	30.057	37668	0.002	37.668
30257	0.002	30.257	37868	0	37.868
30457	0.001	30.457	38068	0.001	38.068
30657	0.00367	30.657	38268	0.001	38.268
30857	0.00467	30.857	38468	0.001	38.468
31059	0.00567	31.059	38668	0.002	38.668
31262	0.004	31.262	38868	0.002	38.868
31463	0.00267	31.463	39068	0.003	39.068
31663	0.00267	31.663	39268	0.00267	39.268
31864	0.00167	31.864	39468	0.00267	39.468
32064	0.002	32.064	39668	0.00167	39.668
32264	0.00167	32.264	39868	0.001	39.868
32464	0.00567	32.464	40068	0.00167	40.068
32667	0.00833	32.667	40268	0.00167	40.268
32867	0.00767	32.867	40468	0.00133	40.468
33067	0.00633	33.067	40669	0.00333	40.669
33267	0.00333	33.267	40876	0.00833	40.876
33468	0.00433	33.468	41086	0.01	41.086
33668	0.00467	33.668	41286	0.008	41.286
33868	0.004	33.868	41486	0.003	41.486
34068	0.003	34.068	41690	0.00167	41.69
34268	0.001	34.268	41890	0.002	41.89
34468	0.001	34.468	42093	0.002	42.093
34668	0.002	34.668	42293	0.002	42.293
34868	0.00167	34.868	42493	0.001	42.493
35068	0.00433	35.068	42693	0.00267	42.693
35268	0.005	35.268	42893	0.00167	42.893
35468	0.00533	35.468	43097	0.004	43.097
35668	0.00367	35.668	43297	0.004	43.297
35868	0.002	35.868	43509	0.005	43.509
36068	0.002	36.068	43716	0.00633	43.716
36268	0.004	36.268	43932	0.00533	43.932
36468	0.00467	36.468	44134	0.00433	44.134
36668	0.00367	36.668	44334	0.00067	44.334
36868	0.00067	36.868	44534	0.00067	44.534
37068	0.001	37.068	44734	0.00433	44.734
37268	0.003	37.268	44943	0.00833	44.943

45145	0.01233	45.145	52915	0.004	52.915
45348	0.01133	45.348	53115	0.002	53.115
45558	0.015	45.558	53316	0.001	53.316
45764	0.01133	45.764	53518	0.008	53.518
45969	0.01133	45.969	53727	0.011	53.727
46169	0.007	46.169	53932	0.012	53.932
46373	0.007	46.373	54132	0.007	54.132
46573	0.01433	46.573	54332	0.008	54.332
46781	0.012	46.781	54532	0.00667	54.532
46999	0.01633	46.999	54732	0.00667	54.732
47209	0.00633	47.209	54932	0.00467	54.932
47409	0.00733	47.409	55132	0.00567	55.132
47613	0.00267	47.613	55332	0.00467	55.332
47813	0.00467	47.813	55562	0.00367	55.562
48013	0.00267	48.013	55762	0.006	55.762
48219	0.004	48.219	56237	0.00567	56.237
48419	0.004	48.419	56439	0.00567	56.439
48619	0.005	48.619	56639	0.00367	56.639
48819	0.004	48.819	56844	0.003	56.844
49019	0.002	49.019	57045	0.002	57.045
49219	0.002	49.219	57245	0.00267	57.245
49419	0.00167	49.419	57445	0.00267	57.445
49644	0.00333	49.644	57857	0.004	57.857
49914	0.00467	49.914	58057	0.002	58.057
50114	0.006	50.114	58257	0.002	58.257
50315	0.00433	50.315	58457	0.00133	58.457
50515	0.006	50.515	58657	0.00067	58.657
50715	0.005	50.715	58859	0.00067	58.859
50915	0.005	50.915	59060	0	59.06
51115	0.002	51.115	59260	0.001	59.26
51315	0.002	51.315	59460	0.00167	59.46
51515	0.004	51.515	59660	0.00367	59.66
51715	0.004	51.715	59860	0.00467	59.86
51915	0.00467	51.915	60060	0.00567	60.06
52115	0.00433	52.115	60263	0.00733	60.263
52315	0.00433	52.315	60463	0.00633	60.463
52515	0.00467	52.515	60663	0.00733	60.663
52715	0.005	52.715	60863	0.00367	60.863

61063	0.00267	61.063	67878	0.006	67.878
61263	0	61.263	68078	0.003	68.078
61463	0.00367	61.463	68278	0.003	68.278
61663	0.00633	61.663	68478	0.002	68.478
61868	0.009	61.868	68678	0.002	68.678
62085	0.00633	62.085	68878	0.002	68.878
62375	0.00533	62.375	69078	0.003	69.078
62576	0.00333	62.576	69278	0.002	69.278
62776	0.004	62.776	69478	0.002	69.478
62976	0.00333	62.976	69678	0.00167	69.678
63176	0.00467	63.176	69878	0.00267	69.878
63382	0.005	63.382	70078	0.00167	70.078
63582	0.004	63.582	70278	0.001	70.278
63782	0.003	63.782	70478	0.001	70.478
64051	0.003	64.051	70678	0.00267	70.678
64258	0.003	64.258	70879	0.00433	70.879
64458	0.002	64.458	71080	0.00333	71.08
64663	0.002	64.663	71280	0.00367	71.28
64866	0.002	64.866	71480	0.002	71.48
65066	0.002	65.066	71680	0.006	71.68
65266	0.00133	65.266	71880	0.00567	71.88
65466	0.00133	65.466	72080	0.00667	72.08
65666	0.00133	65.666	72280	0.00433	72.28
65866	0	65.866	72480	0.00267	72.48
66067	0	66.067	72680	0.00333	72.68
66267	0.003	66.267	72880	0.00167	72.88
66467	0.003	66.467	73080	0.00167	73.08
66668	0.00667	66.668	73280	0.00267	73.28
66875	0.00367	66.875	73488	0.00367	73.488
67075	0.00567	67.075	73688	0.00467	73.688
67275	0.003	67.275	73888	0.00333	73.888
67475	0.007	67.475	74088	0.006	74.088
67678	0.006	67.678	74288	0.007	74.288
67878	0.006	67.878	74494	0.00567	74.494
68078	0.003	68.078	74694	0.002	74.694
68278	0.003	68.278	74894	0.00367	74.894
68478	0.002	68.478	75095	0.00367	75.095
68678	0.002	68.678	75296	0.00567	75.296

75496	0.00533	75.496	83125	0.00167	83.125
75696	0.00533	75.696	83325	0.002	83.325
75896	0.00333	75.896	83525	0.001	83.525
76096	0.002	76.096	83725	0.001	83.725
76299	0.004	76.299	83925	0	83.925
76509	0.005	76.509	84125	0	84.125
76709	0.00467	76.709	84325	0	84.325
76910	0.00367	76.91	84525	0	84.525
77110	0.00367	77.11	84725	0	84.725
77310	0.003	77.31	84925	0.001	84.925
77514	0.006	77.514	85125	0.001	85.125
77714	0.005	77.714	85325	0.001	85.325
77914	0.00567	77.914	85525	0	85.525
78114	0.00367	78.114	85725	0	85.725
78324	0.00367	78.324	85925	0	85.925
78524	0.003	78.524	86125	0	86.125
78724	0.003	78.724	86325	0	86.325
78924	0.00567	78.924	86525	0	86.525
79124	0.00467	79.124	86725	0	86.725
79324	0.00467	79.324	86925	0	86.925
79524	0.00867	79.524	87125	0	87.125
79725	0.00867	79.725	87325	0	87.325
79925	0.00767	79.925	87525	0	87.525
80125	0.001	80.125	87725	0	87.725
80325	0.001	80.325	87925	0	87.925
80525	0	80.525	88125	0	88.125
80725	0	80.725	88325	0	88.325
80925	0	80.925	88525	0	88.525
81125	0.001	81.125	88725	0	88.725
81325	0.001	81.325	88925	0	88.925
81525	0.001	81.525	89125	0.00067	89.125
81725	0	81.725	89325	0.00067	89.325
81925	0.001	81.925	89525	0.00167	89.525
82125	0.003	82.125	89725	0.001	89.725
82325	0.00367	82.325	89925	0.002	89.925
82525	0.00333	82.525	90144	0.002	90.144
82725	0.002	82.725	90344	0.002	90.344
82925	0.00233	82.925	90544	0.001	90.544

90744	0	90.744	98344	0	98.344
90944	0	90.944	98544	0.002	98.544
91144	0	91.144	98744	0.002	98.744
91344	0	91.344	98944	0.002	98.944
91544	0	91.544	99144	0	99.144
91744	0	91.744	99344	0	99.344
91944	0	91.944	99544	0.001	99.544
92144	0	92.144	99744	0.001	99.744
92344	0.001	92.344	99944	0.00367	99.944
92544	0.001	92.544	100145	0.00267	100.145
92744	0.001	92.744	100345	0.00267	100.345
92944	0.00067	92.944	100545	0	100.545
93144	0.00167	93.144	100746	0.002	100.746
93344	0.00167	93.344	100956	0.00367	100.956
93544	0.001	93.544	101156	0.00767	101.156
93744	0	93.744	101356	0.00733	101.356
93944	0	93.944	101556	0.00667	101.556
94144	0	94.144	101756	0.00333	101.756
94344	0.001	94.344	101956	0.00167	101.956
94544	0.001	94.544	102156	0.00333	102.156
94744	0.001	94.744	102356	0.00267	102.356
94944	0	94.944	102556	0.00267	102.556
95144	0	95.144	102756	0.001	102.756
95344	0	95.344	102956	0.00267	102.956
95544	0	95.544	103156	0.007	103.156
95744	0	95.744	103361	0.00967	103.361
95944	0	95.944	103566	0.01233	103.566
96144	0	96.144	103768	0.00867	103.768
96344	0	96.344	103985	0.008	103.985
96544	0	96.544	104185	0.01333	104.185
96744	0	96.744	104385	0.01767	104.385
96944	0	96.944	104614	0.01633	104.614
97144	0	97.144	104820	0.01167	104.82
97344	0	97.344	105020	0.00667	105.02
97544	0	97.544	105220	0.00967	105.22
97744	0	97.744	105420	0.00667	105.42
97944	0	97.944	105620	0.00933	105.62
98144	0	98.144	105820	0.00633	105.82

106026	0.00433	106.026	113653	0	113.653
106226	0.00167	106.226	113853	0.00167	113.853
106426	0.001	106.426	114053	0.00267	114.053
106626	0.002	106.626	114254	0.00367	114.254
106826	0.00267	106.826	114454	0.002	114.454
107026	0.00167	107.026	114654	0.001	114.654
107226	0.00167	107.226	114854	0	114.854
107427	0.001	107.427	115054	0	115.054
107627	0.004	107.627	115254	0.002	115.254
107827	0.00367	107.827	115454	0.002	115.454
108027	0.00533	108.027	115654	0.002	115.654
108227	0.00333	108.227	115854	0	115.854
108427	0.00433	108.427	116054	0	116.054
108627	0.00833	108.627	116254	0	116.254
108827	0.008	108.827	116454	0	116.454
109027	0.00733	109.027	116654	0	116.654
109227	0.00267	109.227	116854	0	116.854
109427	0.004	109.427	117054	0	117.054
109628	0.003	109.628	117254	0	117.254
109828	0.004	109.828	117454	0	117.454
110028	0.003	110.028	117654	0	117.654
110232	0.003	110.232	117854	0	117.854
110432	0.003	110.432	118054	0	118.054
110632	0.005	110.632	118254	0	118.254
110832	0.006	110.832	118454	0	118.454
111037	0.007	111.037	118654	0	118.654
111245	0.004	111.245	118854	0	118.854
111445	0.005	111.445	119054	0	119.054
111646	0.003	111.646	119254	0	119.254
111852	0.00667	111.852	119454	0.001	119.454
112052	0.00467	112.052	119654	0.001	119.654
112252	0.00467	112.252	119854	0.001	119.854
112452	0.002	112.452	120054	0	120.054
112652	0.00267	112.652	120254	0	120.254
112852	0.00467	112.852	120454	0	120.454
113052	0.00567	113.052	120654	0	120.654
113252	0.005	113.252	120854	0.001	120.854
113452	0.002	113.452	121054	0.00167	121.054

121254	0.00167	121.254	129873	0	129.873
121454	0.00067	121.454	130073	0	130.073
121654	0.001	121.654	130273	0	130.273
121854	0.001	121.854	130473	0.001	130.473
122054	0.001	122.054	130673	0.001	130.673
122254	0	122.254	130873	0.002	130.873
122454	0	122.454	131073	0.00167	131.073
122654	0	122.654	131273	0.00233	131.273
122854	0	122.854	131473	0.002	131.473
123054	0	123.054	131673	0.00333	131.673
123254	0	123.254	131873	0.00267	131.873
123454	0	123.454	132073	0.003	132.073
123654	0.001	123.654	132273	0.001	132.273
123854	0.001	123.854	132473	0.001	132.473
124054	0.002	124.054	132673	0.001	132.673
124273	0.002	124.273	132873	0.001	132.873
124473	0.002	124.473	133073	0.001	133.073
124673	0.00167	124.673	133273	0	133.273
124873	0.00067	124.873	133473	0	133.473
125073	0.00067	125.073	133673	0	133.673
125273	0	125.273	133873	0.001	133.873
125473	0	125.473	134073	0.001	134.073
125673	0	125.673	134273	0.001	134.273
125873	0	125.873	134390	0.00924	134.39
126073	0	126.073			
126273	0	126.273			
126473	0	126.473			
126673	0	126.673			
126873	0	126.873			
127073	0	127.073			
127273	0	127.273			
127473	0	127.473			
127673	0	127.673			
127873	0	127.873			
128073	0	128.073			
128273	0	128.273			
128473	0	128.473			
128673	0	128.673			

Table S3.1. USDA PI collections by the country of origins. Accessions used in this study covered 20 countries, with Iran having the most entries.

Location	PI Number					
Bosnia and Herzegovina	251128					
Bulgaria	634302	634303	636567			
Canada	236832					
China	499640	595130	595140	595145	595146	595158
	618975	595167	595170	595178	618972	634304
	655206	W6 23550	W6 23594	W6 23622		
Croatia	251421					
England	595052					
Germany	237708	422463				
Greece	206561	249739				
Hungary	257740	257741				
Iran	227362	227506	227507	229453	229454	229533
	229456	229497	229502	229503	230247	
	251384	251385	268234	330706	380845	380846
	380847	380848	380849	380850	380851	380852
	380853	380854	380855	380856	380857	384860
	384861	384863	547398			
Macedonia	250965	250967				
Montenegro	251123	251125	251126	251127	251131	
Netherlands	237179	315448				
Poland	274619	283320	287541			
Russia	115358	312453	314522	314523	314571	314687
	316249	371896	538933	538934	676177	
Spain	234478	234750	234751	234752	234758	287822
	287823	289652	302898	302899	302900	311403
	311405	318989	318990			
Switzerland	234895	234896	234897	234898	235072	
Turkey	109497	206268	340103	383650	383651	383652
	383653	383654	383655	568183		
United States	578733	537103				
Wales	577098	577099	595049	595050	595051	595059
	595060	595061	595062	595063		
Unknown	189146					

Table S3.2. DNA content estimation and standard deviation of the USDA PI accessions. Data was sorted by the genome size from the smallest to the largest.

PI number	Average Genome Size (pg)	Genome Size SD (pg)	Estimated ploidy level
109497	4.57	0.39	2
115358	3.77	0.19	2
189146	4.61	0.01	2
206268	4.51	0.40	2
206561	4.31	0.20	2
227362	9.77	1.04	4
227506	8.52	0.39	4
227507	8.49	0.27	4
229453	9.54	0.89	4
229454	8.39	0.37	4
229456	14.18	0.46	6
229497	8.72	0.35	4
229502	8.80	1.03	4
229503	14.45	0.64	6
229533	13.02	0.61	6
230247	13.26	0.48	6
234478	4.58	0.33	2
234750	18.07	0.36	8
234751	12.51	0.85	5 (5.32)
234752	8.80	0.70	4
234758	9.74	0.61	4
234895	11.52	0.31	5 (4.90)
234896	14.02	0.44	6
234897	4.20	0.26	2
234898	15.42	0.31	7 (6.56)
236832	13.06	0.31	6
237179	4.22	0.24	2
237708	9.14	0.28	4
249739	4.78	0.18	2
250965	3.94	0.23	2
250967	13.41	0.09	6
251123	9.32	0.18	4
251125	10.18	0.92	4
251126	12.15	0.49	5 (5.17)

251127	10.00	0.71	4
251128	9.14	0.27	4
251131	10.12	0.75	4
251384	4.97	0.15	2
251385	13.96	0.86	6
251421	4.74	0.19	2
257740	4.20	0.21	2
257741	4.64	0.15	2
268234	14.01	1.13	6
274619	12.86	1.25	5 (5.47)
283320	10.14	0.29	4
287541	10.56	0.46	4
287822	9.42	0.92	4
287823	9.36	0.57	4
289652	7.67	0.20	3 (3.27)
302898	9.31	0.32	4
302899	19.66	0.24	8
302900	9.03	0.33	4
311403	18.47	1.18	8
311405	12.61	1.20	5 (5.36)
312453	10.49	0.71	4
314522	12.75	0.74	5 (5.43)
314523	12.71	0.92	5 (5.41)
314571	4.28	0.15	2
314687	4.26	0.32	2
315448	11.98	0.33	5 (5.10)
316249	8.96	0.74	4
318989	19.43	0.76	8
318990	11.05	0.67	5 (4.70)
330706	9.43	0.54	4
340103	9.26	0.27	4
371896	12.57	0.60	5 (5.35)
380845	13.63	0.23	6
380846	13.25	0.30	6
380847	12.24	0.89	5 (5.21)
380848	13.13	0.29	6
380849	15.39	0.37	7 (6.55)
380850	12.31	0.62	5 (5.24)

380851	13.79	0.33	6
380852	14.05	0.80	6
380853	14.35	0.62	6
380854	13.81	0.86	6
380855	14.11	1.12	6
380856	4.77	0.21	2
380857	13.83	0.24	6
383650	9.46	0.36	4
383651	4.87	0.10	2
383652	4.47	0.38	2
383653	8.15	0.23	3 (3.47)
383654	8.38	0.65	4
383655	7.85	0.33	3 (3.34)
384860	11.40	0.60	5 (4.85)
384861	13.08	0.55	6
384863	14.98	0.25	6
422463	4.21	0.36	2
499640	4.35	0.18	2
537103	13.17	1.17	6
538933	11.69	0.32	5 (4.98)
538934	4.17	0.34	2
547398	13.49	0.19	6
568183	4.22	0.23	2
577098	7.62	0.45	3 (3.24)
577099	7.84	0.84	3 (3.33)
578733	4.79	0.21	2
595049	7.62	0.41	3 (3.24)
595050	9.27	0.48	4
595051	9.85	0.30	4
595052	9.34	0.50	4
595059	7.58	0.95	3 (3.23)
595060	8.41	0.51	4
595061	9.36	0.74	4
595062	8.71	1.02	4
595063	8.52	0.43	4
595130	4.10	0.43	2
595140	4.30	0.30	2
595145	4.05	0.21	2

595146	3.82	0.04	2
595158	4.23	0.29	2
595167	4.25	0.04	2
595170	4.47	0.03	2
595178	4.34	0.29	2
618972	5.26	0.24	2
618975	4.08	0.12	2
634302	8.29	0.30	4
634303	4.47	0.17	2
634304	4.51	0.27	2
636567	3.98	0.28	2
655206	5.33	1.00	2
676177	4.63	0.32	2
W6 23550	5.71	0.27	2
W6 23594	5.63	0.09	2
W6 23622	5.40	0.01	2

Table S3.3. The linear regression to associate seed size with ploidy levels. *t* statistics suggested there was a significant difference in seed size between different ploidy levels.

Estimate	Std.	Error	*t* value	Pr (>|*t*|)
(Intercept)	2.1199	0.1079	19.644	< 2.00E-16 ***
Ploidy 4X	0.7584	0.1428	5.312	4.31E-07 ***
Ploidy 6X	2.1432	0.1428	15.013	< 2.00E-16 ***
Ploidy 8X	2.3051	0.1526	15.104	< 2.00E-16 ***

Signif. codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ', 1

Table S3.4. The ANOVA analysis of PI accessions using ploidy levels as the variable. Significant differences were found between seed size and the corresponding ploidy level.

Df	Sum	Sq	Mean	Sq	F- value	Pr(>F)
Ploidy	3	121.102	40.367	1.16E+02	< 2.20E-16	***
Residuals	136	47.514	0.349			

Signif. codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ', 1

Table S4.1. Primer sequences used for ccs file trimming using the lima program

Primer ID	Primer Sequence
NEB_5p	GCAATGAAGTCGCAGGGTTGGG
NEB_Clontech_3p	GTACTCTGCGTTGATACCACTGCTT
Clontech_5p	AAGCAGTGGTATCAACGCAGAGTACATGGGG

Table S4.2. miRNA gene families identified in *F. brevipila* NR transcriptome.

Species	miRNA family
Brachypodium distachyon (5)	bdi-MIR156a
	bdi-MIR171a
	bdi-MIR397a
	bdi-MIR5056
	bdi-MIR7767
Festuca arundinacea (5)	far-MIR1119
	far-MIR1122
	far-MIR1134
	far-MIR159
	far-MIR166
Hordeum vulgare (12)	hvu-MIR5053
	hvu-MIR6178
	hvu-MIR6179
	hvu-MIR6182
	hvu-MIR6187
	hvu-MIR6189
	hvu-MIR6196
	hvu-MIR6199
	hvu-MIR6206
	hvu-MIR6207
	hvu-MIR6214
Oryza sativa (8)	osa-MIR168b
	osa-MIR2927
	osa-MIR396d
	osa-MIR396h
	osa-MIR5075
	osa-MIR5523
	osa-MIR5532
	osa-MIR5538
Zea mays (1)	zma-MIR397a
Triticum aestivum (2)	tae-MIR530
	tae-MIR9773
Populus euphratica (1)	peu-MIR2916
Glycine max (1)	gma-MIR6300
Elaeis guineensis (1)	hpa-MIR156a

Table S5.1. Eight CYP92 genes downloaded from the rice genome database for the phylogenetic tree reconstruction.

Gene ID	Gene Name
LOC_Os02g29960.1	CPY92A15
LOC_Os03g44740.1	CYP92C1
LOC_Os09g26980.1	CYP92A7
LOC_Os09g08920.1	CYP92A13
LOC_Os09g08990.1	CYP92A14
LOC_Os09g26940.1	CYP92A11
LOC_Os09g26960.1	CYP92A9
LOC_Os08g35510.1	CYP92A12

Table S5.2. Selected downregulated genes in propiconazole-treated plants

TRANSCRIPT_17444	AEV91173.1	2.67	MYB-related protein
TRANSCRIPT_45477	AAW88315.1	4.70	Expansin EXPA11
TRANSCRIPT_45494	XP_010228647.1	3.61	Peroxidase
TRANSCRIPT_46420	CAC06433.1	3.23	Expansin
TRANSCRIPT_46942	CAC40806.1	4.69	Beta expansin B3
TRANSCRIPT_50937	CAC40805.1	3.81	Beta expansin B2
TRANSCRIPT_56300	AAT67050.1	3.10	Pathogenesis-related protein 4
TRINITY_DN100223_C1_G3_I3	AJK93406.1	6.23	Cinnamyl alcohol dehydrogenase
TRINITY_DN100821_C1_G2_I9	XP_020168398.1	9.37	Dynamin-like protein ARC5
TRINITY_DN102104_C1_G1_I1	EMS62391.1	9.17	Splicing factor u2af large subunit A
TRINITY_DN104757_C2_G2_I1	NP_037635.1	10.83	RNA-dependent RNA polymerase
TRINITY_DN105176_C4_G2_I3	XP_020172503.1	8.47	Microtubule-associated protein futsch isoform X1
TRINITY_DN105573_C2_G1_I2	XP_020186500.1	8.70	Putative glucuronosyltransferase PGSIP8

TRINITY_DN75902_C1_G1_I6	XP_010230204.1	8.37	DNA (cytosine-5)-methyltransferase CMT1-like
TRINITY_DN76758_C2_G2_I10	ACB45302.1	8.56	Expansin EXPA11
TRINITY_DN79436_C4_G1_I1	XP_020201563.1	2.80	Photosystem II
TRINITY_DN82258_C5_G1_I15	XP_020162795.1	8.05	Transportin MOS14
TRINITY_DN82950_C1_G1_I7	XP_024315534.1	8.00	Serine/threonine-protein phosphatase 6
TRINITY_DN91298_C0_G1_I8	XP_010228785.2	8.38	ABC transporter B family member 6
TRINITY_DN92704_C1_G2_I5	XP_003571639.1	8.09	Exonuclease V, chloroplastic
TRINITY_DN97393_C1_G1_I3	BAD87988.1	2.15	Putative beta-1,3-glucanase precursor
TRINITY_DN97393_C1_G2_I6	AAU11328.1	3.92	Beta-1,3-glucanase 2a
TRINITY_DN98820_C0_G1_I8	XP_020173114.1	7.86	Pentatricopeptide
TRANSCRIPT_23832	XP_004975013.1	-2.29	Alpha-humulene synthase
TRANSCRIPT_24929	Q9FSV7.1	-2.10	Sucrose 1-fructosyltransferase
TRANSCRIPT_33779	XP_020170118.1	-4.37	ALP1-like protein
TRANSCRIPT_34799	AER39773.1	-2.37	Cyp92a44-3

TRANSCRIPT_43011	ACD80366.1	-2.19	WRKY5 transcription factor
TRANSCRIPT_44511	ASU89566.1	-2.11	Drought resistance
TRANSCRIPT_46869	XP_020196033.1	-2.31	RING-H2 finger protein ATL3-like
TRANSCRIPT_47987	XP_020174323.1	-2.29	Histone-lysine N Methyltransferase 2D-like
TRANSCRIPT_53911	XP_020171700.1	-3.44	Adenomatous polyposis
TRANSCRIPT_55428	XP_020198520.1	-3.87	Heavy metal-associated isoprenylated plant protein 47-like
TRANSCRIPT_55823	XP_003563436.1	-3.59	Calcium-binding protein PBP1
TRANSCRIPT_56908	AAF86307.1	-4.05	EF-hand Ca2+ -binding protein CCD1
TRINITY_DN104570_C1_G1_I7	EMS68053.1	-7.80	Eukaryotic
TRINITY_DN105516_C3_G2_I17	XP_003573998.1	-9.81	Flug protein
TRINITY_DN105968_C8_G1_I3	XP_020179664.1	-9.84	RNA uridylyltransferase 1-like
TRINITY_DN71649_C4_G1_I1	XP_020162934.1	-7.68	CO(2)-response secreted protease-like

TRINITY_DN72232_C0_G2_I2	XP_020163886.1	-6.29	Zinc finger protein CONSTANS-LIKE 1-like isoform X2
TRINITY_DN73568_C4_G1_I3	XP_020174579.1	-3.00	Downy mildew resistance protein
TRINITY_DN73598_C0_G1_I4	XP_003577532.1	-9.68	RNA polymerase
TRINITY_DN75106_C1_G2_I9	AQK83574.1	-8.15	NAC domain-containing
TRINITY_DN76083_C0_G1_I4	AFR67777.1	-8.22	AP2 domain CBF protein
TRINITY_DN76820_C2_G1_I3	EMS67097.1	-8.39	Putative 6-phosphogluconolactonase
TRINITY_DN78461_C5_G1_I12	XP_020103799.1	-2.52	Protein ROOT HAIR DEFECTIVE
TRINITY_DN79264_C0_G1_I7	EMS56979.1	-7.94	Protein SPA1-RELATED 3
TRINITY_DN82257_C0_G1_I13	XP_003557393.1	-8.56	Protein glycosyltransferase
TRINITY_DN84047_C5_G2_I3	XP_003557748.1	-8.15	3beta-hydroxysteroid-dehydrogenase/decarboxylase
TRINITY_DN92959_C3_G2_I9	XP_020182066.1	-7.43	Ras-related protein RABA1f-like isoform X2
TRINITY_DN93007_C0_G3_I5	EMS57783.1	-9.02	Naringenin,2-oxoglutarate
TRINITY_DN93029_C0_G1_I8	XP_020193397.1	-3.48	Transcription factor bhlh101-like

TRINITY_DN93083_C0_G1_I13	XP_020164890.1	-11.23	Bromodomain-containing protein
TRINITY_DN95724_C1_G1_I12	XP_020177953.1	-3.80	Long chain acyl-coA synthetase
TRINITY_DN98193_C2_G1_I3	XP_003570531.1	-2.44	Metal tolerance
TRINITY_DN98261_C0_G2_I5	XP_020188012.1	-8.69	AT-hook motif nuclear-localized
TRINITY_DN98980_C0_G1_I9	XP_003558564.1	-8.07	Peptidyl-prolyl cis-trans
TRINITY_DN99735_C0_G1_I2	XP_003563336.1	-9.14	Vacuolar-sorting receptor 6

Table S5.3. Unique metabolite features that were only present in propiconazole-treated plants.

Column	variableMetadata	mz	Rt (min)	isotopes	adduct	PCA_XLOAD-p1	PCA_XLOAD-p2
cHILIC Positive	M483T454	483.184001	7.56	[268][M]+			
C18 Positive	M571T432	571.347308	7.20			-0.0139596	-0.0188306
	M593T441_1	593.355206	7.35			-0.0114425	-0.0174622
	M593T441_2	593.362072	7.35			-0.0115762	-0.0175756
	M593T440	593.368887	7.34			-0.0123899	-0.0168769
	M608T619_2	608.119962	10.32		[M+H]+ 607.11	-0.0208268	-0.0206162
	M608T640_1	608.120887	10.66			-0.0171249	-0.0158836
	M608T619_3	608.127042	10.32		[M+H]+ 607.11	-0.0204855	-0.0222937
	M608T640_2	608.128	10.66			-0.0173484	-0.0158309
	M608T619_4	608.134148	10.32	[678][M]+	[M+K]+ 569.175 [M+Na]+ 585.148 [M+H]+ 607.13	-0.020523	-0.0222508
	M608T619_5	608.141176	10.31		[M+K]+ 569.175 [M+Na]+ 585.148 [M+H]+ 607.13	-0.0204305	-0.0166609
	M615T449_1	615.37466	7.49			-0.0130924	-0.0184822
	M615T449_2	615.381948	7.49			-0.0125187	-0.0186049
	M616T449	615.874948	7.49	[690][M+1]2+		-0.010771	-0.019124
	M637T456_1	637.375532	7.60	[727][M]2+		-0.0113154	-0.0165058
	M637T456_2	637.383195	7.61			-0.0135044	-0.0179125
	M637T456_3	637.390698	7.61	[728][M]2+		-0.0118337	-0.0167707
	M637T457	637.398435	7.61	[729][M]2+		-0.0136284	-0.0175941

	M638T456_2	637.88751	7.60	[728][M+1]2+		-0.0134333	-0.017127
	M638T456_3	637.895186	7.61	[729][M+1]2+		-0.0123065	-0.0171608
	M640T456_1	639.851993	7.60	[737][M]2+		-0.0113178	-0.0178182
	M640T456_2	639.859472	7.61			-0.0104075	-0.0166203
	M640T456_3	639.867468	7.61			-0.0114184	-0.0177042
	M640T457_1	639.875145	7.61			-0.0112872	-0.0166212
	M640T457_2	639.882824	7.61	[738][M]2+		-0.0113314	-0.0177269
	M642T457_1	642.324932	7.61	[746][M]2+		-0.0109258	-0.0179524
	M642T457_2	642.33245	7.61			-0.0108574	-0.0172782
	M642T457_3	642.340093	7.62			-0.0117694	-0.0173735
	M642T457_4	642.347806	7.62	[747][M]2+		-0.0119698	-0.0174446
	M643T457_1	642.834867	7.61	[746][M+1]2+		-0.0101979	-0.0157111
	M659T465_1	659.386497	7.75	[766][M]2+		-0.0117899	-0.0194691
	M659T465_2	659.394458	7.75			-0.0120167	-0.0194397
	M659T464_1	659.402449	7.74			-0.0128677	-0.0183468
	M659T464_2	659.410406	7.74	[767][M]2+		-0.012455	-0.0188563
	M660T465_2	659.90088	7.75			-0.0130456	-0.0188385
	M662T465_6	662.379487	7.75	[771][M+1]2+		-0.0119088	-0.0196229
	M662T465_7	662.387561	7.74			-0.0121793	-0.0180809
	M682T472_1	681.907326	7.86	[786][M+1]2+		-0.0127004	-0.0199883
	M682T472_2	681.915729	7.86	[787][M+1]2+		-0.0125969	-0.0199104
	M682T472_3	681.924185	7.86	[788][M+1]2+		-0.012637	-0.0200158
	M703T479_1	703.410169	7.98		[M+H+NH3]+ 685.388	-0.0113493	-0.0192656
	M703T479_2	703.419118	7.98	[814][M]2+		-0.013806	-0.0168831

	M703T479_3	703.427714	7.98		[M+H+NH3]+ 685.388	-0.0137803	-0.0175943
	M703T479_4	703.436658	7.98		[M+K+NH3]+ 647.448 [M+Na+NH3]+ 663.421 [M+H+NH3]+ 685.403	-0.0134555	-0.0176596
	M704T479_1	703.923621	7.98	[814][M+1]2+		-0.0138542	-0.0182407
	M704T479_2	703.93245	7.98		[M+H+NH3]+ 685.901	-0.0138344	-0.0186102
	M704T479_3	703.941564	7.98	[815][M+1]2+		-0.0111573	-0.0194685
	M706T479_1	705.890313	7.98		[M+Na]+ 682.905 [M+H]+ 704.887	-0.013397	-0.018718
	M706T479_2	705.899108	7.98		[M+Na]+ 682.905 [M+H]+ 704.887	-0.0134922	-0.0189003
	M706T478_1	705.907976	7.97		[M+Na]+ 682.928 [M+H]+ 704.91	-0.0135676	-0.0191987
	M706T478_2	705.916847	7.97		[M+Na]+ 682.928 [M+H]+ 704.91	-0.0136005	-0.019453
	M725T485_1	725.427537	8.09	[830][M]2+		-0.0132702	-0.0176122
	M725T485_2	725.436796	8.09			-0.0136304	-0.017764
	M725T485_3	725.446062	8.09			-0.0137952	-0.0179665
	M725T485_4	725.455263	8.08	[831][M]2+		-0.0130963	-0.0182164
	M726T485_1	725.937464	8.09	[830][M+1]2+		-0.013518	-0.014674
	M726T485_2	725.946659	8.09	[831][M+1]2+		-0.0132776	-0.0156483
	M728T485_1	728.414756	8.09			-0.0135898	-0.0171831

	M728T485_2	728.423954	8.09			-0.0141058	-0.0177795
	M750T491_1	749.928648	8.19			-0.0137401	-0.01693
	M750T491_2	749.938383	8.19			-0.0140348	-0.0171043
C18 Negative	M564T584_1	564.10466	9.74		[M+Cl]- 529.145 [M-H]- 565.121	-0.0172695	0.0153133
C18 Negative	M564T584_2	564.111014	9.74		[M+Cl]- 529.145 [M-H]- 565.121	-0.016888	0.0158583
C18 Negative	M564T584_3	564.117357	9.74		[M+Cl]- 529.145 [M-H]- 565.121	-0.016991	0.01577289
C18 Negative	M564T554	564.117298	9.24			-0.0168645	0.01486005
C18 Negative	M564T584_4	564.123705	9.74		[M+Cl]- 529.145 [M-H]- 565.121	-0.0169758	0.0164383
C18 Negative	M566T575_1	566.111921	9.58		[M+Cl]- 531.146 [M-H]- 567.122	-0.0183743	0.01465833
C18 Negative	M566T584_1	566.111993	9.74			-0.0162502	0.01653315
C18 Negative	M566T575_2	566.118309	9.58		[M+Cl]- 531.146 [M-H]- 567.122	-0.0184635	0.0149058
C18 Negative	M566T584_2	566.118378	9.74			-0.0161048	0.01609361
C18 Negative	M606T592_1	606.11213	9.86		[M+Cl]- 571.146	-0.0125693	0.01342044
C18 Negative	M606T619_1	606.112144	10.32		[M+Cl]- 571.146 [M-H]- 607.123	-0.0128963	0.01596625
C18 Negative	M606T640_1	606.1121	10.66		[M+Cl]- 571.15 [M-H]- 607.126	-0.0127031	0.01535215
C18 Negative	M606T592_2	606.119207	9.86		[M+Cl]- 571.146	-0.0126471	0.01327077

C18 Negative	M606T619_2	606.119222	10.32		[M+Cl]- 571.146 [M-H]- 607.123	-0.0130426	0.01589101
C18 Negative	M606T640_2	606.119177	10.66		[M+Cl]- 571.15 [M-H]- 607.126	-0.0128196	0.01541979
C18 Negative	M606T592_3	606.126284	9.86		[M+Cl]- 571.165	-0.0127462	0.01310416
C18 Negative	M606T619_3	606.1263	10.32		[M-H]- 607.141	-0.0131634	0.01584598
C18 Negative	M606T640_3	606.126248	10.66		[M+Cl]- 571.15 [M-H]- 607.126	-0.0139195	0.01688238
C18 Negative	M606T640_4	606.133324	10.66		[M+Cl]- 571.167 [M-H]- 607.144	-0.0141463	0.01702072
C18 Negative	M606T619_4	606.133378	10.32		[M-H]- 607.141	-0.0134093	0.01423602
C18 Negative	M606T619_5	606.140457	10.32		[M-H]- 607.141	-0.0134335	0.01421024
C18 Negative	M606T640_5	606.140401	10.66		[M+Cl]- 571.167 [M-H]- 607.144	-0.0143307	0.01722117
C18 Negative	M608T592_1	608.112714	9.86		[M-2H+K]- 571.165	-0.0131961	0.01460066
C18 Negative	M608T627_1	608.112681	10.46			-0.0122196	0.01599102
C18 Negative	M608T592_2	608.119827	9.86		[M-2H+K]- 571.165	-0.0135163	0.01497318
C18 Negative	M608T627_2	608.119738	10.46			-0.0125472	0.0184865
C18 Negative	M608T619_2	608.119896	10.32		[M+Cl]- 573.15 [M-H]- 609.127	-0.0119532	0.01817977

C18 Negative	M608T592_3	608.126942	9.86		[M+Cl]- 573.161 [M-H]- 609.138	-0.0136858	0.01482927
C18 Negative	M608T627_3	608.126853	10.46		[M+Cl]- 573.161 [M-H]- 609.138	-0.0122316	0.01653782
C18 Negative	M608T619_3	608.12696	10.32		[M+Cl]- 573.15 [M-H]- 609.127	-0.0129627	0.01584281
C18 Negative	M608T592_4	608.134056	9.86		[M+Cl]- 573.161 [M-H]- 609.138	-0.0139911	0.0126923
C18 Negative	M608T627_4	608.134018	10.46		[M+Cl]- 573.161 [M-H]- 609.138	-0.0123377	0.01602038