MULTIPLE APPROACHES TO MAPPING RESISTANCE TO *MELAMPSORA* LEAF RUST IN SHRUB WILLOW (*SALIX*)

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Dustin Guy Wilkerson

August 2021

MULTIPLE APPROACHES TO MAPPING RESISTANCE TO *MELAMPSORA* LEAF RUST IN SHRUB WILLOW (*SALIX*)

Dustin Guy Wilkerson, Ph. D.

Cornell University 2021

In commercial production, shrub willow grown as a biomass bioenergy crop can potentially remain in the field for more than 20 years. In these dense plantations, biotic stressors can flourish, impacting biomass yield. Among the most impactful are willow leaf rusts caused by *Melampsora* spp. As a macrocyclic and heteroecious rust, breeding for resistance is challenging. Most of the research into the *Melampsora* and *Salix* pathosystem centers around European species, however, revealing a need for the characterization of resistance of their North American counterparts. Two mapping populations have been generated for the species *S. purpurea*, which has high quality reference genomes and is naturalized in North America. The first is an F_2 population, while the second is the *Salix* F_1 hybrid common parent population with eight families. In this population, four females from *S. suchowensis*, *S. viminalis*, and *S. integra* were crossed with a male *S. purpurea*, while four males from *S. suchowensis*, *S. viminalis*, *S. udensis*, and *S. koriyanagi* were crossed with a female *S. purpurea*. The objectives of this research were to 1) identify the temporal, gene-level response to infection and candidate genes associated with resistance to *M. americana* within the *S. purpurea* F_2 population, 2) describe the relatedness within, produce linkage maps, and map the sex determination regions within the *Salix* F_1 hybrid mapping population and within those

same linkage maps 3) identify QTL for leaf rust severity in addition to a multitude of agronomically important traits. Analysis of 3′ RNA-Seq conducted with genotypes resistant to *M. americana* showed a more coordinated response to infection compared to susceptible genotypes leading to the identification of candidate genes involved in plant defense. The production of 16 linkage maps facilitated the mapping of the sex determination regions to the maternal chromosome 15 in each pedigree, revealing that these species use a ZW system. Through the analysis of 41 traits, 87 QTL were mapped in the F_1 hybrid common parent population, including five leaf rust QTL. Future studies should focus on candidate gene and QTL validation in order to expedite the introgression of resistance into North American *Salix*.

BIOGRAPHICAL SKETCH

Dustin grew up in a small town in northeast Missouri. As he got older, he became involved in youth agriculture organizations such as 4H and FFA and learned the importance of agriculture in society. In an introductory plant science class at the University of Missouri – Columbia, he learned about plant breeding and was immediately attracted to its inherent interdisciplinary nature, requiring a functional knowledge of several aspects of agricultural production and plant health. During his undergraduate years, Dustin held internships with MFA, DuPont Pioneer (now Corteva) and Monsanto (now merged with Bayer).

After graduation with a degree in plant science and a minor in agriculture economics, he was accepted into a Master's program at Texas A&M University, majoring in plant breeding and worked on improving drought tolerance in cotton (*Gossypium* spp.), gaining practical knowledge in running a field-based breeding program. He was also able to serve as a teaching assistant and take leadership roles in Texas A&M's Graduate Student Council as the Service Committee chair and helped plan two DuPont Pioneer Plant Science Symposia. His Master's thesis centered on a group of cotton germplasm known as converted race stocks (CRS). These are cotton landraces that had been bred through recurrent selection to flower in Texas but had little to no other form of selection for yield or agronomic traits. Tasked with using handheld phenotyping sensors that measured normalized difference vegetation index (NDVI), leaf surface temperature, stomatal conductance, and chlorophyll content, he hoped to identify novel germplasm among the CRS lines that could bring new alleles

into Texas A&M's breeding program for conferring better yields under dryland/drought conditions. The results showed that though there is potential among the CRS lines for novel alleles that confer drought tolerance, their significant underperformance in fiber yield and quality compared to industry and experimental elite checks lead to a conversation about the cost and benefits of bringing unimproved germplasm into a breeding program.

Upon completing his Master's degree in December 2016, Dustin began his doctoral education at Cornell University under Dr. Lawrence Smart, majoring in plant breeding. Continuing to capitalize on opportunities for leadership and education, Dustin served as the Synapsis (plant breeding graduate student organization) President in 2018 and as a teaching assistant for which he was awarded an *Outstanding College of Agriculture and Life Science Teaching Assistant* award in 2018. Before the end of his first year, Dustin had submitted a collaborative grant proposal to the Schmittau-Novak Small Grants Program offered through the School of Integrative Plant Science at Cornell with two other students. With a budget of $6000, he proposed an experiment that utilized an interspecific polyploidy trio (diploid, tetraploid, and their triploid progeny) grafted in all possible scion/rootstock combinations to investigate if polyploid heterosis (hybrid vigor) could be conveyed across the graft junction and improve the performance of the scion. Using phenotype and 3′ RNA-Seq data, the results showed that rather than the higher ploidy rootstocks improving the performance of the scion, the diploid rootstock significantly stunted the growth of the higher ploidy scions.

During his second year at Cornell, Dustin submitted a grant proposal to the USDA-NIFA predoctoral fellowship program and was recommended for funding. Proposing to use genomic approaches on several existing shrub willow (*Salix* spp.) breeding populations to identify potential sources of resistance to *Melampsora* leaf rust, arguably the most devastating pathogen affecting commercial willow production. This project has several components that include the creation of genetic maps of diverse willow species, quantitative trait loci (QTL) mapping, and time series 3′ RNA-Seq greenhouse inoculation experiment. During his time at Cornell, Dustin was recognized for his outstanding performance in scholarship, research, and service as a recipient of the *Munger/Murphy Award in Plant Breeding* in 2019, named to commemorate the distinguished careers of Professors Henry M. Munger and Royse P. Murphy. Dustin hopes to use his education as a faculty member at an agriculture university, leading a well-rounded breeding program that makes use of both conventional and genomic breeding methods.

I would like to dedicate this document to my grandma, Bonnie Sudsberry, who passed away while I was writing it. She was the kindest, most compassionate person I have ever known and her never ending love and encouragement will be with me always.

ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge Dr. Larry Smart for his guidance and support in helping me to pursue this degree. Larry has been far more patient with me than I feel I deserved. He offered course correcting guidance when I needed it and pointed out my successes when I needed that too. The example he has set has been exemplary. I would also like to acknowledge Drs. Craig Carlson and Eric Fabio. When I first came into the lab, you both would take the time to answer my questions and give advice. I am also extremely grateful to Chase Crowell, my willow leaf rust partner-in-crime. Things were not always easy, but we made it through! The Smart lab has had a remarkable technical staff during my time here. The "WOW" team, Rebecca Wilk, Lauren Carlson, Jane Petzoldt, and Dawn Fishback, was at full strength when I arrived. They were always available and ready to help out or just a small round of gossip. I am both humbled and grateful for the years of support, guidance, and advice from my committee, Dr. Chris Smart, Dr. Bruce Reisch, and Dr. Mike Gore. Funding for the work completed in this document was supported by grants from the United States Department of Agriculture National Institute of Food and Agriculture (USDA-NIFA) #2015-67009-23957 and #2018-68005-27925 and my USDA NIFA predoctoral fellowship program grant #2019-67011-29701.

Finally, I would like to acknowledge the love and support given to me by my family and Dr. Ben Weikert. Without them, this would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

BREEDING SHRUB WILLOW (*SALIX)* FOR RESISTANCE TO *MELAMPSORA* LEAF RUST

1.1 *Abstract*

Shrub willow is commercially grown as a biomass bioenergy crop and is fast-growing, carbon neutral, and can be produced on marginal land. Grown in dense, short rotation coppice (SRC) plantations, diseases can easily take hold and flourish. The most threatening among them is willow leaf rust caused by fungi in the genus *Melampsora*. Considering the yearly potential for sexual recombination in the *Melampsora* spp. life cycle and the more than 25-year duration of SRC production, breeding elite yielding cultivars with durable resistance is challenging. Through the acquisition of diverse genotypes and their characterization for resistance to leaf rust, plant breeders can start the process of using marker-assisted selection for trait introgression and development of durable resistance through the pyramiding of resistance genes. This literature review focuses on breeding for disease resistance and genomics-assisted methods in addition to presenting the results of QTL mapping and RNA-Seq studies into the *Salix* and *Melampsora* pathosystem.

1.2 Introduction

1.2.1 Shrub Willow Breeding

Shrub willow (*Salix* spp.) bioenergy crops, together with poplar, Miscanthus, switchgrass, and Arundo, have many redeeming features as sources of biomass

bioenergy. Willow in particular is a fast-growing perennial that is carbon neutral when grown on land that was previously used for crops and can be grown on poorly drained marginal land unsuitable for many food crops (Kuzovkina & Volk, 2009; Smart et al., 2005; Stoof et al., 2014). Building upon its other uses in ornamental horticulture and phytoremediation, shrub willow is a desirable alternative to other bioenergy crops, especially annuals, that require productive soils for reliable biomass yields. Together with poplar (*Populus* spp.), *Salix* resides within the Salicaceae family, broadly described as trees, shrubs and subshrubs that are dioecious and highly heterozygous. A diverse genus, *Salix* accounts for at least 350 species with ploidies ranging from diploid to dodecaploid (Dickmann & Kuzovkina, 2014; Volk et al., 2006). This innate diversity is beneficial for breeding programs in Europe and the United States that have been active since the 1970s-80s developing cultivars for commercial production.

Commercial producers of shrub willow use a short rotation coppice (SRC) system (Smart et al., 2005). Dormant cuttings are planted in the spring and allowed to grow until the winter when they are cut back (coppiced). In the spring, these plants respond with vigorous regrowth that is then harvested every three to four years until the plants are removed (Figure 1.1). Evaluations into the number of effective harvest cycles before plants reach their peak marginal utility in biomass yield are still ongoing but long-term studies from the late 1990s and early 2000s suggest plants can remain productive for at least 25 years (Gouker et al., 2021; Kopp et al., 2001; Stolarski et al., 2019; Willebrand et al., 1993). Given the long tenure of released cultivars in the field, the accumulation of beneficial alleles through breeding and selection is paramount in developing robust cultivars for commercial production.

Figure 1.1: Single-row shrub willow harvester (Cornell University, 2021).

In beginning the breeding programs in the United States, attempts at introducing elite genotypes of *S. viminalis*, the most popular SRC willow in Europe, failed due to extreme susceptibility to potato leafhopper (*Empoasca fabea*), a widespread North American insect. Since then, mostly Asian species of *Salix* have been bred with naturalized species for cultivar development (Serapiglia et al., 2014b). Most advances occur through the capture of heterosis resulting from species hybridization and clonal propagation. Among these hybrids, triploids formed from diploid by tetraploid crosses have shown improved vigor, yield, and biomass composition relative to their parents (Fabio et al., 2017a; Fabio et al., 2017b; Serapiglia et al., 2014a). Through the intentional combination of species within *Salix*, progress can be made on a litany of breeding objectives beyond biomass yield, including agronomic, physiological and insect and disease resistances (Serapiglia et al., 2013; Smart & Cameron, 2008).

1.2.2 Melampsora *Leaf Rust*

In the dense plantations of commercially produced shrub willow, willow leaf rusts caused by species of the genus *Melampsora* are among the most threatening. Potentially resulting in complete defoliation, yield losses of roughly 50%, and increased chances of secondary infections and herbivory (McCracken & Dawson, 1992). Fungi within the genus *Melampsora* are macrocyclic and heteroecious rusts, with all five stages of the life cycle observed, and require an alternate host for sexual reproduction (Karp et al., 2011; Pei et al., 1996). This sexual recombination drives high levels of population diversity within each rust species, making breeding for

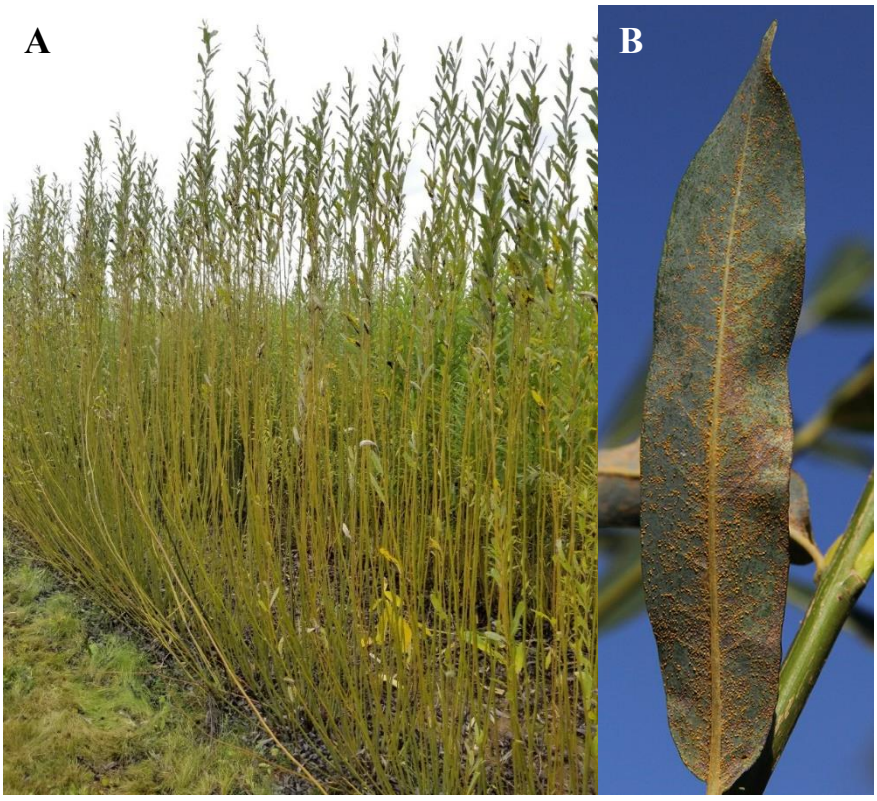


Figure 1.2: Willow leaf rust symptoms caused by *Melampsora* species. A: Leaf defoliation. B: Orange to yellow pustules on the underside of the leaf. Photo Credit: Dustin Wilkerson.

resistance challenging (Crowell et al, in prep). Prevalent across the globe, *Melampsora* spp. can infect willow, poplar, flax, and other hosts (Pei, 2005). Within the Northeast United States, three main taxa were identified as the most prevalent on *Salix*, *M. paradoxa*, *M. americana*, and a yet to be defined species complex (Crowell et al., 2020; Kenaley et al., 2014).

Management of disease severity in SRC willow requires broad spectrum defense, able to neutralize the potential of multiple species and diverse populations of *Melampsora*. Partial control can be achieved through the use of fungicides, but is widely considered to be financially, environmentally, and unsustainable in practice (Dawson & McCracken, 1994; Shield et al., 2015). Planting clonal mixtures comprised of individuals with varying resistance to willow leaf rust, either within the same species or differing species, to control the spread through the field is a far more producer friendly means of control (McCracken & Dawson, 2003; McCracken et al., 2001; McCracken et al., 2005). While there is debate over whether mono- versus polyculture plantings provide more selection pressure for races that defeat resistance, both require characterization of the relative resistances of planted cultivars to a diverse pathogen population. Only through the assessment, selection, and accumulation of resistance (R) genes into high yielding elite cultivars will producers be provided with the best means of control.

1.2.3 Breeding for Disease Resistance

Incorporation of genetic resistance into released cultivars is the environmentally preferred management method that, in conjunction with other

integrative strategies, can reduce the losses caused by plant diseases and the need for control by pesticides (Sánchez-Martín & Keller, 2019; Sharma et al., 2019). Breeding for effective disease resistance relies on the accumulation of R genes; however, the effectors they recognize vary in relative strength. In what is known as the gene-for-gene concept, these R proteins are activated in response to a specific product from the pathogen, a virulence or effector gene (Dangl et al., 2013; Flor, 1971). As mentioned above, monoculture plantings with undefeated R genes increase the selection pressure on pathogens with effectors that avoid R gene recognition, overcoming resistance. Through analysis of this exchange, plant pathologists and breeders can work together to identify the R genes that recognize the largest subset of effectors present in a pathogen population.

Dangl et al. (2013) describes the plant immune system in two tiers. The first is comprised of pattern-recognition receptors (PRR) that are activated when they recognize a pathogen- or microbial-associated molecular pattern (PAMP/MAMP). As PAMP and MAMP recognize evolutionarily conserved regions, resistance achieved at this stage is considered a non-host or incompatible interaction. Pathogens not thwarted by the first tier then encounter the second, mostly made up of intercellular nucleotide-binding leucine-rich repeat (NLR) receptors. This second tier is really where the gene-for-gene interactions are engaged as each NLR only recognizes specific effector protein sequences. The back and forth between the loss of effector recognition and the identification and incorporation of diverse R genes that counter them represent a major challenge for plant breeders and pathologists alike.

Identifying major R genes that confer qualitative or complete resistance is rare although ideal. Typically plant breeders work with quantitative or incomplete resistance that within breeding populations range in the severity of disease that occurs (Sánchez-Martín & Keller, 2019). This distribution of resistance is attributable to the cumulative effects of many genes conferring differing levels of resistance. Tapping into this distribution allows us to take advantage of the genes working in concert to confer durable resistance (Nelson et al., 2018). In shrub willow where plants can remain in fields for decades, bolstering biomass yield with durable forms of resistance that combine R gene and quantitative resistance is the only way forward. Using breeding populations with variable levels of resistance to willow leaf rust in conjunction with genomic data will better enable the identification of R genes and the sources of quantitative resistance.

1.3 Genomics-Assisted Breeding

1.3.1 Genomics

In making selections, plant breeders have a diverse set of tools at their disposal. The most appropriate selection tool varies depending on the targeted trait and each has their advantages and disadvantages. In many circumstances, conventional selection based on phenotypes is the best method, especially if the trait is qualitative and can be described with minimal human error. However, there is a quantitative threshold when traits are controlled by a few large effect genes where employing genomic methods become more effective than conventional selection (Bernardo, 2001). As the complexity of genetic control extends beyond this sweet spot,

conventional selection once again becomes more effective at trait improvement unless using genome-wide markers (Heffner et al., 2009). In making selections for leaf rust resistance, a quantitative trait in *Salix* (Carlson et al., 2019; Hanley et al., 2011), characterization of the genetic underpinning of resistance is essential for determining the best method for accurate and reliable selection.

Genomic resources are becoming increasing available to most breeding programs as sequencing costs decline and throughput increases, resulting in the generation of genome assemblies of increasing quality. Useful for DNA sequence alignment and variant detection, reference genomes allow researchers to use a common genetic sequence across multiple studies. With Zhou et al. (2020) having released the highest-quality, annotated reference in *Salix* (https://phytozome-next.jgi.doe.gov), its subject species, *S. purpurea*, has been used to generate mapping populations. The creation and sequencing of bi- and multi-parental populations followed by alignment to a reference genome facilitates more accurate marker ordering and physical position. Through the advent of genotyping-by-sequencing (GBS) (Elshire et al., 2011), low cost, genome-wide single-nucleotide polymorphisms (SNPs) have become the most commonly used markers as they are co-dominant, abundant, and found throughout the genome (Leng et al., 2017). Through associating these markers with phenotypes, trait variation can be mapped to specific regions of the genome, quantitative trait loci (QTL), revealing potential candidate genes within its boundaries (Paterson et al., 1988).

Another approach to gaining a better understanding of the genetics behind certain traits relies on quantitative analysis of the transcriptome. RNA-Seq allows the

quantification of gene expression, one step closer to the end phenotype than DNA. Effective application of RNA-Seq studies require plants of significant phenotypic differences or the same plants subjected to variable treatments. Through intentional experimental design, genes that are differentially expressed between phenotypic or treatment groups will indicate the complex genetic mechanisms influencing the final phenotype. Gene expression data can even be related directly to the phenotype, mapping an expression QTL (eQTL). eQTL analysis in a segregating population can indicate the regulatory structure controlling gene expression, either as -*cis* (near) or -*trans* (far). However, the largest obstacle to overcome in RNA-Seq experiments is the necessary bioinformatics skills needed to complete them (Costa-Silva et al., 2017).

1.3.2 Marker-Assisted Selection

Once the genes affecting the final phenotype have been identified and validated, the markers associated with them can be targeted for marker-assisted selection (MAS). MAS has the definitive advantage over conventional selection of being time and resource saving, enabling selection during the seedling stage, on single plants, or based on their hetero- or homozygosity at specific markers (Collard & Mackill, 2008). Most importantly, MAS gives plant breeders options in determining the best path toward reaching their breeding objectives. MAS is not without its shortcomings, however. Depending on the genetic control of the trait, the proximity of the associated marker to the causative gene, the size of the breeding population, lack of polymorphism in pedigrees of interest, and the complexity of the genetic background will all play a role in the overall efficacy of MAS (Boopathi, 2013). While

not applicable for every trait, MAS can be adapted to many crossing, selection, and breeding applications. Using multiple previously identified markers, MAS was used to identify the geographic origin and breeding status of 70 Indian genotypes in mango (*Mangifera indica*) (Jena & Chand, 2021), while in hybrid rice (*Oryza* spp.), Yashitola et al. (2002) demonstrated MAS could be used to confirm seed lot purity. More applicable to breeding for willow leaf rust resistance, however, are marker-assisted backcrossing (MAB) and pyramiding.

Backcrossing is not a new concept. It is typically applied when one parent (the recurrent parent) has many desirable characteristics but is lacking in one or a small number of areas while the other parent (the donor parent) is proficient for the genes that the recurrent parent lacks. The recurrent and donor parents are crossed and progeny expressing the target phenotype are selected and crossed back to the recurrent parent. This is repeated until the background of the recurrent parent has been recovered, producing elite cultivars with the best parts of both parents. The application of MAS into this procedure can significantly cut down the time between generations and overall resources as plants not carrying the desirable genetics can be culled during the seedling stage. During the early generations of MAB, foreground selection is applied to select only those carrying the desired region from the donor parent. As generations progress, stricter and stricter background selection is applied to ensure recovery of the recurrent parent. In marker-assisted pyramiding, this concept is applied to multiple parents backcrossed and recovered consecutively or employing foreground selection on multiple markers from the same donor. In dioecious species like shrub willow that suffer from inbreeding depression in advanced generations, using

background selection from the onset reduces the generation requirement, but can also result in fewer selections and possibly incomplete recovery of the recurrent parent.

Marker-assisted selection is useful for the introgression of biotic and abiotic stress resistance into elite cultivars (Anderson & Hubricht, 1938; Das et al., 2017). An implication of trait introgression is that the donor parent is a new introduction to the breeding program, with limited adaptation and agronomic potential. Introgression is often, but not exclusively, reserved for when the functional diversity available in a breeding program is lacking. The constant assessment and introduction of new material into breeding programs is necessary for maintaining a supply of accessible variation (Pratap et al., 2021). Through evaluation of new, diverse genotypes, plant breeders are able to respond more easily to the ever-changing landscape of producer needs. With plant diseases like willow leaf rust, characterizing the resistance available in as many species as possible is important for resistance breeding. Using parents that contrast for resistance and employing the genomics-assisted methods described above offer the greatest potential for genetic gain.

1.4 Identification of Resistance Through Genomics

1.4.1 Populus

In reviewing the literature for studies mapping leaf rust resistance in *Salix*, it is important to consider the macrosynteny between the willow and poplar genomes (Hanley et al., 2006b). Early studies into willow often leveraged the genetic resources of poplar to improve the mapping resolution in *Salix* (Berlin et al., 2010), so consideration should be given to the interaction between *Populus* spp. and

Melampsora spp. in addition to shrub willow. A population of 343 hybrid poplar (*P. deltoides* × *P. trichocarpa)* genotypes was inoculated with seven different isolates of poplar leaf rust (*M. larici-populina*) and mapped QTL based on infection ratings (Jorge et al., 2005). A majority of them co-localized with a previously discovered but defeated resistance gene, theorizing that this gene may still be imparting partial resistance. Among the other QTL, isolate-specific interactions resulted in varying presence and effect. In another study focused on the differential expression of poplar to infection by *M. medusae* f. sp. *deltoidae* and the more pathogenic *M. larici-populina*, Azaiez et al. (2009) found that only 54% of the 416 'rust-response' genes were common among the two species. Of the 46% displaying differential expression, most were associated with the defense response genes triggered by the less pathogenic *M. medusae* f. sp. *deltoidae*.

More recently, the leaf physiological effects of infection by *M. medusae* were compared between resistant and susceptible *P. deltoides*. Measuring photosynthetic rate, intercellular CO_2 concentration, leaf chlorophyll content, and stomatal conductance, Gortari et al. (2018) found inconsistency between their field and greenhouse experiments although concluding that while photosynthetic rate is affected by infection regardless of the level of resistance, the susceptible poplar experienced a greater reduction. Dissecting an R gene supercluster previously found on chr19 in *P. deltoides* for resistance to *M. larici-populina*, Wei et al. (2020a) used differential expression analysis and qRT-PCR to determine that a pathogenesis-related protein (*PR-1)* had the greatest response to infection and therefore was the primary executor of resistance within the supercluster. Although not an exhaustive presentation of the

work completed in poplar, these studies provide insight into working with multiple *Melampsora* species and the effects they can have on identifying resistance genes against diverse pathogen populations.

1.4.2 Salix

A majority of the work mapping resistance genes in *Salix* has largely utilized mapping populations constructed using *S. viminalis* and European species of the willow rust pathogen, but work on *S. purpurea* in North America is expanding. Many of the populations utilizing *S. viminalis* have been crosses with *S. schwerinii*, a known source of resistance to *M. larici-epitea*. Using backcross, *S. viminalis* × (*S. viminalis* × *S. schwerinii*), and F_1, *S. viminalis* × *S. viminalis* mapping populations Samils et al. (2011) used a compartmentalized ratings system to study leaf rust severity. Parsing it into uredinia number, latent period (time between infection and pustule formation) and diameter, they successfully mapped QTL for *M. larici-epitea*. In both populations, they found significance on linkage group 19, having aligned the sequence data to the poplar genome based on *P. trichocarpa*, where a cluster of rust resistance genes was already found. A majority of the other QTL detected were population specific, however, owing to the high diversity of rust genes within and between *Salix* species. In the same backcross population, Martin et al. (2016) mapped QTL for resistance to *M. larici-epitea* finding one of major effect, 56% of the variation, on linkage group 1b of the *S. purpurea* reference genome (v1). Upon further analysis of this region, RGA1, a TIR-NBS-LRR (Toll Interleukin1 Receptor – Nucleotide Bind Site – Leucin-Rich Repeat) was targeted for expression analysis where the authors theorized it played an

important role in pathogen recognition but did not have evidence to suggest that it actively played a role in pathogen defense. While constitutively expressed in both susceptible and resistant genotypes; resistant genotypes expressed more RGA1 supposedly crossing a 'critical threshold' for early pathogen recognition and defense while the susceptible were not able to recognize the pathogen prior to infection. In a separate backcross population using *S. viminalis* and *S. schwerinii*, Sulima et al. (2017) used percentage infected leaf area to map QTL using four separate isolates of *M. larici-populina*. They identified 11 total QTL on six separate linkage groups for rust susceptibility to three out of four isolates. For each isolate, QTL varied in presence and overall effect, suggesting that although collected from the same field each carried differing affinities for infection, congruent with the findings of Ramstedt et al. (2002) that examined 37 over multiple years. This overview of the work that has been done is beneficial for our understanding of disease resistance in *Salix*; however, what it also shows is that the interaction between specific willow and *Melampsora* species has an effect on the discoverable sources of resistance. By focusing on the research within *S. purpurea*, we will be more informed on the QTL associated with North American species of *Melampsora*.

Although there is limited research into resistance in *S. purpurea*, several studies have been published in the past few years and our resources are expanding. Both a *S. purpurea* F_2 population and an association panel of naturalized accessions were used to map a wide range of morphological, physiological, insect and disease resistance and biomass composition traits (Carlson et al., 2019). In mapping leaf rust resistance, they identified QTL on chr01, 05 and 10 each containing numerous NBS-

LRR and PR genes. Although this is the only QTL mapping study currently published, others have added more depth to this interaction. Assessing a subset of diverse genotypes from the breeding program in the northeast United States, Crowell et al. (2020), in addition to identifying *M. americana* as the predominant species present within their field trials, also found that stomatal and trichome density are contributors to resistance and potential breeding objectives. While important first steps, more research is needed before implementation into breeding programs.

1.5 Hypotheses and Expected Outcomes

Given the broad genetic diversity in both *Salix* and *Melampsora*, breeding for resistance to willow leaf rust is challenging. Research into the interaction of resistance genes with specific pathogen isolates is ongoing, but the bulk of this research has been conducted on poplar and European shrub willow species. For focused resistance breeding efforts in North America, there is an immediate need to increase our understanding of willow leaf rust on native and naturalized species, as well as commercial cultivars and pre-commercial breeding lines of shrub willow. Through the application of genomics, the location of R genes and the markers associated with them can be incorporated into resistance breeding strategies. By increasing the diversity within breeding programs, more R genes can be incorporated into elite cultivars through introgression. Given the tenure of willow plants used in SRC production, pyramiding multiple R genes to account for newly evolved effector genes will offer the best path toward durable resistance.

The objectives in this research focus on the use of two breeding populations. The first is the *S. purpurea* F_2 population used in Carlson et al. (2019) and the second is the *Salix* F_1 hybrid common parent mapping population that consists of eight families crossed to a *S. purpurea* common parent. Four males from *S. suchowensis*, *S. viminalis*, *S. udensis*, and *S. koriyanagi* were crossed to the female 94006, while four females from *S. suchowensis*, *S. viminalis* and *S. integra* were crossed to the male 94001. One of the female parents was originally described as *S. alberti*, but my work indicates that it was misidentified and is actually another *S. suchowensis*.

Using selected resistant and susceptible individuals from the F_2 population, I sought to identify the temporal response to infection shortly after inoculation by *M. americana* and identify candidate genes influential in this interaction. Through the use of 3' RNA-Seq on an experimental design with time points, treatments, and resistance type, I employed differential expression and network analysis in addition in mapping of eQTL. By overlapping the results of each of these methods, those genes differently expressed through time and treatment and found to be highly connected within network modules show the best potential for being determinants in the success of infection.

The first of two objectives using the common parent F_1 population relies on the characterization of the relatedness within the population, construction of linkage maps, and mapping the sex determination region (SDR). Principal components, hierarchical clustering, and fastSTRUCTURE analyses were all used to determine the population structure within this population. Then using backcross markers, construction of 16 linkage maps enabled the mapping of the sex determination region in the eight

families. Through population analysis and mapping of the SDR, lesser studied species will become more available to willow breeders. The final objective in this research is to use the linkage maps developed in the prior objective to map QTL for an assortment of important traits. Splitting the traits into four groups, leaf rust, leaf architecture, herbivory resistance, and yield component traits, the genetics of each family were assessed for QTL. By identifying QTL for agronomically important traits, this objective sets up further research into QTL refinement and validation, an important first step into improving the performance of elite shrub willow cultivars through trait introgression.

1.6 REFERENCES

Anderson, E., & Hubricht, L. (1938). Hybridization in tradescantia III. The evidence for introgressive hybridization. *Am. J. Bot., 25*(6), 396-402. doi:https://doi.org/10.1002/j.1537-2197.1938.tb09237.x

Azaiez, A., Boyle, B., Levee, V., & Seguin, A. (2009). Transcriptome profiling in hybrid poplar following interactions with *Melampsora* rust fungi. *Mol Plant Microbe Interact, 22*(2), 190-200. doi:10.1094/MPMI-22-2-0190

Berlin, S., Lagercrantz, U., von Arnold, S., Ost, T., & Ronnberg-Wastljung, A. C. (2010). High-density linkage mapping and evolution of paralogs and orthologs in *Salix* and *Populus*. *BMC Genomics, 11*, 129. doi:10.1186/1471-2164-11-129

Bernardo, R. (2001). What if we knew all the genes for a quantitative trait in hybrid crops? *Crop Sci., 41*(1), 1-4. doi:https://doi.org/10.2135/cropsci2001.4111

Boopathi, N. M. (2013). *Genetic mapping and marker-assisted selection*: Springer.

Carlson, C. H., Gouker, F. E., Crowell, C. R., Evans, L., DiFazio, S. P., Smart, C. D., & Smart, L. B. (2019). Joint linkage and association mapping of complex traits in shrub willow (*Salix purpurea* L.). *Ann. Bot., 124*(4), 701-716. doi:10.1093/aob/mcz047

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos T R Soc B, 363*(1491), 557-572. doi:doi:10.1098/rstb.2007.2170

Cornell University. (2021). Willow energy crop information from Cornell University. Willowpedia. https://willow.cals.cornell.edu/.

Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One, 12*(12), e0190152. doi:10.1371/journal.pone.0190152

Crowell, C. R., Bekauri, M. M., Cala, A. R., Mcmullen, P., Smart, L. B., & Smart, C. D. (2020). Differential susceptibility of diverse *Salix* spp. to *Melampsora americana* and *Melampsora paradoxa*. *Plant Dis., 104*, 2949-2957. doi:10.1094/PDIS-04-20-0718-RE

Dangl, J. L., Horvath, D. M., & Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science, 341*(6147), 746-751. doi:10.1126/science.1236011

Das, G., Patra, J. K., & Baek, K.-H. (2017). Insight into MAS: A molecular tool for development of stress resistant and quality of rice through gene stacking. *Front. Plant Sci., 8*(985). doi:10.3389/fpls.2017.00985

Dawson, W., & McCracken, A. (1994). Effect of *Melampsora* rust on the growth and development of *Salix burjatica* 'Korso'in Northern Ireland. *Eur J Forest Pathol, 24*(1), 32-39.

Dickmann, D. I., & Kuzovkina, J. (2014). Poplars and willows of the world, with emphasis on silviculturally important species. In *Poplars and willows: Trees for society and the environment* (Vol. 22). Rome, Italy: FAO.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One, 6*(5), e19379. doi:10.1371/journal.pone.0019379

Fabio, E. S., Kemanian, A. R., Montes, F., Miller, R. O., & Smart, L. B. (2017a). A mixed model approach for evaluating yield improvements in interspecific hybrids of shrub willow, a dedicated bioenergy crop. *Ind. Crops Prod., 96*, 57-70. doi:10.1016/j.indcrop.2016.11.019

Fabio, E. S., Volk, T. A., Miller, R. O., Serapiglia, M. J., Gauch, H. G., Van Rees, K. C. J., Hangs, R. D., Amichev, B. Y., Kuzovkina, Y. A., Labrecque, M., Johnson, G. A., Ewy, R. G., Kling, G. J., & Smart, L. B. (2017b). Genotype × environment interaction analysis of North American shrub willow yield trials confirms superior performance of triploid hybrids. *GCB Bioenergy, 9*(2), 445-459. doi:10.1111/gcbb.12344

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annu Rev Phytopathol, 9*(1), 275-296.

Gortari, F., Guiamet, J. J., & Graciano, C. (2018). Plant–pathogen interactions: leaf physiology alterations in poplars infected with rust (*Melampsora medusae*). *Tree Physiol, 38*(6), 925-935. doi:10.1093/treephys/tpx174

Gouker, F. E., Fabio, E. S., Serapiglia, M. J., & Smart, L. B. (2021). Yield and biomass quality of shrub willow hybrids in differing rotation lengths and spacing designs. *Biomass Bioenerg, 146*, 105977. doi:https://doi.org/10.1016/j.biombioe.2021.105977

Hanley, S. J., Mallott, M. D., & Karp, A. (2006). Alignment of a *Salix* linkage map to the *Populus* genomic sequence reveals macrosynteny between willow and poplar genomes. *Tree Genet. Genomes, 3*(1), 35-48. doi:10.1007/s11295-006-0049-x

Hanley, S. J., Pei, M. H., Powers, S. J., Ruiz, C., Mallott, M. D., Barker, J. H. A., & Karp, A. (2011). Genetic mapping of rust resistance loci in biomass willow. *Tree Genet. Genomes, 7*(3), 597-608. doi:10.1007/s11295-010-0359-x

Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Sci., 49*(1), 1-12. doi:https://doi.org/10.2135/cropsci2008.08.0512

Jena, R. C., & Chand, P. K. (2021). Multiple DNA marker-assisted diversity analysis of Indian mango (*Mangifera indica* L.) populations. *Scientific Reports, 11*(1), 10345. doi:10.1038/s41598-021-89470-3

Jorge, V., Dowkiw, A., Faivre-Rampant, P., & Bastien, C. (2005). Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytol, 167*(1), 113-127. doi:10.1111/j.1469-8137.2005.01424.x

Karp, A., Hanley, S. J., Trybush, S. O., Macalpine, W., Pei, M., & Shield, I. (2011). Genetic improvement of willow for bioenergy and biofuels. *J. Integr. Plant Biol., 53*(2), 151-165. doi:10.1111/j.1744-7909.2010.01015.x

Kenaley, S. C., Smart, L. B., & Hudler, G. W. (2014). Genetic evidence for three discrete taxa of *Melampsora* (Pucciniales) affecting willows (*Salix* spp.) in New York State. *Fungal Biol, 118*(8), 704-720. doi:10.1016/j.funbio.2014.05.001

Kopp, R. F., Abrahamson, L. P., White, E. H., Volk, T. A., Nowak, C. A., & Fillhart, R. C. (2001). Willow biomass production during ten successive annual

harvests. *Biomass Bioenerg, 20*(1), 1-7. doi:https://doi.org/10.1016/S0961-9534(00)00063-5

Kuzovkina, Y. A., & Volk, T. A. (2009). The characterization of willow (*Salix* L.) varieties for use in ecological engineering applications: Co-ordination of structure, function and autecology. *Ecol. Eng., 35*(8), 1178-1189. doi:https://doi.org/10.1016/j.ecoleng.2009.03.010

Leng, P.-f., Lübberstedt, T., & Xu, M.-l. (2017). Genomics-assisted breeding – A revolutionary strategy for crop improvement. *J. Integr. Agric., 16*(12), 2674-2685. doi:https://doi.org/10.1016/S2095-3119(17)61813-6

Martin, T., Ronnberg-Wastljung, A. C., Stenlid, J., & Samils, B. (2016). Identification of a differentially expressed TIR-NBS-LRR gene in a major QTL associated to leaf rust resistance in *Salix*. *PLoS One, 11*(12), e0168776. doi:10.1371/journal.pone.0168776

McCracken, A. R., & Dawson, M. (1992). Clonal response in *Salix* to *Melampsora* rusts in short rotation coppice plantations. *Eur J Forest Pathol, 22*(1), 19-28. doi:https://doi.org/10.1111/j.1439-0329.1992.tb01332.x

McCracken, A. R., & Dawson, W. M. (2003). Rust disease (*Melampsora epitea*) of willow (*Salix* spp.) grown as short rotation coppice (SRC) in inter- and intra-species mixtures. *Ann Appl Biol, 143*(3), 381-393. doi:DOI 10.1111/j.1744-7348.2003.tb00308.x

McCracken, A. R., Dawson, W. M., & Bowden, G. (2001). Yield responses of willow (*Salix*) grown in mixtures in short rotation coppice (SRC). *Biomass Bioenerg, 21*(5), 311-319. doi:Doi 10.1016/S0961-9534(01)00046-0

McCracken, A. R., Dawson, W. M., & Carlisle, D. (2005). Short-rotation coppice willow mixtures and rust disease development. In M. Pei & A. R. McCracken (Eds.), *Rust diseases of willow and poplar* (pp. 185). Wallingford, UK: CABI Publishing.

Nelson, R., Wiesner-Hanks, T., Wisser, R., & Balint-Kurti, P. (2018). Navigating complexity to breed disease-resistant crops. *Nat. Rev. Genet., 19*(1), 21-33. doi:10.1038/nrg.2017.82

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E., & Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature, 335*(6192), 721-726.

Pei, M. H. (2005). A brief overview of *Melampsora* rusts on *Salix*. In M. H. Pei & A. R. McCracken (Eds.), *Rust Diseases on Willow and Poplar* (pp. 11 - 28). Cambridge, MA: CABI Publishing.

Pei, M. H., Royle, D. J., & Hunter, T. (1996). Pathogenic specialization in *Melampsora epitea* var. *epitea* on *Salix*. *Plant Pathol., 45*(4), 679-690. doi:DOI 10.1046/j.1365-3059.1996.d01-174.x

Pratap, A., Das, A., Kumar, S., & Gupta, S. (2021). Current perspectives on introgression breeding in food legumes. *Front. Plant Sci., 11*(2118). doi:10.3389/fpls.2020.589189

Ramstedt, M., Hurtado, S., & Astrom, B. (2002). Pathotypes of *Melampsora* rust on *Salix* in short-rotation forestry plantations. *Plant Pathol., 51*(2), 185-190. doi:DOI 10.1046/j.1365-3059.2002.00680.x

Samils, B., Rönnberg-Wästljung, A.-C., & Stenlid, J. (2011). QTL mapping of resistance to leaf rust in *Salix*. *Tree Genet. Genomes, 7*(6), 1219-1235. doi:10.1007/s11295-011-0408-0

Sánchez-Martín, J., & Keller, B. (2019). Contribution of recent technological advances to future resistance breeding. *Theor. Appl. Genet., 132*(3), 713-732. doi:10.1007/s00122-019-03297-1

Serapiglia, M. J., Cameron, K. D., Stipanovic, A. J., Abrahamson, L. P., Volk, T. A., & Smart, L. B. (2013). Yield and woody biomass traits of novel shrub willow hybrids at two contrasting sites. *Bioenergy Res., 6*(2), 533-546. doi:10.1007/s12155-012-9272-5

Serapiglia, M. J., Gouker, F. E., Hart, J. F., Unda, F., Mansfield, S. D., Stipanovic, A. J., & Smart, L. B. (2014a). Ploidy level affects important biomass traits of novel shrub willow (*Salix*) hybrids. *Bioenergy Res., 8*(1), 259-269. doi:10.1007/s12155-014-9521-x

Serapiglia, M. J., Gouker, F. E., & Smart, L. B. (2014b). Early selection of novel triploid hybrids of shrub willow with improved biomass yield relative to diploids. *BMC Plant Biol, 14*, 74. doi:10.1186/1471-2229-14-74

Sharma, A., Jones, J. B., & White, F. F. (2019). Recent advances in developing disease resistance in plants. *F1000 Research, 8*, F1000 Faculty Rev-1934. doi:10.12688/f1000research.20179.1

Shield, I., Macalpine, W., Hanley, S., & Karp, A. (2015). Breeding willow for short rotation coppice energy cropping. In *Industrial crops* (pp. 67-80): Springer.

Smart, L. B., & Cameron, K. D. (2008). Genetic improvement of willow (*Salix* spp.) as a dedicated bioenergy crop. In *Genetic improvement of bioenergy crops* (pp. 377-396): Springer.

Smart, L. B., Volk, T., Lin, J., Kopp, R. F., Phillips, I. S., Cameron, K. D., White, E. H., & Abrahamson, L. (2005). Genetic improvement of shrub willow (*Salix* spp.) crops for bioenergy and environmental applications in the United States. *Unasylva, 56*, 51-55.

Stolarski, M. J., Szczukowski, S., Tworkowski, J., Krzyżaniak, M., & Załuski, D. (2019). Willow production during 12 consecutive years—The effects of harvest rotation, planting density and cultivar on biomass yield. *GCB Bioenergy, 11*(4), 635-656. doi:https://doi.org/10.1111/gcbb.12583

Stoof, C. R., Richards, B. K., Woodbury, P. B., Fabio, E. S., Brumbach, A. R., Cherney, J., Das, S., Geohring, L., Hansen, J., Hornesky, J., Mayton, H., Mason, C., Ruestow, G., Smart, L. B., Volk, T. A., & Steenhuis, T. S. (2014). Untapped potential: opportunities and challenges for sustainable bioenergy production from marginal lands in the Northeast USA. *Bioenergy Res., 8*(2), 482-501. doi:10.1007/s12155-014-9515-8

Sulima, P., Przyborowski, J. A., Kuszewska, A., Zaluski, D., Jedryczka, M., & Irzykowski, W. (2017). Identification of quantitative trait loci conditioning the main biomass yield components and resistance to *Melampsora* spp. in *Salix viminalis* x *Salix schwerinii* hybrids. *Int. J. Mol. Sci., 18*(3). doi:10.3390/ijms18030677

Volk, T., Abrahamson, L., Nowak, C., Smart, L., Tharakan, P., & White, E. (2006). The development of short-rotation willow in the northeastern United States for bioenergy and bioproducts, agroforestry and phytoremediation. *Biomass Bioenerg, 30*(8-9), 715-727. doi:10.1016/j.biombioe.2006.03.001

Wei, S., Wu, H., Li, X., Chen, Y., Yang, Y., Dai, M., & Yin, T. (2020). Identification of genes underlying the resistance to *Melampsora larici-populina* in an R Gene supercluster of the *Populus deltoides* genome. *Plant Dis., 104*(4), 1133-1143. doi:10.1094/pdis-08-19-1699-re

Willebrand, E., Ledin, S., & Verwijst, T. (1993). Willow coppice systems in short rotation forestry: Effects of plant spacing, rotation length and clonal composition on biomass production. *Biomass Bioenerg, 4*(5), 323-331. doi:https://doi.org/10.1016/0961-9534(93)90048-9

Yashitola, J., Thirumurugan, T., Sundaram, R. M., Naseerullah, M. K., Ramesha, M. S., Sarma, N. P., & Sonti, R. V. (2002). Assessment of purity of rice hybrids using microsatellite and STS markers. *Crop Sci., 42*(4), 1369-1373. doi:https://doi.org/10.2135/cropsci2002.1369

Zhou, R., Macaya-Sanz, D., Carlson, C. H., Schmutz, J., Jenkins, J. W., Kudrna, D., Sharma, A., Sandor, L., Shu, S., Barry, K., Tuskan, G. A., Ma, T., Liu, J., Olson, M., Smart, L. B., & DiFazio, S. P. (2020). A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biol., 21*(1), 38. doi:10.1186/s13059-020-1952-4

CHAPTER 2

COMPARATIVE TRANSCRIPTOMICS AND eQTL MAPPING OF RESPONSE TO *MELAMPSORA AMERICANA* IN SELECTED *SALIX PURPUREA* F_2 PROGENY

Submitted for publication as: Wilkerson, DG, Crowell, CR, Carlson, CH, McMullen, PW, Smart, CD, Smart, LB. 2021. *BMC Genomics*.

2.1 *Abstract*

Melampsora spp. rusts are the greatest pathogen threat to shrub willow (*Salix* spp.) bioenergy crops. Genetic resistance is key to limiting the effects of these foliar diseases on host response and biomass yield, however, the genetic mechanism of host resistance has not been characterized. The addition of new genomic resources for *Salix* provides greater power to investigate the interaction between *S. purpurea* and *M. americana*, species commonly found in the Northeast US. Here, I utilize 3' RNA-seq to investigate host-pathogen interactions following controlled inoculations of *M. americana* on resistant and susceptible F_2 *S. purpurea* genotypes that were used in a recent study to identify QTL associated with leaf rust resistance. Differential gene expression, network analysis, and eQTL mapping was used to contrast the response to inoculation and to identify associated candidate genes. Controlled inoculation in a replicated greenhouse study identified 19 and 105 differentially expressed genes between resistant and susceptible genotypes at 42 and 66 HPI, respectively. Defense response gene networks were activated in both resistant and susceptible genotypes and enriched for many of the same defense response genes, yet the hub genes of these common response modules showed greater mean expression among the resistant

plants. Further, eight and six eQTL hotspots were identified at 42 and 66 HPI, respectively. The combined results of the three analyses highlight 124 candidate genes in the host for further analysis while analysis of pathogen RNA showed differential expression of 22 genes, two of which are candidate pathogen effectors. We identified two differentially expressed *M. americana* transcripts and 124 *S. purpurea* genes that are good candidates for future studies to confirm their role in conferring resistance.

2.2 Introduction

Shrub willow (*Salix* spp.) are fast-growing perennials that can be grown as a sustainable source of bioenergy, in riparian buffers, or as ornamentals (Kuzovkina & Volk, 2009). *Salix* is incredibly diverse, comprised of over 350 species, with a native range that primarily spans the northern hemisphere, but is cultivated around the world (Dickmann & Kuzovkina, 2014). Of the species found in the northeastern US, naturalized *S. purpurea* has been the focus of bioenergy breeding programs for its high yield, vertical growth habit, and broad resistance to pests and pathogens (Serapiglia et al., 2014b; Smart & Cameron, 2008; Smart et al., 2005). Genomic resources have been developed for the establishment of *S. purpurea* as a model bioenergy crop, which includes high-quality, annotated reference genomes (Zhou et al., 2020) (https://phytozome-next.jgi.doe.gov). In addition, genetic resources have been generated to better understand the inheritance of key traits used in breeding and selection.

The plant pathogen that is the greatest threat to shrub willow grown in commercial production is willow leaf rust (*Melampsora spp.*) (McCracken & Dawson,

1998, 2003; Pei & McCracken, 2005). *Melampsora* rusts infecting willow are lesser-known members of the order Pucciniales that includes wheat stem rust (*Puccinia graminis*), coffee rust (*Hemileia vastatrix*) and over 7,000 other species (Aime et al., 2018; Pei & McCracken, 2005). Previous work has identified *M. americana* as the primary contributor to disease epidemics on *S. purpurea* in the northeastern US (Crowell et al., 2020; Kenaley et al., 2014). Defined as a macrocyclic and heteroecious obligate biotroph, *M. americana* requires an alternate host to complete all five spore stages in its life cycle and cannot be cultured outside of its living host (Karp et al., 2011; Pei et al., 1996). Aeciospores produced on *Abies balsamea* are the primary source of inoculum, traveling to susceptible willow hosts via winds in the late spring and early summer months (Kenaley et al., 2014). Rapid host disease development caused by the production and spread of asexual uredospores on willow which are then spread to nearby leaves via wind, lead to significant yield losses (Verwijst, 1990). Given the prolific nature of this disease, durable genetic resistance is essential to achieving sustained shrub willow biomass yield. Recent investigations have identified morphological characteristics that may impact rust infection, including stomatal density and trichome density (Crowell et al., 2020), however, the genetic basis for *M. americana* rust resistance in willow is not well understood.

In closely related pathosystems, including poplar rust caused by *M. larici-populina* or flax rust caused by *M. lini*, research has identified specific quantitative and qualitative rust resistance loci using candidate gene analysis and QTL mapping approaches (Bresson et al., 2011; Dodds et al., 2004; Lorrain et al., 2015; Petre et al., 2016; Petre et al., 2012; Rönnberg-Wästljung et al., 2008). Most research in the *Salix*

– *Melampsora* pathosystem has focused on *S. viminalis* and *M. larici-epitea* (Martin et al., 2016; Samils et al., 2011; Sulima et al., 2017). While *S. viminalis* is well-adapted and popular in European bioenergy willow breeding programs, *S. purpurea* is most commonly used in the US. Carlson et al., 2019 identified quantitative trait loci (QTL) on chromosomes (chr) 1, 5, and 10 associated with leaf rust resistance in a *S. purpurea* F_2 population. Hanley et al., 2011 also described a rust resistance QTL, *Salix Rust Resistance 1* (SRR1), on chr01. Although genetic mapping studies have identified major effect loci involved in rust resistance, specific genes responsible for host resistance in these populations were not characterized.

RNA-seq has been used to demonstrate that differentially expressed genes coincide with the resistance response in many pathosystems, including potato-*Phytophthora infestans* (Gao et al., 2013), soybean-*Xanthomonas axonopodis* (Kim et al., 2011), and Verticillium wilt in cotton (Xu et al., 2011). Network analysis of 3' RNA-seq data from resistant *S. purpurea* and susceptible *S. viminalis* parents and their segregating F_1 hybrid progeny identified key regulatory hub genes involved in the defense response potato leafhopper (*Empoasca fabea*) (Wang et al., 2020). Hub genes are the most connected genes within a co-expression module that are predicted to be highly influential in regulating the expression of the other genes within their module. Applying expression QTL (eQTL) analysis in a segregating pedigree enables the identification of local *cis* and distant *trans* factors in the genome that regulate the expression levels of key genes correlated with traits of interest. For instance, Mähler et al., 2020 used eQTL analysis to identify a key set of candidate genes that determine leaf shape characteristics in *Populus*.

While much has been learned about willow leaf rust over the past decades (Bennett et al., 2011; Kenaley et al., 2014; Smith et al., 2004), no study has specifically investigated the transcriptomes of *M. americana* and *S. purpurea* shortly after inoculation. This project uses 3' RNA-seq to investigate the post-inoculation expression profiles in resistant and susceptible progeny in a *S. purpurea* F_2 mapping population (Carlson et al., 2019), as well as in the pathogen, *M. americana*.

2.3 Materials and Methods

2.3.1 Inoculation of Salix purpurea *Leaves with* Melampsora americana *Uredospores*

Plants were established from dormant stem cuttings in the greenhouse and grown for two months before inoculation. For the inoculated treatment, 1 mg of uredospores of *M. americana* rust isolate R15-033-03 was applied to each of five mature leaves per plant of each genotype using a paintbrush as previously described (Crowell et al., 2020). Plants were incubated for 12 h in dark mist chambers at 20°C with 100% humidity, then returned to a greenhouse under 14:10 L:D photoperiod at 24ºC:18ºC respectively. Leaf discs (6.4 mm) were collected using a leaf disc puncher (BioSpec Products, Bartlesville, OK) from two leaves starting with the first fully developed mature leaf on each of three shoots. The inoculated shoots were flagged to help identify inoculated leaves at later time points. To determine the optimal time for tissue collection, two replicated greenhouse inoculation experiments were completed on two ‘Fish Creek’ and two 94006 *S. purpurea* plants (treatments = inoculated and control) and leaf discs were collected every 24 h over the course of 5 d. Based on the analysis of that pilot study data, leaf discs were collected from the full study of 60

genotypes (see below) at 42 and 66 HPI. Each time, the leaf discs were collected between 11 am and 2 pm EST then immediately frozen in liquid nitrogen and stored at -80°C until RNA was extracted. Leaf rust severity ratings of the inoculated treatments were visually assessed based on the percentage leaf area covered in uredospore pustules at 9 DPI for comparison between the greenhouse replication and field survey data.

2.3.2 RNA Extraction and 3' RNA-Seq Analysis

Frozen leaf disc tissue was disrupted using a GenoGrinder 2000 (SPEX CertiPrep, Metuchen, NJ) and RNA was isolated using Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO). Resulting RNA was quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA) and quality was assessed using an Experion (Bio-Rad, Hercules, CA). Libraries for 3' RNA-Seq were constructed by the Cornell Institute for Biotechnology (Ithaca, NY) using the Lexogen QuantSeq 3' mRNA-seq Library Prep Kit (Greenland, NH) and sequencing was completed using Illumina (San Diego, CA) NextSeq500 (1x75 bp) technology. Sequencing reads were checked for quality using FastQC Version 0.11.8 (Andrews, 2010) and trimmed using Trimmomatic (Bolger et al., 2014) to remove the polyA tail. The RNA-seq data from host genotype 10X-317-029 collected at 42 HPI in Rep 1 was overrepresented as compared to the other samples sequenced on the same lane. Resulting reads from this sample were randomly subsampled to match the mean read depth of all sequenced samples to 125,000 total reads. Trimmed raw reads were aligned to the *S. purpurea* 94006 v5.1 reference genome (Zhou et al., 2020) using the STAR aligner v2.7.5a

(Dobin et al., 2012). Read counts were generated using HTSeq v0.11.1 (Anders et al., 2014) and differential expression was determined using the R package DEseq2 (Love et al., 2014). Total number of differentially expressed genes was calculated using a direct contrast of the inoculated and control shrub-X-replicate-X-time.

2.3.3 Selection of F_2 Genotypes for eQTL Mapping

This study relied on a *S. purpurea* F_2 population previously reported in Carlson et al., 2019 that was generated by crossing female clone 94006 and male clone 94001. Two F_1 individuals from that cross, 'Fish Creek' and 'Wolcott', selected based on their vertical growth habit, superior yield, and differing resistance to leaf rust, were crossed to generate the F_2 population. The F_2 population is comprised of 485 individuals and is planted in randomized complete blocks in Geneva, NY at Cornell AgriTech. The ratings from 2015 and 2017 (Carlson et al., 2019) were used to identify 28 susceptible and 28 resistant F_2 genotypes by sorting each year by percent severity and identifying genotypes with either consistently high or consistently low severity in both years. Among these 56 genotypes and the two parents and two grandparents of the F_2 population, the correlation between the 2015 and 2017 surveys was 0.86 with a p-value of 8.9e-16. Two plants of each of these 60 genotypes (28 resistant, 28 susceptible, 2 parents and 2 grandparents) were established in 11.4 L pots from dormant stem cuttings planted on June 18 and September 20, 2018 for each of two greenhouse inoculation experiments conducted in separate greenhouse rooms using the inoculation and leaf disc collection procedure described above.

2.3.4 Differential Expression Analysis of S. purpurea *Transcripts*

Analysis of differential expression was conducted to achieve two aims. First was to identify the differential expression between the resistant and susceptible genotypes through a direct contrast by splitting the samples into six time point by treatment groups (0 HPI-INOC, 42 HPI-INOC, 66 HPI-INOC, etc.). The other was to investigate the differential response to infection by contrasting the control and inoculated treatments within the resistant and susceptible genotypes separately by splitting the samples into six time point by type groups (0 HPI-Resistant, 0 HPI-Susceptible, 42 HPI-Resistant, etc.). After genes with low counts (< 10 across all samples) were removed, each group was normalized independently in DESeq2 v1.26 (Love et al., 2014). Sample outliers were then identified and removed in R through PCA and hierarchical clustering. Differentially expressed genes were obtained through the 'DESeq' function using the designs, gene counts ~ TYPE and gene counts ~ TREATMENT to isolate the contrast of 'susceptible' vs 'resistant' and 'inoculated' vs 'control', respectively. Significance was determined based on DESeq2's adjusted p-value, a modified Benjamini-Hochberg false discovery rate, of less than 0.05 and surpassing a log-fold change cutoff of ±1.

To isolate the 42 and 66 HPI inoculated specific DEG in the contrast of 'susceptible' vs 'resistant', DEG were removed from either 42 or 66 HPI if that same gene was differentially expressed either at 0 HPI or within each time point's control treatment. Concomitantly, the contrast of treatments, 'inoculated' vs 'control', results in the identification of type specific and not type specific DEG by first removing genes that were also differentially expressed at 0 HPI and then grouping the remaining

genes into resistant, susceptible, or not type specific DEG for 42 HPI and 66 HPI separately. For the purposes of clarity, ‘not type specific DEG’ refers to differentially expressed genes that were detected in both the ‘resistant’ and ‘susceptible’ genotypes. The resulting gene lists for both contrast groups were divided based on the direction of their LFC. Each contrast group was subjected to GO analysis in agriGO v2.0 (Tian et al., 2017) using a custom background. As the available *Salix* background on agriGO is based on the *S. purpurea* v1.0 reference genome rather than the current v5.1, a customized reference was created that utilized the *Arabidopsis* homologs included in the v5.1 reference annotation file to translate the *Salix* gene ids into *Arabidopsis* gene ids. Significant terms were determined using an FDR of 0.05.

2.3.5 Network Analysis of S. purpurea *Transcripts*

Network analysis is used to identify groups of genes that co-express and are often involved in similar biological processes (van Dam et al., 2017). To focus on the transcriptome-scale differences in response to infection, network analysis was only performed on the inoculated treatment. Samples from all time points from the inoculated treatment were then divided based on type, susceptible or resistant. After counts were filtered and normalized in DESeq2 and outlier samples were identified using PCA and hierarchical clustering and removed, network analysis was performed using a weighted gene co-expression network analysis (WGCNA) in the R package WGCNA (Langfelder & Horvath, 2008). The function ‘blockwiseModules’ was used with the following parameters for both networks; ‘power’ = 12, ‘networkType’ = ‘signed’, ‘minmodsize’ = 20, ‘deepsplit’ = 3, and ‘mergecutheight’ = 0.25. Each

module was analyzed for enriched GO terms using agriGO v2.0 (Tian et al., 2017) as described above in the differential expression of *S. purpurea* transcripts section.

A hypergeometric test using the susceptible network modules as the background was used to compare gene placement across the two networks using a p-value of 0.05. Modules found to be enriched for defense related terms or showed a significant relationship with time point were targeted for hub gene analysis. Selected modules were loaded into Cytoscape (Shannon et al., 2003) and analyzed using the plug-in cytoHubba (Chin et al., 2014). Module hub genes were identified based on the overlap of greater than 0.8 module membership and greater than 1.5 standard deviations above the mean of log transformed maximum clique centrality (MCC) from cytoHubba.

2.3.6 eQTL Mapping of S. purpurea *Transcripts*

Similar methods to Carlson et al., 2019 (Carlson et al., 2019) were used to identify SNPs within the *S. purpurea* F_2 population for eQTL analysis. Briefly, the TASSEL v5 GBS Discovery Pipeline was used on the full 485 individual population for the initial variant discovery and quality filtering (Bradbury et al., 2007). Reads were aligned to a modified *S. purpurea 94006* v5.1 reference genome (Zhou et al., 2020), [DOE-JGI, http://phytozome.jgi.doe.gov/] with the 15Z chromosome removed using the Burrows-Wheeler algorithm (BWA) (Li & Durbin, 2009). The resulting 191,650 SNPs were filtered for minor allele frequency greater than 0.01 and 80% missing tolerance before input into LinkImputeR (Money et al., 2017). Setting SNP calls with a depth less than 5 to missing, LinkImputeR's estimated imputation

accuracy of 97.5% was selected, resulting in 47,221 imputed SNPs. Deriving consensus genotypes from multiple sequencing runs of the parents enabled classification of marker types as female or male backcross and intercross markers. Expected segregation ratios based on marker types were tested using a Chi-square test and a Bonferroni corrected p-value of 1.7e-6 resulted in 22,570 SNP markers. The 56 selected F_2 individuals were then isolated and filtered for minor allele frequency > 0.05, with a final marker count of 22,068. Final markers were coded as 0, 1, 2 based on the occurrence of the minor allele.

Sample expression count data were divided into six groups based on the sample's time point and treatment (0 HPI-INOC, 0 HPI-CTRL, 42 HPI-INOC, 42 HPI-CTRL, 66 HPI-INOC, and 66 HPI-CTRL). Genes with raw counts <10 across all samples were removed from the analysis then each group was normalized separately in DESeq2 (Love et al., 2014) using the 'estimateSizeFactors' function and log transformed to account for outlier counts. eQTL detection was performed in MatrixEQTL (Shabalin, 2012) with the cut-off being *cis*- and *trans*- acting eQTL set at 1 Mb, 'useModel' set to modelANOVA with no covariates. eQTL significance for both cis and trans eQTL was determine based on a false discovery rate of 0.05 for *cis*- and 0.1 for *trans*- as calculated by MatrixEQTL. The 42 HPI and 66 HPI inoculated specific eQTL were isolated by comparing the lists of significant eQTL, removing those eQTL from 42 HPI and 66 HPI that were present during 0 HPI and those that were detected at 42 HPI or 66 HPI but in the control treatment. eQTL hotspots were determined based on a 1000 iteration permutation analysis where the number of eQTL per gene was fixed and place randomly among the SNPs without replacement (West et

al., 2007). The maximum number of eQTL occurring on a single SNP by chance was saved from each iteration to form a distribution. The distributions for both 42 and 66 HPI showed that 95% of the maximum eQTL per SNP occurring by chance are less than a threshold of 14 eQTL. To better describe the composition of the genes associating with each hotspot, GO analysis was performed using agriGO v2.0 (Tian et al., 2017) as described above in the differential expression of *S. purpurea* transcripts section.

2.3.7 Differential Expression Analysis of M. americana *Transcripts*

RNA extractions, sequencing, and data analysis was performed as described above with the following deviations. Trimmed 3' RNA-seq reads of the inoculated treatment from Rep 1 and Rep 2 were aligned to the *M. americana* reference genome R15-033-03 v1.0 (https://mycocosm.jgi.doe.gov/Melame1/Melame1.home.html) using the STAR aligner V2.7.5a (Dobin et al., 2012). A simple contrast was performed for each timepoint by combining RNASeq reads from both replicates of all susceptible genotypes and contrasting that with the combined RNASeq reads from both replicates of all resistant genotypes. *In silico* effector prediction was determined by generation of a predicted secretome using SignalP V5.0 using default settings (Zhang & Henzel, 2004). The resulting secretome was analyzed using EffectorP V2.0 (Sperschneider et al., 2018) for fungal effector prediction, run with default settings. Resulting transcripts were cross referenced to differential expression data.

2.4 Results

2.4.1 Preliminary Study of Differential Expression

We conducted a preliminary RNA-seq study by inoculating *M. americana* on reference *Salix* genotypes to determine the optimum time post-inoculation to observe differential expression. We inoculated *S. purpurea* hosts 'Fish Creek' and 94006 with uredospores of *M. americana* isolate R15-033-03 and then extracted RNA at 0, 18, 42, 66, 90, and 114 HPI from inoculated leaves and un-inoculated control leaves. These two host genotypes were selected because 94006 is the maternal grandparent and resistant while 'Fish Creek' is the paternal parent of the F_2 mapping population and susceptible to willow leaf rust. A direct contrast between the inoculated and control treatment for each genotype-by-time was performed to generate a total number of differentially expressed genes (DEGs) up-regulated and down-regulated for each host genotype.

The total number of DEGs ($p \leq 0.05$) for 'Fish Creek' were 0 (0 HPI), 0 (24 HPI), 5,589 (48 HPI), 562 (72 HPI), 1637 (96 HPI), and 3061 (120 HPI) (Appendix Fig. A.1), whereas DEGs for parent 94006 was 0, 0, 3796, 948, 597, and 1,293 for each ascending time-point, respectively. Neither parent displayed signs nor symptoms of infection during the experiment; however, signs of rust were visibly detected at 210 HPI. While uredospore sporulation appeared greater on 'Fish Creek' by 258 HPI, both genotypes were susceptible to the pathogen. The greatest number of DEGs was observed in both genotypes around 48 HPI. Thus, time-points 42 and 66 HPI were selected for the full experiment to capture the maximum host and pathogen response after inoculation.

2.4.2 Greenhouse Inoculation of Selected Resistant and Susceptible F_2 Genotypes

Based on field ratings of rust severity conducted in 2015 and 2017 in a replicated trial of an F_2 *S. purpurea* QTL mapping population (Carlson et al., 2019), 28 resistant and 28 susceptible genotypes were selected for controlled inoculation and 3' RNA-seq. At 42 and 66 HPI, leaf discs were collected from six leaves of the 56 F_2 genotypes, the two parents, and two grandparents of the population, as well as from uninoculated control plants. This experiment was conducted twice in separate greenhouses. Leaf rust severity was assessed in the inoculated treatment at 9 DPI as total percent leaf area coverage of uredospore pustules. The greenhouse ratings were moderately correlated with the 2015 and 2017 field ratings, with Pearson's correlation values of 0.48 (p-value = 9.4×10^{-5}) and 0.53 (p-value = 1.6×10^{-5}), respectively. The susceptible genotypes had a significantly greater mean rust severity (44.8% - CV: 17%) than the resistant genotypes (28.1% - CV: 54%) based on a t-test (CI = 95%) despite considerably more variability among the resistant genotypes (Fig. 2.1).

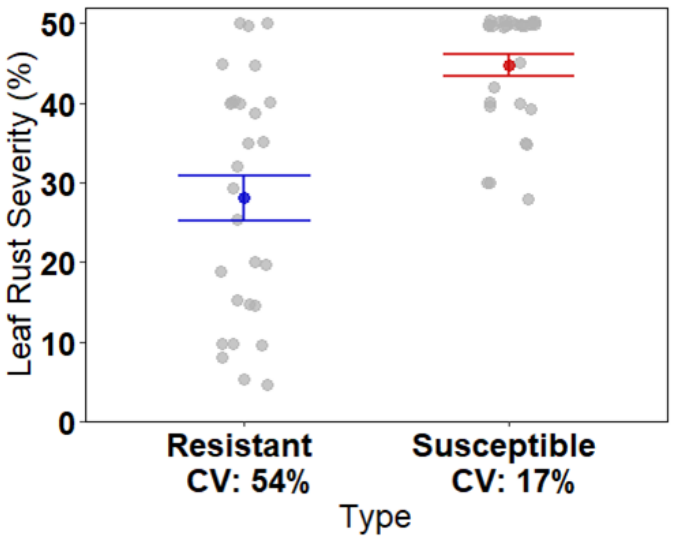


Figure 2.1: Greenhouse leaf rust severity (%) collected 9 DPI for the resistant and susceptible groups of willow genotypes. Each grey point represents an individual genotype severity while the blue and red points are the mean severity for the resistant and susceptible groups, respectively. Error bars are the standard error of the mean. (CV – Coefficient of Variation)

2.4.3 Differential Expression Analysis of S. purpurea *Transcripts*

Two separate contrasts in DESeq2 were used to identify differentially expressed genes in this study. In the direct contrast between inoculated susceptible and resistant groups, there were 19 and 105 differentially expressed genes at time points 42 HPI and 66 HPI, respectively (Fig. 2.2A). Of the 19 DEGs at 42 HPI, six were up-regulated in the resistant genotypes, including a polyubiquitin protein (*UBQ10*), a plasma membrane intrinsic protein (*PIP2;8*), a phosphoglycerate kinase 1 (*PGK1*), a chaperone DnaJ-domain superfamily protein, and two genes of unknown function (DUF). The remaining 13 differentially expressed genes at 42 HPI were up-regulated in the susceptible genotypes and included several genes associated with the flavanone synthesis pathway. The 105 DEGs at 66 HPI consisted of 35 genes up-regulated in the resistant group, while the remaining 70 were up-regulated in the susceptible group. Genes up-regulated at 66 HPI in the resistant group include several involved in defense response such as: wall-associated kinase 2 (*WAK2*), WRKY DNA-binding protein 51, CAP superfamily protein, cytochrome P450, and chitinase A, but as a group, were not significantly enriched for any GO terms. Gene enrichment of the up-regulated susceptible genes were responsive to heat, stress, and reactive oxygen species.

The contrast of inoculated treatments versus uninoculated controls highlighted the response to infection. By performing separate paired analyses for both the resistant and susceptible groups, then intersecting DEGs, variable responses to inoculation were identified at each time-point. We classified DEGs as susceptible-specific, resistant-

specific, and not-type-specific (common response between the resistant and susceptible groups). At both time points, the largest group of DEGs was the not-type-specific, positive log_2-fold change (LFC) group, with 990 and 1862 genes at 42 HPI and 66 HPI respectively (Fig. 2.2B). All groups of DEGs that were up-regulated after inoculation were enriched for defense response at 42 HPI. However, only the resistant-specific and not-type-specific groups retained enrichment of upregulated defense response genes at 66 HPI. At 66 HPI, the susceptible-specific group lacked genes associated with defense response, but instead displayed upregulation of heat response genes. The resistant-specific and the susceptible-specific groups that were down-regulated at 42 HPI were both enriched for chloroplast components, with the susceptible-specific category also enriched for down-regulated 'response to heat' genes. There was no significant GO term enrichment at 66 HPI for genes down-regulated in the susceptible-specific category, while both the down-regulated resistant-specific and not-type-specific categories were enriched for genes associated with photosynthesis.

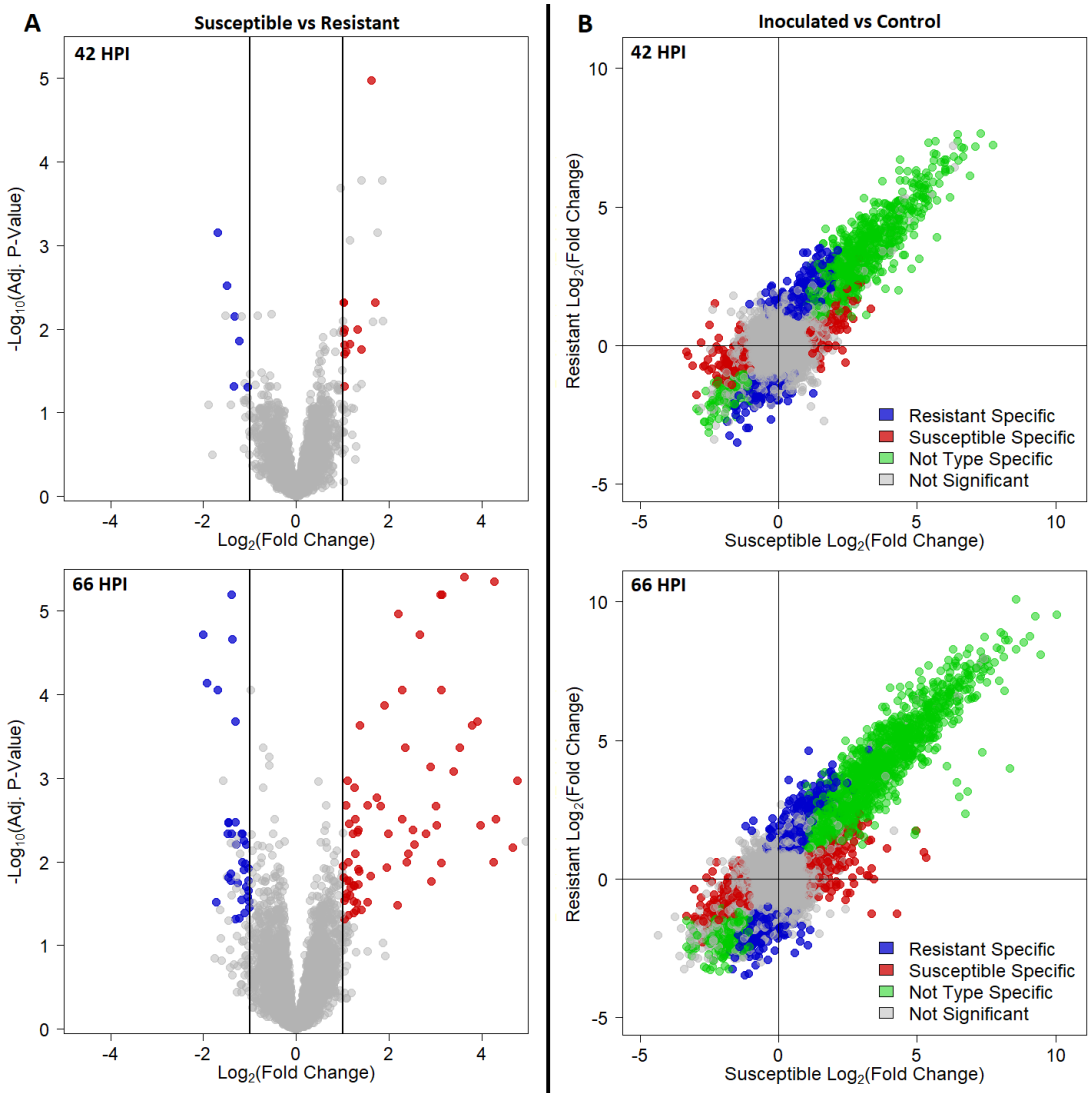


Figure 2.2: (A) Volcano plots depicting differential expression analysis between inoculated resistant and susceptible groups. Each point represents a gene. Positive LFC indicate upregulation in the susceptible genotypes (red points) while negative LFC are up-regulated in the resistant genotypes (blue points). (B) Differential expression in inoculated treatments compared with controls plotted as the LFC in expression of the susceptible genotypes on the x axis versus the LFC in expression of the resistant genotypes on the y axis.

2.4.4 Network Analysis of S. purpurea *Transcripts*

A comparison between transcriptome-wide expression in the inoculated resistant and susceptible groups was performed in WGCNA, which defined co-

expression modules based on correlated gene expression. Each module was randomly assigned a color name by the R package and is only relevant in distinguishing modules within networks, not in making comparisons between them. In this study, modules are referred to either as 'R-module' or 'S-module' to distinguish between those associated with resistant (R-) or susceptible (S-) plant networks. After removal of outlier samples and genes with low counts, the resistant network retained 75 samples and 16,410 genes, while the susceptible network retained 73 samples and 16,427 genes.

Of the 16,410 genes expressed in the resistant network, 10,176 genes were assigned to 14 modules, while the other 6,234 genes were assigned to the 'grey' module (unassigned genes). Modules sizes ranged from 33 to 5,085 genes, of which nine modules were correlated with time-point (Appendix Fig. A.2). The largest module 'R - turquoise' (n = 5,085) was positively correlated with time point ($r = 0.92$) and was the only module enriched for defense-related GO terms in the resistant network. The 'R-blue' module (n = 1,853) was negatively correlated with time point ($r = –0.89$) and enriched for photosynthesis-related GO terms. A total of 10,977 genes in the susceptible network were placed into 15 modules, with the remaining 5,450 placed within the 'grey' module. Co-expression modules ranged in size from 25 to 4,661 genes, of which 12 were correlated with time point (Appendix Fig. A.2).

A hypergeometric test ($p \leq 0.05$) facilitated a direct comparison between the resistant and susceptible networks to identify significant representation of the susceptible network modules within the 'R-turquoise' and 'R-blue' resistant modules. The 'R-turquoise' and 'R-blue' modules shared significant portions of four and six modules, respectively (Fig. 2.3A). Two modules correlated with time point in the

susceptible network with significant ‘R-turquoise’ module representation were ‘S-turquoise’ (n = 4,661, *r* = 0.88) and ‘S-salmon’ (n = 89, *r* = 0.51), and were the only susceptible modules enriched for defense-related GO terms. Concomitantly, among the six susceptible modules represented within the ‘R-blue’ module and correlated with time point, only the ‘S-brown’ (n = 1,258, *r* = –0.83) and ‘S-red’ (n = 264, *r* = –0.57) modules were enriched for photosynthetic genes.

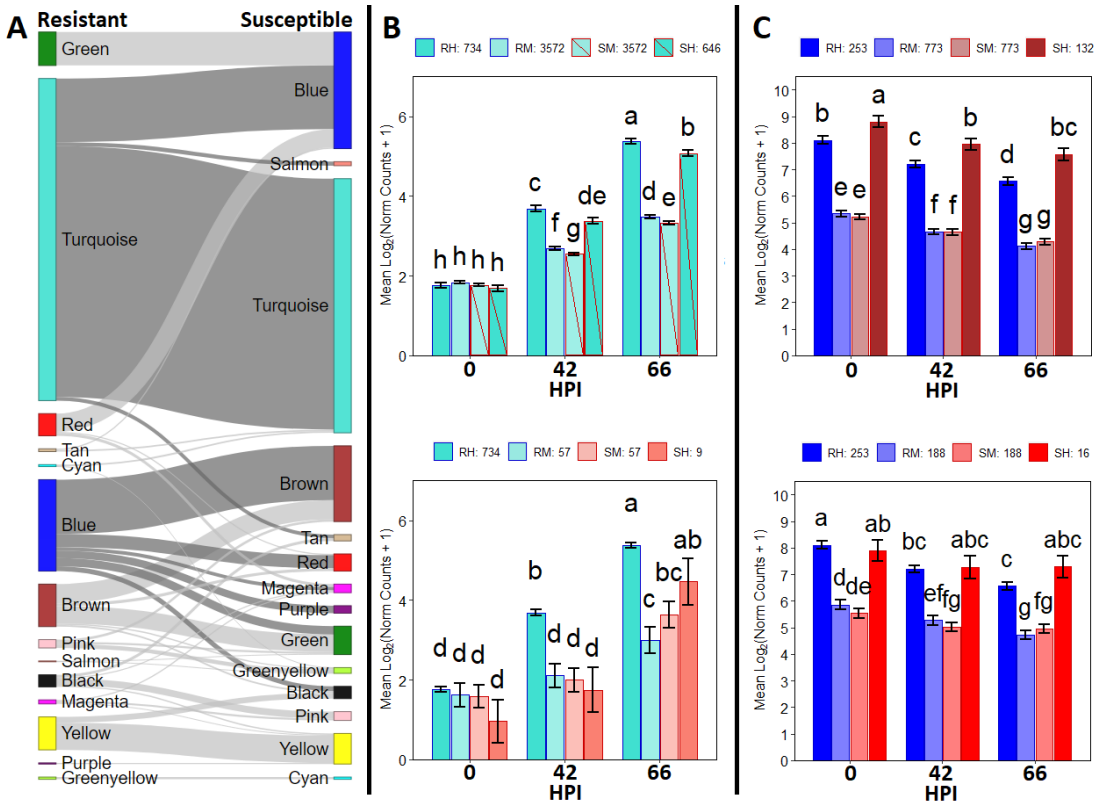


Figure 2.3. Comparison between the gene expression networks in inoculated resistant and susceptible groups of willow genotypes. A) Sankey plot of the modules from the resistant network on the left and the susceptible network on the right. Colors represent modules of co-expressed genes. Each connection is significant at the 0.05 level. B and C) RH – Resistant Hub; RM – Resistant Module; SM – Susceptible Module; SH – Susceptible Hub. Numbers of genes in each group are indicated in the legends above each graph. B) Mean expression, standard errors, and Fisher’s LSD group for the ‘R-turquoise’ module compared to the ‘S-turquoise’ and ‘S-salmon’ modules. (C) Mean expression, standard errors, and Fisher’s LSD group for the ‘R-blue’ modules compared to the ‘S-brown’ and ‘S-red’ modules.

To gain insight in the role of hub genes in module composition, hub gene analysis was performed on the ‘R-turquoise’ and ‘R-blue’ modules, in addition to the ‘S-turquoise’, ‘S-salmon’, ‘S-brown’, and ‘S-red’ modules from the susceptible network (Chin et al., 2014; Shannon et al., 2003). Significant differences in mean expression of each module's hub genes and genes commonly co-expressed across networks were determined using Fisher’s least significant difference ($p < 0.05$). The ‘R-turquoise’ and ‘S-turquoise’ modules had 3,572 genes in common, yet at 42 HPI and 66 HPI the mean expression of these genes was greater among resistant genotypes (Fig. 2.3B). This trend persisted at 42 and 66 HPI among their respective hub genes, whose expression exceeded that of the shared genes. There were only 57 genes shared between the ‘R-turquoise’ and ‘S-salmon’ modules and were not differentially expressed throughout the experiment. However, the expression of ‘S-salmon’ hub genes did not significantly increase until 66 HPI, while ‘R-turquoise’ hub gene expression increased over time.

The ‘R-blue’ module from the resistant network was enriched for photosynthesis-related GO terms and shared commonly co-expressed genes with the ‘S-brown’ and ‘S-red’ modules from the susceptible network, similarly, enriched for photosynthesis (Fig. 2.3C). There were 773 shared genes in ‘R-blue’ and ‘S-brown’ modules with similar patterns of decreased expression over time, however, the mean expression of corresponding ‘R-blue’ hub genes was lower at each time point. The genes commonly co-expressed in ‘R-blue’ and ‘S-red’ only accounted for 188 genes that gradually decreased expression through time. Their hub genes, however, show

that while the ‘R-blue’ genes decreased after 0 HPI and were beginning to level off by 42 HPI, the ‘S-red’ genes held similar expression throughout.

2.4.5 eQTL Analysis of S. purpurea *Transcripts*

Mapping of eQTL was performed using 22,068 SNPs and 16,270 genes to interrogate eQTL associated with the response to inoculation, removing those that were detected either at T0 or within the control treatment at the same time point. A total of 38,480 *cis* and 9,460 *trans* eQTL were identified at 42 HPI, 45,148 *cis* and 10,638 *trans* eQTL at 66 HPI, and 13,860 *cis* and 1,839 *trans* eQTL at both time points (Fig. 2.4A). Any SNP with more than 14 eQTL, the 95% confidence threshold

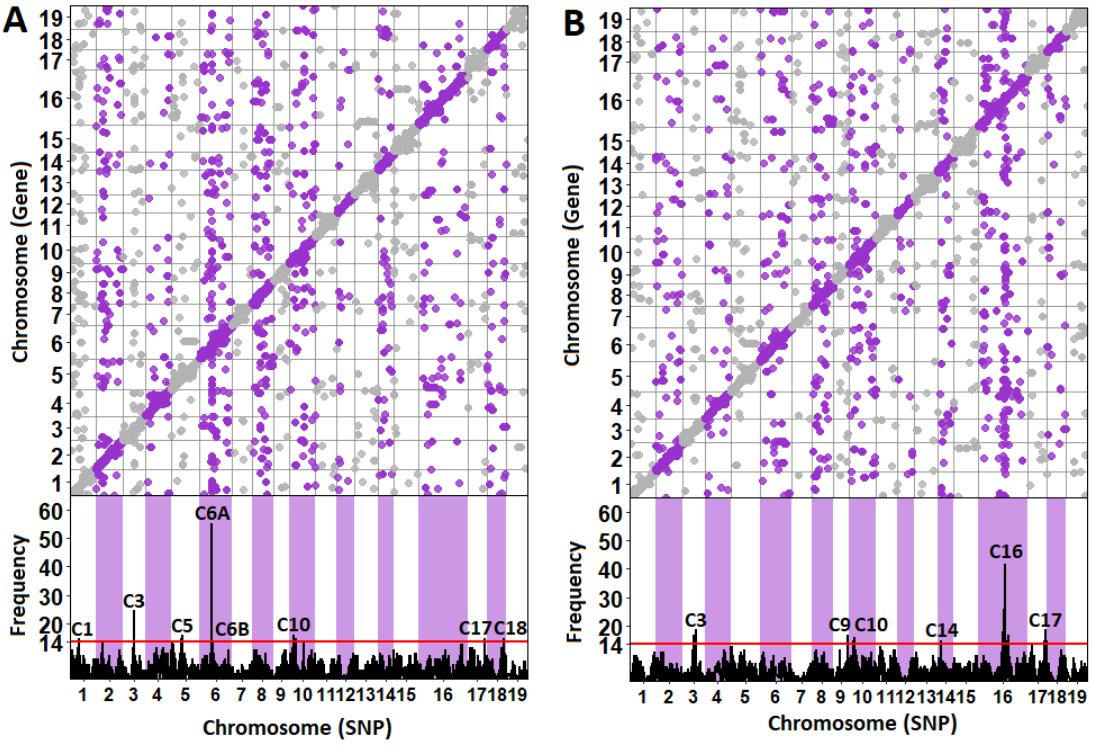


Figure 2.4. eQTL mapping by time points 42 HPI (A) and 66 HPI (B). SNPs are sorted by chromosome across the x-axis. The y-axis of the top panels represents genes mapped to chromosomes. The y-axis of the bottom panels indicates eQTL frequency. The red line indicates the threshold for hotspots set at 14 eQTL.

identified through permutation, was identified as a *trans* eQTL hotspot. A hotspot is considered to be a locus influencing the regulation of multiple genes related to allelic genotype. Simple correlation analysis ($p < 0.05$) condensed the significant eSNPs into eight eQTL hotspots at 42 HPI and six at 66 HPI (Fig. 2.4B). Hotspot sizes ranged from 14 to 55 eQTL associations and only three hotspots were enriched for any GO terms. The chr03 hotspot at 42 HPI (C3) was enriched for cell communication and signaling while the chr06 hotspot at 42 HPI (C6A) was enriched for chloroplast components. The only hotspot at 66 HPI showing GO enrichment was located on chr16 for photosynthesis and chloroplast components.

2.4.6 Candidate Genes for S. purpurea *Resistance to* M. americana

Candidate genes which potentially determine a compatible interaction (successful infection) between *S. purpurea* and *M. americana* were identified using the intersection of network analysis, differential expression, and eQTL mapping. Candidate genes were defined as the hub genes of modules found to be enriched for plant defense-related terms and differentially expressed either between resistant and susceptible genotypes or between the inoculated and control treatments. While associations with an eQTL hotspot for response to inoculation were not required for identification as candidate genes, it does aid in prioritization for further research. We identified candidate genes associated with the defense response enriched 'R-turquoise' module at 42 HPI (n = 31) and 66 HPI (n = 69) (Appendix Table A.1), of which 18 and 20 genes were correlated with leaf rust severity, respectively. Hub genes from the 'R-blue' module were associated with a reduction in photosynthesis through GO

enrichment analysis. From these hub genes only 3 (42 HPI) and 21 (66 HPI) met our criteria for candidate gene selection, with all three genes at 42 HPI and one gene at 66 HPI having a significant correlation with leaf rust severity (Appendix Table A.2).

2.4.7 Differential Expression Analysis of M. americana *Transcripts*

Total raw reads of the inoculated treatments for each of the 60 willow genotypes (two replicates) were aligned to the *M. americana* reference genome R15-033-03 v1.0 (Crowell et al. 2021 - submitted). A direct contrast between genotypes previously identified as resistant and susceptible was performed at each time point (42 HPI and 66 HPI). A total of 22 *M. americana* genes were differentially expressed (FDR = 0.1) between the resistant and susceptible willow genotypes at 42 HPI, yet none at 66 HPI (Fig. 2.5). The majority of differentially expressed genes were up-regulated in the resistant group (20 genes) as compared to the susceptible (2 genes). A BLAST search of these 22 DEGs was queried against the NCBI nt database (Altschul et al., 1990). One transcript sequence (CDS_5062) was homologous to a known effector ubiquitin carboxyl extension protein in the plant parasitic nematode *Globodera rostochiensis* (Chronis et al., 2013).

The *in-silico* proteome of the *M. americana* reference genome was analyzed using SignalPv5.0 using default settings to generate an *in silico* secretome, which resulted in 1,779 predicted secreted proteins and analyzed for effector prediction using EffectorPv2.0 using the default settings. These proteins were then cross-referenced to the list of differentially expressed fungal transcripts between resistant and susceptible host groups. One (CDS_12834) of the 22 transcripts differentially

expressed between the resistant and susceptible hosts was identified as a potential effector.

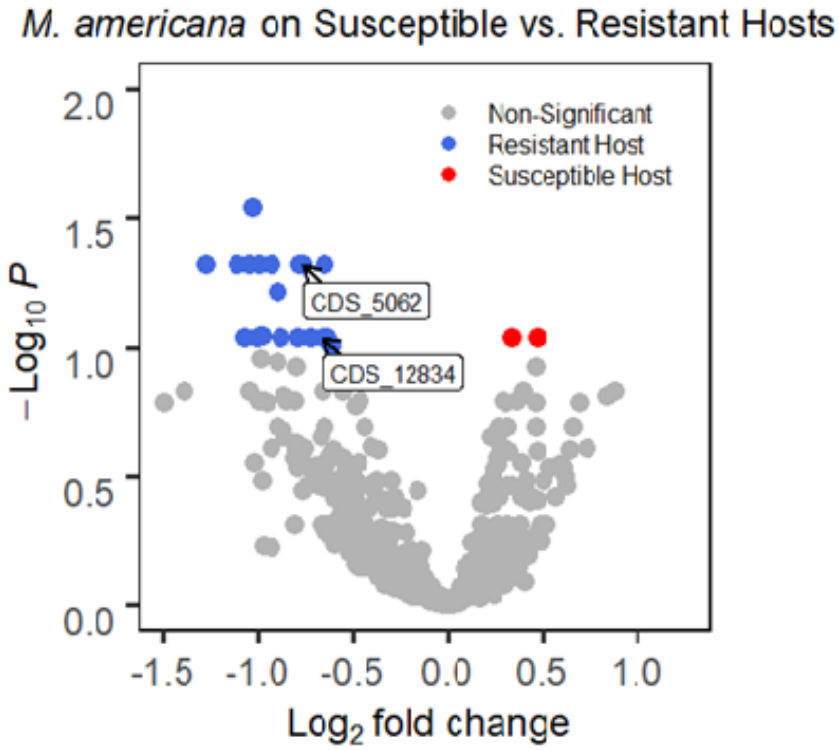


Figure 2.5. Volcano plot of differentially expressed transcripts of willow rust pathogen *M. americana* at 42 HPI. LFC indicate up-regulation of *M. americana* genes when grown on the susceptible genotypes (red) while negative LFC indicate up-regulation of *M. americana* genes when grown the resistant genotypes (blue). Transcripts identified by arrows were predicted to play a role in fungal infection based on *in silico* effector prediction software Effector P 2.0 (CDS_12834) or through sequence homology to known effectors (CDS_5062). A modified Benjamini-Hochberg adjusted p-value cutoff of <0.1 with no log fold change cutoff was used to determine significance. [Created by Chase Crowell]

2.5 Discussion

Melampsora americana has previously been shown to be the dominant rust species infecting *S. purpurea* willow in the northeast United States, yet little is known about the mechanisms of pathogen virulence or host resistance (Crowell et al., 2020;

Kenaley et al., 2014). By applying transcriptomics on willow leaf rust and the *S. purpurea* host over a time series post-inoculation, we defined gene networks associated with a reduction in photosynthesis and an increase in defense response, while simultaneously revealing candidate effectors for pathogenicity. By leveraging network analysis, differential expression, and eQTL mapping of the host transcriptome we identified 124 candidate genes associated with a compatible interaction between *M. americana* and *S. purpurea* for future functional characterization.

2.5.1 Willow Transcriptomics

Through the combined use of differential expression, network analysis, and eQTL mapping, this study demonstrated that layering the strengths of each highlights the early response of *S. purpurea* to inoculation by *M. americana* and the varied response between resistant and susceptible genotypes. The contrast between the resistant and susceptible host genotypes produced a moderate number of DEGs. This could be, in part, attributed to the level of resistance observed in the greenhouse compared to the field. Genotypes were selected for this experiment based on disease ratings over two field seasons, but disease development in the greenhouse environment was rapid and severe. Rust severity at 9 DPI in susceptible genotypes ranged from 25-50% diseased leaf area at while resistant genotypes ranged from 0-50%. There were greater numbers of genes up-regulated in the susceptible genotypes, many of which were heat shock proteins. Heat shock proteins have been implicated as molecular chaperones that target misfolded proteins for proteolysis and are thought to prevent cell death (Park & Seo, 2015). The not-type-specific groups in inoculated vs

control treatment contrasts show many genes in both the resistant and susceptible genotypes that responded similarly to inoculation. Network and differential expression analyses revealed that a defense response was triggered in both susceptible and resistant genotypes by the number of shared genes within the 'R-turquoise' and 'S-turquoise' modules and up-regulated genes in the not-type-specific group, all enriched for defense response. While the differential expression of defense response genes was common, the fold-change in expression of hub genes coordinating the resistant response was greater than those of the susceptible group (Fig. 2.3). By comparing the networks from the inoculated resistant and susceptible genotypes, changes in gene coordination were found that would otherwise be difficult to resolve through a direct contrast, given the sample size. Network hub genes are often found to have regulatory control over the other genes in the module, suggesting that small changes in their expression will cascade and resolve in larger changes downstream. This was not unexpected, because prior research suggested that control of leaf rust severity in the *S. purpurea* F_2 population used in this study was multi-genic and quantitative in nature (Carlson et al., 2019) and would translate into many genes at lower fold-changes that could be difficult to detect.

The inoculated versus control contrasts revealed that host genes associated with photosynthesis were down-regulated in the resistant group faster than in the susceptible group. A reduction in photosynthesis has been shown in other systems to be an initial response to pathogen attack by redirecting resources toward defense response (Lu & Yao, 2018). Here, co-expression modules enriched for photosynthesis and related terms were negatively correlated with time post inoculation. While the 'R-

blue' module was the only one in the resistant network enriched for photosynthesis-related genes, it was split into six separate modules within the susceptible genotypes. In combination with the differential expression results, hub genes of 'R-blue' were better able to coordinate resources away from photosynthesis and toward defense response. A faster, more coordinated response in the resistant interaction has similarly been found in the interaction between *Populus* and *P. larici-populina* (Hacquard et al., 2011).

Three of the 14 eQTL hotspots detected in this study were enriched for chloroplast components and photosynthesis (C6A at 42 HPI and C16 at 66 HPI) and communication and signaling (C3 at 42 HPI). Although not significantly enriched for GO terms, several defense response genes were associated with all eQTL hotspots. It is likely that the results were influenced by sample size, as the power of the eQTL analysis was limited. Despite that, many differentially expressed genes and co-expression module hub genes were connected to an eQTL hotspot, either by direct association or genomic proximity. Based on the intersection of all three analyses, 124 genes predicted to be associated with promoting the defense response and aiding in the coordination of photosynthesis that should be targeted for future studies.

2.5.2 Rust Transcriptomics

As *M. americana* is an obligate biotroph, *in silico* techniques can narrow down candidate effector genes that are most likely to modulate host immunity. Effector prediction has been a successful initial strategy in the poplar rust pathogen *M. larici-populina* (Hacquard et al., 2012; Hacquard et al., 2011; Lorrain et al., 2015; Petre et

al., 2012) and has led to functional assays that further validate candidate effector function (Petre et al., 2016; Petre et al., 2015). After *in silico* effector prediction, Petre et al., 2016 was able to utilize live cell imaging by laser-scanning confocal microscopy in combination with florescent tagged candidate effector chloroplast-targeted protein 1 (CTP1) in *Nicotiana benthamiana* to track cellular localization of the translocated protein. To begin the process of effector discovery and validation of effectors in *M. americana*, we identified two candidate fungal effectors that were differentially expressed between resistant and susceptible hosts. These candidates were discovered based on direct homology to a known effector in nematode (CDS_5062) by using an effector prediction software (CDS_12834). Both transcripts were identified when grown on resistant hosts, possibly indicating that corresponding R-genes exist in the susceptible pool that recognize these transcripts. It was surprising that CDS_5062 showed strong homology to a ubiquitin carboxyl extension effector protein in the nematode *Globadera rostochiensis*, which may be evidence of convergent evolution. In this nematode, it was shown by Chronsis et al., 2013 (Chronis et al., 2013) that the peptide is cleaved into a ubiquitin subunit involved in suppression of immunity and a carboxyl extension subunit involved in promoting feeding cell formation. Perhaps the translated CDS_5062 transcripts function similarly, utilizing free ubiquitin as an immunity suppressor. Future proteomic studies will determine if the protein product is similarly cleaved, and functional studies may reveal what role it plays in parasitism.

Both identified candidate effector sequences show promise for future studies, however the overall number of differentially expressed pathogen transcripts identified between the resistant and susceptible groups was quite small. It is possible that this

accurately reflects a small number of differentially expressed transcripts and that most of the identified differentially expressed genes play an unknown role in infection. It is also possible that we lacked the proper statistical power to capture the true number of differentially expressed fungal genes, and since less than 0.5% of transcripts aligned to the fungal genome, we likely only captured those with the greatest abundance. This could be due to the overrepresentation of willow RNA extracted from the leaf punch samples in the greenhouse experiment resulting in a low number of total genes aligning to the *M. americana* reference genome. This overrepresentation may be due to a deficit of fungal infection structures at these early infection time points, a phenomenon observed in similar rust pathosystems targeting early infection (Petre et al., 2012; Stergiopoulos et al., 2012; Wang et al., 2017). Additionally, it is possible that the plant RNA extraction kit we used may not have been optimal for extracting fungal transcripts. Future studies utilizing highly-sensitive RNA extraction strategies like laser capture microdissection or haustoria extraction coupled with fungal specific RNA extraction chemistry may achieve greater sensitivity for differential expression studies of *M. americana*. Regardless, *in silico* prediction of rust effectors remains a challenging task. There are a few species-specific rust effector motifs, but these have not been proven to be suitable for universal predictions across all rust species (Duplessis et al., 2011; Zhao et al., 2020). As a result, general peptide characteristics such as length, amino acid proportions, and predicted secretion are used as indicators of putative effectors (Hacquard et al., 2011; Lorrain et al., 2015; Petre et al., 2014; Sperschneider et al., 2018).

2.6 Conclusions

This study described the complex changes in the transcriptomes of both the pathogen and host in the *S. purpurea – M. americana* pathosystem using differential expression, network analysis, and eQTL mapping. Differential expression analysis of fungal RNA produced a short list of genes of interest, with two candidate effector genes that were highly expressed when grown on the resistant hosts. Analysis of host gene expression revealed 124 candidate genes that were differentially expressed co-expression module hub genes associated with an eQTL hotspot. Future research could use qRT-PCR to validate differential expression of listed candidate genes produced through this RNA-seq approach. This study represents a step toward developing true understanding of this pathosystem and unlocking the key to breeding shrub willow resistant to this devastating pathogen.

2.7 REFERENCES

Aime, M. C., Bell, C. D., & Wilson, A. W. (2018). Deconstructing the evolutionary complexity between rust fungi (Pucciniales) and their plant hosts. *Studies in Mycology, 89*, 143-152. doi:10.1016/j.simyco.2018.02.002

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*, 403-410.

Anders, S., Pyl, P. T., & Huber, W. (2014). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics, 31*(2), 166-169. doi:10.1093/bioinformatics/btu638

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. In.

Bennett, C., Aime, M. C., & Newcombe, G. (2011). Molecular and pathogenic variation within *Melampsora* on *Salix* in western North America reveals numerous cryptic species. *Mycologia, 103*, 1004-1018. doi:10.3852/10-289

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633-2635. doi:10.1093/bioinformatics/btm308

Bresson, A., Jorge, V., Dowkiw, A., Guerin, V., Bourgait, I., Tuskan, G. A., Schmutz, J., Chalhoub, B., Bastien, C., & Faivre Rampant, P. (2011). Qualitative and quantitative resistances to leaf rust finely mapped within two nucleotide-binding site leucine-rich repeat (NBS-LRR)-rich genomic regions of chromosome 19 in poplar. *New Phytol, 192*(1), 151-163. doi:10.1111/j.1469-8137.2011.03786.x

Carlson, C. H., Gouker, F. E., Crowell, C. R., Evans, L., DiFazio, S. P., Smart, C. D., & Smart, L. B. (2019). Joint linkage and association mapping of complex traits in shrub willow (*Salix purpurea* L.). *Ann. Bot., 124*(4), 701-716. doi:10.1093/aob/mcz047

Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., & Lin, C.-Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol, 8 Suppl 4*(Suppl 4), S11-S11. doi:10.1186/1752-0509-8-S4-S11

Chronis, D., Chen, S., Lu, S., Hewezi, T., Carpenter, S. C. D., Loria, R., Baum, T. J., & Wang, X. (2013). A ubiquitin carboxyl extension protein secreted from a plant-parasitic nematode *Globodera rostochiensis* is cleaved in planta to promote plant parasitism. *Plant J, 74*(2), 185-196. doi:https://doi.org/10.1111/tpj.12125

Crowell, C. R., Bekauri, M. M., Cala, A. R., Mcmullen, P., Smart, L. B., & Smart, C. D. (2020). Differential susceptibility of diverse *Salix* spp. to *Melampsora americana* and *Melampsora paradoxa*. *Plant Dis., 104*, 2949-2957. doi:10.1094/PDIS-04-20-0718-RE

Dickmann, D. I., & Kuzovkina, J. (2014). Poplars and willows of the world, with emphasis on silviculturally important species. In *Poplars and willows: Trees for society and the environment* (Vol. 22). Rome, Italy: FAO.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2012). STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Ayliffe, M. A., & Ellis, J. G. (2004). The *Melampsora lini AvrL567* avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell, 16*(3), 755. doi:10.1105/tpc.020040

Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B. L., Chiu, R., Coutinho, P. M., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M. G., Kodira, C. D., Kohler, A., Kües, U., Lindquist, E. A., Lucas, S. M., Mago, R., Mauceli, E., Morin, E., Murat, C., Pangilinan, J. L., Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Salamov, A. A., Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G. A., Henrissat, B., Van de Peer, Y., Rouzé, P., Ellis, J. G., Dodds, P. N., Schein, J. E., Zhong, S., Hamelin, R. C., Grigoriev, I. V., Szabo, L. J., & Martin, F. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *P Natl Acad Sci USA, 108*(22), 9166-9171. doi:10.1073/pnas.1019315108

Gao, L., Tu, Z. J., Millett, B. P., & Bradeen, J. M. (2013). Insights into organ-specific pathogen defense responses in plants: RNA-Seq analysis of potato tuber-

Phytophthora infestans interactions. *BMC Genomics, 14*(1), 340. doi:10.1186/1471-2164-14-340

Hacquard, S., Joly, D. L., Lin, Y.-C., Tisserant, E., Feau, N., Delaruelle, C., Legué, V., Kohler, A., Tanguay, P., Petre, B., Frey, P., Van de Peer, Y., Rouzé, P., Martin, F., Hamelin, R. C., & Duplessis, S. (2012). A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol Plant Microbe Interact, 25*(3), 279-293. doi:10.1094/mpmi-09-11-0238

Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., & Duplessis, S. (2011). The poplar-poplar rust interaction: insights from genomics and transcriptomics. *J Pathog, 2011*.

Hanley, S. J., Pei, M. H., Powers, S. J., Ruiz, C., Mallott, M. D., Barker, J. H. A., & Karp, A. (2011). Genetic mapping of rust resistance loci in biomass willow. *Tree Genet. Genomes, 7*(3), 597-608. doi:10.1007/s11295-010-0359-x

Karp, A., Hanley, S. J., Trybush, S. O., Macalpine, W., Pei, M., & Shield, I. (2011). Genetic improvement of willow for bioenergy and biofuels. *J. Integr. Plant Biol., 53*(2), 151-165. doi:10.1111/j.1744-7909.2010.01015.x

Kenaley, S. C., Smart, L. B., & Hudler, G. W. (2014). Genetic evidence for three discrete taxa of *Melampsora* (Pucciniales) affecting willows (*Salix* spp.) in New York State. *Fungal Biol, 118*(8), 704-720. doi:10.1016/j.funbio.2014.05.001

Kim, K. H., Kang, Y. J., Kim, D. H., Yoon, M. Y., Moon, J.-K., Kim, M. Y., Van, K., & Lee, S.-H. (2011). RNA-Seq analysis of a soybean near-isogenic line

carrying bacterial leaf pustule-resistant and -susceptible alleles. *DNA Res, 18*(6), 483-497. doi:10.1093/dnares/dsr033

Kuzovkina, Y. A., & Volk, T. A. (2009). The characterization of willow (*Salix* L.) varieties for use in ecological engineering applications: Co-ordination of structure, function and autecology. *Ecol. Eng., 35*(8), 1178-1189. doi:https://doi.org/10.1016/j.ecoleng.2009.03.010

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics, 9*(1), 559. doi:10.1186/1471-2105-9-559

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

Lorrain, C., Hecker, A., & Duplessis, S. (2015). Effector-mining in the poplar rust fungus *Melampsora larici-populina* secretome. *Front. Plant Sci., 6*, 1051.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol., 15*(12), 550. doi:10.1186/s13059-014-0550-8

Lu, Y., & Yao, J. (2018). Chloroplasts at the crossroad of photosynthesis, pathogen infection and plant defense. *Int. J. Mol. Sci., 19*(12). doi:10.3390/ijms19123900

Mähler, N., Schiffthaler, B., Robinson, K. M., Terebieniec, B. K., Vučak, M., Mannapperuma, C., Bailey, M. E. S., Jansson, S., Hvidsten, T. R., & Street, N.

R. (2020). Leaf shape in *Populus tremula* is a complex, omnigenic trait. *Ecol Evol, 10*(21), 11922-11940. doi:https://doi.org/10.1002/ece3.6691

Martin, T., Ronnberg-Wastljung, A. C., Stenlid, J., & Samils, B. (2016). Identification of a differentially expressed TIR-NBS-LRR gene in a major QTL associated to leaf rust resistance in *Salix*. *PLoS One, 11*(12), e0168776. doi:10.1371/journal.pone.0168776

McCracken, A. R., & Dawson, W. M. (1998). Short rotation coppice willow in Northern Ireland since 1973: development of the use of mixtures in the control of foliar rust (*Melampsora* spp.). *Eur J Forest Pathol, 28*(4), 241-250.

McCracken, A. R., & Dawson, W. M. (2003). Rust disease (*Melampsora epitea*) of willow (*Salix* spp.) grown as short rotation coppice (SRC) in inter- and intra-species mixtures. *Ann Appl Biol, 143*(3), 381-393. doi:DOI 10.1111/j.1744-7348.2003.tb00308.x

Money, D., Migicovsky, Z., Gardner, K., & Myles, S. (2017). LinkImputeR: user-guided genotype calling and imputation for non-model organisms. *BMC Genomics, 18*(1), 523. doi:10.1186/s12864-017-3873-5

Park, C.-J., & Seo, Y.-S. (2015). Heat shock proteins: A review of the molecular chaperones for plant immunity. *Plant Pathol J., 31*(4), 323-333. doi:10.5423/PPJ.RW.08.2015.0150

Pei, M. H., & McCracken, A. R. (2005). *Rust diseases of willow and poplar*. Cambridge, MA: CABI Publishing.

Pei, M. H., Royle, D. J., & Hunter, T. (1996). Pathogenic specialization in *Melampsora epitea* var. *epitea* on *Salix*. *Plant Pathol., 45*(4), 679-690. doi:DOI 10.1046/j.1365-3059.1996.d01-174.x

Petre, B., Joly, D. L., & Duplessis, S. (2014). Effector proteins of rust fungi. *Front. Plant Sci., 5*, 416. doi:10.3389/fpls.2014.00416

Petre, B., Lorrain, C., Saunders, D. G. O., Win, J., Sklenar, J., Duplessis, S., & Kamoun, S. (2016). Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. *Cell Microbiol, 18*, 453-465. doi:10.1111/cmi.12530

Petre, B., Morin, E., Tisserant, E., Hacquard, S., Da Silva, C., Poulain, J., Delaruelle, C., Martin, F., Rouhier, N., Kohler, A., & Duplessis, S. (2012). RNA-Seq of early-infected poplar leaves by the rust pathogen *Melampsora larici-populina* uncovers *PtSultr3;5*, a fungal-induced host sulfate transporter. *PLoS One, 7*(8), e44408. doi:10.1371/journal.pone.0044408

Petre, B., Saunders, D. G. O., Sklenar, J., Lorrain, C., Win, J., Duplessis, S., & Kamoun, S. (2015). Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. *Mol Plant Microbe Interact, 28*(6), 689-700. doi:10.1094/mpmi-01-15-0003-r

Rönnberg-Wästljung, A. C., Samils, B., Tsarouhas, V., & Gullberg, U. (2008). Resistance to *Melampsora larici-epitea* leaf rust in *Salix*: analyses of quantitative trait loci. *J. Appl. Genet., 49*, 321-331. doi:10.1007/BF03195630

Samils, B., Rönnberg-Wästljung, A.-C., & Stenlid, J. (2011). QTL mapping of resistance to leaf rust in *Salix*. *Tree Genet. Genomes, 7*(6), 1219-1235. doi:10.1007/s11295-011-0408-0

Serapiglia, M. J., Gouker, F. E., & Smart, L. B. (2014). Early selection of novel triploid hybrids of shrub willow with improved biomass yield relative to diploids. *BMC Plant Biol, 14*, 74. doi:10.1186/1471-2229-14-74

Shabalin, A. A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics, 28*(10), 1353-1358. doi:10.1093/bioinformatics/bts163

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res, 13*(11), 2498-2504. doi:10.1101/gr.1239303

Smart, L. B., & Cameron, K. D. (2008). Genetic improvement of willow (*Salix* spp.) as a dedicated bioenergy crop. In *Genetic improvement of bioenergy crops* (pp. 377-396): Springer.

Smart, L. B., Volk, T., Lin, J., Kopp, R. F., Phillips, I. S., Cameron, K. D., White, E. H., & Abrahamson, L. (2005). Genetic improvement of shrub willow (*Salix* spp.) crops for bioenergy and environmental applications in the United States. *Unasylva, 56*, 51-55.

Smith, J. A., Blanchette, R. A., & Newcombe, G. (2004). Molecular and morphological characterization of the willow rust fungus, *Melampsora epitea*,

from arctic and temperate hosts in North America. *Mycologia, 96*, 1330-1338. doi:10.2307/3762149

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol, 19*, 1-17. doi:10.1111/mpp.12682

Stergiopoulos, I., Kourmpetis, Y. A. I., Slot, J. C., Bakker, F. T., De Wit, P. J. G. M., & Rokas, A. (2012). In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. *Mol. Biol. Evol., 29*, 3371-3384. doi:10.1093/molbev/mss143

Sulima, P., Przyborowski, J. A., Kuszewska, A., Zaluski, D., Jedryczka, M., & Irzykowski, W. (2017). Identification of quantitative trait loci conditioning the main biomass yield components and resistance to *Melampsora* spp. in *Salix viminalis* x *Salix schwerinii* hybrids. *Int. J. Mol. Sci., 18*(3). doi:10.3390/ijms18030677

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., & Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res, 45*(W1), W122-W129. doi:10.1093/nar/gkx382

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., & de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene–disease predictions. *Brief Bioinform, 19*(4), 575-592. doi:10.1093/bib/bbw139

Verwijst, T. (1990). Clonal differences in the structure of a mixed stand of *Salix viminalis* in response to *Melampsora* and frost. *Can J For Res, 20*(5), 602-605. doi:DOI 10.1139/x90-079

Wang, N., Cao, P., Xia, W., Fang, L., & Yu, H. (2017). Identification and characterization of long non-coding RNAs in response to early infection by *Melampsora larici-populina* using genome-wide high-throughput RNA sequencing. *Tree Genet. Genomes, 13*(2). doi:10.1007/s11295-017-1116-1

Wang, W., Carlson, C. H., Smart, L. B., & Carlson, J. E. (2020). Transcriptome analysis of contrasting resistance to herbivory by *Empoasca fabae* in two shrub willow species and their hybrid progeny. *PLoS One, 15*(7), e0236586. doi:10.1371/journal.pone.0236586

West, M. A., Kim, K., Kliebenstein, D. J., van Leeuwen, H., Michelmore, R. W., Doerge, R. W., & St Clair, D. A. (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. *Genetics, 175*(3), 1441-1450. doi:10.1534/genetics.106.064972

Xu, L., Zhu, L., Tu, L., Liu, L., Yuan, D., Jin, L., Long, L., & Zhang, X. (2011). Lignin metabolism has a central role in the resistance of cotton to the wilt fungus *Verticillium dahliae* as revealed by RNA-Seq dependent transcriptional analysis and histochemistry. *J Exp Bot, 62*(15), 5607-5621. doi:10.1093/jxb/err245

Zhang, Z., & Henzel, W. J. (2004). Signal peptide prediction based on analysis of experimentally verified cleavage sites. *Protein Sci, 13*(10), 2819-2824.

Zhao, S., Shang, X., Bi, W., Yu, X., Liu, D., Kang, Z., Wang, X., & Wang, X. (2020). Genome-wide identification of effector candidates with conserved motifs from the wheat leaf rust fungus *Puccinia triticina*. *Front Microbiol, 11,* 1-15. doi:10.3389/fmicb.2020.01188

Zhou, R., Macaya-Sanz, D., Carlson, C. H., Schmutz, J., Jenkins, J. W., Kudrna, D., Sharma, A., Sandor, L., Shu, S., Barry, K., Tuskan, G. A., Ma, T., Liu, J., Olson, M., Smart, L. B., & DiFazio, S. P. (2020). A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biol., 21*(1), 38. doi:10.1186/s13059-020-1952-4

CHAPTER 3

MAPPING THE SEX DETERMINATION REGION IN EIGHT DIVERSE *SALIX* F_1 HYBRID FAMILIES

Formatted for submission as: Wilkerson, DG, Taskiran, B, Carlson, CH, Smart, LB. 2021. *G3*.

3.1 Abstract

Within the genus *Salix*, there are roughly 350 species native primarily to the northern hemisphere and adapted to a wide range of habitats. This diversity provides an opportunity to identify different alleles conferring traits important for production as a bioenergy crop, but also evolutionarily important genes, such as those regulating sex determination. I created mapping populations based on crosses with common male and female parents to leverage the genomic resources developed for *Salix purpurea*, while identifying unique alleles from related species. Eight F_1 hybrid mapping populations were created by crossing *S. viminalis*, *S. suchowensis*, *S. integra*, *S. koriyanagi*, *S. udensis*, and *S. alberti* with either *S. purpurea* 94006 (female) or 94001 (male). Each family was genotyped using genotyping-by-sequencing library preparation and sequencing. The relationship between the parents was assessed using principal component analysis, hierarchical clustering, and fastSTRUCTURE. Lastly, linkage map construction using female and male backcross markers and QTL detection were conducted in R using the packages R/qtl and ASMap. These analyses resolved the parents and F_1 progeny consistent with their phylogenetic section, while

fastSTRUCTURE results indicated that the *S. alberti* parent was misidentified and was most like *S. suchowensis*. Sixteen linkage maps with 19 linkage groups each were constructed. A QTL for sex in each family was identified solely in the female map, suggesting that the female is the heterogametic sex in a ZW sex determination system. This study provides an avenue to identify beneficial alleles that can be quickly introgressed into elite bioenergy cultivars.

3.2 Introduction

The establishment of genomic resources is an important step in developing a fully realized breeding program, reinforced by modern tools for trait mapping, candidate gene identification, and marker informed selection. *Salix*, along with the genus *Populus*, make up the majority of species in the Salicaceae family encompassing trees, shrubs and subshrubs that are dioecious and highly heterozygous. Shrub willow is grown in northern latitudes as a high yielding, carbon neutral, bioenergy crop that can grow on marginal land and provide multiple ecosystem services (Clifton-Brown et al., 2019; Fabio & Smart, 2020; Smart et al., 2005; Stoof et al., 2014). While shrub willow breeding has been underway in the United States since the 1980's, the genus *Salix* includes more than 350 species (Dickmann & Kuzovkina, 2014; Stanton et al., 2014) many of which have not been tapped as sources of diverse alleles. Due to the high degree of synteny between poplar and shrub willow (Berlin et al., 2010; Hanley et al., 2006a), genomic resources developed for poplar were used in early genomic studies in *Salix*.

Genomic resources for *Salix* are currently centered around a few key species. In Europe, *Salix viminalis* is an important bioenergy crop with a recently published, high quality genome assembly (Almeida et al., 2020). *Salix viminalis* has been used in several QTL mapping studies for resistance to *Melampsora larici-epitea* (Rönnberg-Wästljung et al., 2008; Samils et al., 2011; Sulima et al., 2017), drought tolerance (Rönnberg-Wästljung et al., 2005), and growth and phenology (Hallingbäck et al., 2019; Hallingbäck et al., 2016). While in the United States, *S. purpurea* is a model species for willow bioenergy crop breeding, genetics, and genomics. The US Department of Energy Joint Genome Institute has produced the highest quality annotated *Salix* reference genome assemblies of a male and a female *S. purpurea* available on Phytozome (Zhou et al., 2020; Zhou et al., 2018) (https://phytozome-next.jgi.doe.gov/info/Spurpurea_v5_1; https://phytozome-next.jgi.doe.gov/info/SpurpureaFishCreek_v3_1). Carlson et al. (2019) used both linkage mapping of a *S. purpurea* F_2 population and an association panel of naturalized *S. purpurea* to map a wide range of morphological, physiological, insect and disease resistance and biomass composition traits. There is value in studying the genomes of lesser researched *Salix* species for phylogenomic analysis and to discover diverse sources of alleles for introgression into elite yielding cultivars.

Here, I introduce the *Salix* F_1 hybrid common parent mapping population. The parents, described by Fabio et al. (2019) and Crowell et al. (2020), represent a diverse subset of species from the subgenus Vetrix (Dickmann & Kuzovkina, 2014). Female *S. viminalis*, *S. integra*, *S. alberti*, and *S. suchowensis* were crossed with male *S. purpurea* 94001, while male *S. viminalis*, *S. udensis*, *S. koriyanagi*, and *S.*

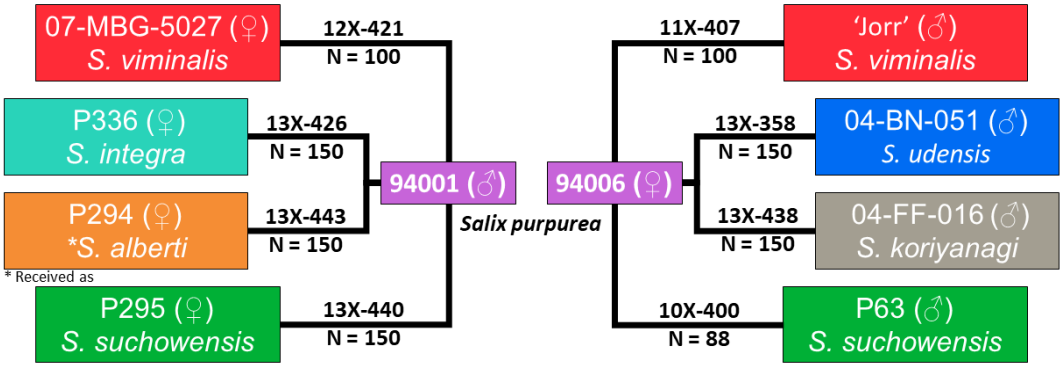


Figure 3.1: Pedigrees of *Salix* F_1 hybrid common parent mapping populations. There were four full-sib families with 94001 as a parent and four full-sib families with 94006 as a parent. Reciprocal crosses were made with male and female *S. viminalis* and *S. suchowensis* while *S. integra*, *S. alberti*, *S. udensis*, and *S. koriyanagi* were crossed in only one male-female arrangement. P294 was received with the identification of *S. alberti*.

suchowensis were crossed with female *S. purpurea* 94006 creating eight F_1 hybrid families (Fig. 3.1). Literature describing these species ranges from the high-quality reference genomes available for *S. purpurea* and *S. viminalis* to the scarcely studied *S. alberti*. *Salix suchowensis* is native to China and has been used recently to generate a chromosome scale genome assembly (Wei et al., 2020b). This species has been assessed for its response to drought stress (Jia et al., 2020) and was one of the first *Salix* used to map the sex determination region (SDR) (Liu et al., 2013). *Salix udensis*, formally known as *S. sachalinensis*, has been described as a Japanese riparian willow species that acts as a natural nest cavity for fish owls (Niiyama, 2008; Slaght et al., 2018) and is suggested to have sexually dimorphic characteristics (Ueno et al., 2006; Ueno & Seiwa, 2003). Some genomic resources are available for the Korean *S. koriyanagi*, as its chloroplast genome has been sequenced (Kim et al., 2019; Park et al., 2019) while *S. integra* has been assessed for its phytoremediation potential (Cao et

al., 2020; Yang et al., 2020b). The *Salix* F_1 hybrid common parent mapping population was developed to interrogate the genetics of several understudied *Salix* species.

Salicaceae represents an interesting family for the study of the evolution of dioecy and the mechanisms of sex determination. Mapping of the sex determination regions (SDR) in Salicaceae has revealed variability not only in the locations of the SDR in different species, but also in the sex determination systems (Yang et al., 2021). In *Populus*, a ZW system SDR was mapped to chromosome (chr) 14 in *P. euphratica*, while both *P. alba* and *P. trichocarpa* use an XY system with an SDR on chr19 (Geraldes et al., 2015; Paolucci et al., 2010; Yang et al., 2021). The SDR in *S. purpurea* was mapped primarily to a single location on chr15, but there appear to be regions of similarity to the *P. trichocarpa* SDR region on chr19 (Zhou et al., 2018); however, it is not known whether dioecy evolved first in willow or poplar (Hou et al., 2015). Like *S. purpurea*, *S. suchowensis* (Chen et al., 2016; Hou et al., 2015) and *S. viminalis* (Pucholt et al., 2015; Pucholt et al., 2017b) use ZW systems with SDRs that map to chr15. However, the tree-form species *S. nigra* uses an XY system with an SDR that maps to chr07 (Sanderson et al., 2021). Considering the variability in the SDR already discovered within *Salix*, it is compelling to elucidate the mechanisms of sex determination in more species of *Salix* to better understand its role in speciation.

To establish the F_1 hybrid common parent mapping population as a genetic resource, this study sought to describe the relationship between the parent species, develop linkage maps for each parent in each family, and map their SDR. Berlin et al. (2014) and Gouker et al. (2019) used a combination of principal component analysis, hierarchical clustering, and STRUCTURE to characterize populations of *S. viminalis*

and *S. purpurea*. Following similar methods, we employed fastSTRUCTURE (Raj et al., 2014), a version of STRUCTURE specially designed to handle genotyping-by-sequencing (GBS) datasets and recently used for studies of cassava (dos Santos Silva et al., 2021), rice (Thapa et al., 2021) and poplar (Yao et al., 2021). While all of the species utilize a ZW sex determination system with an SDR that maps to chr15, these genetic resources provide a foundation for further characterization of the mechanism of sex determination and mapping of other key traits in these related species.

3.3 Materials and Methods

3.3.1 Germplasm and DNA Extraction

Crosses were conducted by forcing floral catkins from dormant shoots in a greenhouse, pollen was extracted from males using toluene essentially as described in Kopp et al. (2002) and was applied to receptive female catkins. Seedlings were established in potting mix in a greenhouse, then transplanted to nursery beds in the field at Cornell AgriTech in Geneva, NY. Mapping populations were established in replicated field trials from dormant cuttings collected from one-year old stems in nursery beds. The *S. purpurea* common parents, 94006 (female) and 94001 (male) were chosen based upon their differential resistance to willow leaf rust (Crowell et al., 2020) and the availability of a high-quality reference genome (Zhou et al., 2020). Shoot tips for DNA extraction were collected from plants in nursery beds and stored in desiccant until DNA extraction. Dried shoot tips were ground to a fine powder with a Geno/Grinder (SPEX SamplePrep, Metuchen, NJ, USA) prior to genomic DNA extraction using the DNeasy Plant Mini Kit (QIAGEN Inc., Valencia, CA, USA).

After checking the DNA quality using gel electrophoresis and estimated quantity using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA), the genomic DNA was submitted to the University of Wisconsin Biotechnology Center (Madison, WI) for 96-plex GBS library preparation using *ApeK* I and sequencing using 1x 100bp Illumina HiSeq 2500 (Illumina, Inc., San Diego, CA, USA).

3.3.2 Variant Discovery and Imputation

Initial variant discovery and filtering followed the TASSEL-5 GBSv2 Discovery/Production Pipeline (Glaubitz et al., 2014). Reads were trimmed to 64kb and aligned using BWA mem (Li & Durbin, 2009) under default parameters to the *S. purpurea* reference genome (Zhou et al., 2020) (https://phytozome-next.jgi.doe.gov/info/Spurpurea_v5_1) with chr15Z removed (available on www.github.com/willowpedia). This process was repeated once with all eight families and parents and then for all eight of the F_1 hybrid families separately. The resulting VCF files contained roughly 500,000 SNPs for the bulk analysis and between 174,000 and 267,000 depending on the family. Imputation was only performed on the family specific SNPs not the bulk analysis. Prior to imputation, SNPs with more than 70% missing data and minor allele frequency less than 0.01 were removed. F_1 individuals were dropped if they had more than 80% missing data and/or were divergent based on principal component analysis. Using LinkImputeR (Money et al., 2017) to impute missing genotypes, genotype calls with a depth less than five were set to missing before filtering for missingness greater than 70%. LinkImputeR imputed missing

genotypes among 127,000 and 200,000 SNPs with accuracies between 0.84 and 0.93, depending on the family.

3.3.3 Population Structure

In the Tassel 5 GUI (Bradbury et al., 2007), SNPs called on the bulk analysis of all eight families and the parents were filtered to retain markers and individuals with no more than 20% missing data and SNPs with a minor allele frequency greater than 0.01 resulting in 55,398 SNPs. Principal components and kinship matrix were derived in Tassel and visualized in R. Multiple runs of the parents and 10 randomly selected F_1 progeny were analyzed using fastSTRUCTURE (Raj et al., 2014), an algorithm designed to determine the population structure within large genotypic datasets.

3.3.4 Linkage Map Construction and QTL Mapping

Using custom R (R Core Team 2020) code (available on Github/willowpedia), multiple runs of each parent were combined to form consensus genotypes for each SNP. In cases where a single parent was unable to reach a consensus genotype with confidence, the genotype of the missing parents was inferred based on the genotype of the known parent and the segregation ratio of the F_1. The consensus genotypes were used to identify the female backcross (AB x AA) and male backcross (AA x AB) markers that will be used to generate linkage maps for each parent. For each marker set, alleles were coded and formatted for R packages R/qtl (Broman et al., 2003) and ASMap (Taylor & Butler, 2017). Once in these packages, SNPs were filtered for co-

location and extreme segregation distortion prior to the formation of linkage groups. Once linkage groups and marker order had been established in ASMap, a custom R function (available on Github/willowpedia) was used to perform simple error correction to reduce the number of double crossovers and deflate map distances before once again forming linkage groups and removing co-located markers. Overall map quality was checked using the function heatmap in ASMap and plotting the cM by physical distance for each chromosome. I verified the sex of each F_1 individual within these families by observing floral morphology for multiple years. Using the constructed linkage maps with sex as the trait, QTL mapping was performed in R/qtl using the scanone function. QTL significance was determined based on the results of 1000 iteration permutation tests performed on each map separately.

3.4 Results

The *Salix* F_1 hybrid common parent mapping population was designed to capture alleles from what I initially thought were six diverse *Salix* species crossed to common *S. purpurea* male and female parents. I used genome-wide GBS markers to characterize the structure of the population, the relationship of the parents, and generate linkage maps for each parent. Upon analysis of these relationships, we questioned the original identification of P294, and thus the crossing design does not capture as much species diversity as we had hoped. Regardless, the mapping of the SDR in each species resulted in the verification of ZW sex determination system located on chr15 in each species.

3.4.1 Population Structure

A combination of PCA, hierarchical clustering and fastSTRUCTURE was used to describe the population structure within the F_1 hybrid common parent mapping populations. A PCA of just the parents revealed three distinct groupings, formed by two principal components accounting for 36.8% and 22.5% of the total genetic variation (Fig. 3.2A). The two *S. purpurea* parents, 94006 and 94001, formed one group and were resolved from the other parents by PC2. The two *S. viminalis* parents, 'Jorr' and 07-MBG-5027 formed a group with *S. udensis*, 04-BN-051, differentiated from the remaining species by PC1. Adding in the F_1 progeny from each family resolved these groups with each PC accounting for 26% and 10.1% of the total variation (Fig. 3.2B). Specifically, PC2 split *S. udensis* from the two *S. viminalis* parents and resolved *S. koriyanagi* from *S. suchowensis*, *S. integra*, and *S. alberti*. As expected, each F_1 individual was intermediate between the *S. purpurea* common parent and their other species parent, the F_1 progeny of females *S. suchowensis* P295, *S. alberti* P294, and *S. integra* P336 crossed with male *S. purpurea* 94001 co-localize.

Hierarchical clustering of the parent marker data grouped the two *S. viminalis* parents with *S. udensis*, the two *S. purpurea* parents together, and *S. koriyanagi* with *S. integra* (Fig. 3.2C). *Salix suchowensis* was at the opposite end of the hierarchy from *S. viminalis* and *S. udensis* with *S. purpurea* in the middle similar to PC1 from Fig. 2A. At the top of the hierarchy, the female *S. suchowensis* P295 was branched apart from the male *S. suchowensis* P63, yet it was closer to *S. alberti* P294. Using a subset of the full population that included multiple runs of each parent along with 10 randomly selected F_1 from each family, multiple analyses of fastSTRUCTURE were

completed at levels of K (number of populations) 3-10. Analysis with K = 6 represented the model complexity that maximized the marginal likelihood suggesting six separate populations within the sample subset represented visually using the fastSTRUCTURE python script 'distruct.py' (Fig. 3.2D). The admixture analysis for

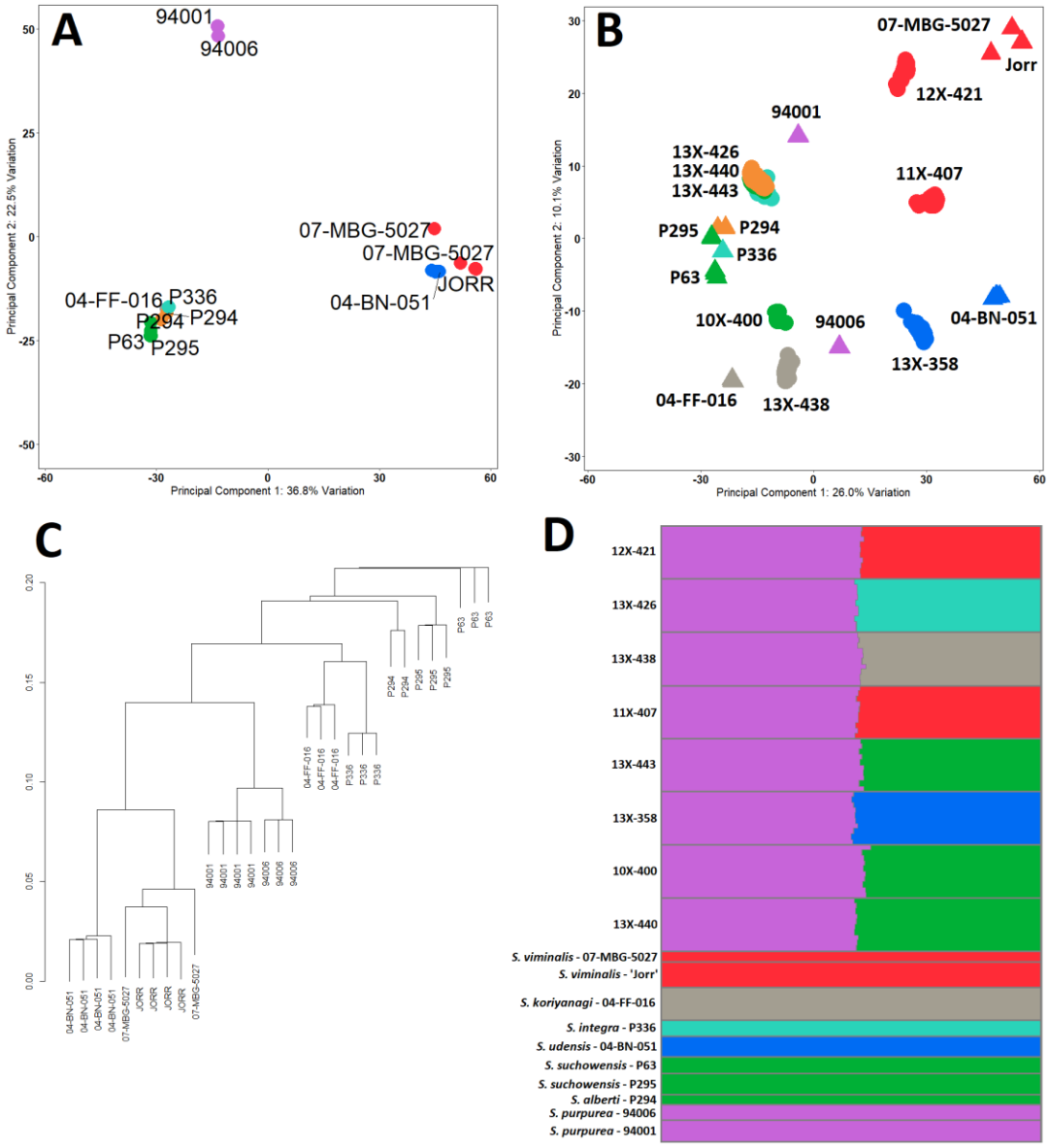


Figure 3.2: Results of PCA and fastStructure analysis of the full population. For panels A, B, and D: purple - *S. purpurea*, red - *S. viminalis*, green - *S. suchowensis*, blue - *S. udensis*, teal – *S. integra*, grey – *S. koriyanagi*, orange - *S. alberti*. A: PCA of the parents; B: PCA of the F_1 and the parents; C: Hierarchical clustering of the parents; D: Distruct plot using fastSTRUCTURE results.

each of the eight families reflected roughly half *S. purpurea* genetic background (purple) and half from the other species parent as anticipated for F_1 individuals. *Salix viminalis, S. integra, S. koriyanagi, S. udensis* and *S. suchowensis* form distinct populations while *S. alberti* P294 grouped together with the *S. suchowensis* population.

3.4.2 Linkage Map Construction and QTL Mapping

Variant discovery and marker filtration was performed for each family separately. Using consensus genotypes derived from multiple sequence runs of the parents, markers were split into female (AB x AA) and male (AA x AB) backcross markers for linkage map construction, resulting in 16 separate linkage maps (Fig. 3.3). Each linkage map consisted of 19 linkage groups with total map lengths ranging from 3939.9 – 6957.3 cM and total marker counts between 2035 – 3852 markers (Appendix Table A.3). Recombination frequency and centiMorgan by physical distance plots generated for each linkage map revealed that marker order within each linkage group was linear with respect to their physical position within the reference genome (Appendix Fig. A.3).

Sex phenotypes were used for QTL mapping of the SDR. Six of the eight families displayed significant sex ratio bias towards females based on a simple chi-square test ($p < 0.05$), while the *S. integra* x *S. purpurea* family was entirely female (Table 3.1). In seven of the eight families, a single QTL for sex was found on chr15 but only in the maternal maps (Fig. 3.4). No QTL for sex were detected on any chromosome in the paternal maps.

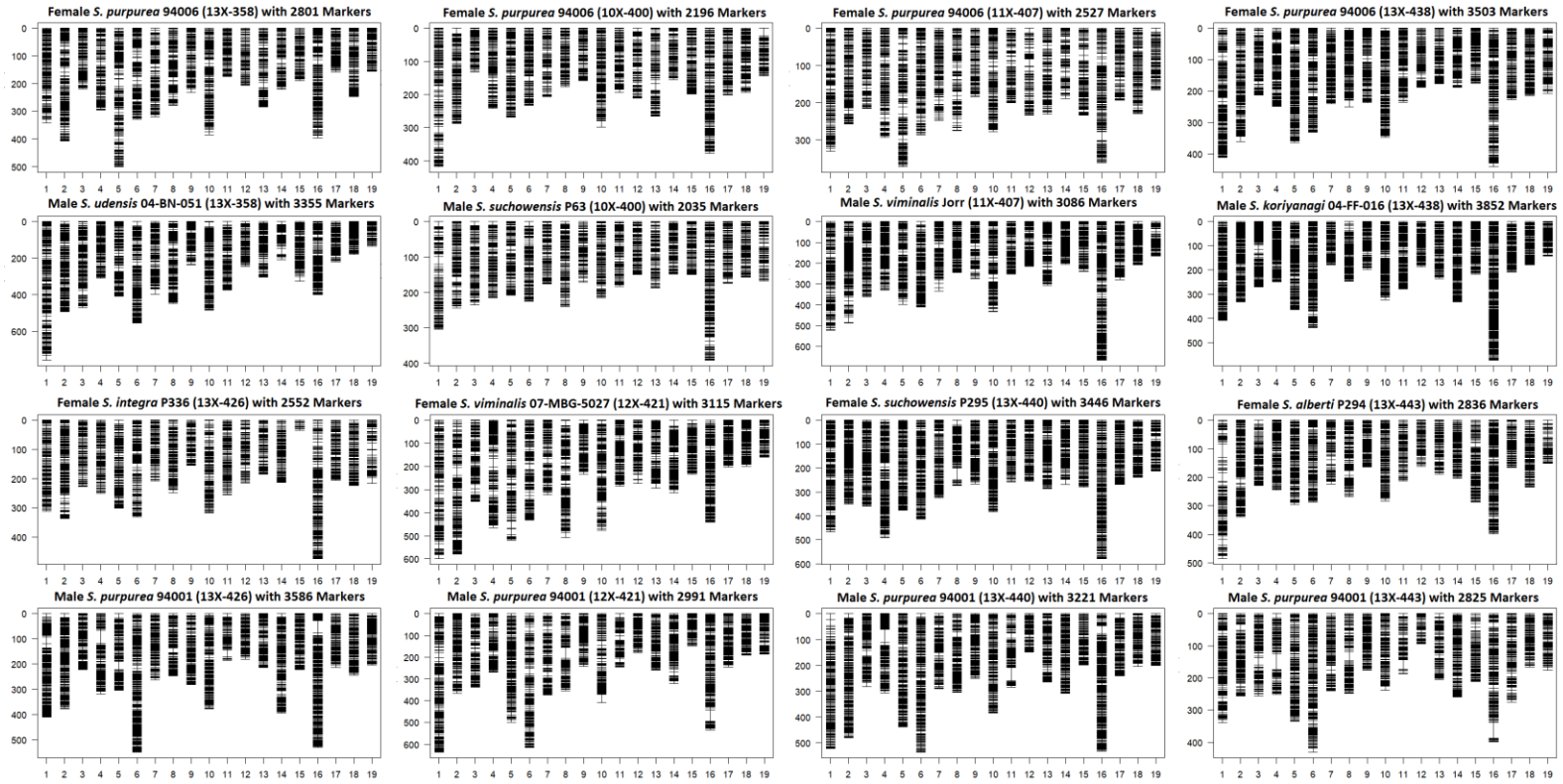


Figure 3.3: Linkage maps for each of the parents within the *Salix* F_1 hybrid common parent mapping populations. Female maps (first and third rows) were constructed using female informative markers (AB x AA), while male maps (second and forth rows) were constructed using male informative markers (AA x AB).

Table 3.1: Sex phenotype statistics for the eight families in the *Salix* F_1 hybrid common parent mapping populations. P-value significance of chi-square tests for no sex ratio bias was 0.05.

Family	Female Parent	Male Parent	No. Female	No. Male	P-Value
10X-400	*S. purpurea*	*S. suchowensis*	53	33	0.03*
11X-407	*S. purpurea*	*S. viminalis*	60	40	0.05
12X-421	*S. viminalis*	*S. purpurea*	53	47	0.55
13X-358	*S. purpurea*	*S. udensis*	91	59	0.01*
13X-426	*S. integra*	*S. purpurea*	150	0	0.00*
13X-438	*S. purpurea*	*S. koriyanagi*	93	57	0.00*
13X-440	*S. suchowensis*	*S. purpurea*	91	58	0.01*
13X-443	*S. alberti*	*S. purpurea*	87	62	0.04*

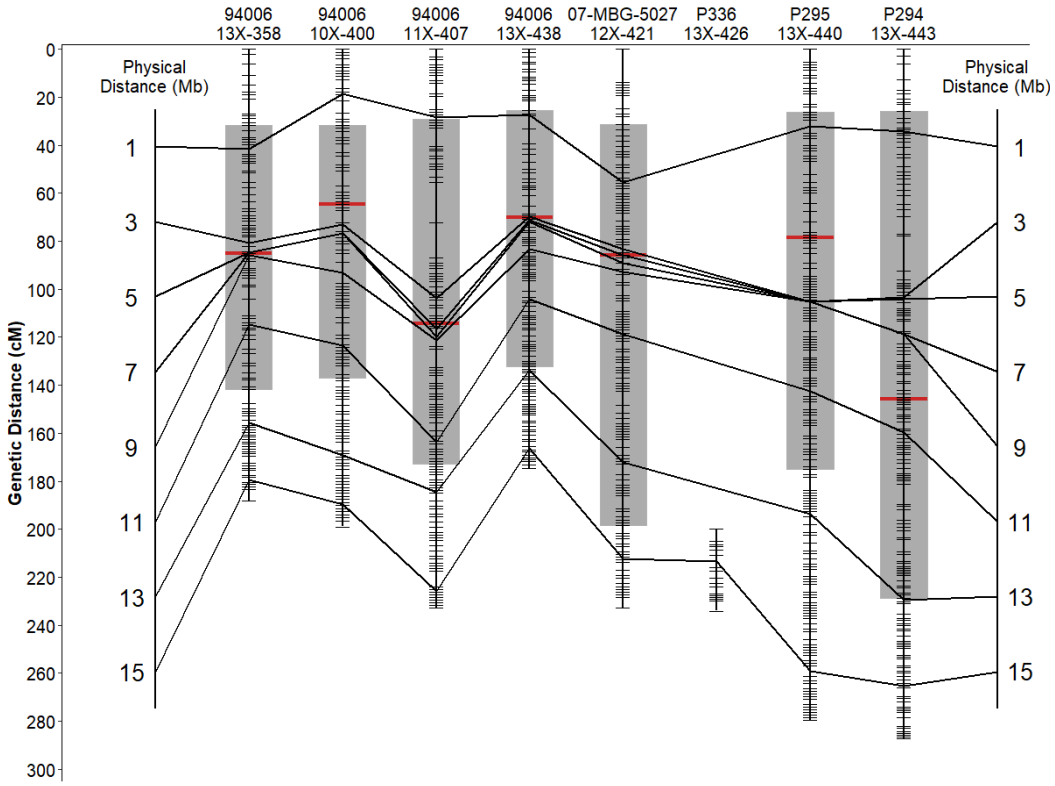


Figure 3.4: Side by side comparison of chr15 from each female parent linkage map with physical distance and sex QTL overlay. The small horizontal lines denote individual markers while the connecting lines show the approximate physical position of each Mb based on alignment to the *S. purpurea* reference genome. Genetic (cM) and physical distance (Mb) both start at zero. The grey shaded area represents the area within the QTL for sex while the red bar shows the position of the peak marker.

3.5 Discussion

Salix is a very diverse genus, consisting of more than 350 species. As the genomic tools available for shrub willow breeders increase in number and quality, so does the ability to characterize and employ *Salix's* innate diversity in improving high yielding cultivars. *Salix* has already been shown to have variability between species even in traits as evolutionarily important as sex determination (Yang et al., 2021). By generating mapping populations that include characterized species crossed with those less studied, I will increase the number of *Salix* species available for trait introgression. Through this research, we introduced the *Salix* F_1 hybrid common parent mapping population. Female and male *S. purpurea*, 94006 and 94001, were crossed to male and female *S. viminalis* and *S. suchowensis* in addition to male *S. udensis* and *S. koriyanagi* and female *S. integra* and *S. alberti*. Using GBS, I analyzed described the population structure among the eight families, generated linkage maps of each of the parents using backcross markers and mapped the sex determination region in each family using field surveyed sex data. PCA, hierarchical clustering and fastSTRUCTURE analysis used multiple runs of the parents and their F_1 progeny to describe the relationships within the common parent F_1 hybrid population. PCA and hierarchical clustering predominately separated the population by section with the F_1 splitting the distance between the parents as expected (Fig. 3.2ABC). *Salix viminalis* and *S. udensis* from Section Vimen while the remaining species are from Section Helix (Dickmann & Kuzovkina, 2014). In each analysis, the *S. alberti* parent P294 was found to be closely related to the *S. suchowensis* parents, grouping with them in fastSTRUCTURE (Fig. 3.2D). Given these results and the limited presence of this

species in the literature, P294 is likely a true or hybrid *S. suchowensis*. Based on my analysis of P294, I believe it should be classified as *S. suchowensis*, so I will describe it as such from this point on, abandoning the incorrect designation as *S. alberti*.

Among the 16 linkage maps produced, QTL for sex were detected in only seven of the eight families, all on chr15 and only on the maternal maps. As seen in a recent study mapping sex in *Salix* using backcross markers, Li et al. (2020) were only able to detect QTL within the maternal map, suggesting that *S. triandra* uses a ZW sex determination system on chr15. While this had been known for *S. purpurea* (Zhou et al., 2018), *S. viminalis* (Pucholt et al., 2017b), and *S. suchowensis* (Chen et al., 2016), this is the first study to identify the ZW sex determination system and region on chr15 in *S. koriyanagi* and *S. udensis*.

Since the entire *S. integra* x *S. purpurea* family was phenotyped as female, I did not expect to map QTL for sex. Although not to the extreme seen in this family, the occurrence of sex ratio bias and sexual dimorphism is prevalent in *Salix*. Of those studied here, *S. purpurea* (Gouker et al., 2020), *S. viminalis* (Alström-Rapaport et al., 1997), *S. suchowensis* (Yang et al., 2020a), *S. udensis* (Ueno & Seiwa, 2003), and *S. integra* (Tozawa et al., 2009) have documented cases of sex ratio bias, but this is the first time sex ratio bias has been reported in *S. koriyanagi*. Predictions for the potential causes of sex ratio bias from those cited above include higher mortality rates in males, increased herbivory and pathogen resistance in females, and the presence of a sex distorter locus (Pucholt et al., 2017a).

The location of the SDR in this study shows that the QTL on chr15 accounted for a majority of the chr. For the four families where *S. purpurea* 94006 is the female

parent, each QTL started from 0.9-1.5 Mb and ended between 11.7 – 12.8 Mb (Fig 3.4). These QTL included more of the two pericentromeric pseudoautosomal regions than the SDR found by Zhou et al. (2020) (2.3 to 9.1 Mb) while in some cases included the 11.4 to 12.3 Mb SDR proposed by Carlson et al. (2017) through analysis of differential gene expression. Both of these papers used individuals from a *S. purpurea* F_2 population described fully in Carlson et al. (2019) that defined the boundary of their SDR to 4.5-11.4 Mb. This variation in the size of the SDR could be attributed to the analysis methods used to derive them.

The most recent delimitation of the SDR in *S. viminalis* spanned roughly 3.4 Mb (approx. 2.3 – 5.7 Mb) (Almeida et al., 2020), considerably narrower than the region found in the female 07-MBG-5027 linkage map, 0.3 Mb – 13.9 Mb. Concomitantly, Almeida et al. (2020) aligned chr15 of the *S. purpurea* and *S. viminalis* references and found synteny between the SDRs, yet with several chromosomal rearrangements. This is in contrast to what is seen in my results, where each chr15 is co-linear, likely due to each family's alignment to the *S. purpurea* v.5.1 reference genome. The chr15 from *S. integra* P336 is considerably smaller than the other families, only aligning to the 14.2 – 15.5 Mb region of the *S. purpurea* reference genome. This family's chr15 is the only chromosome among the 16 maps that had less than near complete coverage. While no QTL was expected for this family, this region is outside the QTL intervals of the other female maps, making it impossible to draw conclusions about its' extreme sex bias within this study. Based on annotations from the *Populus* reference, differential gene expression analysis between male and female plants led to predictions that the SDR in *S. suchowensis* was on chr14 (Liu et al.,

2013); however, later work repositioned the SDR to the centromeric region of chr15 when based on *Salix* alignment (Chen et al., 2016). This study was able to provide a defined physical distance for the SDR of 0.8 – 12.3 Mb in P295 and 0.5 – 12.8 Mb in P294

Suppressed recombination is a hallmark of chromosomes containing an SDR. Comparing the map of chr15 from each family with a mapped SDR (Fig. 3.4), it extends across a region of sparse marker density, roughly 3 Mb to 9 Mb, and is flanked by those with greater marker density. As described above, this centromeric region of repressed recombination is often associated with the SDR in *Salix*. In three of the four *S. purpurea* 94006 maps and the *S. viminalis* 07-MBG-5207 map, the peak QTL marker is located within this region. The *S. suchowensis* P295 map and the remaining *S. purpurea* map (which is crossed to the male *S. suchowensis*) both peak just prior to this region around 2.6 and 2.9 Mb, respectively while the peak marker in P294 was located just after this region at 9.7 Mb. Although the confidence intervals of the QTL presented in this paper are wider than those presented previously the region of repressed recombination and closely associated peak markers are consistent with the literature.

This study successfully described the population structure among the eight families within the *Salix* F_1 hybrid common parent mapping population, constructed linkage maps for each parent, and mapped the SDR in each family. QTL mapping of the SDR to chr15 was successful for the maternal maps of seven of the eight families, proving that all the species within this population use a ZW sex determination system. The introduction of the F_1 hybrid common parent mapping population provides the

opportunity to map QTL for phenotypic traits beyond sex determination while the linkage maps themselves could be used to anchor contigs in the generation of new reference genomes for each of the parents.

3.6 REFERENCES

Almeida, P., Proux-Wera, E., Churcher, A., Soler, L., Dainat, J., Pucholt, P., Nordlund, J., Martin, T., Rönnberg-Wästljung, A.-C., Nystedt, B., Berlin, S., & Mank, J. E. (2020). Genome assembly of the basket willow, *Salix viminalis*, reveals earliest stages of sex chromosome expansion. *BMC Biol.*, *18*(1), 78. doi:10.1186/s12915-020-00808-1

Alström-Rapaport, C., Lascoux, M., & Gullberg, U. (1997). Sex determination and sex ratio in the dioecious shrub *Salix viminalis* L. *Theor. Appl. Genet.*, *94*(3-4), 493-497.

Berlin, S., Lagercrantz, U., von Arnold, S., Ost, T., & Ronnberg-Wastljung, A. C. (2010). High-density linkage mapping and evolution of paralogs and orthologs in *Salix* and *Populus*. *BMC Genomics, 11*, 129. doi:10.1186/1471-2164-11-129

Berlin, S., Trybush, S. O., Fogelqvist, J., Gyllenstrand, N., Hallingbäck, H. R., Åhman, I., Nordh, N. E., Shield, I., Powers, S. J., Weih, M., Lagercrantz, U., Rönnberg-Wästljung, A. C., Karp, A., & Hanley, S. J. (2014). Genetic diversity, population structure and phenotypic variation in European *Salix viminalis* L. (Salicaceae). *Tree Genet. Genomes, 10*(6), 1595-1610. doi:10.1007/s11295-014-0782-5

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633-2635. doi:10.1093/bioinformatics/btm308

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics, 19*(7), 889-890. doi:10.1093/bioinformatics/btg112

Cao, Y., Ma, C., Chen, H., Zhang, J., White, J. C., Chen, G., & Xing, B. (2020). Xylem-based long-distance transport and phloem remobilization of copper in *Salix integra* Thunb. *J. Hazard. Mater., 392*, 122428. doi:https://doi.org/10.1016/j.jhazmat.2020.122428

Carlson, C. H., Choi, Y., Chan, A. P., Serapiglia, M. J., Town, C. D., & Smart, L. B. (2017). Dominance and sexual dimorphism pervade the *Salix purpurea* L. transcriptome. *Genome Biol. Evol., 9*(9), 2377-2394. doi:10.1093/gbe/evx174

Carlson, C. H., Gouker, F. E., Crowell, C. R., Evans, L., DiFazio, S. P., Smart, C. D., & Smart, L. B. (2019). Joint linkage and association mapping of complex traits in shrub willow (*Salix purpurea* L.). *Ann. Bot., 124*(4), 701-716. doi:10.1093/aob/mcz047

Chen, Y., Wang, T., Fang, L., Li, X., & Yin, T. (2016). Confirmation of single-locus sex determination and female heterogamety in willow based on linkage analysis. *PLoS One, 11*(2), e0147671. doi:10.1371/journal.pone.0147671

Clifton-Brown, J., Harfouche, A., Casler, M. D., Dylan Jones, H., Macalpine, W. J., Murphy-Bokern, D., Smart, L. B., Adler, A., Ashman, C., & Awty-Carroll, D. (2019). Breeding progress and preparedness for mass-scale deployment of perennial lignocellulosic biomass crops switchgrass, miscanthus, willow and poplar. *GCB Bioenergy, 11*(1), 118-151.

Crowell, C. R., Bekauri, M. M., Cala, A. R., Mcmullen, P., Smart, L. B., & Smart, C. D. (2020). Differential susceptibility of diverse *Salix* spp. to *Melampsora americana* and *Melampsora paradoxa*. *Plant Dis., 104*, 2949-2957. doi:10.1094/PDIS-04-20-0718-RE

Dickmann, D. I., & Kuzovkina, J. (2014). Poplars and willows of the world, with emphasis on silviculturally important species. In *Poplars and willows: Trees for society and the environment* (Vol. 22). Rome, Italy: FAO.

dos Santos Silva, P. P., Sousa, M. B. e., de Oliveira, E. J., Morgante, C. V., de Oliveira, C. R. S., Vieira, S. L., & Borel, J. C. (2021). Genome-wide association study of drought tolerance in cassava. *Euphytica, 217*(4), 60. doi:10.1007/s10681-021-02800-4

Fabio, E. S., Leary, C. J., & Smart, L. B. (2019). Tolerance of novel inter-specific shrub willow hybrids to water stress. *Trees, 33*(4), 1015-1026. doi:10.1007/s00468-019-01835-4

Fabio, E. S., & Smart, L. B. (2020). Genetic and environmental influences on first rotation shrub willow (*Salix* spp.) bark and wood elemental composition. *Bioenergy Res., 13*(3), 797-809. doi:10.1007/s12155-020-10122-x

Geraldes, A., Hefer, C. A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R. Y., Stanton, B., Guy, R. D., Mansfield, S. D., Douglas, C. J., & Cronk, Q. C. B. (2015). Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Mol. Ecol., 24*(13), 3243-3256. doi:https://doi.org/10.1111/mec.13126

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire. R. J., Sun, Q., Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One.* 9(2): e90346. https://doi.org/10.1371/journal.pone.0090346

Gouker, F. E., Carlson, C. H., Zou, J., Evans, L., Crowell, C. R., Smart, C. D., DiFazio, S. P., & Smart, L. B. (2020). Sexual dimorphism and sex ratio bias in the dioecious willow *Salix purpurea* L. *bioRxiv*, 2020.2004.2005.026427. doi:10.1101/2020.04.05.026427

Gouker, F. E., DiFazio, S. P., Bubner, B., Zander, M., & Smart, L. B. (2019). Genetic diversity and population structure of native, naturalized, and cultivated *Salix purpurea*. *Tree Genet. Genomes*, *15*(3), 47. doi:10.1007/s11295-019-1359-0

Hallingbäck, H. R., Berlin, S., Nordh, N.-E., Weih, M., & Rönnberg-Wästljung, A.-C. (2019). Genome wide associations of growth, phenology, and plasticity traits in willow [*Salix viminalis* (L.)]. *Front. Plant Sci.*, *10*(753). doi:10.3389/fpls.2019.00753

Hallingbäck, H. R., Fogelqvist, J., Powers, S. J., Turrion-Gomez, J., Rossiter, R., Amey, J., Martin, T., Weih, M., Gyllenstrand, N., Karp, A., Lagercrantz, U., Hanley, S. J., Berlin, S., & Rönnberg-Wästljung, A.-C. (2016). Association mapping in *Salix viminalis* L. (Salicaceae) – identification of candidate genes associated with growth and phenology. *GCB Bioenergy*, *8*(3), 670-685. doi:https://doi.org/10.1111/gcbb.12280

Hanley, S., Mallott, M., & Karp, A. (2006). Alignment of a *Salix* linkage map to the *Populus* genomic sequence reveals macrosynteny between willow and poplar genomes. *Tree Genet. Genomes*, *3*(1), 35-48.

Hou, J., Ye, N., Zhang, D., Chen, Y., Fang, L., Dai, X., & Yin, T. (2015). Different autosomes evolved into sex chromosomes in the sister genera of *Salix* and *Populus*. *Scientific Reports, 5*(1), 9076. doi:10.1038/srep09076

Jia, H., Wang, L., Li, J., Sun, P., Lu, M., & Hu, J. (2020). Physiological and metabolic responses of *Salix sinopurpurea* and *Salix suchowensis* to drought stress. *Trees*, *34*(2), 563-577. doi:10.1007/s00468-019-01937-z

Kim, J., Kim, Y., & Park, J. (2019). Complete chloroplast genome sequence of the *Salix koriyanagi* Kimura ex Goerz (Salicaceae). *Mitochondrial DNA B: Resour.*, *4*(1), 549-550. doi:10.1080/23802359.2018.1553521

Kopp, R., Maynard, C., Rocha De Niella, P., Smart, L., & Abrahamson, L. (2002). Collection and storage of pollen from *Salix* using organic solvents. *Am. J. Bot.*, *89*, 248-252.

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

Li, W., Wu, H., Li, X., Chen, Y., & Yin, T. (2020). Fine mapping of the sex locus in *Salix triandra* confirms a consistent sex determination mechanism in genus *Salix*. *Hortic. Res.*, *7*(1), 64. doi:10.1038/s41438-020-0289-1

Liu, J., Yin, T., Ye, N., Chen, Y., Yin, T., Liu, M., & Hassani, D. (2013). Transcriptome analysis of the differentially expressed genes in the male and

female shrub willows (*Salix suchowensis*). *PLoS One, 8*(4), e60181. doi:10.1371/journal.pone.0060181

Money, D., Migicovsky, Z., Gardner, K., & Myles, S. (2017). LinkImputeR: user-guided genotype calling and imputation for non-model organisms. *BMC Genomics, 18*(1), 523. doi:10.1186/s12864-017-3873-5

Niiyama, K. (2008). Coexistence of *Salix* species in a seasonally flooded habitat. In *Ecology of Riparian Forests in Japan* (pp. 165-174): Springer.

Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Mugnozza, G. S., & Sabatti, M. (2010). Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genet. Genomes, 6*(6), 863-875.

Park, J., Kim, Y., & Xi, H. (2019). The complete chloroplast genome sequence of male individual of Korean endemic willow, *Salix koriyanagi* Kimura ex Goerz (Salicaceae). *Mitochondrial DNA B: Resour., 4*(1), 1619-1621. doi:10.1080/23802359.2019.1602012

Pucholt, P., Hallingbäck, H. R., & Berlin, S. (2017a). Allelic incompatibility can explain female biased sex ratios in dioecious plants. *BMC Genomics, 18*(1), 251. doi:10.1186/s12864-017-3634-5

Pucholt, P., Rönnberg-Wästljung, A. C., & Berlin, S. (2015). Single locus sex determination and female heterogamety in the basket willow (*Salix viminalis* L.). *Heredity, 114*(6), 575-583. doi:10.1038/hdy.2014.125

Pucholt, P., Wright, A. E., Conze, L. L., Mank, J. E., & Berlin, S. (2017b). Recent sex chromosome divergence despite ancient dioecy in the willow *Salix viminalis*. *Mol. Biol. Evol., 34*(8), 1991-2001. doi:10.1093/molbev/msx144

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics, 197*(2), 573-589. doi:10.1534/genetics.114.164350

Rönnberg-Wästljung, A. C., Glynn, C., & Weih, M. (2005). QTL analyses of drought tolerance and growth for a *Salix dasyclados* x *Salix viminalis* hybrid in contrasting water regimes. *Theor. Appl. Genet., 110*(3), 537-549. doi:10.1007/s00122-004-1866-7

Rönnberg-Wästljung, A. C., Samils, B., Tsarouhas, V., & Gullberg, U. (2008). Resistance to *Melampsora larici-epitea* leaf rust in *Salix*: analyses of quantitative trait loci. *J. Appl. Genet., 49*, 321-331. doi:10.1007/BF03195630

Samils, B., Rönnberg-Wästljung, A.-C., & Stenlid, J. (2011). QTL mapping of resistance to leaf rust in *Salix*. *Tree Genet. Genomes, 7*(6), 1219-1235. doi:10.1007/s11295-011-0408-0

Sanderson, B. J., Feng, G., Hu, N., Carlson, C. H., Smart, L. B., Keefover-Ring, K., Yin, T., Ma, T., Liu, J., DiFazio, S. P., & Olson, M. S. (2021). Sex determination through X–Y heterogamety in *Salix nigra*. *Heredity, 126*(4), 630-639. doi:10.1038/s41437-020-00397-3

Slaght, J. C., Surmach, S. G., & Kisleiko, A. A. (2018). Ecology and conservation of Blakiston's fish owl in Russia. In *Biodiversity Conservation Using Umbrella Species* (pp. 47-70): Springer.

Smart, L. B., Volk, T., Lin, J., Kopp, R. F., Phillips, I. S., Cameron, K. D., White, E. H., & Abrahamson, L. (2005). Genetic improvement of shrub willow (*Salix* spp.) crops for bioenergy and environmental applications in the United States. *Unasylva, 56*, 51-55.

Stanton, B. J., Serapiglia, M. J., & Smart, L. B. (2014). The domestication and conservation of *Populus* and *Salix* genetic resources. In *Poplars and willows: trees for society and the environment.* (pp. 124-199). Wallingford, UK: CAB International.

Stoof, C. R., Richards, B. K., Woodbury, P. B., Fabio, E. S., Brumbach, A. R., Cherney, J., Das, S., Geohring, L., Hansen, J., Hornesky, J., Mayton, H., Mason, C., Ruestow, G., Smart, L. B., Volk, T. A., & Steenhuis, T. S. (2014). Untapped potential: opportunities and challenges for sustainable bioenergy production from marginal lands in the Northeast USA. *Bioenergy Res., 8*(2), 482-501. doi:10.1007/s12155-014-9515-8

Sulima, P., Przyborowski, J. A., Kuszewska, A., Zaluski, D., Jedryczka, M., & Irzykowski, W. (2017). Identification of quantitative trait loci conditioning the main biomass yield components and resistance to *Melampsora* spp. in *Salix viminalis* x *Salix schwerinii* hybrids. *Int. J. Mol. Sci., 18*(3). doi:10.3390/ijms18030677

Taylor, J., & Butler, D. (2017). R Package ASMap: efficient genetic linkage map construction and diagnosis. *J. Stat. Softw., 79*(6), 1-29. doi:10.18637/jss.v079.i06

Thapa, R., Tabien, R. E., & Septiningsih, E. M. (2021). Genome-wide association study to identify chromosomal regions related to panicle architecture in rice (*Oryza sativa*). *Genet. Resour. Crop Evol.* doi:10.1007/s10722-021-01159-8

Tozawa, M., Ueno, N., & Seiwa, K. (2009). Compensatory mechanisms for reproductive costs in the dioecious tree *Salix integra*. *Botany, 87*(3), 315-323. doi:10.1139/b08-125

Ueno, N., Kanno, H., & Seiwa, K. (2006). Sexual differences in shoot and leaf dynamics in the dioecious tree *Salix sachalinensis*. *Botany, 84*(12), 1852-1859.

Ueno, N., & Seiwa, K. (2003). Gender-specific shoot structure and functions in relation to habitat conditions in a dioecious tree, *Salix sachalinensis*. *J. For. Res., 8*(1), 0009-0016.

Wei, S., Yang, Y., & Yin, T. (2020). The chromosome-scale assembly of the willow genome provides insight into Salicaceae genome evolution. *Hortic. Res., 7*(1), 45. doi:10.1038/s41438-020-0268-6

Yang, G., Xu, Q., Li, W., Ling, J., Li, X., & Yin, T. (2020a). Sex-related differences in growth, herbivory, and defense of two *Salix* species. *Forests, 11*(4), 450.

Yang, W., Wang, D., Li, Y., Zhang, Z., Tong, S., Li, M., Zhang, X., Zhang, L., Ren, L., & Ma, X. (2021). A general model to explain repeated turnovers of sex determination in the Salicaceae. *Mol. Biol. Evol., 38*(3), 968-980.

Yang, W., Wang, Y., Liu, D., Hussain, B., Ding, Z., Zhao, F., & Yang, X. (2020b). Interactions between cadmium and zinc in uptake, accumulation and bioavailability for *Salix integra* with respect to phytoremediation. *Int. J. Phytoremediation, 22*(6), 628-637. doi:10.1080/15226514.2019.1701981

Yao, L., Li, P., Du, Q., Quan, M., Li, L., Xiao, L., Song, F., Lu, W., Fang, Y., & Zhang, D. (2021). Genetic architecture underlying the metabolites of chlorogenic acid biosynthesis in *Populus tomentosa*. *Int. J. Mol. Sci., 22*(5), 2386.

Zhou, R., Macaya-Sanz, D., Carlson, C. H., Schmutz, J., Jenkins, J. W., Kudrna, D., Sharma, A., Sandor, L., Shu, S., Barry, K., Tuskan, G. A., Ma, T., Liu, J., Olson, M., Smart, L. B., & DiFazio, S. P. (2020). A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biol., 21*(1), 38. doi:10.1186/s13059-020-1952-4

Zhou, R., Macaya-Sanz, D., Rodgers-Melnick, E., Carlson, C. H., Gouker, F. E., Evans, L. M., Schmutz, J., Jenkins, J. W., Yan, J., Tuskan, G. A., Smart, L. B., & DiFazio, S. P. (2018). Characterization of a large sex determination region in *Salix purpurea* L. (Salicaceae). *Mol. Genet. Genomics, 293*(6), 1437-1452. doi:10.1007/s00438-018-1473-y

CHAPTER 4

QTL MAPPING OF *MELAMPSORA* LEAF RUST RESISTANCE, INSECT DAMAGE, AND YIELD COMPONENT TRAITS IN EIGHT *SALIX* F_1 HYBRID COMMON PARENT FAMILIES

Formatted for submission as: Wilkerson, DG, Crowell, CR, Smart, CD, Smart, LB. 2021.

4.1 Abstract

Trait introgression requires the assessment and identification of novel sources of variation among targeted traits. For shrub willow (*Salix*) breeders, there exists a plethora of understudied species within a genus with more than 350 species. Shrub willow is primarily bred for biomass yield for bioenergy, therefore typical breeding objectives include insect and pathogen resistance, crown architecture and yield component traits. Among these, breeding for durable resistance to willow leaf rust (*Melampsora americana*) is of particular importance as the pathogen can significantly reduce biomass yields in commercial production. A *Salix* F_1 hybrid common parent mapping population was created to characterize the variation among eight species-hybrid families and map QTL. With female and male *S. purpurea* as the common parents, crosses were made to male *S. suchowensis*, *S. viminalis*, *S. koriyanagi*, and *S. udensis* and female *S. viminalis*, *S. integra*, and two *S. suchowensis*. All eight families were planted in field trials at Cornell AgriTech in Geneva, NY. Family mean separation, testing for sexual dimorphism, and QTL mapping were conducted on each

trait. I collected 41 phenotypes and mapped 87 separate QTL onto 16 existing backcross linkage maps. This currently represents the largest QTL study in *Salix* as a foundation for future research into fine mapping, identification of candidate genes and development of markers for marker-assisted selection and trait introgression.

4.2 Introduction

Shrub willows (*Salix* spp.) are woody perennials with ornamental, riparian, and bioenergy applications. Consisting of over 350 species, the genus *Salix* can be found worldwide in northern latitudes and grown on marginal land unsuitable for food crops (Smart et al., 2005). *Salix* species have been targeted as a bioenergy feedstock, producing fast-growing, carbon-neutral biomass without occupying land needed for food crops with minimal herbicide and fertilizer applications (Pacaldo et al., 2014; Volk et al., 2016). *Salix* is taxonomically challenging, however (Sanderson et al., 2020), as species are often misidentified or reassigned. Skvortsov (1999) proposed the most commonly used classification system, dividing Salicaceae into three genera, *Salix, Populus,* and *Chosenia*, further dividing *Salix* into three subgenera: *Salix, Vetrix,* and *Chamaetia* (Wu et al., 2015). The species targeted for potential as bioenergy feedstocks are found in the subgenus *Vetrix*, including the two most common in commercial production in the United States and Europe, *S. purpurea* (Section: *Helix*) and *S. viminalis* (Section: *Viminella*) (Wagner et al., 2018).

Active breeding began in earnest in Europe and the United States in the 1970s and 1990s, employing the innate diversity within *Salix* to improve shrub willow for commercial biomass production. In addition to biomass yield, breeding objectives

have included agronomic traits, insect and pathogen resistance, and leaf morphology. A boon to breeders, many species in *Salix* readily hybridize increasing their pools of functional diversity available for trait improvement. While considerable progress has been made through species hybridization, formation of polyploids, and clonal propagation (Serapiglia et al., 2014b), there remains potential for targeted trait improvement through the introgression of diverse alleles into elite cultivars.

Introgression is defined as the transfer of genetic material from a donor parent through hybridization and repeated backcrossing to the recipient parent (Anderson & Hubricht, 1938; Harrison & Larson, 2014). Not without its disadvantages, introgression is typically employed after the available diversity is exhausted for the target trait (Pratap et al., 2021). However, the incorporation of diverse alleles in the early stages of a breeding program reduces the time lost in backcross generations as introgression and elite selection can occur concurrently as opposed to consecutively. Among simply inherited traits, introgression can occur through visual phenotyping and selection while more complex traits benefit from the use of molecular markers.

Mapping quantitative trait loci (QTL) identifies the genomic regions most associated with specific traits and is a natural starting point for marker-assisted selection (Collard & Mackill, 2008). Through the use of associated markers in high density linkage maps that define smaller target regions, the efficiency of trait introgression can be improved (Hernandez et al., 2020). Given their popularity in breeding programs, published mapping populations in *Salix* largely utilize *S. viminalis* and *S. purpurea*. Traits mapped in *S. viminalis* populations include: drought tolerance (Rönnberg-Wästljung et al., 2005), growth and phenology (Hallingbäck et al., 2019;

Hallingbäck et al., 2016), and resistance to *Melampsora larici-epitea* (Hanley et al., 2011; Ronnberg-Wastljung et al., 2008; Samils et al., 2011; Sulima et al., 2017). *Salix viminalis* is extremely susceptible to potato leafhopper (PLH, *Empoasca fabae*) (Wang et al., 2020), so it serves as a potential trait donor when crossed to other species, since many hybrids are largely resistant to PLH. Recently, Carlson et al. (2019) published an extensive QTL and GWAS study mapping a multitude of QTL in *S. purpurea* for morphological, physiological, pest and disease resistance, and wood chemical composition traits. Leveraging these results in conjunction with new mapping populations between *S. purpurea* and other diverse species will provide plant breeders a more complete representations of the QTL available for trait improvement.

Among the traits targeted by plant breeders in North America is the improvement of resistance to willow leaf rust, *Melampsora americana*. Willow leaf rust is the most devastating pathogen to shrub willow grown in commercial production, potentially reducing yields by 50 percent and increasing susceptibility to other pathogens (McCracken et al., 2001). While management strategies include planting clonal mixtures and herbicide applications, these methods do not provide a sustainable level of resistance. The previous chapter introduced the *Salix* F_1 hybrid common parent mapping population consisting of eight families. Four families each were crossed to common female and male *S. purpurea* parents. The locus controlling sex was determined for 16 linkage maps, one for each parent of the eight families produced through that study. This study includes a more extensive analysis of phenotypes measured in the mapping population, including resistance to willow leaf rust, insect damage, and key components traits for biomass yield. These results

establish targets for introgression of advantageous alleles into new, improved hybrid cultivars.

4.3 Materials and Methods

4.3.1 Germplasm

The *Salix* F1 hybrid common parent mapping population is described in Chapter 3. Briefly, this population is comprised of eight hybrid families, four of which share *S. purpurea* 94006 as the female parent crossed to male *S. viminalis*, *S. suchowensis*, *S. udensis*, and *S. koriyanagi*. The other four share the male *S. purpurea* 94001 and are crossed to female *S. viminalis*, *S. integra*, and two *S. suchowensis* (Fig 4.1.). One parent was originally described as *S. alberti*, but genome-wide marker analysis confirmed it to be *S. suchowensis* (Chapter 3). All eight families were planted from dormant stem cuttings in different years in randomized complete blocks with

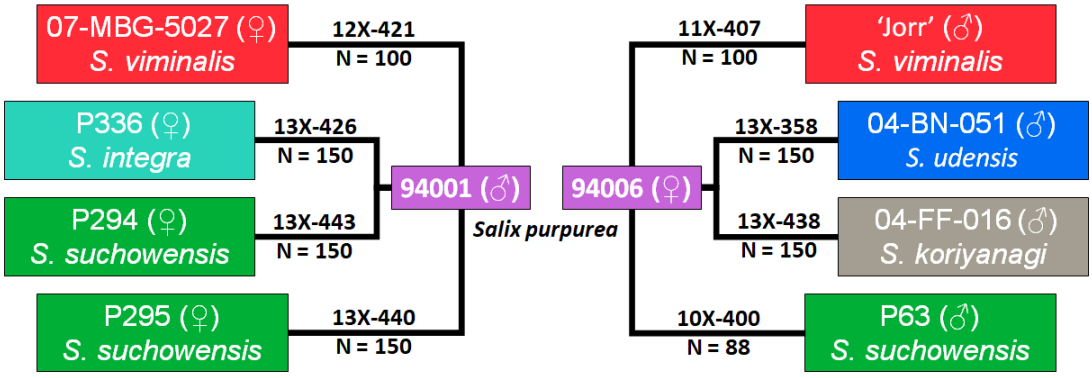


Figure 4.1: Updated pedigrees of the *Salix* F1 hybrid common parent mapping population. There were four full-sib families with 94001 as a parent and four full-sib families with 94006 as a parent. Reciprocal crosses were made with male and female *S. viminalis* and *S. suchowensis* while *S. integra*, *S. alberti*, *S. udensis*, and *S. koriyanagi* were crossed in only one male-female arrangement. P294 was received as *S. alberti* but shown in Chapter 3 to be *S. suchowensis*.

four replications in three adjacent fields at Cornell AgriTech in Geneva, NY. Regrowth was synchronized via coppicing prior to initial data collection.

4.3.2 Phenotype Collection

Phenotypic traits collected in this study were split into four groups, leaf architecture, leaf rust, herbivory, and yield components. All of the traits in this study were collected following the methodologies described by Carlson et al. (2019). Traits classified as leaf architecture include leaf length (LL, cm), width (LW, cm), area, (LA, cm^2), perimeter (LP, cm), aspect ratio (LFR), and shape factor (LFF). Fully mature leaves were dried at 65°C and weighed (DLW, g). DLW was then divided by LA to produce specific leaf area (SLA, g cm^{-2}). All leaf architecture traits were collected during the summers of 2017 and 2018. Leaf rust severity (RST) ratings were collected in 2017 and 2019 as a 0 – 50% rating scale based on percent infected leaf area per plot. 50% being the highest rating due to extensive defoliation at that level of severity. No ratings were conducted in 2018 due to lack of rust incidence in the field trials. Ratings in each year were completed in all families within a seven-day period in September, when severity was at its highest level.

Herbivory ratings focused on PLH and imported willow leaf beetle (WLB, *Plagiodera versicolor*). Potato leafhopper ratings were split into three categories of damage, leaf curl (PLHC), necrosis (PLHN), and shoot tip death (PLHS) since variation exists among the three damage types that one rating could not capture (Wang et al., 2020). Ratings were collected in 2017 and 2019. Imported willow leaf beetle damage was identified as small pinholes (1-2 mm diameter) in the leaves observed

concomitant with abundance of WLB populations. Ratings were based on percent damaged leaf area and were collected in 2017, 2018, and 2019. Yield component traits collected during plant dormancy include plant height (HT, m) and average stem diameter (AVGDIAM, cm) measured in 2017 and 2018, and crown form (CROWN, degrees) as described in 2018. Plant area (PLTAREA) and volume (PLTVOL) were calculated as in Carlson et al. (2019) and PLTVOL was used as a proxy for biomass yield. Leaf chlorophyll content (SPAD) was estimated using SPAD-502 meters (Minolta Osaka Co., Ltd, Japan) in August and October 2018 and August 2019.

4.3.3 Statistical Analysis

All statistical analyses were conducted in R (R Core Team 2020). Unless otherwise stated, custom R code (available on Github.com/Willowpedia) was used to execute the below analyses. Outliers were identified and removed on a per family basis based upon a point's Cook's distance and interquartile. Then, using Fisher's LSD ($p < 0.05$) family mean separations were based upon the regression model, trait = FIELD + FIELD/REP + FIELD/FAMILY. Whether or not a trait was sexually dimorphic ($p < 0.05$) was determined using a Student's-t if the males and females within a family were normally distributed and had equal variance. If not normally distributed, then a Wilcoxon signed rank test was used while if variances were unequal then Welch's t-test was used to compare the male and female plants within a family.

4.3.4 Linkage Map Construction and QTL Mapping

The linkage maps used for QTL mapping in this study were described in Chapter 3. QTL determination in this study used genotype level means in the R/qtl function ‘cim’ to perform composite interval mapping. Significance was determined based upon the results of a 1000 iteration permutation test for each trait on each linkage map. QTL visualizations were created in R using custom code.

4.4 Results

The *Salix* F_1 hybrid common parent mapping population was used in this study to map 41 traits, grouped into four main categories, leaf rust resistance, herbivory, leaf architecture, and yield components in additional to comparing the family means and determining which families have sexually dimorphic traits. While there were differences between the family means for every trait, sexual dimorphism was more varied within traits and families. Using the linkage maps previously introduced in Chapter 3, this study identified 87 QTL on 55 separate linkage groups across 16 linkage maps. These QTL accounted for an average of 26.9 percent of the total phenotypic variation with three QTL greater than 50 percent.

4.4.1 Leaf Rust

Leaf rust severity (RST, %) was surveyed in all eight families in 2017 and 2019. While overall mean RST was greater in 2019, there were significant differences between the mean RST of the families in both years (Appendix Table A.4). While the progeny means in 2017 were greater than either of the parents, the means from 2019

were closer to the means of the parents (Appendix Fig. A.4). The mean RST of the *S. purpurea* parent in 2019 was consistently greater than the other parent species with the exception of *S. suchowensis* P295, female parent of the 13X-440 family. Consistently, the 13X-440 family (*S. suchowensis* × *S. purpurea*) had the greatest mean RST (4.47% in 2017, 11.03% in 2019) with 13X-438 (*S. purpurea* × *S. koriyanagi*) a close second (3.11% and 10.64%), while the 11X-407 (*S. purpurea* × *S. viminalis*) was among the lowest in both years (0.22% and 4.08%). The RST for 12X-421 in 2017 and 2019 and for 11X-407 and 13X-443 (*S. suchowensis* × *S. purpurea*) families displayed significant sexual dimorphism in 2019. While the female F_1 progeny in the 13X-443 and 12X-421 families had greater mean RST, the male F_1 progeny of the 11X-407 family had significantly greater RST means.

Five QTL for RST were mapped in this study to CHR 16, 18, and 19 in five separate linkage maps (Fig. 4.2). One QTL each was mapped to CHR 16 and 18 of the 11X-407 94006 *S. purpurea* and 13X-358 04-BN-051 *S. udensis* linkage maps using the 2017 and 2019 surveys, respectively. Of the remaining three QTL on CHR 19, the 10X-400 P63 *S. suchowensis* QTL was from the 2017, while the 13X-440 94001 *S. purpurea* and the 13X-443 94001 *S. purpurea* were from the 2019 survey. The 10X-400 QTL range in physical position was 10.9 – 11 Mb, while the 13X-440 and 13X-443 QTL was 1.95 – 2.1 and 0.15 - 1.95 Mb (Table 4.1). Two of the QTL, one on P63 CHR 19 and another on 04-BN-051 CHR 18 explained 99.4 and 99.8 percent of the total variation in rust severity within their families. The remaining three QTL account for 30.6, 55.2, and 34.6 percent of variation within the 11X-407, 13X-440, and 13X-443 families. In three of the five QTL, the heterozygous genotype conferred increased

susceptibility to leaf rust, while in the two 11X-407 and 13X-443 QTL the heterozygous genotype conferred resistance (Appendix Fig. A.5).

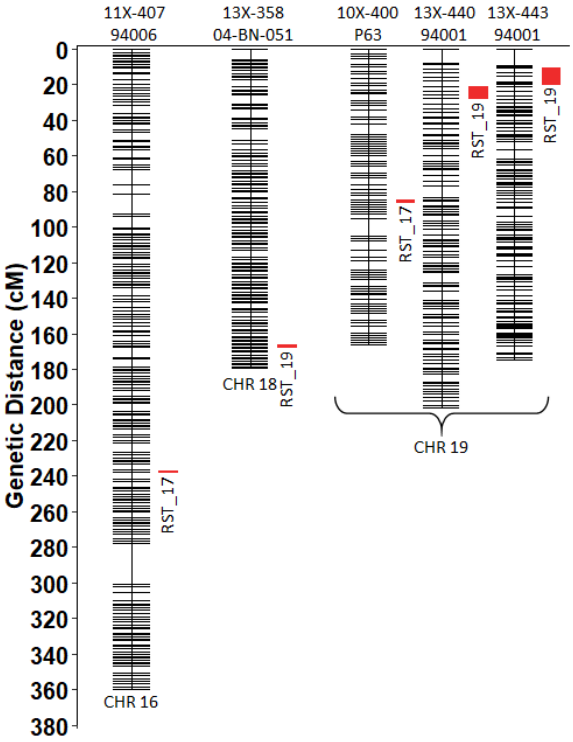


Figure 4.2: Linkage groups (LG) and QTL associated with leaf rust severity (RST). Only the LG with QTL are shown. The source linkage map is listed above, while the CHR based on alignment to the *S. purpurea* reference is listed below. Each QTL interval (red block) is positioned to the right of its LG and is labeled with its respective trait (RST) and year.

Table 4.1: Leaf Rust QTL from 2017 and 2019.

FAMILY	PARENT	SPECIES	TRAIT	CHR	MIN (Mb)	MAX (Mb)	PEAK (Mb)	PVE (%)
10X-400	P63	*S. such.*	RST_17	19	10.9	11.0	10.9	99.4
11X-407	94006	*S. purp.*	RST_17	16	21.2	21.2	21.2	30.6
13X-358	04-BN-051	*S. uden.*	RST_19	18	10.7	10.8	10.7	99.8
13X-440	94001	*S. purp.*	RST_19	19	2.0	2.1	2.0	55.2
13X-443	94001	*S. purp.*	RST_19	19	0.1	2.0	1.6	34.6

MIN and MAX: the confidence interval of the QTL; PEAK: position of the peak marker in physical distance per alignment to the *S. purpurea* reference; PVE: Percentage of Variation Explained by the QTL.

4.4.2 Leaf Architecture

Traits in leaf architecture describe the size and shape of the leaves within each family and were collected in 2017 and 2018. Leaf length (LL, cm), width (LW, cm), area (LA, cm^2), perimeter (LP, cm), factor (LFF) and ratio (LFR) were determined using a CID CI-203 laser leaf area meter, while dry leaf weight (DLW, grams) was determined gravimetrically, and specific leaf area (SLA, grams/cm^2) was derived from leaf area and leaf weight. All eight traits showed significant differences between the family means for both years (Appendix Table A.4). In terms of leaf area, the 10X-400 (*S. purpurea* × *S. suchowensis*) (19.53 cm^2), 11X-407 (19.79 cm^2), and 12X-421 (19.65 cm^2) families had the greatest average leaf area in 2017, while the 12X-421 (13.18 cm^2) and 13X-440 (12.93 cm^2) families were largest in 2019 (Appendix Fig. A.4). The 13X-358 (*S. purpurea* × *S. udensis*) (12.25 cm^2) and 13X-438 (7.84 cm^2) families had the smallest average leaf area in 2017 and 2018, respectively.

These traits displayed sexual dimorphism in at least one year and one family. DLW and LA were only dimorphic among the 12X-421 family in 2017 and 13X-443 family in 2018, while SLA was dimorphic among the 11X-407 family in 2017 and the 13X-358 and 13X-438 families in 2018. Within four of the eight families, 10X-400, 12X-421, 13X-440, and 13X-443, LW was dimorphic in both years. Both 2017 LL and LP were dimorphic among the 13X-358 family, while just LP was dimorphic with the 12X-421 family. LFF and LFR were the most prolific dimorphic leaf architecture traits. LFR was dimorphic in all families except with 12X-421, 13X-438, and 13X-426 families while LFF was lacking in the 13X-358 and 13X-438 families.

QTL mapping of leaf architecture traits detected 30 QTL on 21 linkage groups from 12 linkage maps (Fig. 4.3). DLW was the only leaf architecture trait without a significant QTL in this study. There were seven QTL for LL in 2017 and four in 2018, mapping to CHR 2, 3, 11, 12, 16, 17, 18, and 19 depending on the family (Table 4.2). Four LA QTL were identified on CHR 11, 12, and 18, while SLA QTL were detected on CHR 7, 9, and 16. LP QTL were found on CHR 11, 12, 13, 17 and 19, while LW QTL were found on CHR 1, 11 and 12. Aside from the QTL hotspots on CHR 11 and 12 in specific families, most were distributed across multiple CHR. The 30 QTL associated with leaf architecture had an average percent variation explained (PVE) of 23.3 percent, ranging from 14.8 to 32.6 percent (Appendix Fig. A.5).

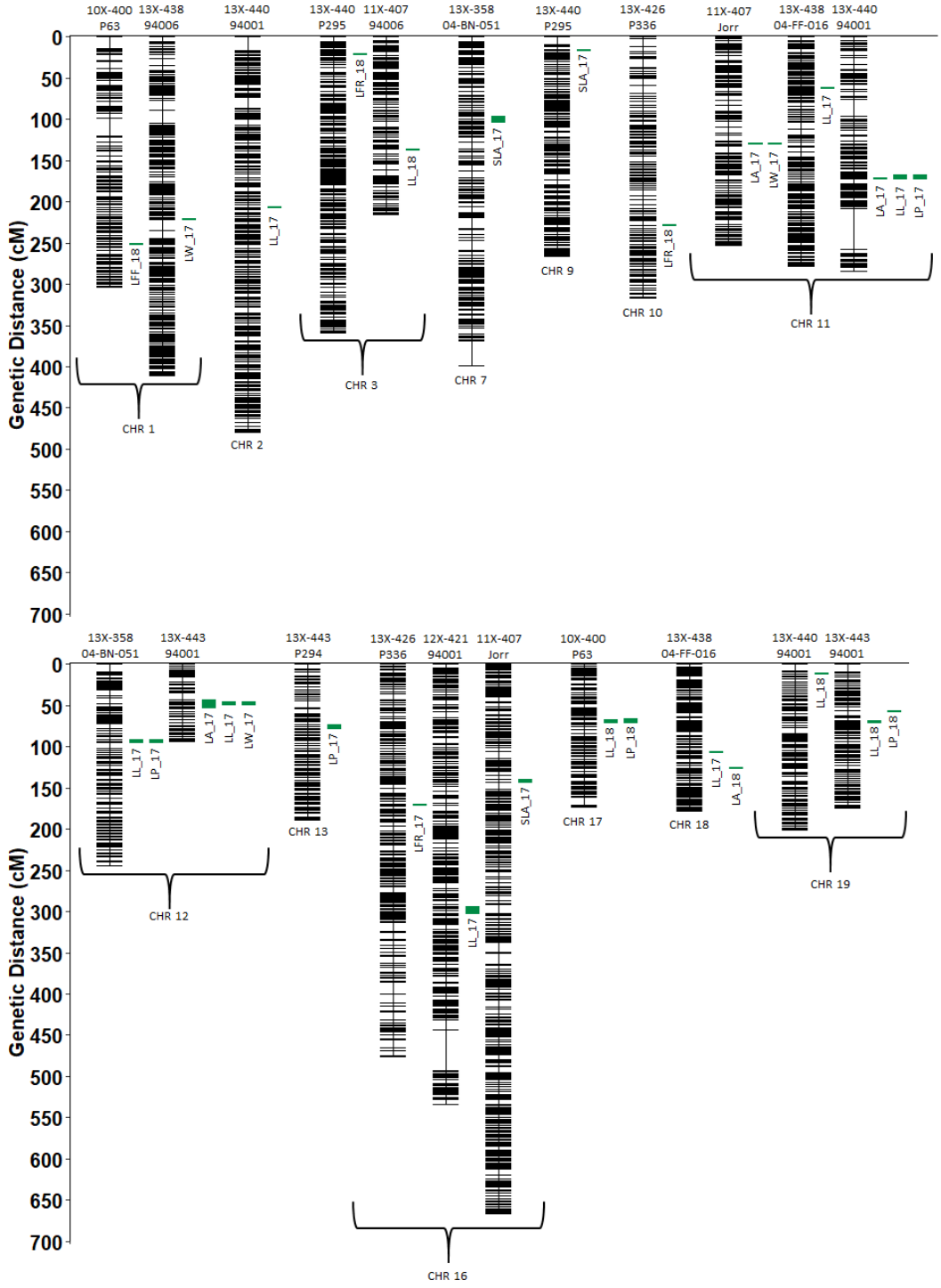


Figure 4.3: Linkage groups (LG) and QTL associated with leaf architecture. Only the LG with QTL are shown. The source linkage map is listed above, while the CHR based on alignment to the *S. purpurea* reference is listed below. Each QTL (green block) is positioned to the right of its LG and is labeled with its respective trait and year.

Table 4.2: Leaf architecture QTL from 2017 and 2018.

FAMILY	PARENT	SPECIES	TRAIT	CHR	MIN (Mb)	MAX (Mb)	PEAK (Mb)	PVE (%)
10X-400	P63	*S. such*	LFF_18	1	13.2	13.2	13.2	29.1
10X-400	P63	*S. such*	LL_18	17	5.0	5.3	5.3	27.0
10X-400	P63	*S. such*	LP_18	17	5.0	5.3	5.3	28.3
11X-407	94006	*S. purp*	LL_18	3	9.5	9.5	9.5	23.3
11X-407	Jorr	*S. vimin*	LA_17	11	5.9	5.9	5.9	26.1
11X-407	Jorr	*S. vimin*	LW_17	11	5.9	5.9	5.9	25.7
11X-407	Jorr	*S. vimin*	SLA_17	16	12.1	12.1	12.1	26.4
12X-421	94001	*S. purp*	LL_17	16	20.7	21.2	21.1	32.6
13X-358	04-BN-051	*S. uden*	SLA_17	7	0.9	0.9	0.9	26.3
13X-358	04-BN-051	*S. uden*	LL_17	12	3.3	3.6	3.5	28.3
13X-358	04-BN-051	*S. uden*	LP_17	12	3.3	3.6	3.5	30.2
13X-426	P336	*S. integ*	LFR_18	10	11.3	11.3	11.3	17.9
13X-426	P336	*S. integ*	LFR_17	16	13.4	13.5	13.5	14.8
13X-438	94006	*S. purp*	LW_17	1	5.5	5.5	5.5	16.8
13X-438	04-FF-016	*S. kori*	LL_17	11	3.4	3.4	3.4	16.3
13X-438	04-FF-016	*S. kori*	LA_18	18	8.6	8.6	8.6	16.3
13X-438	04-FF-016	*S. kori*	LL_17	18	7.9	7.9	7.9	17.7
13X-440	P295	*S. such*	LFR_18	3	1.2	1.3	1.2	26.3
13X-440	P295	*S. such*	SLA_17	9	2.7	2.7	2.7	25.2
13X-440	94001	*S. purp*	LL_17	2	9.6	9.6	9.6	26.1
13X-440	94001	*S. purp*	LA_17	11	6.5	6.1	6.5	23.8
13X-440	94001	*S. purp*	LL_17	11	5.9	6.1	5.9	28.7
13X-440	94001	*S. purp*	LP_17	11	5.9	6.1	6.5	24.9
13X-440	94001	*S. purp*	LL_18	19	0.3	0.3	0.3	29.0
13X-443	P294	*S. such*	LP_17	13	2.0	2.2	2.0	18.2
13X-443	94001	*S. purp*	LA_17	12	2.5	2.7	2.7	21.4
13X-443	94001	*S. purp*	LL_17	12	2.6	2.7	2.7	17.7
13X-443	94001	*S. purp*	LW_17	12	2.6	2.7	2.7	18.3
13X-443	94001	*S. purp*	LL_18	19	5.7	7.0	7.0	18.6
13X-443	94001	*S. purp*	LP_18	19	4.0	4.0	4.0	19.1

MIN and MAX: the confidence interval of the QTL; PEAK: position of the peak marker in physical distance per alignment to the *S. purpurea* reference; PVE: Percentage of Variation Explained by the QTL.

4.4.3 Herbivory

All traits in herbivory were phenotyped based on percent severity of damage caused by WLB and PLH. In the case of PLH, ratings were divided into three distinct types of damage, leaf curl (PLHC), leaf necrosis (PLHN) and shoot tip death (PLHS), designed to capture precise variation within a plant's response. WLB ratings were collected in 2017, 2018, and 2019, while PLH ratings were conducted in 2017 and 2019. For both insect damage ratings, there were significant differences among family means in all years (Appendix Table A.4). Throughout the three years of WLB ratings, the 13X-358 (7.28, 4.53, and 6.07%), 12X-421 (5.12, 4.69, and 4.99%), and 11X-407 (4.3, 5.42, and 7.14%) families had the greatest mean severity, while the 13X-438 (0.02, 0.83, and 2.62%) and 13X-426 (0, 0.68, 4.2%) families typically had the lowest (Appendix Fig. A.4). The PLH ratings had similar trends. The 13X-358, 11X-407, and 12X-421 families had the greatest mean severity for both years for all three damage phenotypes. The 13X-438 family had the most consistently low PLH ratings of the eight families, occasionally matched by the 13X-443 and 10X-400 families.

Five of the eight families were sexually dimorphic for either WLB or PLH severity ratings. The 10X-400 and 11X-407 families were dimorphic for WLB severity in 2017 and 2019, respectively. The 11X-407 family was the only family to be dimorphic for PLHS, while family 13X-440 was dimorphic for PLHN. Family 13X-440 was dimorphic for PLHC in both years; however, families 13X-438 and 13X-443 were only dimorphic in 2017. QTL analysis of herbivory identified 20 QTL on 19 linkage groups from 14 linkage maps (Fig. 4.4). There were two, one, and four QTL associated with the 2017, 2018, and 2019 WLB surveys. The two 2017 QTL

mapped to CHR 5 from the 13X-440 94001 *S. purpurea* map and CHR 18 in the 12X-421 07-MBG-5027 *S. viminalis* map (Table 4.3). The lone QTL from 2018 mapped to CHR 4 from 13X-443 *S. suchowensis* map. Of the four QTL from 2019, three were from the 13X-440 family. Two QTL from *S. purpurea* 94001 mapped to CHR 6 and 9, and the last from *S. suchowensis* P295 mapped to CHR 9. The remaining WLB 2019 QTL was also mapped to CHR 9, but in 13X-443 94001 *S. purpurea*.

Table 4.3: Herbivory QTL from 2017 - 2019.

FAMILY	PARENT	SPECIES	TRAIT	CHR	MIN (PD)	MAX (PD	PEAK (PD)	PVE (%)
10X-400	94006	*S. purp*	PLHS_17	18	1.1	1.1	1.1	28.7
10X-400	P63	*S. such*	PLHS_19	18	3.3	3.7	3.6	31.1
11X-407	94006	*S. purp*	PLHN_17	4	12.3	12.3	12.3	24.8
11X-407	Jorr	*S. vimin*	PLHC_17	8	6.5	6.9	6.5	25.7
12X-421	07-MBG-5027	*S. vimin*	WLB_17	18	11.5	11.9	11.5	24.1
13X-358	94006	*S. purp*	PLHN_17	18	1.8	1.8	1.8	27.0
13X-358	94006	*S. purp*	PLHS_17	13	0.3	0.4	0.3	24.6
13X-358	04-BN-051	*S. uden*	PLHC_17	15	15.0	15.1	15.0	26.0
13X-358	04-BN-051	*S. uden*	PLHN_17	5	17.1	17.1	17.1	24.3
13X-426	P336	*S. integ*	PLHC_19	11	3.5	3.5	3.5	16.2
13X-426	94001	*S. purp*	PLHC_19	9	3.6	3.6	3.6	15.9
13X-426	94001	*S. purp*	PLHS_19	11	4.0	4.0	4.0	22.4
13X-438	04-FF-016	*S. kori*	PLHC_17	4	13.2	13.4	13.4	20.0
13X-438	04-FF-016	*S. kori*	PLHN_17	4	14.0	14.1	14.1	27.2
13X-440	P295	*S. such*	WLB_19	9	6.2	6.5	6.3	32.0
13X-440	94001	*S. purp*	WLB_17	5	14.3	14.3	14.3	25.7
13X-440	94001	*S. purp*	WLB_19	6	19.2	19.2	19.2	24.5
13X-440	94001	*S. purp*	WLB_19	9	6.4	6.8	6.6	30.8
13X-443	P294	*S. such*	WLB_18	4	3.2	3.9	3.7	29.4
13X-443	94001	*S. purp*	WLB_19	9	6.3	6.6	6.6	20.6

MIN and MAX: the confidence interval of the QTL; PEAK: position of the peak marker in physical distance per alignment to the *S. purpurea* reference; PVE: Percentage of Variation Explained by the QTL.

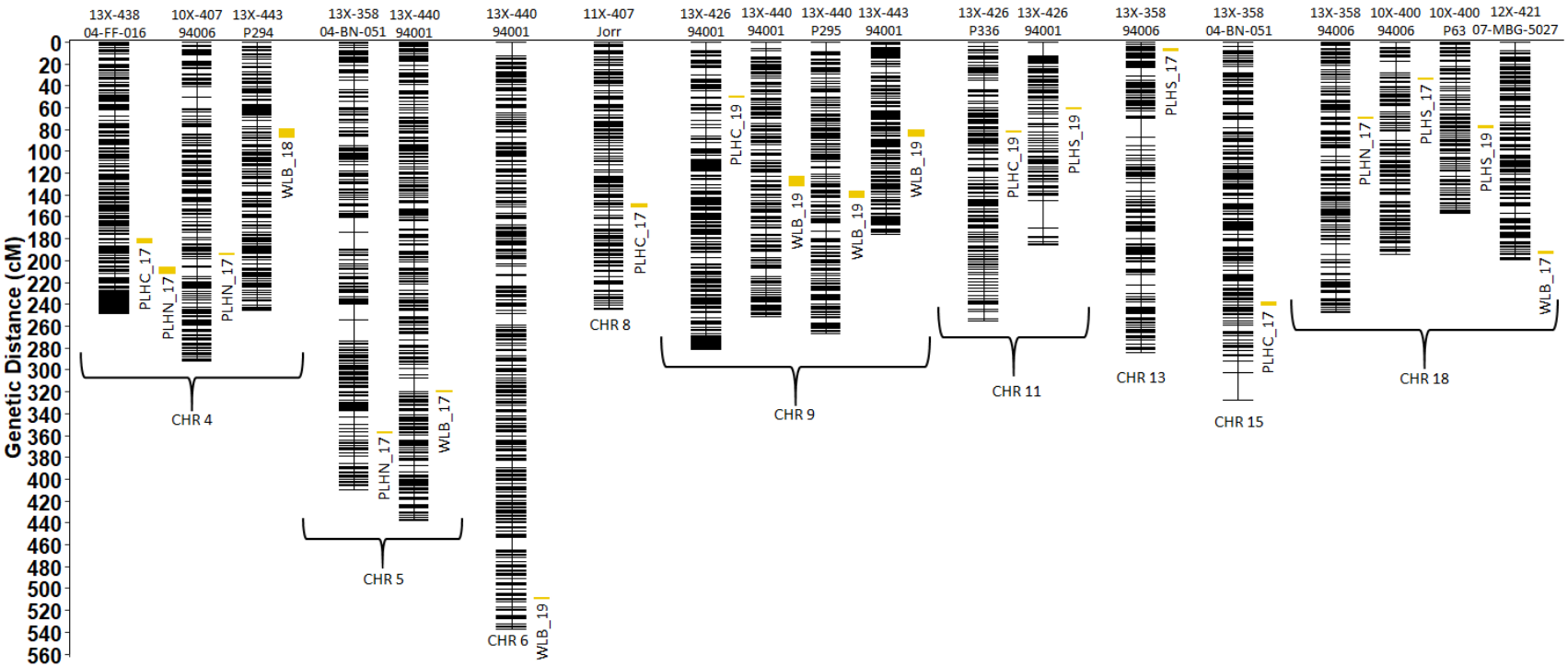


Figure 4.4: Linkage groups (LG) and QTL associated with herbivory. Only the LG with QTL are shown. The source linkage map is listed above, while the CHR based on alignment to the *S. purpurea* reference is listed below. Each QTL (yellow block) is positioned to the right of its LG and is labeled with its respective trait and year.

There were five QTL associated with PLHC, three from 2017 and two from 2019, all mapping to different CHR. The three from 2017 mapped to CHR 4, 8, and 15 on the 13X-438 04-FF-016 *S. koriyanagi*, 11X-407 Jorr *S. viminalis*, and 13X-358 04-BN-051 *S. udensis* maps. The PLHC QTL from 2019 both mapped to the 13X-426 family, 94001 *S. purpurea* CHR 9 and P336 *S. integra* CHR 11. PLHN surveyed in 2017 identified four QTL. Two of them were mapped in the 13X-358 family on 04-BN-051 *S. udensis* CHR 5 and 94006 *S. purpurea* CHR 18. The remaining two PLHN QTL were both mapped to CHR 4, but on separate maps, 11X-407 94006 *S. purpurea* and 13X-438 04-FF-016 *S. koriyanagi*. There were two QTL each for PLHS in 2017 and 2019. The two QTL from 2017 were both mapped to 94006 *S. purpurea* in the 13X-358 and 10X-400 families. The two QTL from 2019 however were mapped to 10X-400 P63 *S. suchowensis* CHR 18 and 13X-426 94001 *S. purpurea* CHR 11. PVE for the 20 herbivory QTL average to 25.1 percent, ranging from 15.9 to 32 percent.

4.4.4 Yield Components

All yield component traits other than SPAD measurements were collected on dormant plants during the 2017-2018 (17) and 2018-2019 (18) winters. Yield component traits included AVGDIAM, STEMCT, HT, PLTAREA, and PLTVOL as a proxy for biomass yield. SPAD measurements were collected in August (SPAD_818) and October (SPAD_1018) of 2018 and August in 2019 (SPAD_819). For every trait, there were significant differences between family means (Appendix Table A.4). In both years, the 13X-438 family grew to the tallest height (292.9 and 417.14 cm), while the 13X-358 (143.4 and 220.4 cm) and 12X-421 (135.1 and 215.3 cm) families were

the shortest (Appendix Fig. A.4). Following one year of coppice regrowth (2017), the 10X-400 family had the greatest estimated PLTVOL (1427.5 cm^3) followed by the 13X-443 (1117.3 cm^3), 13X-438 (1081.5 cm^3), 13X-426 (1080.5 cm^3) families. The 13X-440 (839.3 cm^3), 11X-407 (521.1 cm^3), 12X-421 (420 cm^3), and 13X-358 (360.5 cm^3) families comprised the bottom four families. After another year of growth, the 13X-443 (3565.9 cm^3) family was the top estimated biomass producer followed by the 10X-400 (3142.9 cm^3) family. The 13X-440 family had the greatest mean SPAD measurements during all collection times with comparable means for family 10X-400 in October 2018. Consistently among the lowest mean SPAD measurements, the 13X-426 family was often comparable to the 12X-421 and 13X-438 families.

Six of the eight families in this study were sexually dimorphic for at least one yield component trait. Among these, STEMCT and AVGDIAM were the most common, both years of AVGDIAM were dimorphic among the 11X-407, 13X-358, and the 13X-443 families. While STEMCT was only dimorphic among the 13X-440s in 2017, it was dimorphic in both years among the 10X-400 and 13X-443 families. PLTAREA was dimorphic in the 13X-440 family in 2017, while both PLTAREA and PLTVOL were dimorphic in 2018 in the 13X-358 family. HT was dimorphic in both years among the 13X-358 family. However, only in 2017 among the 11X-407. Finally, August SPAD collected in 2018 was dimorphic among the 13X-358 and, with the addition of August 2019, 11X-407 families.

I identified 32 QTL among these traits, located on 20 separate linkage groups from 14 linkage maps (Fig. 4.5). QTL for AVGDIAM in 2017 and 2018, HT and PLTAREA in 2018 and PLTVOL in 2017 mapped to identical regions on CHR 19 in

two 94001 *S. purpurea* linkage maps constructed from the 13X-440 and 13X-443 families (Table 4.4). Seven SPAD QTL, one from August 2018, four from October 2018 and two from August 2019, all mapped to different CHR from different families. PVE within the 32 QTL for yield component traits averaged 25.7 percent, ranging from 15.6 to 39.4 percent of the total phenotype variation (Appendix Fig. A.5).

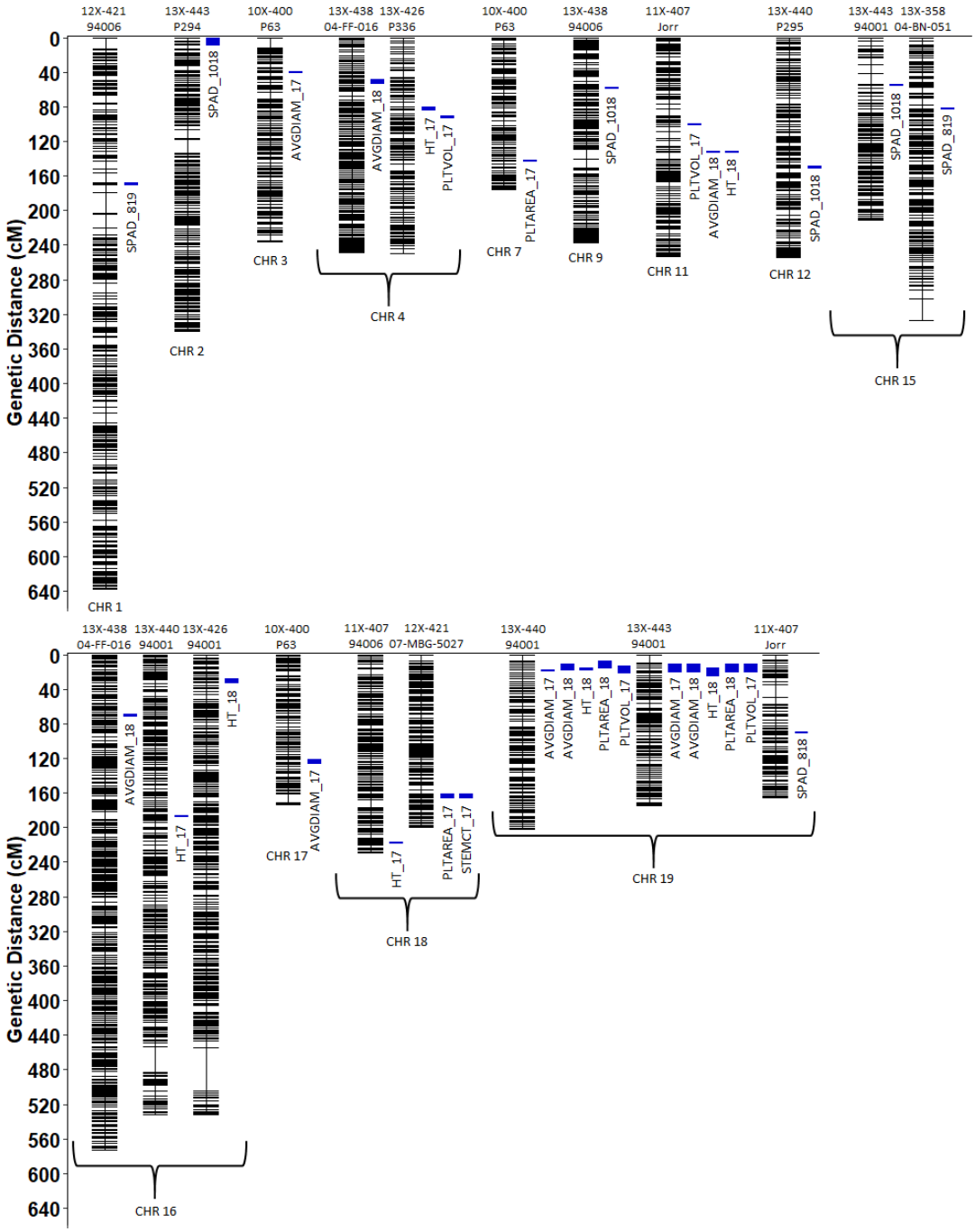


Figure 4.5: Linkage groups (LG) and QTL associated with yield components. Only the LG with QTL are shown. The source linkage map is listed above, while the CHR based on alignment to the *S. purpurea* reference is listed below. Each QTL (blue blocks) is positioned to the right of its LG and is labeled with its respective trait and year.

Table 4.4: Yield Components QTL from 2017 and 2018.

FAM	PAR	SPECIES	TRAIT	CHR	MIN (Mb)	MAX (Mb)	PEAK (Mb)	PVE (%)
10X-400	P63	*S. such*	AVGDIAM_17	3	4.3	4.3	4.3	26.6
10X-400	P63	*S. such*	PLTAREA_17	7	10.7	10.7	10.7	28.3
10X-400	P63	*S. such*	AVGDIAM_17	17	10.1	10.3	10.1	26.6
11X-407	94006	*S. purp*	HT_17	18	10.7	10.7	10.7	28.9
11X-407	Jorr	*S. vimin*	AVGDIAM_18	11	5.9	7.1	5.9	25.3
11X-407	Jorr	*S. vimin*	HT_18	11	7.1	7.1	7.1	24.9
11X-407	Jorr	*S. vimin*	PLTVOL_17	11	4.4	4.4	4.4	27.6
11X-407	Jorr	*S. vimin*	SPAD_818	19	5.4	5.2	5.2	25.7
12X-421	07-MBG-5027	*S. vimin*	PLTAREA_17	18	10.0	9.5	10.1	25.3
12X-421	07-MBG-5027	*S. vimin*	STEMCT_17	18	10.0	9.5	10.1	26.1
12X-421	94001	*S. purp*	SPAD_819	1	2.8	2.8	2.7	29.4
13X-358	04-BN-051	*S. uden*	SPAD_819	15	4.6	4.6	4.6	21.8
13X-426	P336	*S. integ*	HT_17	4	4.9	5.1	4.9	18.1
13X-426	P336	*S. integ*	PLTVOL_17	4	5.5	5.6	5.6	15.6
13X-426	94001	*S. purp*	HT_18	16	2.3	2.6	2.3	24.2
13X-438	94006	*S. purp*	SPAD_1018	9	4.5	4.5	4.5	16.8
13X-438	04-FF-016	*S. kori*	AVGDIAM_18	4	2.6	2.9	2.6	17.2
13X-438	04-FF-016	*S. kori*	AVGDIAM_18	16	4.5	4.5	4.5	16.3
13X-440	P295	*S. such*	SPAD_1018	12	7.8	7.8	7.8	28.5
13X-440	94001	*S. purp*	HT_17	16	14.3	14.3	14.3	24.5
13X-440	94001	*S. purp*	AVGDIAM_17	19	1.7	1.7	1.7	32.6
13X-440	94001	*S. purp*	AVGDIAM_18	19	0.3	1.7	1.6	24.9
13X-440	94001	*S. purp*	HT_18	19	1.6	1.7	1.6	30.1
13X-440	94001	*S. purp*	PLTAREA_18	19	0.1	1.6	1.6	27.2
13X-440	94001	*S. purp*	PLTVOL_18	19	1.6	2.0	1.7	39.4
13X-443	P294	*S. such*	SPAD_1018	2	0.0	0.3	0.1	21.0
13X-443	94001	*S. purp*	SPAD_1018	15	1.2	1.3	1.2	24.9
13X-443	94001	*S. purp*	AVGDIAM_17	19	0.1	2.0	1.6	31.8
13X-443	94001	*S. purp*	AVGDIAM_18	19	0.1	2.0	1.6	31.8
13X-443	94001	*S. purp*	HT_18	19	1.6	2.0	2.0	29.3
13X-443	94001	*S. purp*	PLTAREA_18	19	0.1	2.0	1.6	32.3
13X-443	94001	*S. purp*	PLTVOL_18	19	0.1	2.0	1.6	18.6

MBG: 07-MBG-5027; MIN and MAX: the confidence interval of the QTL; PEAK: position of the peak marker in physical distance per alignment to the *S. purpurea* reference; PVE: Percentage of Variation Explained by the QTL.

4.5 Discussion

Shrub willow (*Salix* spp.) breeding for biomass bioenergy in the northeast United States has relied heavily on unpredictable hybrid vigor resulting from species

hybridization. In order to precisely map the genes conferring trait advantages and expand the pool of diverse alleles available for introgression to breed better performing cultivars, female and male *S. purpurea* were crossed to five other *Salix* species. The resulting eight F_1 hybrid families were subjected to extensive phenotyping during multiple years in the field. Using these data, we performed mean separations, tested for sexual dimorphism, and mapped QTL using existing linkage maps. This study represents the largest *Salix* QTL study to date, mapping 87 separate QTL for phenotypic traits, and makes significant progress towards identifying alleles that can be introgressed via marker-assisted selection in future studies.

Among the species included here, *S. purpurea* and *S. viminalis* are the only ones with a prior history in QTL mapping studies therefore all QTL detected within *S. koriyanagi*, *S. suchowensis*, *S. integra*, and *S. udensis* were novel. Forty-two out of 87 QTL were found on *S. purpurea* linkage maps, nine on 94006 and 33 on 94001. Carlson et al. (2019) phenotyped a similar set of traits within an F_2 mapping population using 94006 and 94001 as grandparents and a *S. purpurea* association panel comprised of North American accessions. While the GWAS was underpowered, they identified 99 significant marker associations, many of which were tied to biomass traits spread throughout the genome, while their QTL study placed these traits on CHR 4, 5, 6, and 10 in QTL hotspots. In mapping leaf rust severity, their study detected QTL on CHR 1, 5, and 10 compared to mine on CHR 16 and 19. While this study's *S. purpurea* QTL share no overlap with their results, these QTL are visible based on the genetic differences between 94006, 94001, and the other species. In QTL mapping with backcross markers, significant QTL are associated with only the informative

(heterozygous) parent whereas when using intercross markers, both parents are informative. This allows additive and dominance affects to be estimated for a given trait and highlights different sets of alleles influencing the final phenotype. Combining their results with mine showed that analysis of *S. purpurea* can offer insight into traits genetics that can accelerate breeding in North America and in other parts of the world.

QTL studies including *S. viminalis* are more prolific than those in *S. purpurea*, although commonality between traits measured in those studies and those measured here are limited. Crossing *S. viminalis* to *S. schwerinii*, a known source of resistance to European willow leaf rust (*Melampsora larici-epitea*), resulted in mapping rust resistance QTL on different CHR, with consistent QTL mapping to CHR 1 (Hanley et al., 2011; Ronnberg-Wastljung et al., 2008; Samils et al., 2011; Sulima et al., 2017). Although I was not able to map significant QTL in either *S. viminalis* parent for leaf rust resistance in this study, I was able to map a QTL on CHR 8 for PLHC, a potential susceptibility locus. *Salix viminalis* individuals have a remarkable susceptibility to PLH, widespread across North America, that is thought to be caused by PLH-induced cell wall modifications (Wang et al., 2020). It is likely that this susceptibility to PLH is the cause of the limited number of QTL, severely affecting the ability to measure other phenotypes to the same standards as the other families.

Sexual dimorphism, referring to sex-specific phenotypic differences (Poissant et al., 2010) in dioecious plants is common (Ågren et al., 1999), and is frequently documented in *Salix* species. Tissue specific sexual differences in gene expression not directly associated with the sex determination region on CHR 15 in *S. purpurea* (Carlson et al., 2017) provide a potential upstream cause for the litany of dimorphic

traits described by Gouker et al. (2020). Of the species studied here, *S. viminalis* has also been found to be dimorphic, with males possessing improved tolerance to cadmium lending to increased ability over females for phytoremediation (Zou et al., 2021). *Salix udensis* (syn. *S. sachalinensis*), was found to be dimorphic for the number of reproductive shoots (Ueno et al., 2006), but not for plant height and stem diameter (Ueno et al., 2007). All eight of the species hybrid families included here were sexually dimorphic for at least one trait, ranging from the three in the 13X-438 to the 13 in the 11X-407 family. Comparatively, the previously cited works focus on single species families, while I report data for species hybrids. It is unclear how one species prone to sexual dimorphism interacts with another in a species hybrid. Although the genetic underpinnings of sexual dimorphism remain unclear, *Salix* offers remarkable potential for future studies in this area.

This study mapped QTL and described the occurrence of sexual dimorphism within the eight families comprising the *Salix* F_1 hybrid common parent mapping population. Across the 16 parental backcross linkage maps, I mapped 87 separate QTL onto 55 different linkage groups. Identifying sexually dimorphic traits within every family, this study also contributed to the burgeoning research focused on this phenomenon. An important beginning step in trait introgression, this research will help focus future works into QTL refinement and identification of candidate markers for marker-assisted selection with the intent of producing superior shrub willow cultivars.

4.6 REFERENCES

Ågren, J., Danell, K., Elmqvist, T., Ericson, L., & Hjältén, J. (1999). Sexual dimorphism and biotic interactions. In M. A. Geber, T. E. Dawson, & L. F. Delph (Eds.), *Gender and Sexual Dimorphism in Flowering Plants* (pp. 217-246). Berlin, Heidelberg: Springer Berlin Heidelberg.

Anderson, E., & Hubricht, L. (1938). Hybridization in tradescantia III. The evidence for introgressive hybridization. *Am. J. Bot.*, *25*(6), 396-402. doi:https://doi.org/10.1002/j.1537-2197.1938.tb09237.x

Carlson, C. H., Choi, Y., Chan, A. P., Serapiglia, M. J., Town, C. D., & Smart, L. B. (2017). Dominance and sexual dimorphism pervade the *Salix purpurea* L. transcriptome. *Genome Biol. Evol.*, *9*(9), 2377-2394. doi:10.1093/gbe/evx174

Carlson, C. H., Gouker, F. E., Crowell, C. R., Evans, L., DiFazio, S. P., Smart, C. D., & Smart, L. B. (2019). Joint linkage and association mapping of complex traits in shrub willow (*Salix purpurea* L.). *Ann. Bot.*, *124*(4), 701-716. doi:10.1093/aob/mcz047

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos T R Soc B*, *363*(1491), 557-572. doi:doi:10.1098/rstb.2007.2170

Gouker, F. E., Carlson, C. H., Zou, J., Evans, L., Crowell, C. R., Smart, C. D., DiFazio, S. P., & Smart, L. B. (2020). Sexual dimorphism and sex ratio bias in the dioecious willow *Salix purpurea* L. *bioRxiv*, 2020.2004.2005.026427. doi:10.1101/2020.04.05.026427

Hallingbäck, H. R., Berlin, S., Nordh, N.-E., Weih, M., & Rönnberg-Wästljung, A.-C. (2019). Genome wide associations of growth, phenology, and plasticity traits in willow [*Salix viminalis* (L.)]. *Front. Plant Sci., 10*(753). doi:10.3389/fpls.2019.00753

Hallingbäck, H. R., Fogelqvist, J., Powers, S. J., Turrion-Gomez, J., Rossiter, R., Amey, J., Martin, T., Weih, M., Gyllenstrand, N., Karp, A., Lagercrantz, U., Hanley, S. J., Berlin, S., & Rönnberg-Wästljung, A.-C. (2016). Association mapping in *Salix viminalis* L. (Salicaceae) – identification of candidate genes associated with growth and phenology. *GCB Bioenergy, 8*(3), 670-685. doi:https://doi.org/10.1111/gcbb.12280

Hanley, S. J., Pei, M. H., Powers, S. J., Ruiz, C., Mallott, M. D., Barker, J. H. A., & Karp, A. (2011). Genetic mapping of rust resistance loci in biomass willow. *Tree Genet. Genomes, 7*(3), 597-608. doi:10.1007/s11295-010-0359-x

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *J. Hered, 105*(S1), 795-809. doi:10.1093/jhered/esu033

Hernandez, J., Meints, B., & Hayes, P. (2020). Introgression breeding in barley: perspectives and case studies. *Front. Plant Sci., 11*(761). doi:10.3389/fpls.2020.00761

McCracken, A. R., Dawson, W. M., & Bowden, G. (2001). Yield responses of willow (*Salix*) grown in mixtures in short rotation coppice (SRC). *Biomass Bioenerg, 21*(5), 311-319. doi:Doi 10.1016/S0961-9534(01)00046-0

Pacaldo, R. S., Volk, T. A., & Briggs, R. D. (2014). Carbon sequestration in fine roots and foliage biomass offsets soil CO_2 effluxes along a 19-year chronosequence

of shrub willow (*Salix x dasyclados*) biomass crops. *Bioenergy Res., 7*(3), 769-776. doi:10.1007/s12155-014-9416-x

Poissant, J., Wilson, A. J., & Coltman, D. W. (2010). Sex-specific genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. *Evol, 64*(1), 97-107. doi:https://doi.org/10.1111/j.1558-5646.2009.00793.x

Pratap, A., Das, A., Kumar, S., & Gupta, S. (2021). Current perspectives on introgression breeding in food legumes. *Front. Plant Sci., 11*(2118). doi:10.3389/fpls.2020.589189

Rönnberg-Wästljung, A. C., Glynn, C., & Weih, M. (2005). QTL analyses of drought tolerance and growth for a *Salix dasyclados* x *Salix viminalis* hybrid in contrasting water regimes. *Theor. Appl. Genet., 110*(3), 537-549. doi:10.1007/s00122-004-1866-7

Ronnberg-Wastljung, A. C., Samils, B., Tsarouhas, V., & Gullberg, U. (2008). Resistance to *Melampsora larici-epitea* leaf rust in *Salix*: analyses of quantitative trait loci. *J. Appl. Genet., 49*(4), 321-331. doi:10.1007/BF03195630

Samils, B., Rönnberg-Wästljung, A.-C., & Stenlid, J. (2011). QTL mapping of resistance to leaf rust in *Salix*. *Tree Genet. Genomes, 7*(6), 1219-1235. doi:10.1007/s11295-011-0408-0

Sanderson, B. J., DiFazio, S. P., Cronk, Q. C. B., Ma, T., & Olson, M. S. (2020). A targeted sequence capture array for phylogenetics and population genomics in

the Salicaceae. *Appl Plant Sci, 8*(10), e11394. doi:https://doi.org/10.1002/aps3.11394

Serapiglia, M. J., Gouker, F. E., & Smart, L. B. (2014). Early selection of novel triploid hybrids of shrub willow with improved biomass yield relative to diploids. *BMC Plant Biol, 14*, 74. doi:10.1186/1471-2229-14-74

Skvortsov, A. (1999). Willows of Russia and adjacent countries (Translated by N. Kadis, 1999). *Joensuu University, Finland*.

Smart, L. B., Volk, T., Lin, J., Kopp, R. F., Phillips, I. S., Cameron, K. D., White, E. H., & Abrahamson, L. (2005). Genetic improvement of shrub willow (*Salix* spp.) crops for bioenergy and environmental applications in the United States. *Unasylva, 56*, 51-55.

Sulima, P., Przyborowski, J. A., Kuszewska, A., Zaluski, D., Jedryczka, M., & Irzykowski, W. (2017). Identification of quantitative trait loci conditioning the main biomass yield components and resistance to *Melampsora* spp. in *Salix viminalis* x *Salix schwerinii* hybrids. *Int. J. Mol. Sci., 18*(3). doi:10.3390/ijms18030677

Ueno, N., Kanno, H., & Seiwa, K. (2006). Sexual differences in shoot and leaf dynamics in the dioecious tree *Salix sachalinensis*. *Botany, 84*(12), 1852-1859.

Ueno, N., Suyama, Y., & Seiwa, K. (2007). What makes the sex ratio female-biased in the dioecious tree *Salix sachalinensis*? *J Ecol, 95*(5), 951-959. doi:https://doi.org/10.1111/j.1365-2745.2007.01269.x

Volk, T. A., Heavey, J. P., & Eisenbies, M. H. (2016). Advances in shrub-willow crops for bioenergy, renewable products, and environmental benefits. *Food Energy Secur, 5*(2), 97-106. doi:https://doi.org/10.1002/fes3.82

Wagner, N. D., Gramlich, S., & Hörandl, E. (2018). RAD sequencing resolved phylogenetic relationships in European shrub willows (*Salix* L. subg. *Chamaetia* and subg. *Vetrix*) and revealed multiple evolution of dwarf shrubs. *Ecol Evol, 8*(16), 8243-8255. doi:https://doi.org/10.1002/ece3.4360

Wang, W., Carlson, C. H., Smart, L. B., & Carlson, J. E. (2020). Transcriptome analysis of contrasting resistance to herbivory by *Empoasca fabae* in two shrub willow species and their hybrid progeny. *PLoS One, 15*(7), e0236586. doi:10.1371/journal.pone.0236586

Wu, J., Nyman, T., Wang, D.-C., Argus, G. W., Yang, Y.-P., & Chen, J.-H. (2015). Phylogeny of *Salix* subgenus *Salix* s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. *BMC Evolutionary Biology, 15*(1), 31. doi:10.1186/s12862-015-0311-7

Zou, J., Zhang, Y., Li, X., Ma, X., Liu, J., Peng, X., & Sun, Z. (2021). Sexual differences in root growth and antioxidant characteristics in *Salix viminalis* exposed to cadmium stress. *Int. J. Phytoremediation*, 1-10. doi:10.1080/15226514.2021.1904825

CHAPTER 5

CONCLUSION

5.1 Introduction

The over-arching goal in this dissertation was to investigate the interaction between shrub willow species and willow leaf rust in the context of identifying sources of resistance that could be used for breeding. In this conclusion, I will discuss the objectives and main points from each chapter in the context of this goal. Additionally, I will conclude by proposing future studies that could build upon this work. By the end of this chapter, I hope to have positioned this research in how it can benefit those seeking to breed for resistance to *Melampsora* leaf rust.

5.2 Revisiting the Research Objectives

The three research chapters comprising this dissertation used two mapping populations involving *S. purpurea* individuals 94006, female and 94001, male. In addition to 94006 being the focus of the current reference genome, these individuals were chosen for their vertical growth habit, superior agronomic appeal, and variable resistance to willow leaf rust. They were crossed to form an F_1 then two F_1 full-siblings were crossed to form an F_2 for use in genetic mapping. A known source of QTL for leaf rust severity, the objective of Chapter 2 was to use a subset of this population to identify the temporal, gene-level response to infection and candidate genes associated with resistance to *M. americana*. Using two years of leaf rust ratings, resistant and susceptible genotypes to leaf rust were selected and inoculated in the

greenhouse alongside an uninoculated control. Although the study resulted in a list of candidate genes associated with the up-regulation of defense and down-regulation of photosynthesis in a coordinated response, some of the resistant genotypes had severity ratings similar to the susceptible genotypes, affecting my ability to resolve more subtle differences in gene expression due to the extraneous variation. Despite this, using multiple RNA-Seq analysis methods to highlight different characteristics of gene expression made it possible to identify a much larger list of candidate genes than though differential expression alone. A likely cause of the inconsistency within the resistant genotypes could have originated from the method used to select them. In the two years of field rust data, the 2017 ratings had limited penetrance compared to the 2015 ratings as the overall level of disease in 2017 was much lower than 2015 probably due to environmental conditions that were more favorable to disease development in 2015. Although the 2015 ratings were more heavily weighted in the selection, individual's resistant based on the 2017 ratings could have simply never come into contact with the pathogen leading to an incorrect identification. Future studies of this type should do controlled inoculations with the planned isolate on a larger subset of samples in same environment as the full experiment, then select within those for the verified resistant and susceptible genotypes before proceeding with the larger experiment.

For the third and fourth chapters, 94006 and 94001 were crossed to five other *Salix* species to form the *Salix* F_1 hybrid common parent mapping population. These eight families consisted of a balance of well-studied species, *S. purpurea* and *S. viminalis,* and lesser-known species such as *S. udensis* and *S. koriyanagi*. This

population was created with the mandate of representing the diversity within *Salix* in genomic studies. In chapter three, the objective was to describe the relatedness within the member species, create linkage maps, and map their SDR. Exemplifying the taxonomic difficulty present in *Salix,* the population analysis demonstrated that the individual identified as *S. alberti* was a misidentified *S. suchowensis*. Additionally, through the creation on backcross linkage maps, the SDR were mapped to large pericentromeric regions within the maternal parent's chromosome 15, indicating that as the female is the heterogametic sex, the member species used a ZW sex determination system with the exception of *S. integra*. Obviously, in a family that is completely female, mapping a QTL for sex is impossible. Adding to the complexity, however, was that the only stable piece of *S. integra's* chromosome 15 aligned outside of the LOD support intervals for the other family's SDR. Breaking down exactly what is happening with sex determination in *S. integra* is challenging. Given the variability in *Salix* for not only the genomic region of the SDR but also the system, it should be given priority from those interested in this area. P294 and *S. integra* both provide ample evidence for the need to be constantly adding diversity into breeding populations in *Salix*. With over 350 species, most of which readily hybridize, integrating diversity has both the potential to resolve taxonomic differences and to discover unique phenotypes that can further the understanding of evolutionarily important traits like sex determination.

The fourth chapter of this research sought to fulfill the objective of identifying QTL for leaf rust severity in addition to a multitude of agronomically important traits. In addition to mapping QTL for the 41 collected phenotypes, mean separations and

tests for sexual dimorphism provided a picture of the variation within and between the eight families that comprise the *Salix* F_1 hybrid common parent mapping population. While every trait showed significant mean separation between the families, sexually dimorphic traits were more varied. 87 QTL in total were mapped during the course of this study and every linkage map contained at least one. Despite identifying some QTL with a PVE over 50 percent, there was very little consistency within traits, species, or years. While this does not necessarily negate the QTL identified in this chapter, congruency across years or populations is one means of QTL validation. Despite this, several of these species have never before been included in a QTL mapping study, adding to the possibility for trait improvement.

5.3 Future Work

The combination of these three objectives demonstrated the variation available within *Salix* and how it can be characterized and employed for the breeding of improved resistance to willow leaf rust. Moving forward, potential next steps to fully capitalize on the work completed here centers on validation. For the candidate genes identified in chapter 2, qRT-PCR could be used to validate their differential expression between resistant and susceptible individuals although the results from RNASeq are well accepted as a valid reflection of expression. In chapter 3, further refinement of the SDR's is needed to identify the causal genes in each species. Lastly in chapter 4, working to validate the QTL through the generation of new mapping populations could demonstrate that some of the QTL are stable enough to be transferred through trait introgression.

APPENDIX

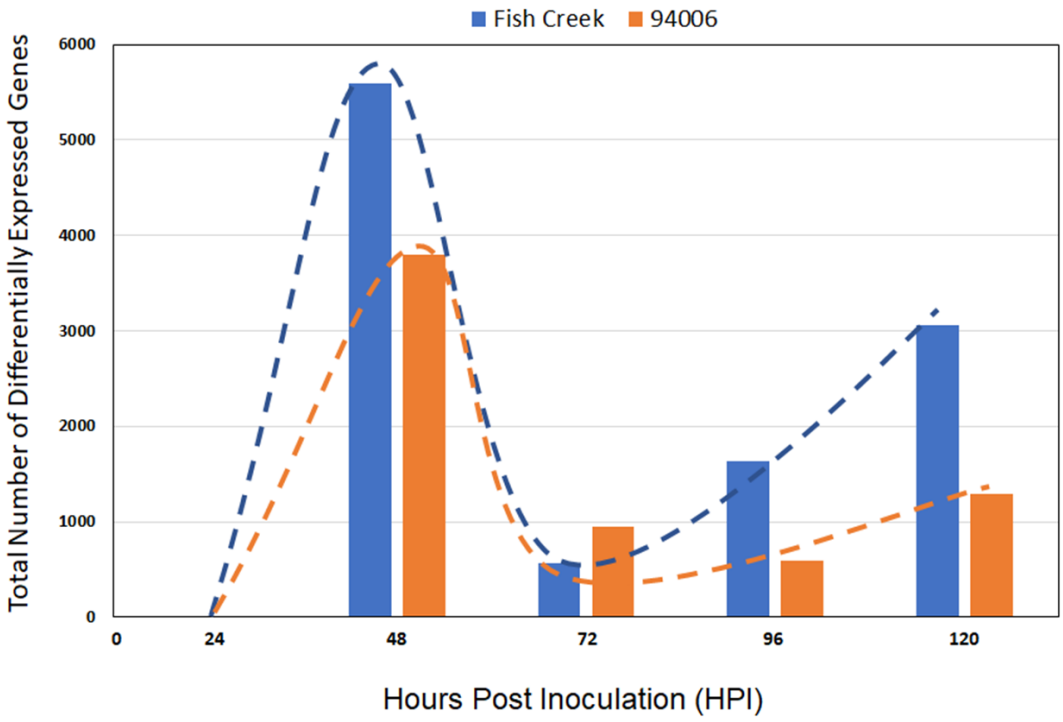


Figure A.1: Bar graph of total number of differentially expressed transcripts between inoculated and control treatments of Fish Creek (blue) and 94006 (orange) at 1-day intervals for 5 days. Dotted lines represent approximated trends of expression over duration of experiment. [Created by Chase Crowell]

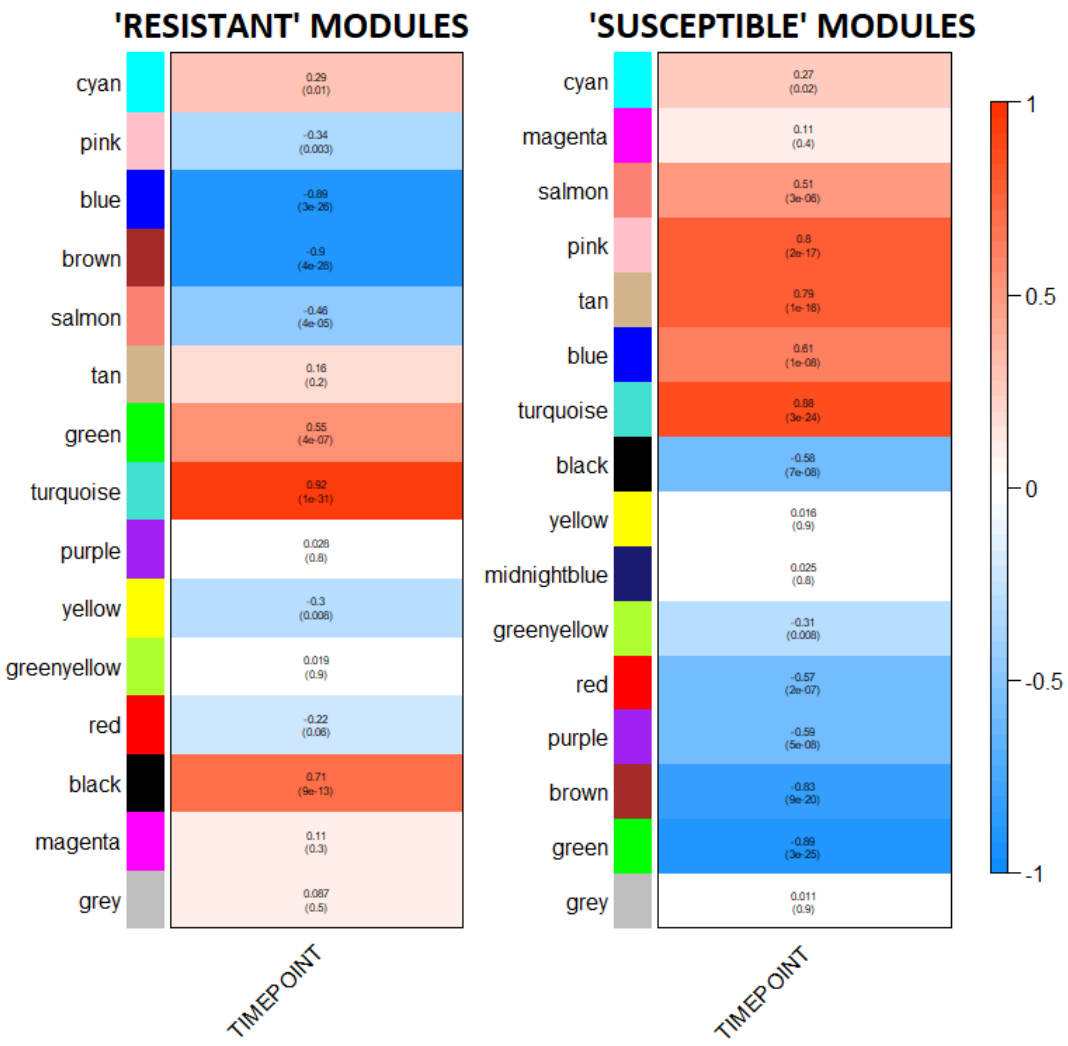


Figure A.2: Module eigengene correlations with time point as calculated in WGCNA. Time point was coded as 0, 2, 3. The modules from the resistant network are on the left while the susceptible network modules are on the right. Significance was determined at the 0.05 value. Positive correlations become deeper red while negative correlations become blue.

Table A.1: Candidate genes associated with the R-Turquoise module.

TIME POINT	TRANSCRIPT	RUST CORR	CORR PVAL	BEST HIT ARABI. NAME
42 HPI	Sapur.017G089700.1	0.448	0.001	AT5G14700.1
42 HPI	Sapur.005G090900.1	0.444	0.001	AT3G51240.1
42 HPI	Sapur.005G090800.1	0.385	0.005	AT3G51240.1
42 HPI	Sapur.017G110600.1	0.379	0.006	AT5G40990.1
42 HPI	Sapur.007G123400.1	-0.375	0.006	AT5G44210.1
42 HPI	Sapur.15WG126700.1	-0.364	0.008	AT5G61790.1
42 HPI	Sapur.019G061200.1	0.363	0.008	AT5G05270.1
42 HPI	Sapur.014G013100.1	0.354	0.010	AT3G16150.1
42 HPI	Sapur.019G112500.1	0.334	0.016	AT2G30490.1
42 HPI	Sapur.014G117800.1	0.331	0.017	AT5G13930.1
42 HPI	Sapur.010G161500.1	-0.328	0.018	AT5G02500.1
42 HPI	Sapur.014G102200.1	-0.327	0.018	AT4G02380.1
42 HPI	Sapur.006G106600.1	-0.322	0.020	AT4G05320.4
42 HPI	Sapur.001G071400.1	-0.321	0.020	AT5G42020.1
42 HPI	Sapur.010G100300.1	0.314	0.023	AT1G75290.1
42 HPI	Sapur.017G017500.2	-0.309	0.026	AT1G04260.1
42 HPI	Sapur.001G050500.1	0.305	0.028	AT2G33520.1
42 HPI	Sapur.018G074400.1	-0.283	0.042	AT4G30380.1
42 HPI	Sapur.009G070000.1	-0.228	0.104	AT1G27730.1
42 HPI	Sapur.010G040400.1	-0.198	0.160	
42 HPI	Sapur.006G165300.1	-0.194	0.169	AT3G11820.1
42 HPI	Sapur.010G049600.1	0.138	0.328	AT2G29420.1
42 HPI	Sapur.010G040200.1	-0.120	0.395	AT3G23240.1
42 HPI	Sapur.008G163800.1	-0.095	0.505	AT3G16510.1
42 HPI	Sapur.010G040800.1	0.093	0.510	AT2G38940.1
42 HPI	Sapur.004G162100.1	-0.070	0.623	AT5G60800.2
42 HPI	Sapur.010G039500.1	0.048	0.735	AT1G06620.1
42 HPI	Sapur.010G049400.1	0.039	0.785	AT2G29420.1
42 HPI	Sapur.018G112700.1	-0.013	0.929	AT1G54470.2
42 HPI	Sapur.010G037700.1	0.004	0.976	AT5G43580.1
66 HPI	Sapur.003G113500.1	-0.433	0.001	
66 HPI	Sapur.003G112800.1	-0.403	0.003	AT4G34131.1
66 HPI	Sapur.009G065100.1	-0.395	0.003	AT4G33720.1
66 HPI	Sapur.005G090900.1	0.393	0.003	AT3G51240.1
66 HPI	Sapur.005G090800.1	0.377	0.005	AT3G51240.1
66 HPI	Sapur.013G002700.1	-0.358	0.008	AT5G61190.1
66 HPI	Sapur.006G095600.1	0.358	0.008	AT2G36530.1
66 HPI	Sapur.016G287900.1	-0.329	0.015	AT4G21390.1
66 HPI	Sapur.003G133600.1	0.328	0.015	AT5G13930.1

Table A.1: Candidate genes associated with the R-Turquoise module.

TIME POINT	TRANSCRIPT	RUST CORR	CORR PVAL	BEST HIT ARABI. NAME
66 HPI	Sapur.014G060300.1	-0.327	0.016	AT2G45760.1
66 HPI	Sapur.001G062500.1	-0.327	0.016	
66 HPI	Sapur.15WG018900.1	-0.320	0.018	AT5G24090.1
66 HPI	Sapur.009G088600.1	-0.310	0.022	AT4G34320.1
66 HPI	Sapur.002G167000.1	-0.308	0.024	AT1G02450.1
66 HPi	Sapur.002G059400.1	-0.299	0.028	AT1G21270.1
66 HPI	Sapur.002G057900.1	-0.282	0.039	AT1G31130.1
66 HPI	Sapur.014G013100.1	0.281	0.040	AT3G16150.1
66 HPI	Sapur.007G075100.1	-0.280	0.040	AT5G64810.1
66 HPI	Sapur.016G085600.1	0.278	0.042	AT2G37040.1
66 HPI	Sapur.004G139800.1	-0.277	0.043	AT2G36430.1
66 HPI	Sapur.001G041600.1	0.265	0.053	AT5G13930.1
66 HPI	Sapur.008G028000.1	0.258	0.060	AT2G37040.1
66 HPI	Sapur.001G041500.1	0.235	0.087	AT5G13930.1
66 HPI	Sapur.016G151200.1	0.231	0.093	AT4G05320.2
66 HPI	Sapur.017G110600.1	0.214	0.120	AT5G40990.1
66 HPI	Sapur.003G049200.1	0.209	0.130	AT4G17500.1
66 HPI	Sapur.016G173000.1	-0.204	0.138	AT3G49780.1
66 HPI	Sapur.017G124600.1	0.187	0.177	AT1G65930.1
66 HPI	Sapur.003G078300.1	-0.165	0.233	AT4G22620.1
66 HPI	Sapur.010G040400.1	0.154	0.266	
66 HPI	Sapur.009G096300.1	0.149	0.283	AT4G39090.1
66 HPI	Sapur.013G140300.1	0.141	0.311	AT5G05340.1
66 HPI	Sapur.014G029000.1	0.113	0.416	AT5G42650.1
66 HPI	Sapur.003G065200.1	-0.108	0.439	AT3G07040.1
66 HPI	Sapur.003G078900.1	0.107	0.440	AT1G62790.2
66 HPI	Sapur.017G118400.1	0.102	0.463	AT3G16520.3
66 HPI	Sapur.016G145300.1	0.101	0.468	AT5G12470.1
66 HPI	Sapur.017G116100.1	0.092	0.510	AT5G40380.1
66 HPI	Sapur.016G121900.1	0.087	0.532	AT4G12735.1
66 HPI	Sapur.017G120700.1	-0.087	0.533	AT1G01490.1
66 HPI	Sapur.016G188400.1	-0.086	0.538	AT1G17840.1
66 HPI	Sapur.018G074400.1	-0.078	0.575	AT4G30380.1
66 HPI	Sapur.016G154700.1	0.073	0.601	
66 HPI	Sapur.016G122000.1	0.071	0.612	AT4G12735.1
66 HPI	Sapur.003G088500.1	-0.071	0.612	AT3G47670.1
66 HPI	Sapur.016G161000.1	0.062	0.654	AT5G12340.1
66 HPI	Sapur.016G146200.1	-0.062	0.654	AT4G16260.1
66 HPI	Sapur.004G035000.1	0.054	0.698	AT4G17500.1

Table A.1: Candidate genes associated with the R-Turquoise module.

TIME POINT	TRANSCRIPT	RUST CORR	CORR PVAL	BEST HIT ARABI. NAME
66 HPI	Sapur.014G014700.1	-0.053	0.704	AT4G06534.1
66 HPI	Sapur.003G093000.1	-0.048	0.732	AT4G23420.1
66 HPI	Sapur.003G039300.1	0.047	0.735	AT1G72450.1
66 HPI	Sapur.010G039500.1	0.040	0.774	AT1G06620.1
66 HPI	Sapur.003G084900.1	0.038	0.783	AT4G11820.2
66 HPI	Sapur.016G170900.1	-0.033	0.814	AT2G14580.1
66 HPI	Sapur.016G170900.1	-0.033	0.814	AT2G14580.1
66 HPI	Sapur.019G022000.1	0.031	0.821	AT3G56710.1
66 HPI	Sapur.009G112900.1	0.023	0.868	AT3G12500.1
66 HPI	Sapur.003G069600.1	-0.022	0.874	AT1G32350.1
66 HPI	Sapur.016G155100.1	0.021	0.880	AT4G37980.1
66 HPI	Sapur.010G040200.1	0.017	0.902	AT3G23240.1
66 HPI	Sapur.003G086300.1	-0.016	0.909	AT1G12640.1
66 HPI	Sapur.010G037700.1	-0.015	0.914	AT5G43580.1
66 HPI	Sapur.016G175300.1	0.012	0.930	AT3G19615.1
66 HPI	Sapur.003G089000.1	-0.009	0.950	AT1G63245.1
66 HPI	Sapur.016G156400.1	0.007	0.957	AT1G12740.1
66 HPI	Sapur.017G110900.1	0.007	0.958	AT3G14470.1
66 HPI	Sapur.010G040800.1	-0.006	0.967	AT2G38940.1
66 HPI	Sapur.016G155000.1	0.004	0.977	

Table A.2: Candidate genes associated with the R-Blue module.

TIME POINT	TRANSCRIPT	RUST CORR	CORR PVAL	BEST HIT ARABI. NAME
42 HPI	Sapur.011G018600.1	0.315	0.023	AT4G05180.1
42 HPI	Sapur.007G010800.1	0.315	0.023	AT5G59690.1
42 HPI	Sapur.004G019000.1	0.280	0.045	AT4G05180.1
66 HPI	Sapur.003G045400.1	0.275	0.044	AT5G47500.1
66 HPI	Sapur.15WG080500.1	0.159	0.250	AT1G74470.1
66 HPI	Sapur.009G103000.1	0.157	0.257	AT2G21170.1
66 HPI	Sapur.016G136300.1	0.131	0.346	AT3G46780.1
66 HPI	Sapur.009G116200.1	0.130	0.350	AT3G44620.1
66 HPI	Sapur.15WG080200.1	0.123	0.377	AT3G47470.1
66 HPI	Sapur.009G039600.1	0.113	0.414	AT1G51400.1
66 HPI	Sapur.016G158400.1	0.109	0.433	AT1G08500.1
66 HPI	Sapur.002G197400.1	0.096	0.490	AT2G20260.1
66 HPI	Sapur.016G144800.1	0.090	0.520	AT1G51400.1
66 HPI	Sapur.002G044200.1	0.076	0.584	AT1G06680.1
66 HPI	Sapur.010G019800.1	0.075	0.590	AT1G67740.1
66 HPI	Sapur.003G086200.1	0.057	0.683	AT4G20360.1
66 HPI	Sapur.009G122700.1	0.041	0.766	AT2G21530.1
66 HPI	Sapur.016G177700.1	-0.022	0.872	AT2G14880.1
66 HPI	Sapur.016G132700.1	-0.020	0.885	AT2G28900.1
66 HPI	Sapur.016G146600.1	0.015	0.912	AT2G01870.1
66 HPI	Sapur.016G139900.1	0.013	0.927	AT3G47070.1
66 HPI	Sapur.018G101300.1	0.010	0.941	AT4G02530.1
66 HPI	Sapur.017G103400.1	0.007	0.962	AT5G16710.1
66 HPI	Sapur.016G131100.1	-0.005	0.972	AT2G28800.1

Table A.3: Marker count and total cM lengths for each linkage map, grouped by family.

FAMILY	10X - 400			
PARENT	Female *S. purpurea* 94006		Male *S. suchowensis* P63	
LINKAGE GROUP	Markers	Length (cM)	Markers	Length (cM)
1	169	418.6	123	303.4
2	144	286.9	126	243.8
3	76	131.8	120	235.9
4	125	242.2	110	215.5
5	150	268.7	112	208.1
6	127	232.3	117	224.6
7	97	206.9	95	176.1
8	102	177.4	125	239.5
9	82	157.2	85	171.2
10	163	298.5	117	215.4
11	99	194	93	182.6
12	89	210.8	86	149.5
13	115	264.9	87	187.7
14	82	154.7	87	148.2
15	108	199	82	149.5
16	191	377.8	210	392
17	103	201.5	92	173.7
18	106	194.1	88	157.1
19	68	143.4	80	166.2
TOTAL	2196	4360.7	2035	3939.9

FAMILY	11X - 407			
PARENT	Female *S. purpurea* 94006		Male *S. viminalis* 'Jorr'	
LINKAGE GROUP	Markers	Length (cM)	Markers	Length (cM)
1	187	330	225	521.8
2	152	256.9	215	487.1
3	115	215.6	184	362
4	155	292	168	328.5
5	176	372.2	183	398.4
6	174	286.2	216	409.9
7	130	247.2	140	333.7
8	120	274	137	244.7
9	103	183.9	125	273.9
10	161	276.8	194	433.2
11	106	200.5	127	252.7
12	97	233.2	118	213.7

13	121	230.9	136	305.9
14	94	188.7	113	204.7
15	118	232.9	124	240.5
16	187	359.8	342	667.3
17	106	192.7	145	278.4
18	134	229.6	117	207
19	91	165.5	77	165.1
TOTAL	2527	4768.4	3086	6328.5

FAMILY	**13X - 358**			
PARENT	**Female *S. purpurea* 94006**		**Male *S. udensis* 04-BN-051**	
LINKAGE GROUP	**Markers**	**Length (cM)**	**Markers**	**Length (cM)**
1	183	341.4	305	760.9
2	211	407.9	252	494.5
3	125	219.1	217	469.8
4	145	294.9	175	310.6
5	225	501.5	188	409.6
6	172	329.5	246	557.2
7	164	320	177	398.4
8	143	279.2	214	447
9	122	232.4	126	235.7
10	203	385.2	236	485.1
11	107	173.4	178	375
12	97	205.3	124	244
13	140	284.5	144	303.1
14	121	221.8	91	208.3
15	102	188.4	167	327.1
16	217	396.4	204	400.5
17	101	157.6	116	218.2
18	134	247	122	179.4
19	89	155.6	73	133
TOTAL	2801	5341.1	3355	6957.3

FAMILY	**13X - 438**			
PARENT	**Female *S. purpurea* 94006**		**Male *S. koriyanagi* 04-FF-016**	
LINKAGE GROUP	**Markers**	**Length (cM)**	**Markers**	**Length (cM)**
1	258	411.5	284	408.7
2	242	361.1	258	332.1
3	141	213.5	168	269.4
4	180	248.8	200	248.6
5	272	364.2	254	362.6

6	224	331.5	304	437.7
7	177	241	129	179
8	187	250.5	175	246.8
9	162	237.4	127	196.5
10	265	348	230	323.1
11	163	235.8	202	278.4
12	118	188.8	130	186.6
13	118	177.1	149	235.1
14	143	187.9	256	332.1
15	120	174.8	171	216.1
16	295	442.6	430	573.2
17	150	227.2	140	208.6
18	163	214	144	178.1
19	125	208.2	101	141.8
TOTAL	3503	5063.9	3852	5354.5

FAMILY	**13X - 440**			
PARENT	**Female S. suchowensis P295**		**Male *S. purpurea* 94001**	
LINKAGE GROUP	**Markers**	**Length (cM)**	**Markers**	**Length (cM)**
1	240	466.5	221	415
2	222	350.3	257	480.2
3	209	359.1	163	281.6
4	247	489.6	146	307.2
5	207	376.5	240	437.5
6	219	412.5	282	536.8
7	170	325.5	139	290.9
8	124	272.9	159	304.6
9	151	266.7	142	251.4
10	222	383.1	191	383.9
11	136	258	108	284.8
12	137	254.7	84	149.1
13	153	285.3	142	263.9
14	151	267.9	193	309.1
15	157	279.7	109	198
16	306	580.1	278	531.4
17	154	268.7	130	241
18	139	237.3	120	204.8
19	102	211.6	102	201.4
TOTAL	3446	6345.8	3206	6072.7

FAMILY	**12X - 421**	
PARENT	**Female *S. viminalis***	**Male *S. purpurea***

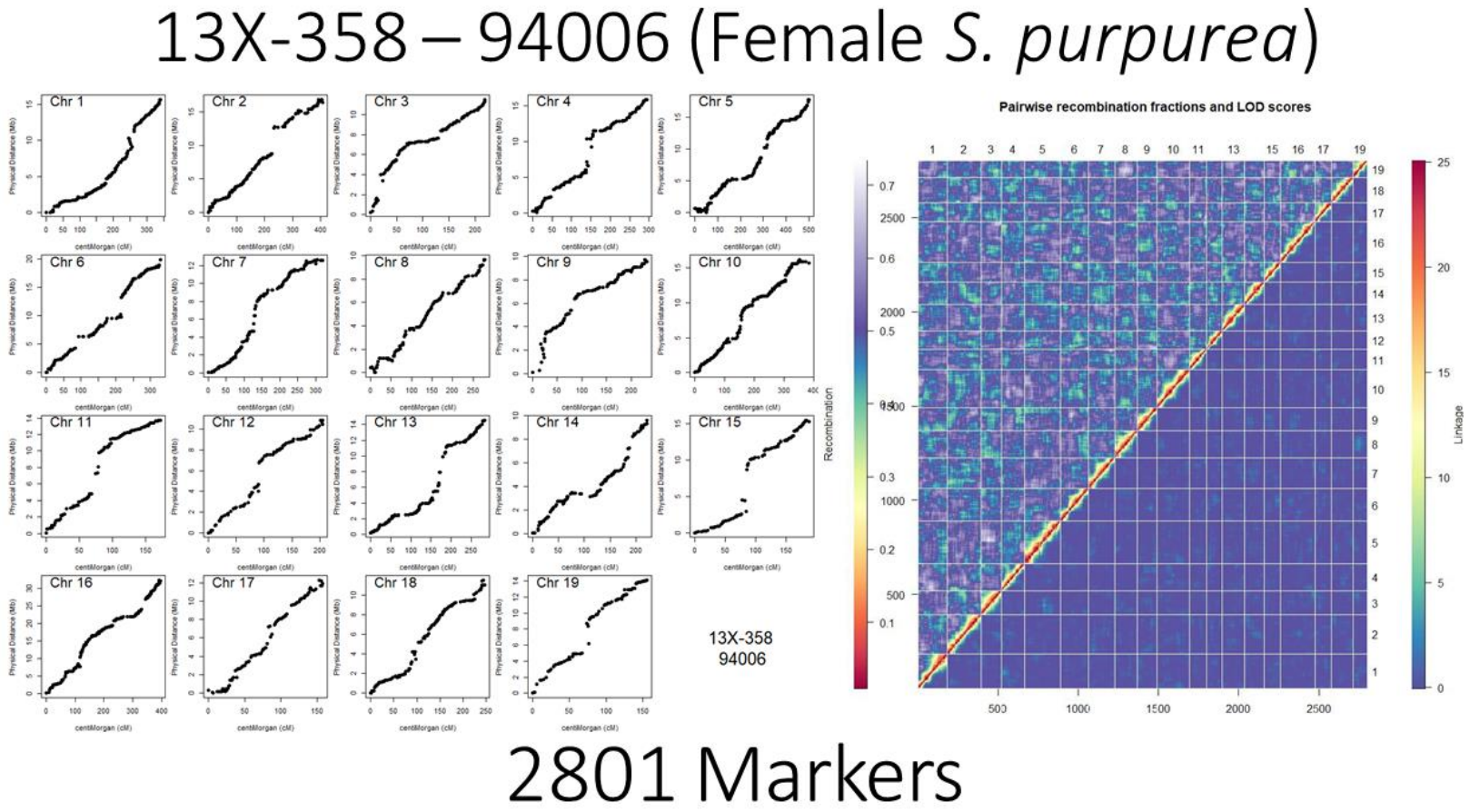
	07-MBG-5027		94001	
LINKAGE GROUP	**Markers**	**Length (cM)**	**Markers**	**Length (cM)**
1	212	600.3	270	637.7
2	238	580	186	366.6
3	156	352.4	147	338.2
4	188	465.7	120	269.2
5	188	518.1	236	499.3
6	190	431.3	255	614.1
7	136	323	155	372.9
8	217	508.7	146	354.2
9	118	234.1	127	243.9
10	207	476.9	168	409.7
11	159	283.8	101	246.5
12	124	273.4	107	179.1
13	135	293.8	131	260.5
14	159	313.5	156	320.6
15	126	232.8	88	148
16	239	439.7	250	534.3
17	119	202.4	136	246
18	121	199.1	114	190.2
19	83	159.1	98	186
TOTAL	3115	6888.3	2991	6417.2

FAMILY	**13X - 426**			
PARENT	**Female *S. integra* P336**		**Male *S. purpurea* 94001**	
LINKAGE GROUP	**Markers**	**Length (cM)**	**Markers**	**Length (cM)**
1	169	311.4	252	410.7
2	200	337.1	251	376.9
3	123	227.9	145	222.8
4	144	249.7	174	318.6
5	154	300.8	173	303.2
6	178	332.2	343	551
7	123	206.9	169	263.5
8	149	249	154	247.3
9	100	155.3	192	280.9
10	149	316.8	229	378.3
11	147	254.7	93	185.9
12	108	214.4	109	182.1
13	101	184.4	140	213.5
14	129	213.3	238	394.5
15	17	34.1	141	221.2

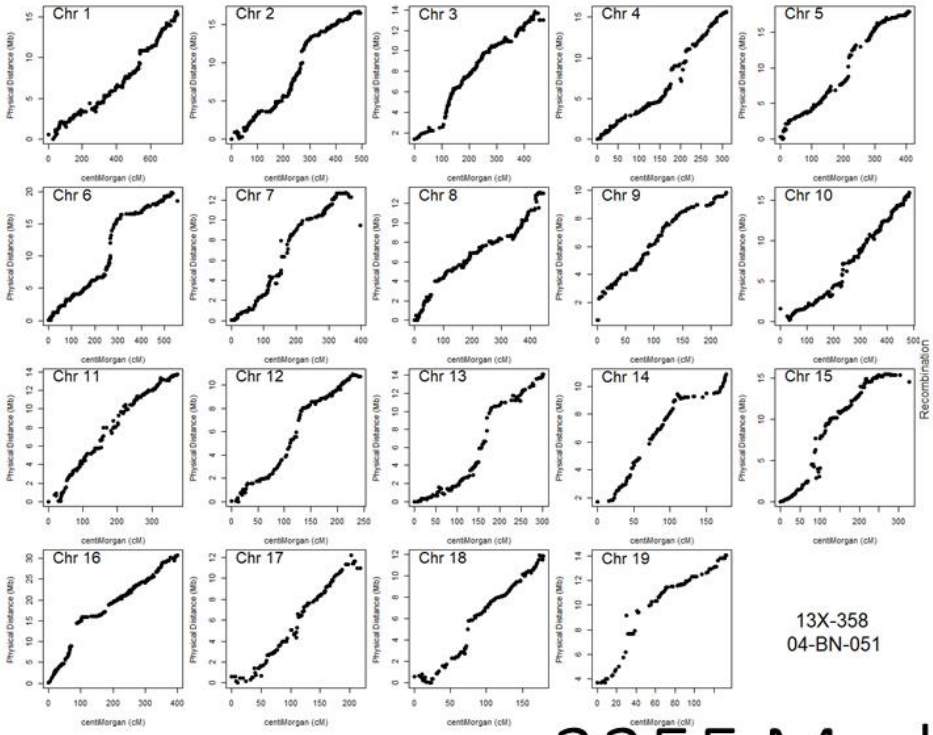
16	236	475.6	329	532.5
17	124	205.3	144	213.4
18	121	223.5	169	245.2
19	80	216.8	141	204.5
TOTAL	2552	4709.4	3586	5746.1

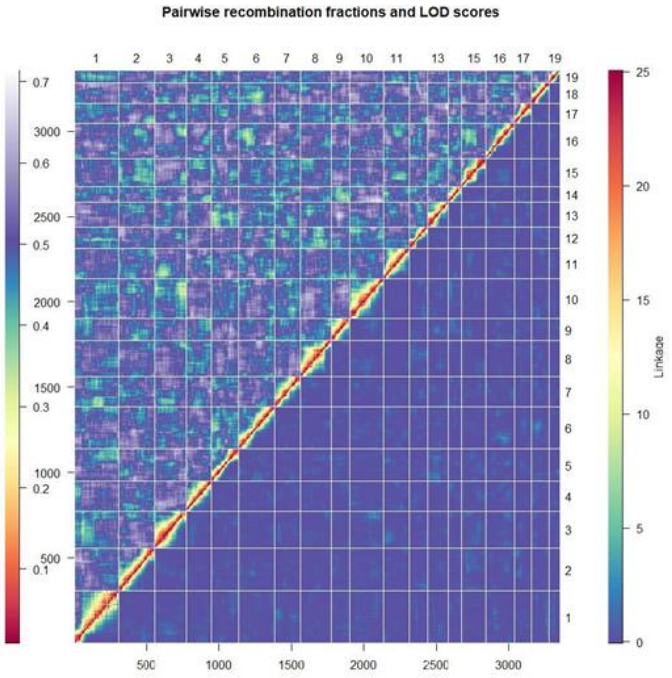
FAMILY	**13X - 443**			
PARENT	**Female *S. suchowensis* P294**		**Male *S. purpurea* 94001**	
LINKAGE GROUP	**Markers**	**Length (cM)**	**Markers**	**Length (cM)**
1	163	453.6	212	339.4
2	224	339.3	167	256.6
3	138	229.5	131	255.6
4	158	245.4	116	251.3
5	178	295.3	209	334.3
6	182	287	275	431.6
7	121	224.1	150	239.2
8	157	267.6	138	248.5
9	115	163.1	138	176.3
10	186	283.2	156	238.2
11	116	212.9	81	187.7
12	99	160.9	54	93.2
13	116	188.4	125	205.1
14	135	203.1	188	258.7
15	162	287.4	113	210.7
16	273	398.2	208	399.4
17	96	165.4	141	274.6
18	145	233.6	117	166.3
19	67	150.2	106	174.9
TOTAL	2831	4788.4	2825	4741.5

Figure A.3: Figures for assessing the quality and stability of the linkage maps. cM x physical position (Mb) based on alignment to the *S. purpurea* reference genome and heatmaps showing estimated recombination and linkage in the upper and lower portions, respectively.



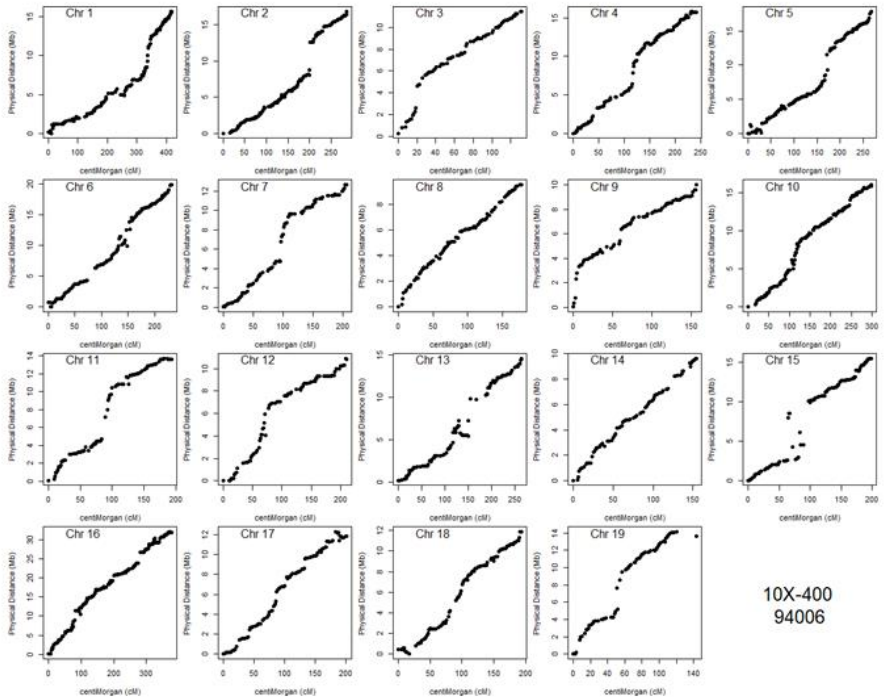
13X-358 – 04-BN-051 (Male *S. udensis*)

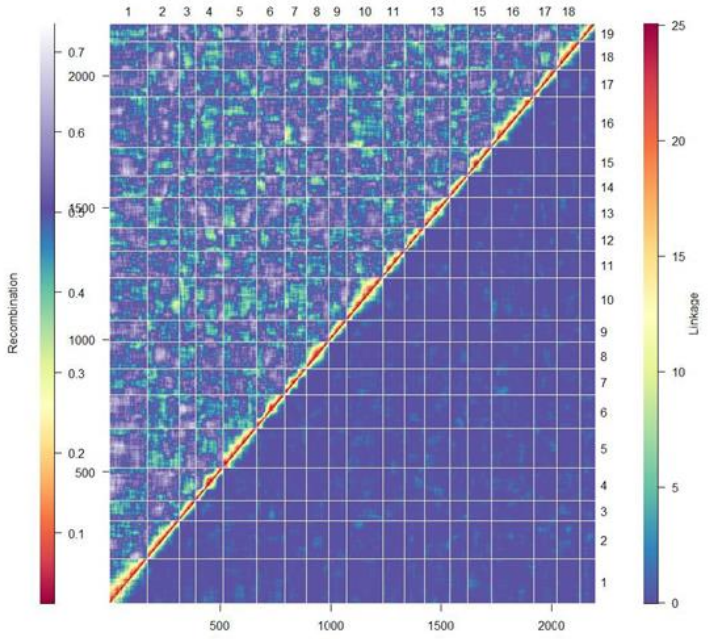




3355 Markers

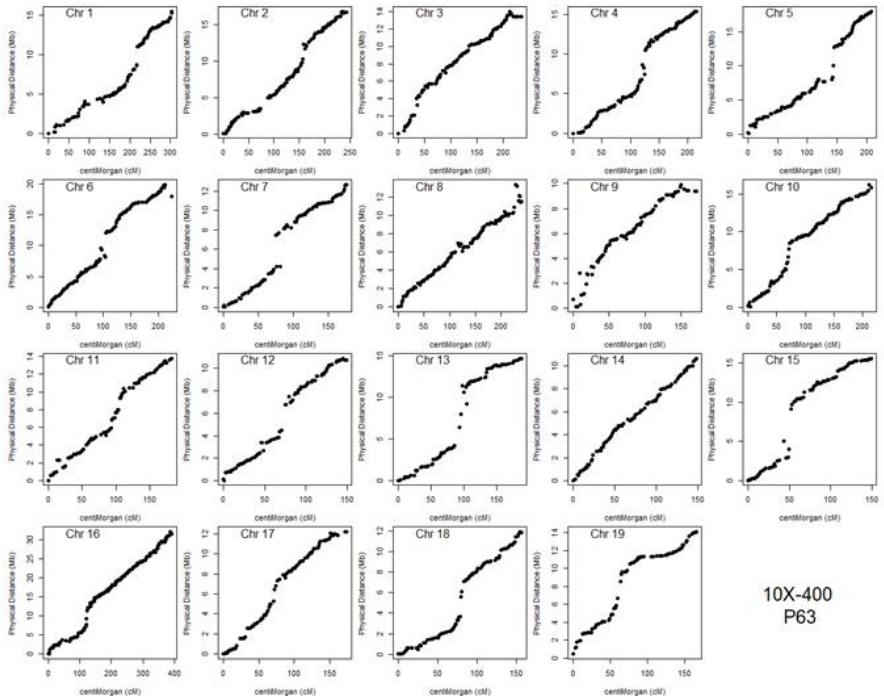
10X-400 – 94006 (Female *S. purpurea*)

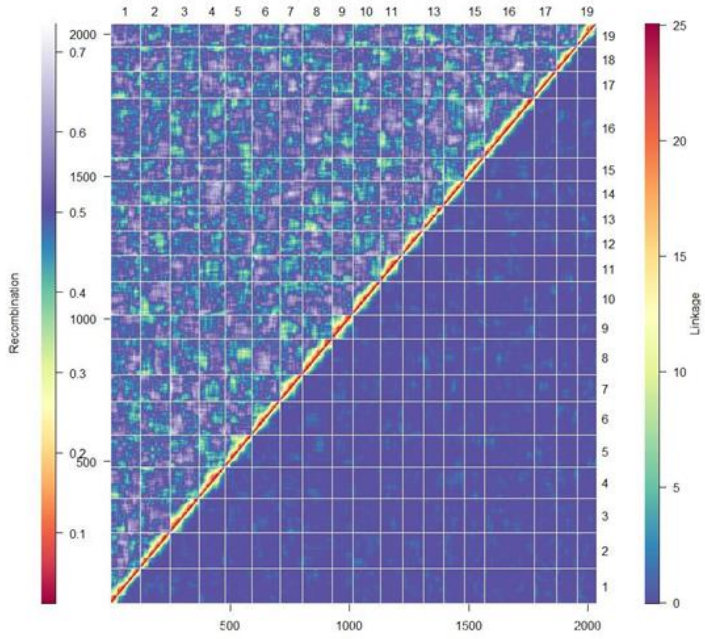




2196 Markers

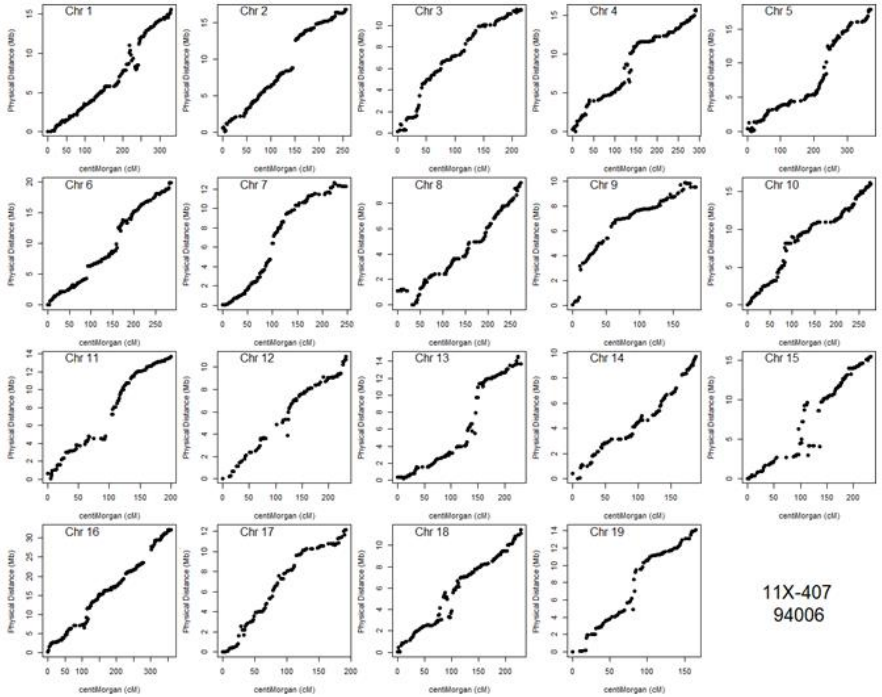
10X-400 – P63 (Male *S. suchowensis*)

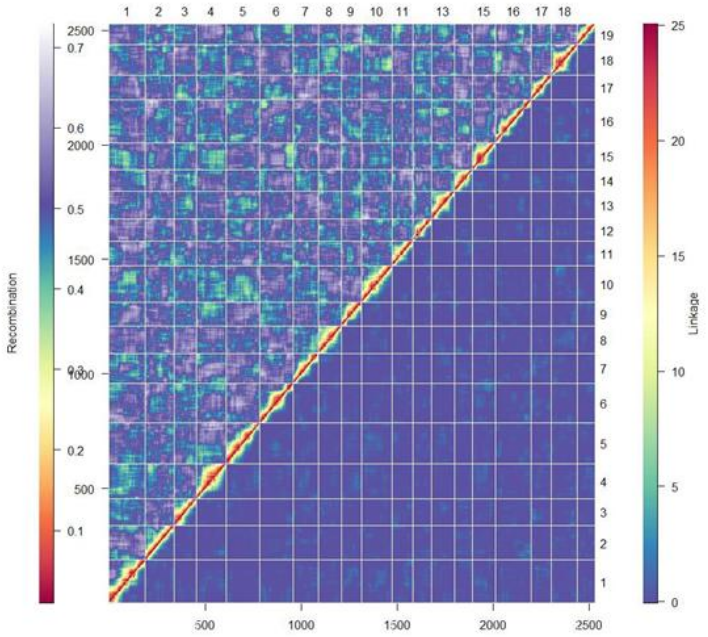




2035 Markers

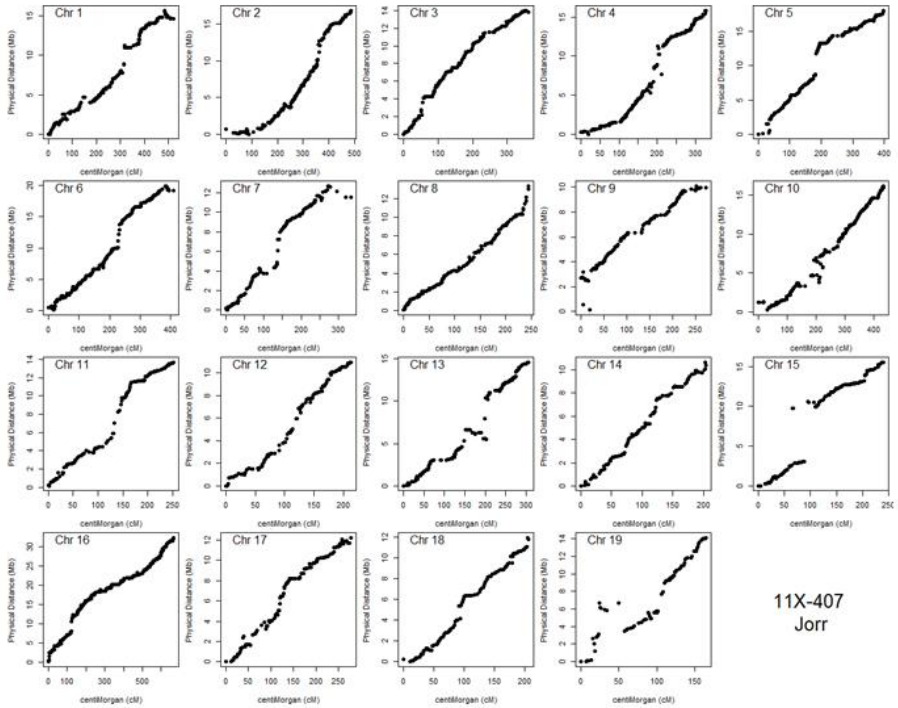
11X-407 – 94006 (Female *S. purpurea*)



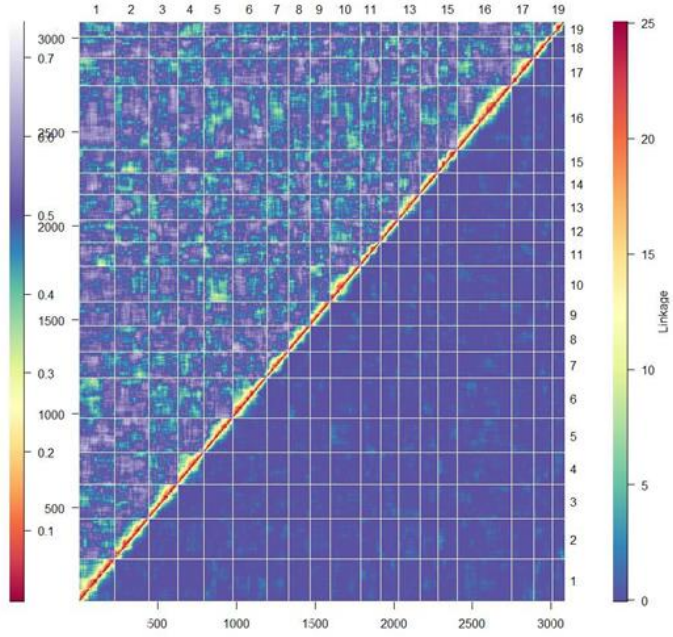


2527 Markers

11X-407 – Jorr (Male *S. viminalis*)

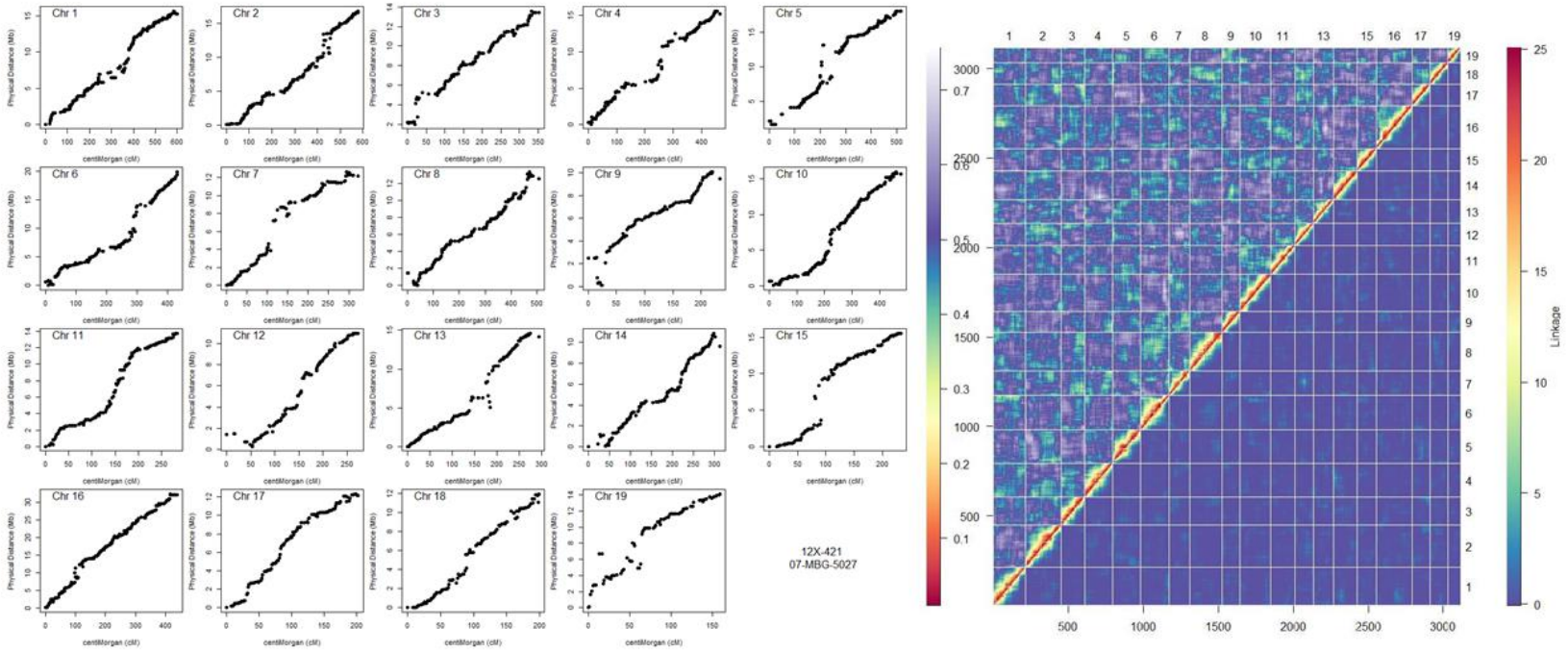


11X-407
Jorr



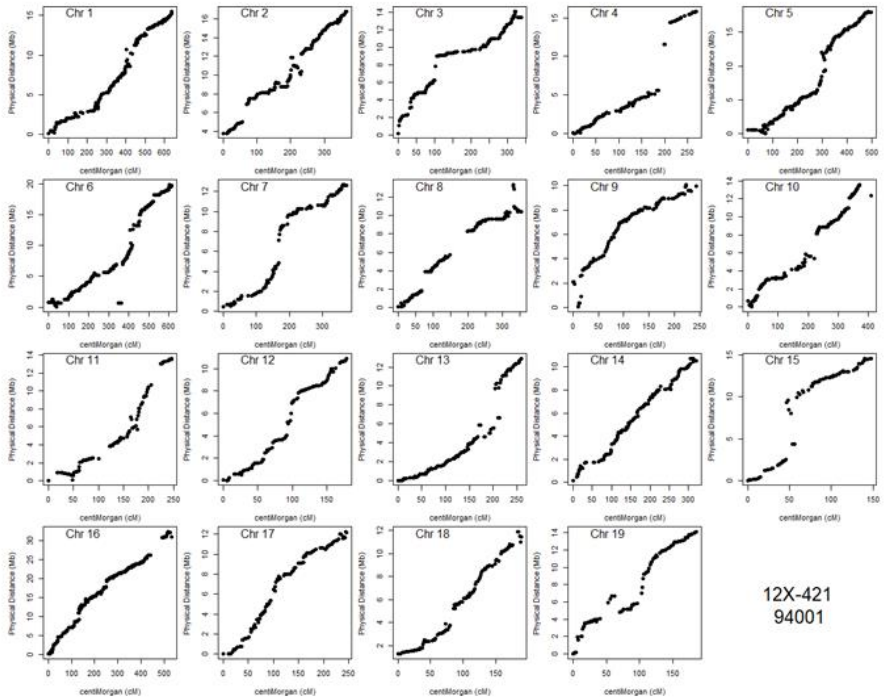
3086 Markers

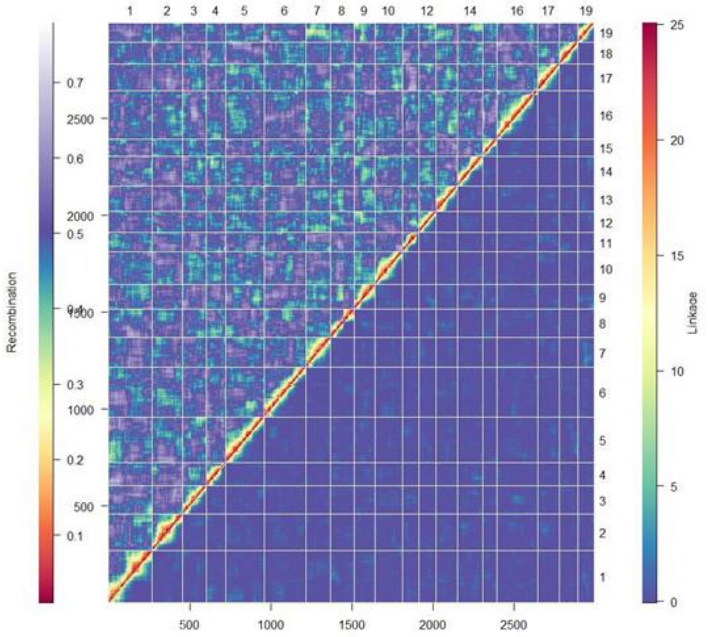
12X-421 – 07-MBG-5027 (Female *S. viminalis*)



3115 Markers

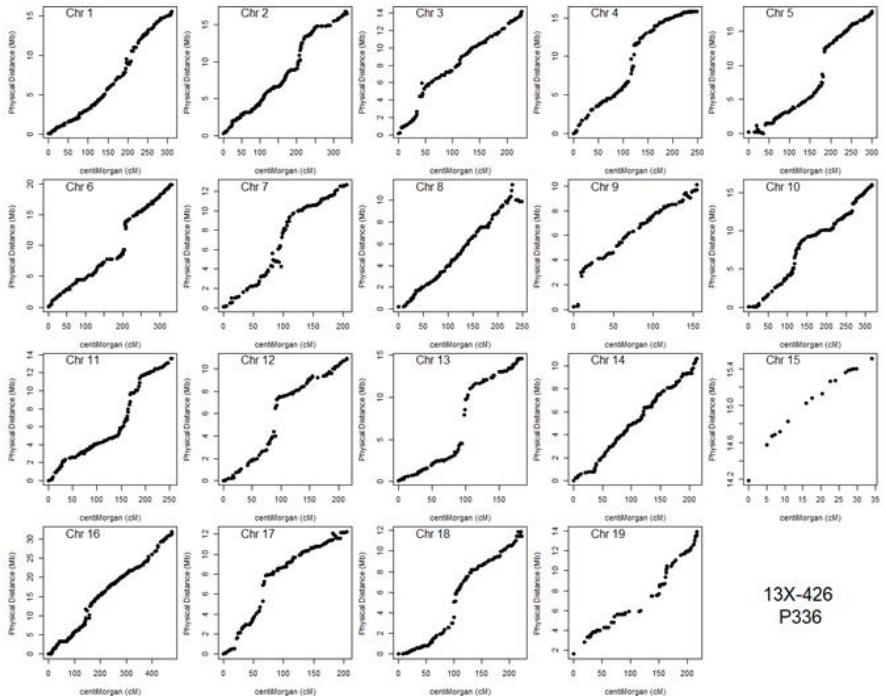
12X-421 – 94001 (Male *S. purpurea*)

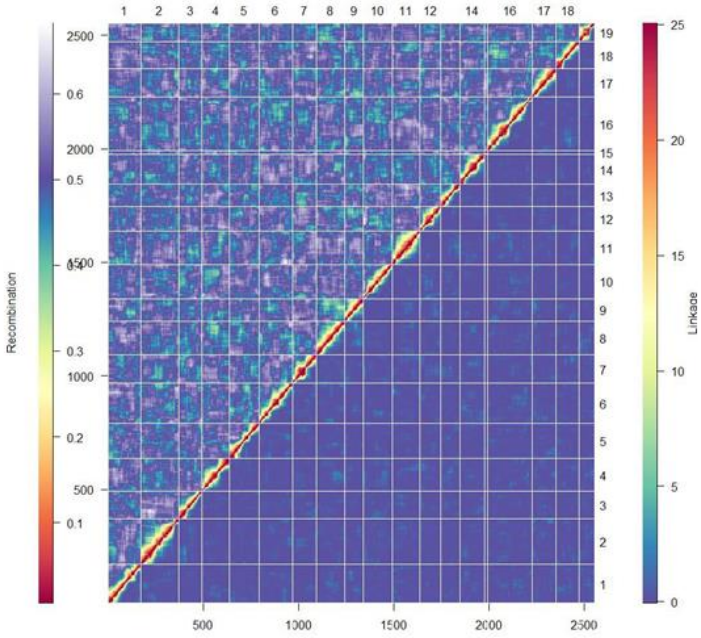




2991 Markers

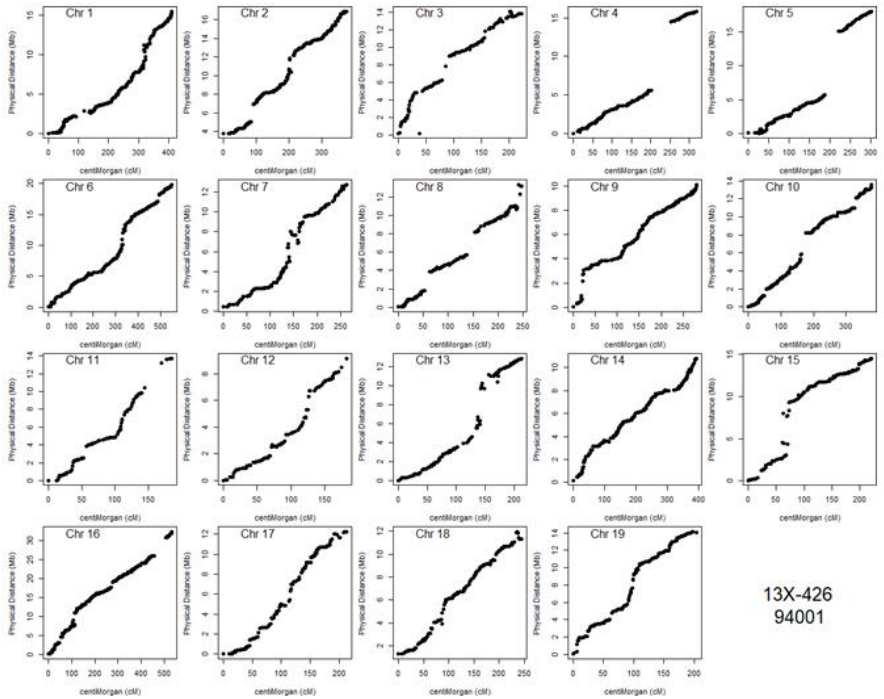
13X-426 – P336 (Female *S. integra*)

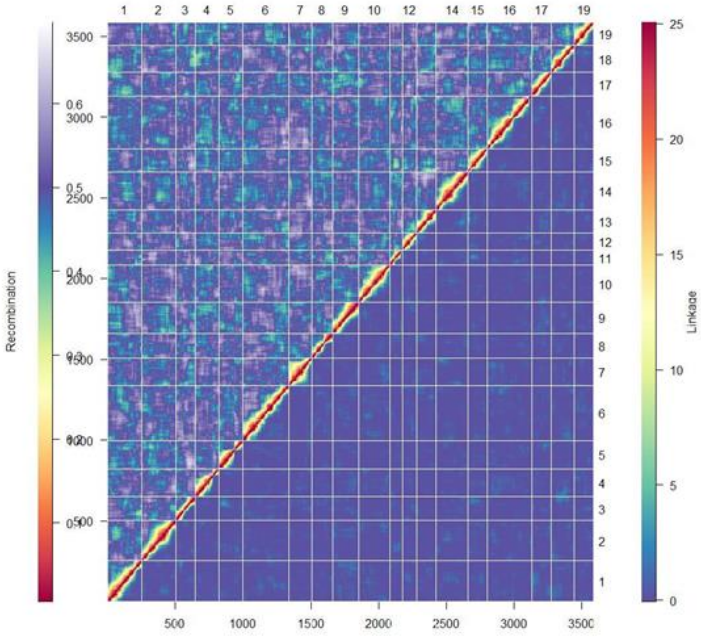




2552 Markers

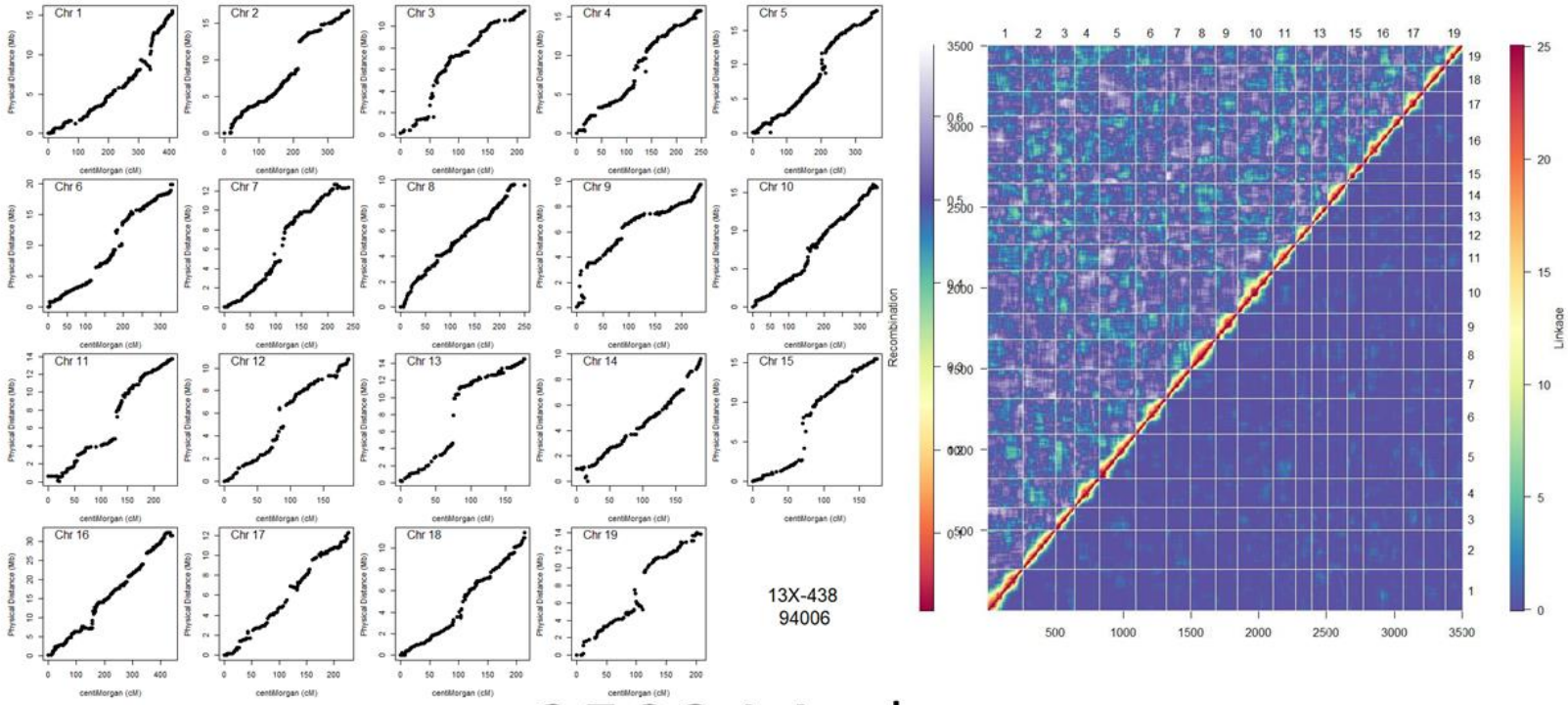
13X-426 – 94001 (Male *S. purpurea*)





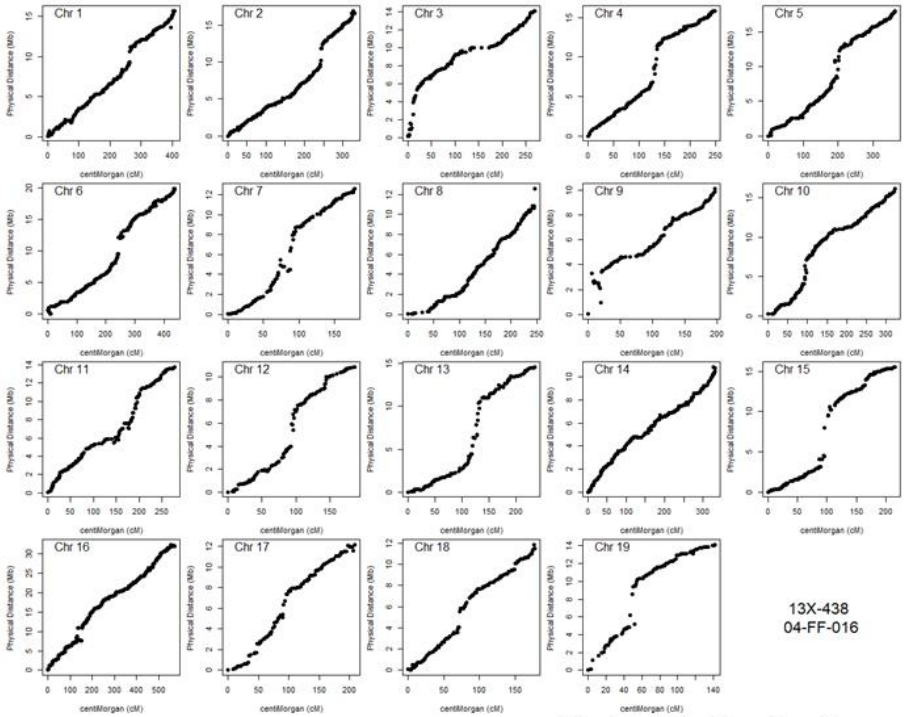
3586 Markers

13X-438 – 94006 (Female *S. purpurea*)

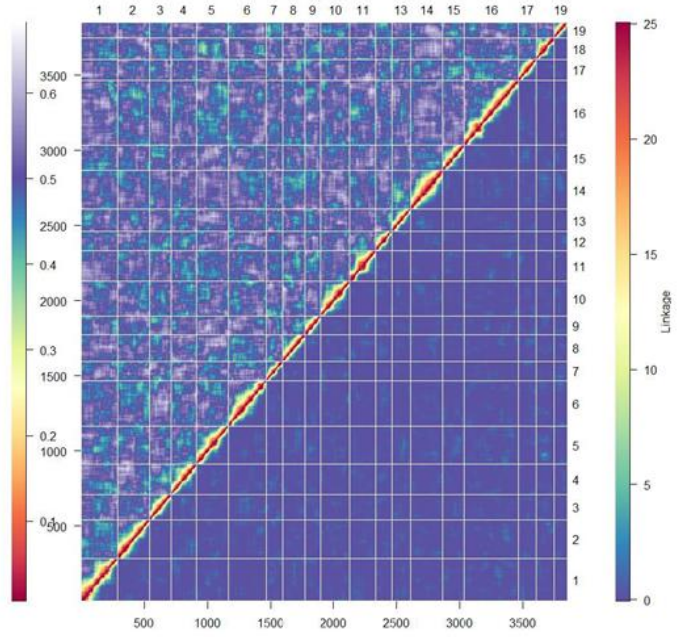


3503 Markers

13X-438 – 04-FF-016 (Male *S. koriyanagi*)

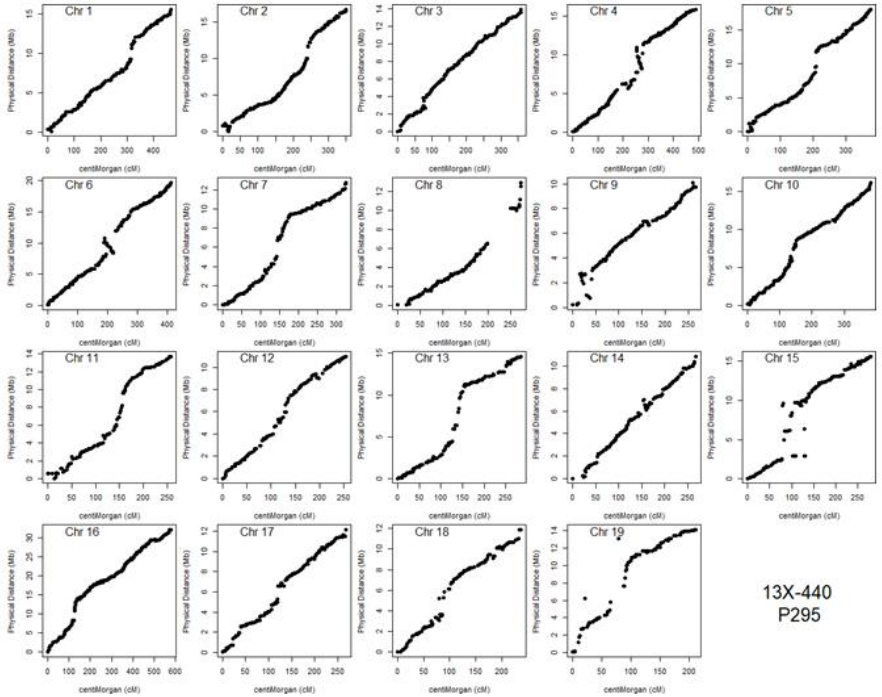


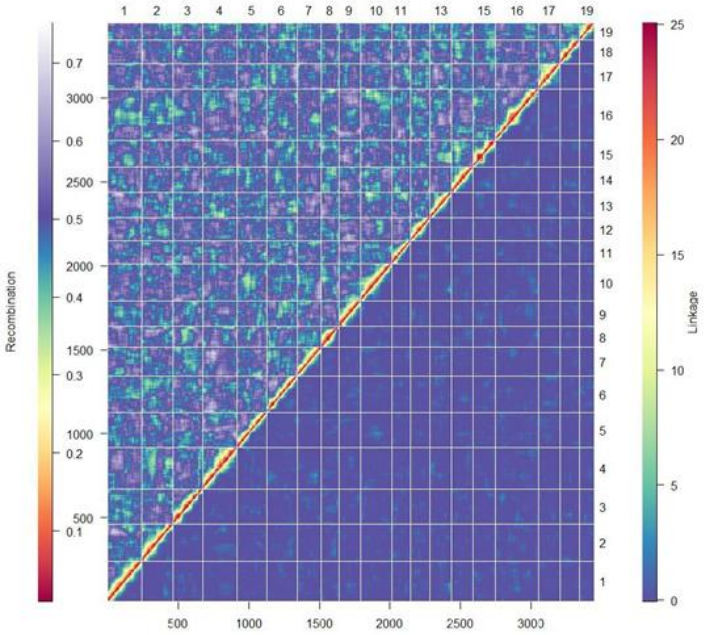
13X-438
04-FF-016



3852 Markers

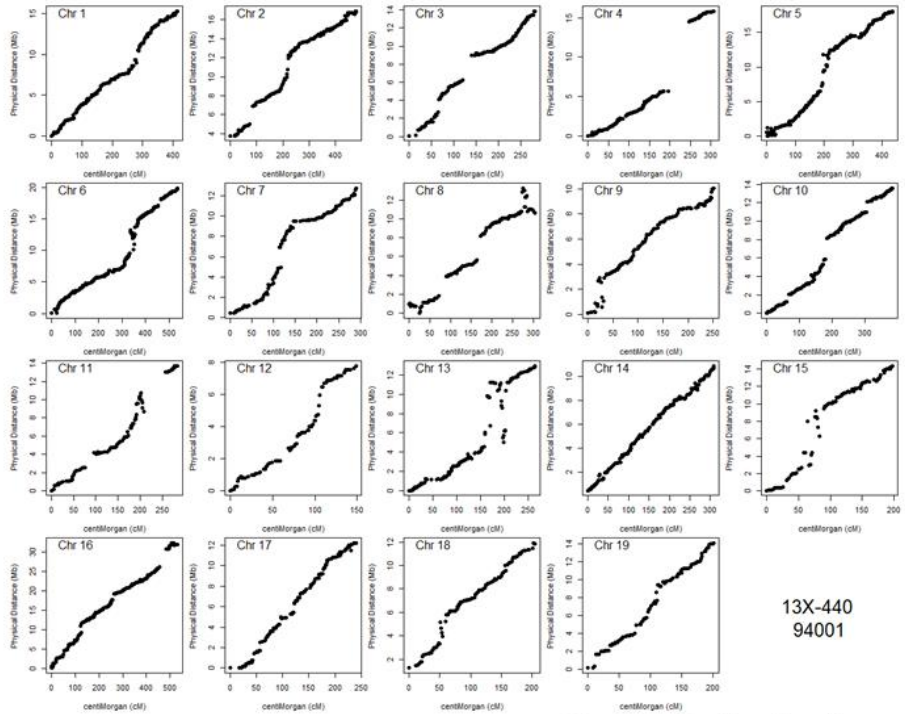
13X-440 – P295 (Female *S. suchowensis*)



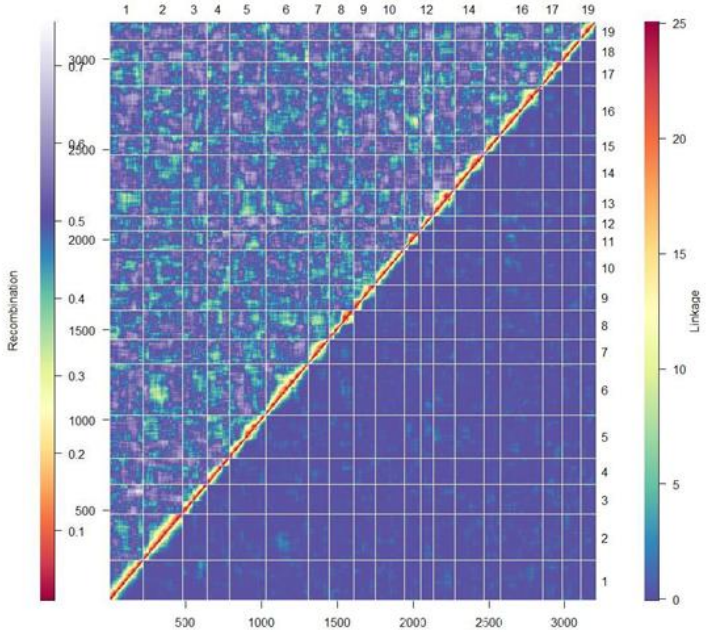


3446 Markers

13X-440 – 94001 (Male *S. purpurea*)

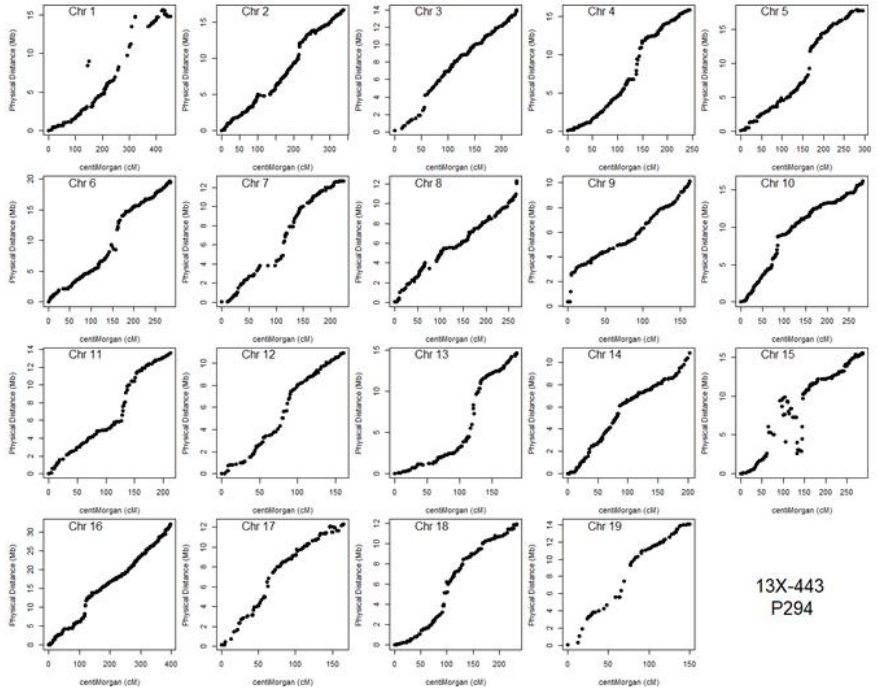


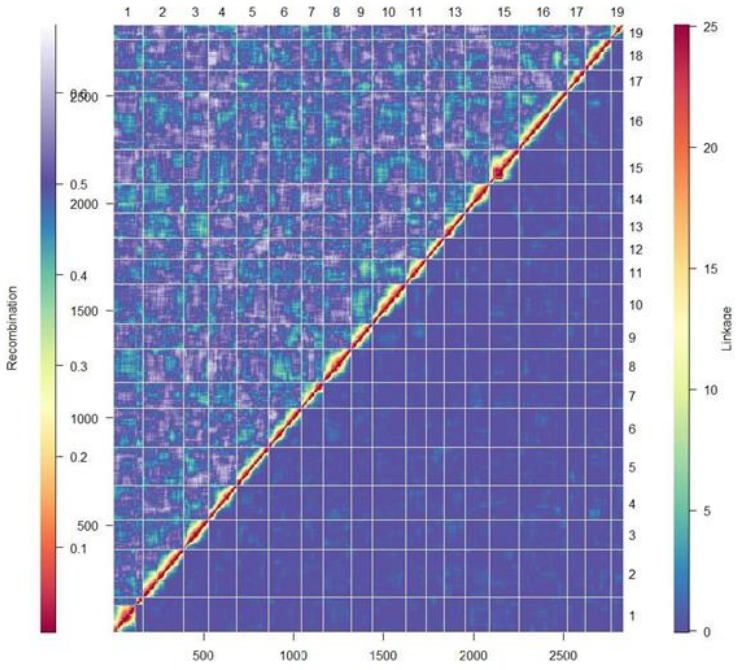
13X-440
94001



3221 Markers

13X-443 – P294 (Female *S. suchowensis*)





2836 Markers

Table A.4: Trait regression and Fisher's LSD results from the eight F_1 hybrid families.

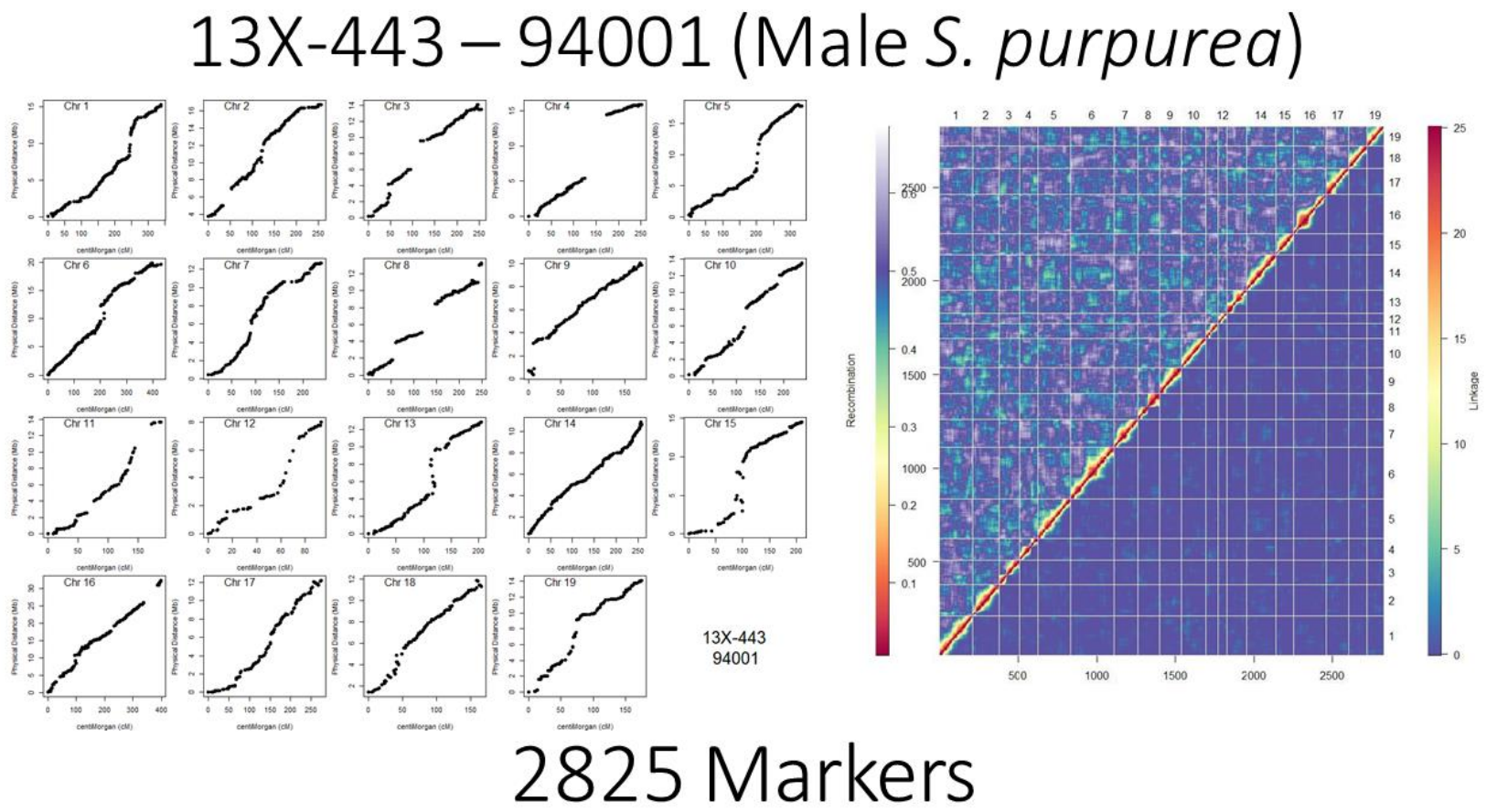


Table A.4: Trait regression, family-level means, and Fisher's LSD results from the eight F_1 hybrid families.

	WLB_17	**WLB_18**	**WLB_19**	**PLHC_17**	**PLHC_19**	**PLHN_17**	**PLHN_19**
FIELD	7.15E-56	0	4.60E-144	0	2.10E-187	1.77E-232	0
FIELD:REP	2.59E-21	3.47E-26	9.87E-22	7.75E-231	1.90E-132	1.04E-217	1.08E-40
FIELD:FAMILY	0	0	3.43E-195	0	0	0	0
10X-400	2.5f	1.6c	2.6g	6.93g	10.88g	1.86ef	1.61d
11X-407	4.3d	5.42a	7.14a	50.9c	39.69c	19.22b	5.44a
12X-421	5.12c	4.69b	4.99c	61.78a	47.6a	14.06c	4.03c
13X-358	7.28a	4.53b	6.07b	53.61b	37.04d	21.63a	4.54b
13X-426	0g	0.68d	4.2d	35.1d	45.27b	2.52e	1.64d
13X-438	0.02g	0.83d	2.62g	6.64g	11.64g	0.94f	1.16f
13X-440	3.9e	1.58c	2.97f	27.27e	23.02e	5.45d	1.64d
13X-443	5.46b	1.67c	3.34e	9.38f	14.21f	2.2e	1.37e
CV (%)	69.6	56.6	51.6	58.7	59.6	98.6	60.3
	PLHS_17	**PLHS_19**	**DLW_17**	**DLW_18**	**LA_17**	**LA_18**	**SLA_17**
FIELD	5.37E-135	8.35E-241	1.05E-29	2.72E-05	1.91E-64	4.71E-14	6.06E-44
FIELD:REP	2.53E-203	5.41E-18	2.68E-15	1.29E-79	2.93E-18	6.85E-72	3.04E-07
FIELD:FAMILY	1.26E-125	0	9.45E-17	3.61E-116	5.89E-156	1.05E-122	1.36E-36
10X-400	10.64c	5.37h	0.18a	0.09c	19.53a	11.12c	112.54d
11X-407	18.71a	51.23b	0.13e	0.08de	19.79a	10.71c	152.97a
12X-421	20.29a	42.74c	0.14e	0.1b	19.65a	13.18a	149.55a
13X-358	15.62b	65.24a	0.14e	0.08e	12.25d	11.74b	105.15e
13X-426	3.27d	23.76d	0.15d	0.07f	18.38b	9.97d	139.52b
13X-438	1.31e	10.22g	0.14e	0.06g	14.58c	7.84e	114.39d
13X-440	10.99c	21.63e	0.16b	0.11a	18.7b	12.93a	129.27c
13X-443	2.01de	13.18f	0.16c	0.09cd	18.79b	11.13c	133.97bc
CV (%)	131.4	65.2	38.2	47.2	29.3	35.9	40.1

	SLA_18	LL_17	LL_18	LW_17	LW_18	LP_17	LP_18
FIELD	2.70E-07	1.51E-76	3.09E-25	4.06E-58	5.67E-26	6.70E-112	6.85E-130
FIELD:REP	7.38E-33	6.46E-25	1.55E-78	4.32E-09	3.38E-57	2.99E-36	5.73E-95
FIELD:FAMILY	5.33E-73	0	6.93E-265	6.06E-277	1.28E-199	8.52E-238	5.34E-156
10X-400	128.61d	11.85b	9.28bc	2.11d	1.67d	24.8a	19.59de
11X-407	140.24c	11.23d	8.28d	2.27c	1.75c	25.18a	22.29b
12X-421	139.63c	10.9e	9.1c	2.36b	1.96a	24.67a	24.21a
13X-358	158.87a	8.34h	8.19d	1.98e	1.94ab	18.03e	20.14cd
13X-426	149.63b	8.87g	6.64f	2.54a	1.91b	19.79d	14.85f
13X-438	137.57c	10.39f	7.4e	1.74f	1.41f	21.45c	15.27f
13X-440	120.68e	11.45c	9.81a	2.08d	1.73c	23.79b	20.6c
13X-443	135.92c	12.2a	9.41b	1.99e	1.59e	25.16a	19.53e
CV (%)	26	16.2	20.3	17.3	19.5	17.6	25.1

	LFF_17	LFF_18	LFR_17	LFR_18	RST_17	RST_19	AVGDIAM_17
FIELD	1.41E-62	5.20E-221	7.03E-75	3.01E-53	3.85E-45	5.61E-29	0
FIELD:REP	6.13E-08	2.12E-32	1.53E-31	1.97E-07	3.61E-06	0.002759599	1.49E-10
FIELD:FAMILY	0	0	0	0	4.71E-34	9.75E-63	0
10X-400	5.71c	0.36e	0.4d	5.65b	2.32cd	3.85e	0.98d
11X-407	4.97e	0.31f	0.4d	4.77d	0.22f	4.08de	0.72f
12X-421	4.67f	0.29g	0.41c	4.68d	0.32f	5.19cd	0.66g
13X-358	4.23g	0.39c	0.48b	4.24e	1.91de	5.29c	0.8e
13X-426	3.52h	0.57a	0.58a	3.48f	1.61e	4.57cde	1.01c
13X-438	6.02b	0.42b	0.4d	5.29c	3.11b	10.64a	1.08a
13X-440	5.58d	0.38d	0.41c	5.74b	4.47a	11.03a	1.06b
13X-443	6.19a	0.36e	0.37e	6.01a	2.74bc	7.22b	1.08a
CV (%)	15.7	21.3	16	17.5	186.3	116.4	13.5

	AVGDIAM_18	STEMCT_17	STEMCT_18	HT_17	HT_18	PLTAREA_17	PLTAREA_18
FIELD	0	3.69E-100	5.80E-201	0	0	1.88E-79	6.98E-247
FIELD:REP	8.03E-44	5.64E-22	5.01E-180	1.18E-18	2.41E-27	9.02E-14	6.63E-19
FIELD:FAMILY	0	1.54E-33	1.36E-65	0	0	1.65E-94	1.62E-297
10X-400	1.09d	17.84a	18.88a	288.05b	389.33d	14.14a	22.44b
11X-407	0.89f	16.03b	12.07c	155.96e	256.17e	9.25de	7.88e
12X-421	0.79g	16.7b	12.85b	135.09g	215.27f	8.81e	6.8ef
13X-358	0.98e	11.42f	8.31g	143.36f	220.44f	7.19f	6.6f
13X-426	1.36c	14.73c	12.6bc	274.33c	436.75b	11.63b	19.99c
13X-438	1.47b	12.66e	10.69e	292.88a	417.14c	10.8c	20.39c
13X-440	1.51a	11.4f	9.16f	257.03d	419.39c	9.53d	18.77d
13X-443	1.53a	13.44d	11.34d	288.06b	444.88a	11.43b	23.54a
CV (%)	16.8	38.9	40.2	16.6	15.8	40.1	47

	PLTVOL_17	PLTVOL_18	SPAD_818	SPAD_1018	SPAD_819	CROWN_18
FIELD	1.94E-250	7.43E-285	2.71E-32	9.36E-45	2.73E-14	5.33E-87
FIELD:REP	8.62E-14	5.56E-11	0.000366573	2.08E-28	3.40E-09	1.19E-12
FIELD:FAMILY	4.95E-259	0	4.53E-298	1.40E-91	3.36E-91	9.61E-50
10X-400	1427.5a	3142.87b	41.84b	45.11a	42.7c	43.45d
11X-407	521.12d	770.96e	38.16c	39.67b	42.33cd	50.92a
12X-421	419.99e	538.24f	37d	37.38d	40.98ef	51.86a
13X-358	360.52f	520.17f	35.61f	40.55b	44.23b	43.05d
13X-426	1080.51b	2918.31c	36.37de	38.22cd	40.86f	45.96c
13X-438	1081.45b	2905.71c	36.01ef	38.31c	41.54de	42.05e
13X-440	839.26c	2721.45d	45.58a	45.46a	47.35a	45.7c
13X-443	1117.25b	3565.87a	42.26b	40.59b	44.35b	48.17b
CV (%)	48.2	54.5	13.8	16.8	13.7	16.5

CV: Coefficient of Variation

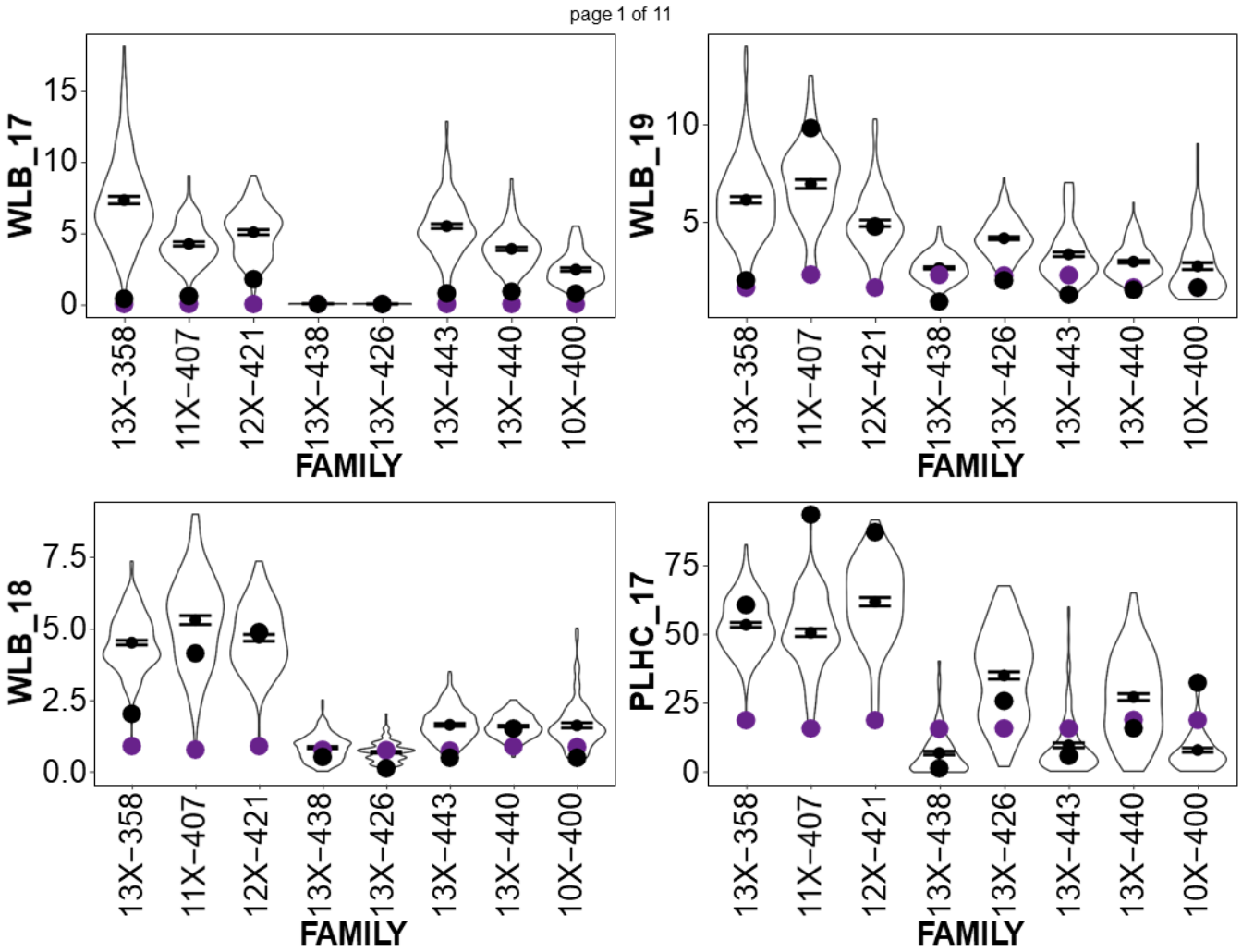
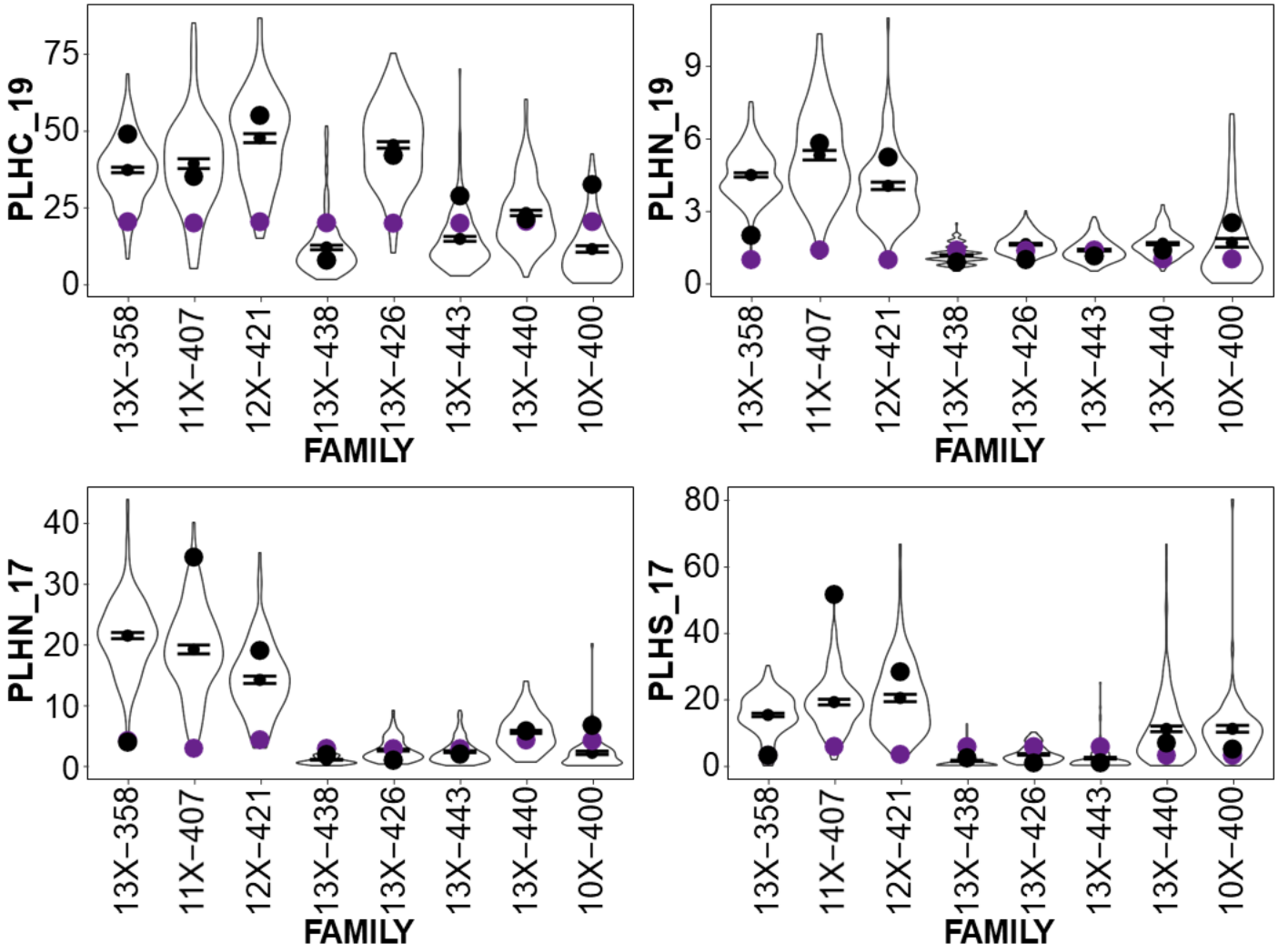
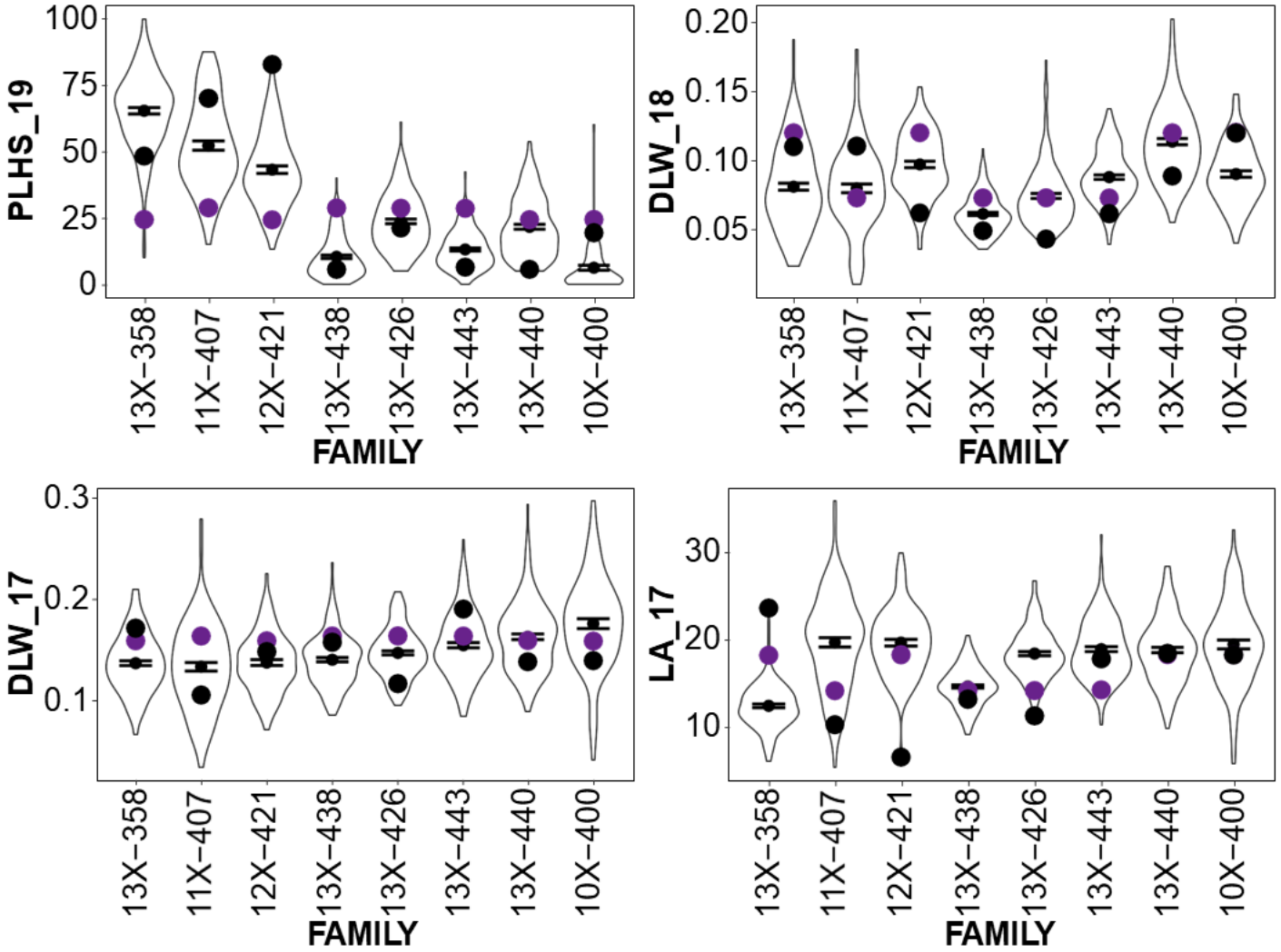
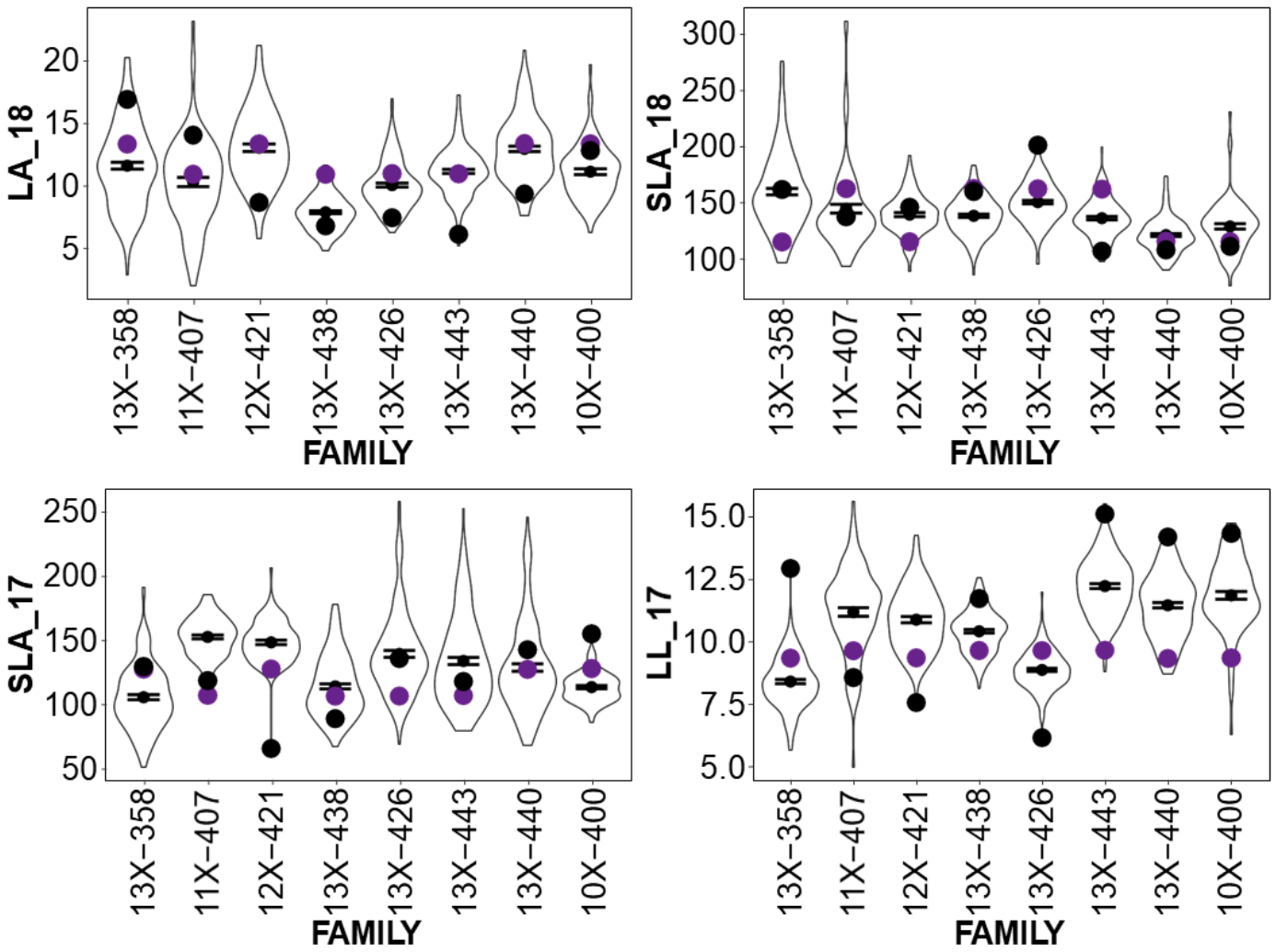
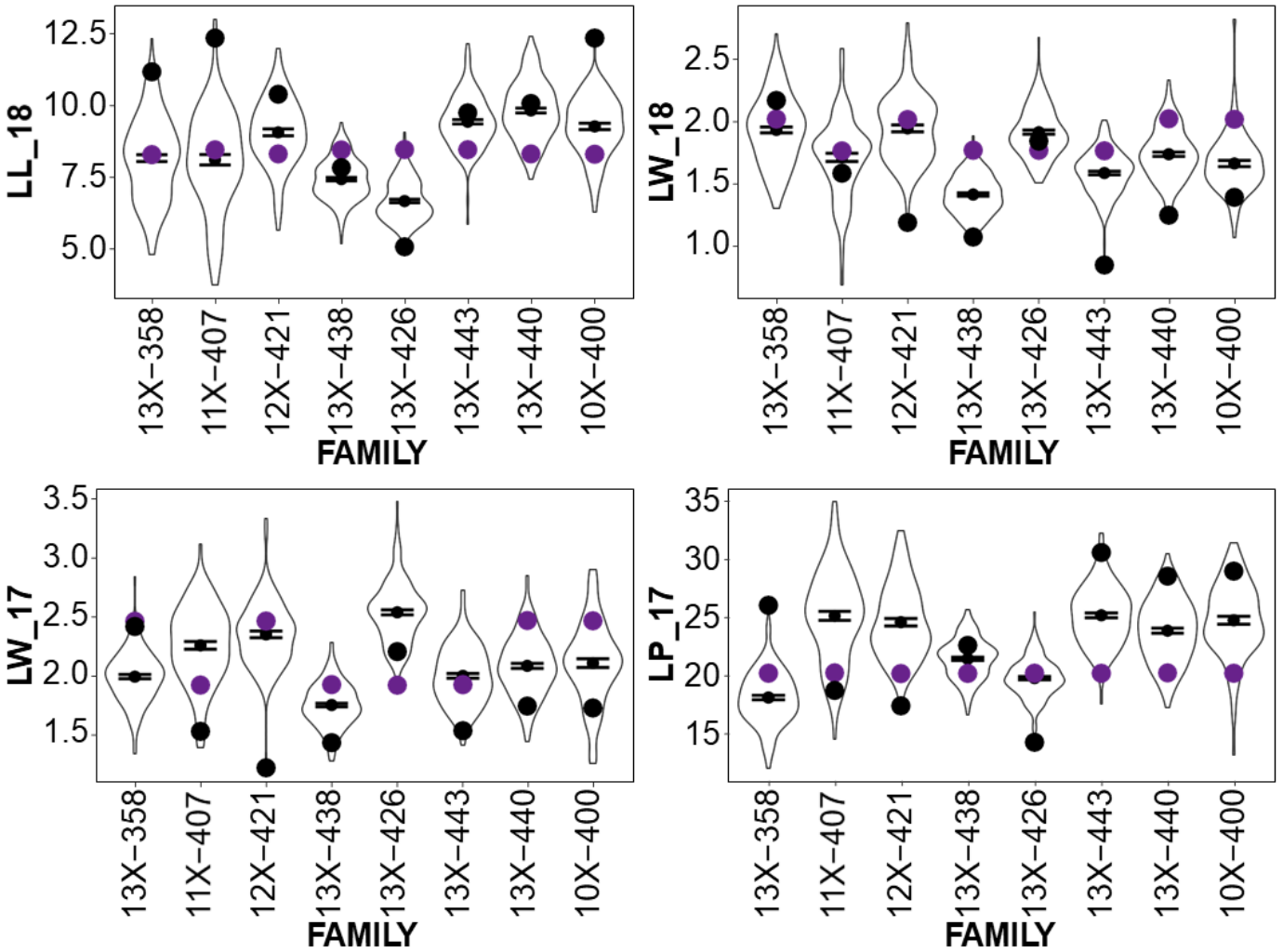


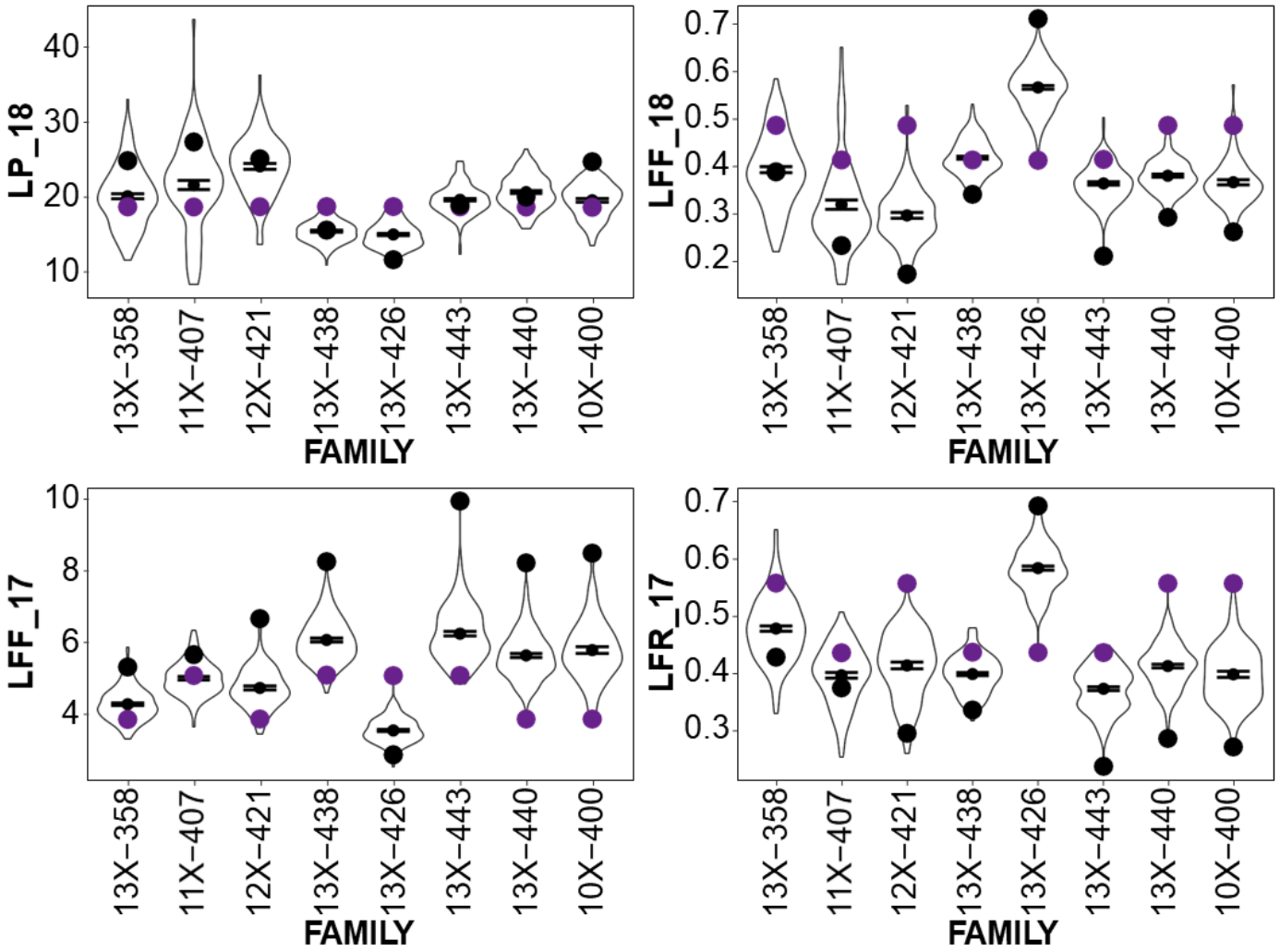
Figure A.4: Violin plots showing the variation within and between the F_1 hybrid families. The purple dots are the *S. purpurea* common parent, while the black dots are the family specific parents. The smaller black dots and error bars are the family mean and standard error.

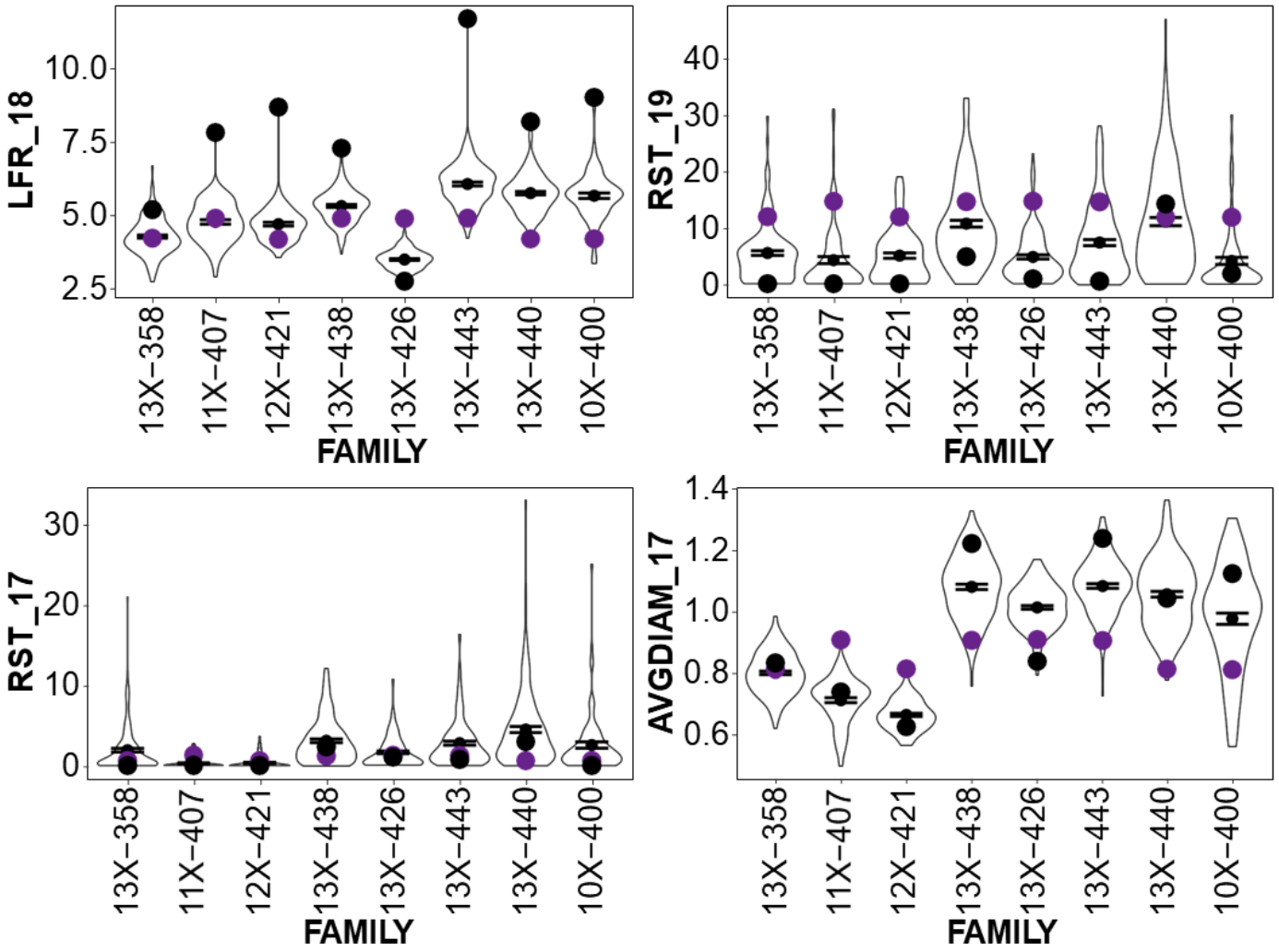

PLHC_19
75
50
25
0
PLHN_19
9
6
3
0
PLHN_17
40
30
20
10
0
PLHS_17
80
60
40
20
0
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY

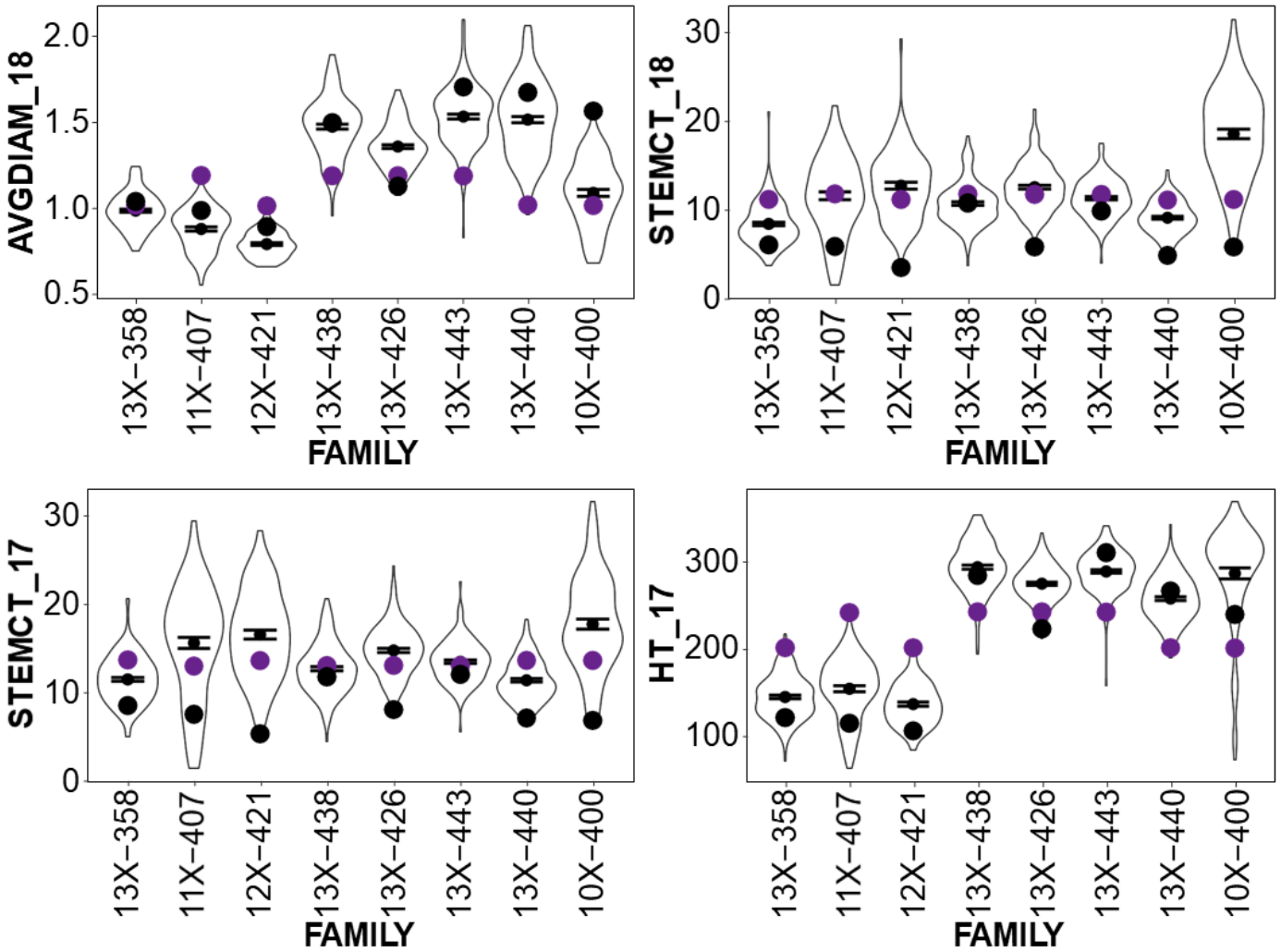

PLHS_19
100
75
50
25
0
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY
DLW_18
0.20
0.15
0.10
0.05
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY
DLW_17
0.3
0.2
0.1
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY
LA_17
30
20
10
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY

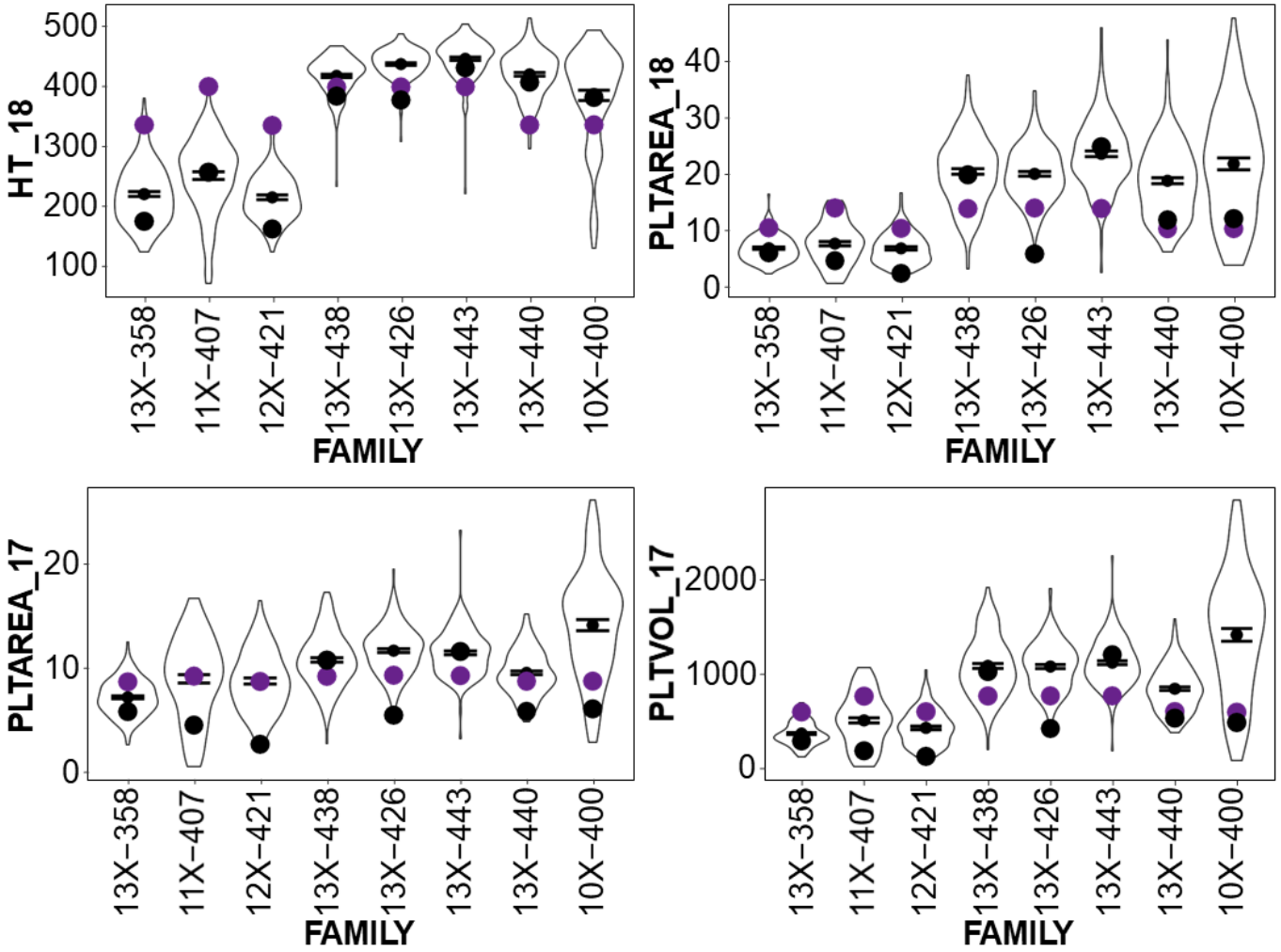

LA_18
SLA_18
SLA_17
LL_17
FAMILY
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400

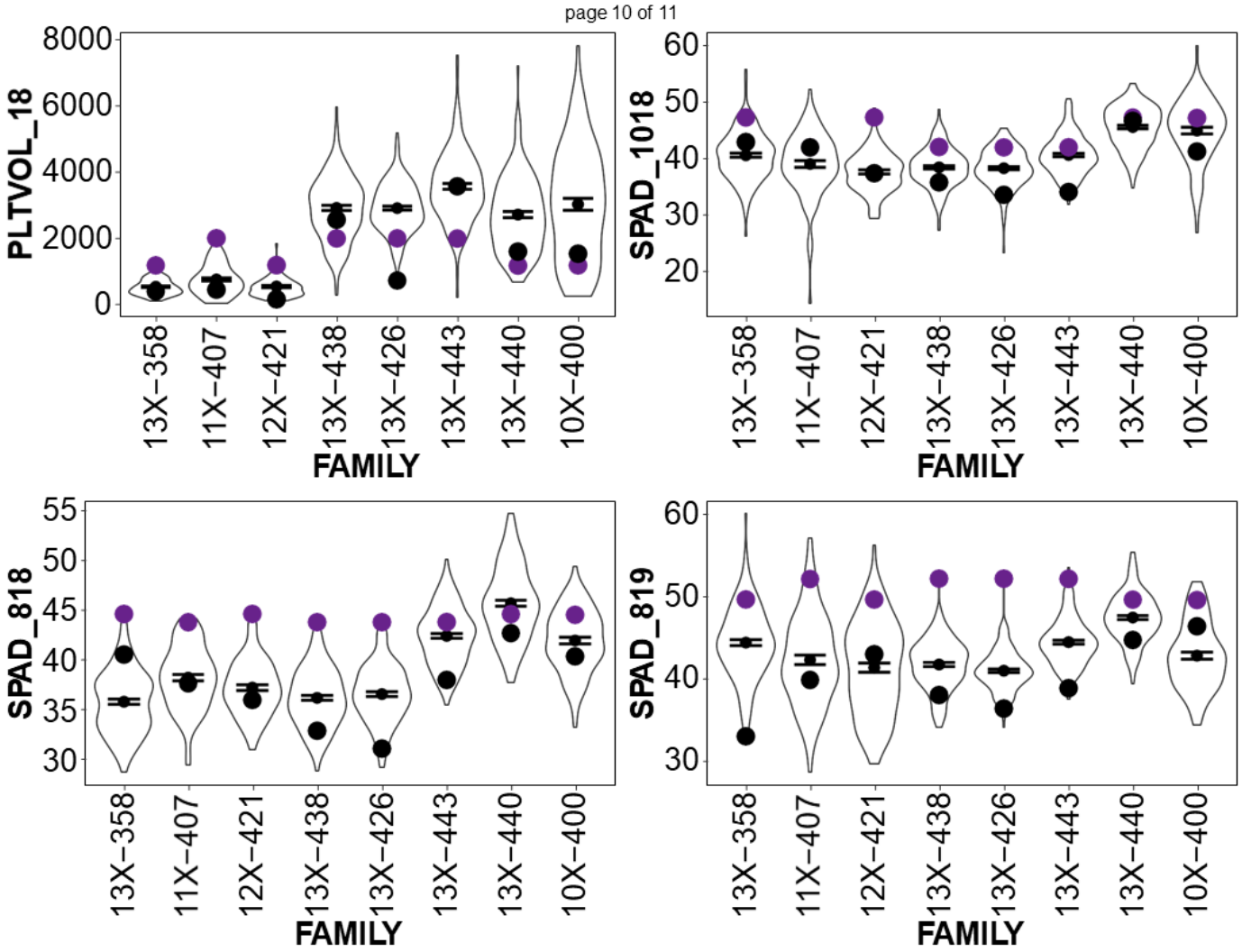

LL_18
12.5
10.0
7.5
5.0
LW_18
2.5
2.0
1.5
1.0
LW_17
3.5
3.0
2.5
2.0
1.5
LP_17
35
30
25
20
15
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY


LP_18
40
30
20
10
LFF_18
0.7
0.6
0.5
0.4
0.3
0.2
LFF_17
10
8
6
4
LFR_17
0.7
0.6
0.5
0.4
0.3
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY


LFR_18
10.0
7.5
5.0
2.5
RST_19
40
30
20
10
0
RST_17
30
20
10
0
AVGDIAM_17
1.4
1.2
1.0
0.8
0.6
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY


AVGDIAM_18
2.0
1.5
1.0
0.5
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY
STEMCT_18
30
20
10
0
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY
STEMCT_17
30
20
10
0
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY
HT_17
300
200
100
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY


HT_18
500
400
300
200
100
PLTAREA_18
40
30
20
10
0
PLTAREA_17
20
10
0
PLTVOL_17
2000
1000
0
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY


PLTVOL_18
8000
6000
4000
2000
0
SPAD_1018
60
50
40
30
20
SPAD_818
55
50
45
40
35
30
SPAD_819
60
50
40
30
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY

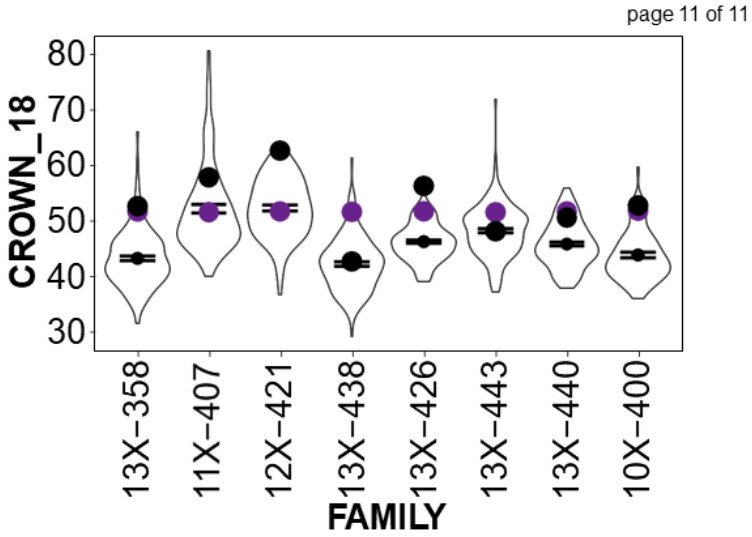
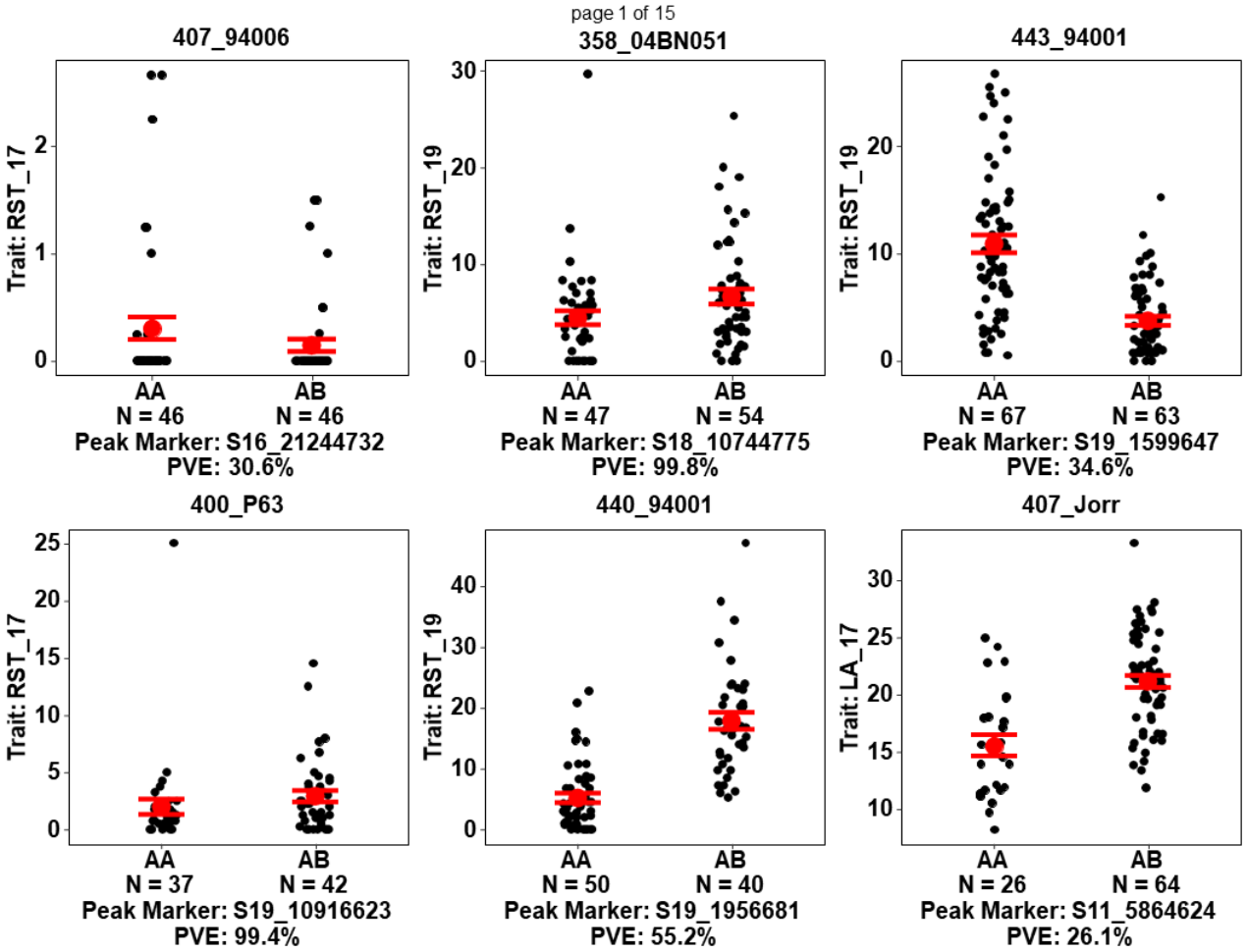
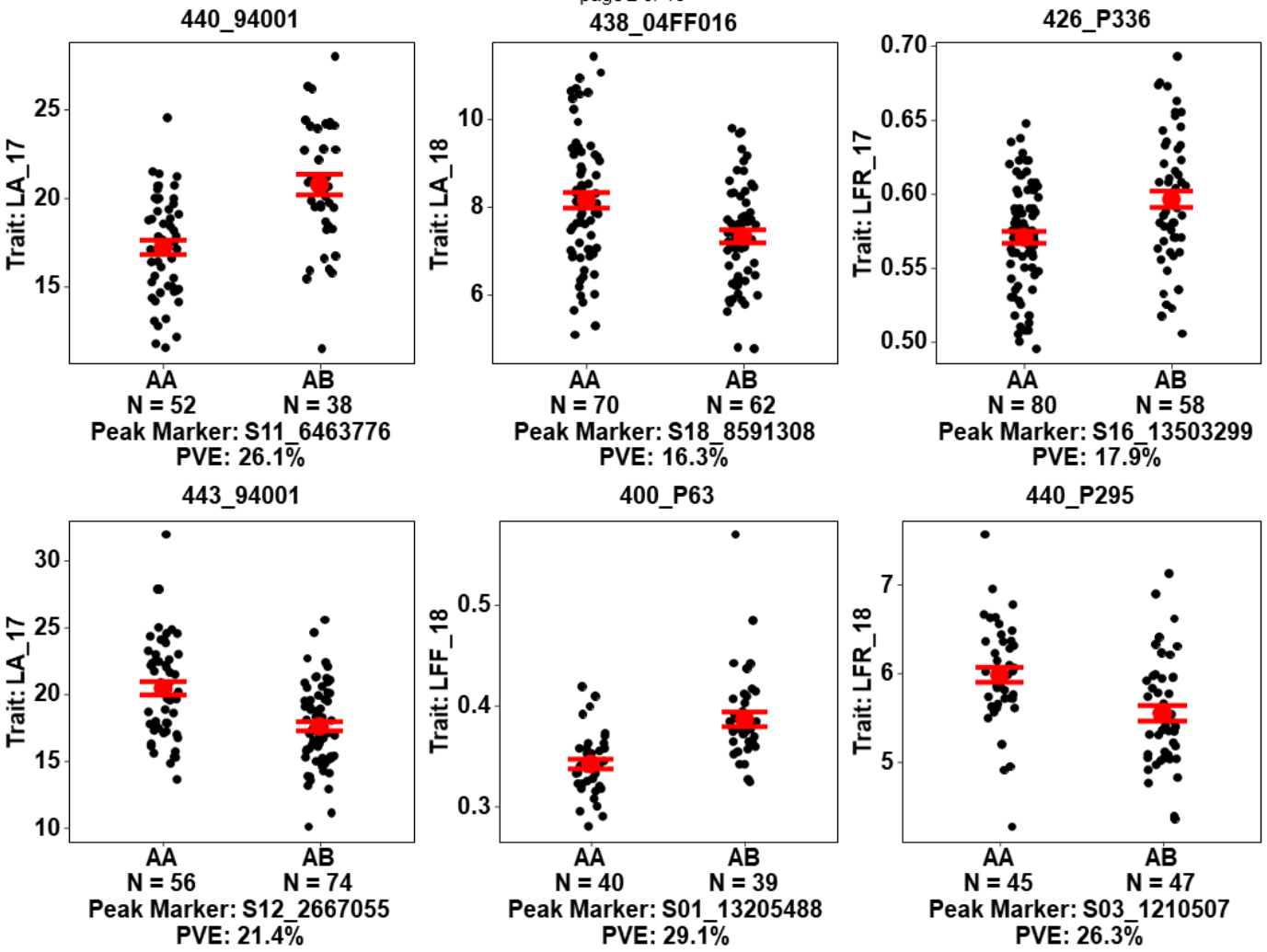
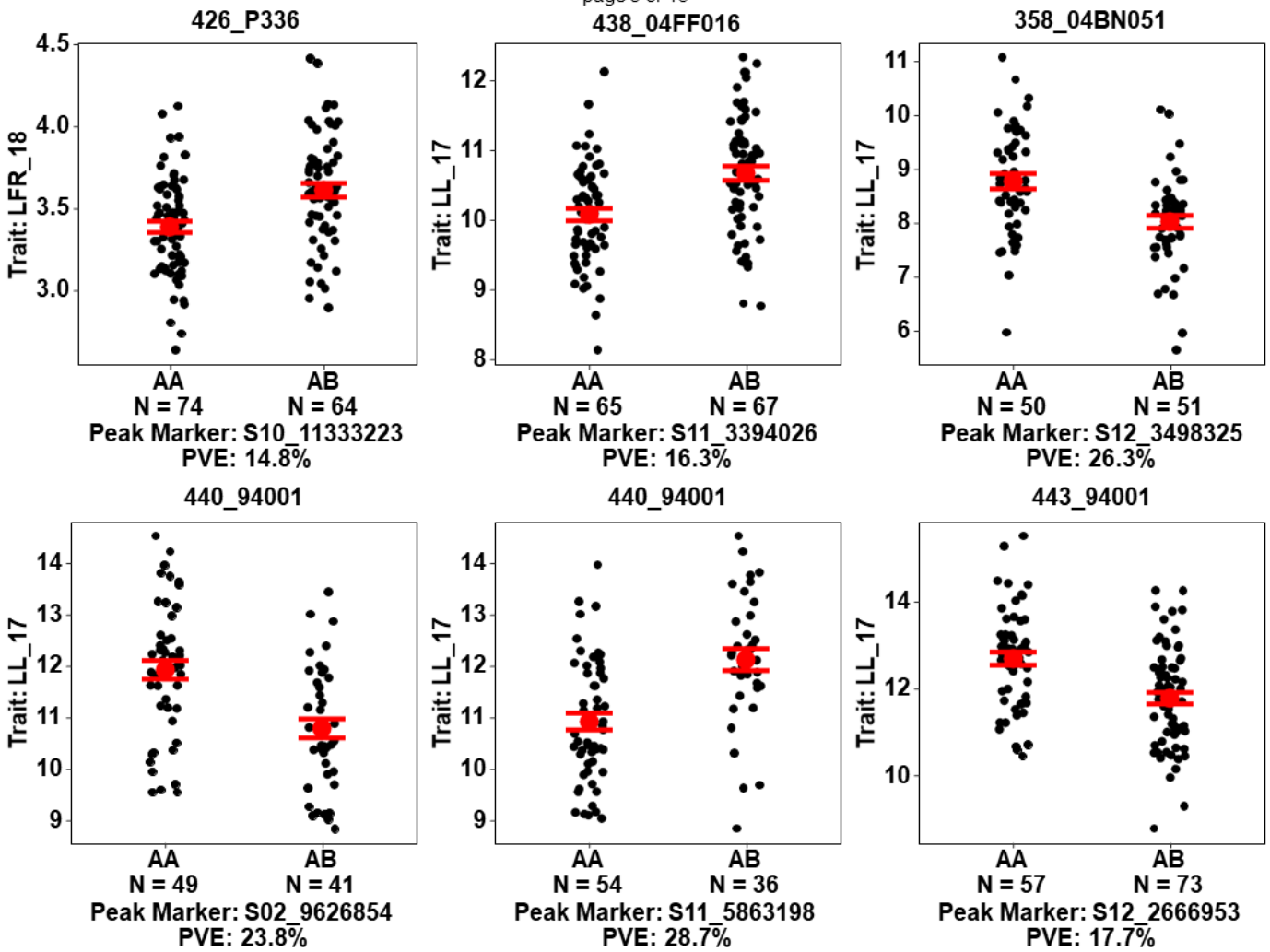
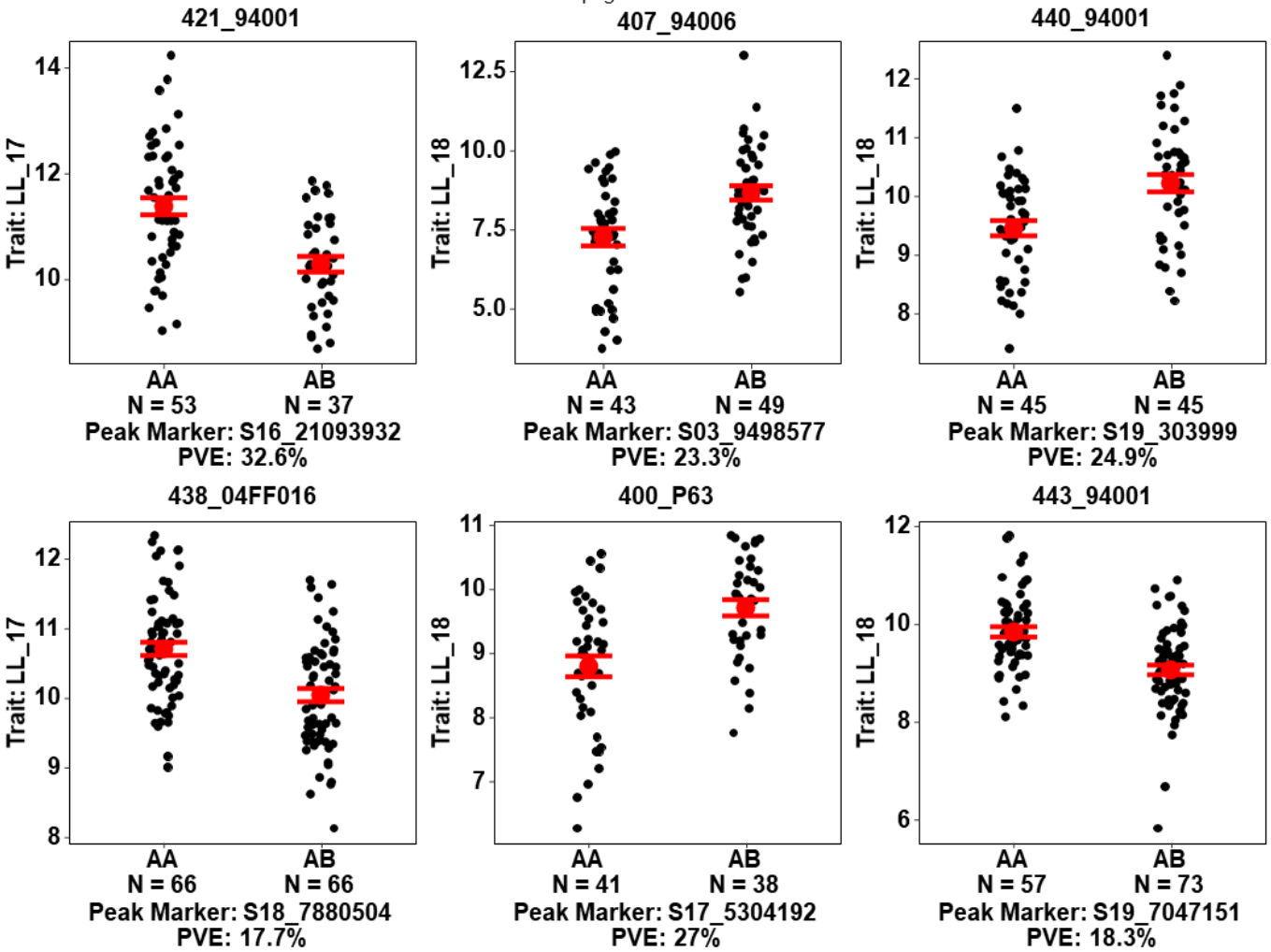

CROWN_18
80
70
60
50
40
30
13X−358
11X−407
12X−421
13X−438
13X−426
13X−443
13X−440
10X−400
FAMILY

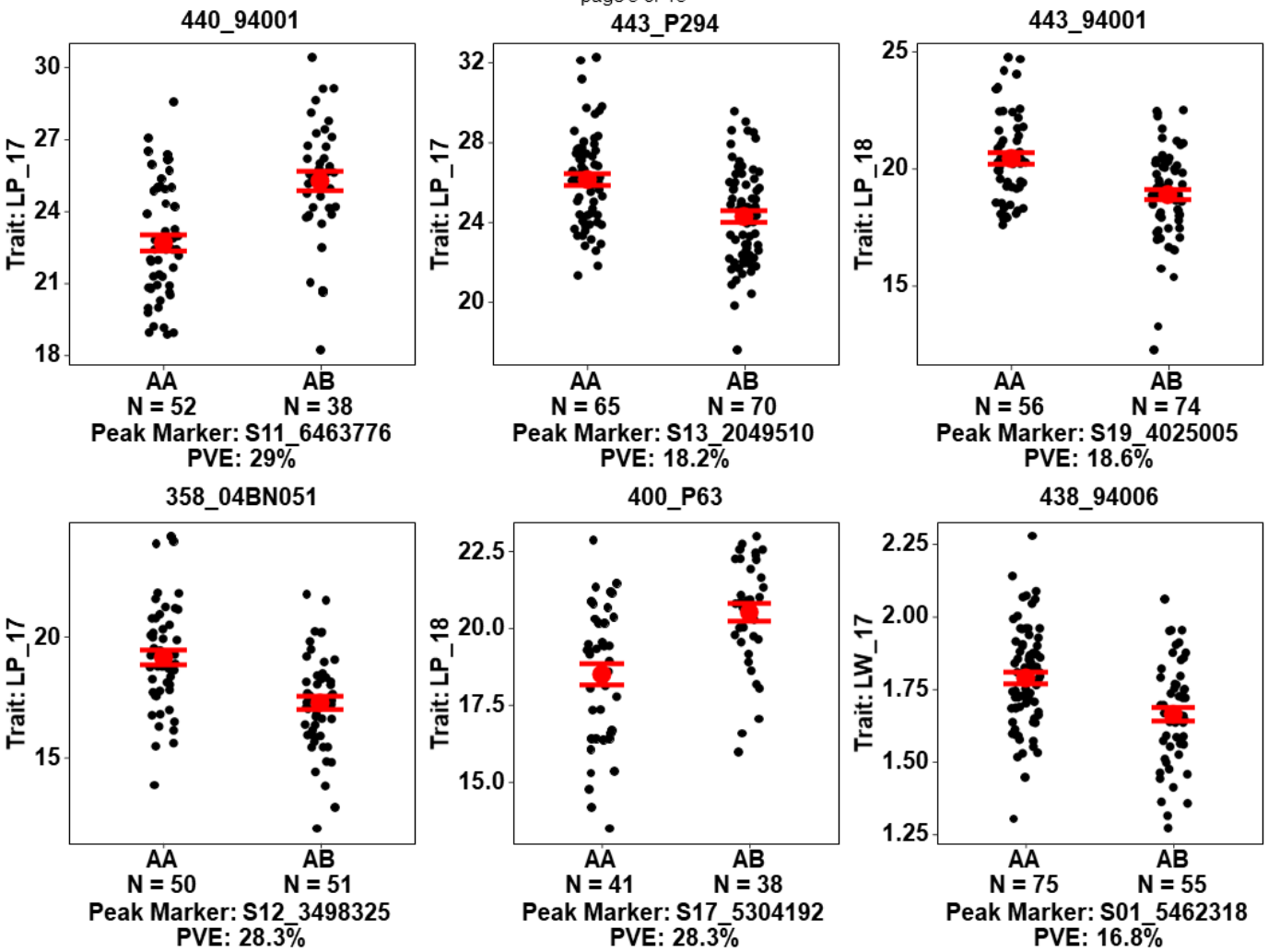
Figure A.5: The genotype by trait at the peak marker for each QTL. Red dots and brackets show the trait mean and standard errors. PVE; Percentage Variation Explained.

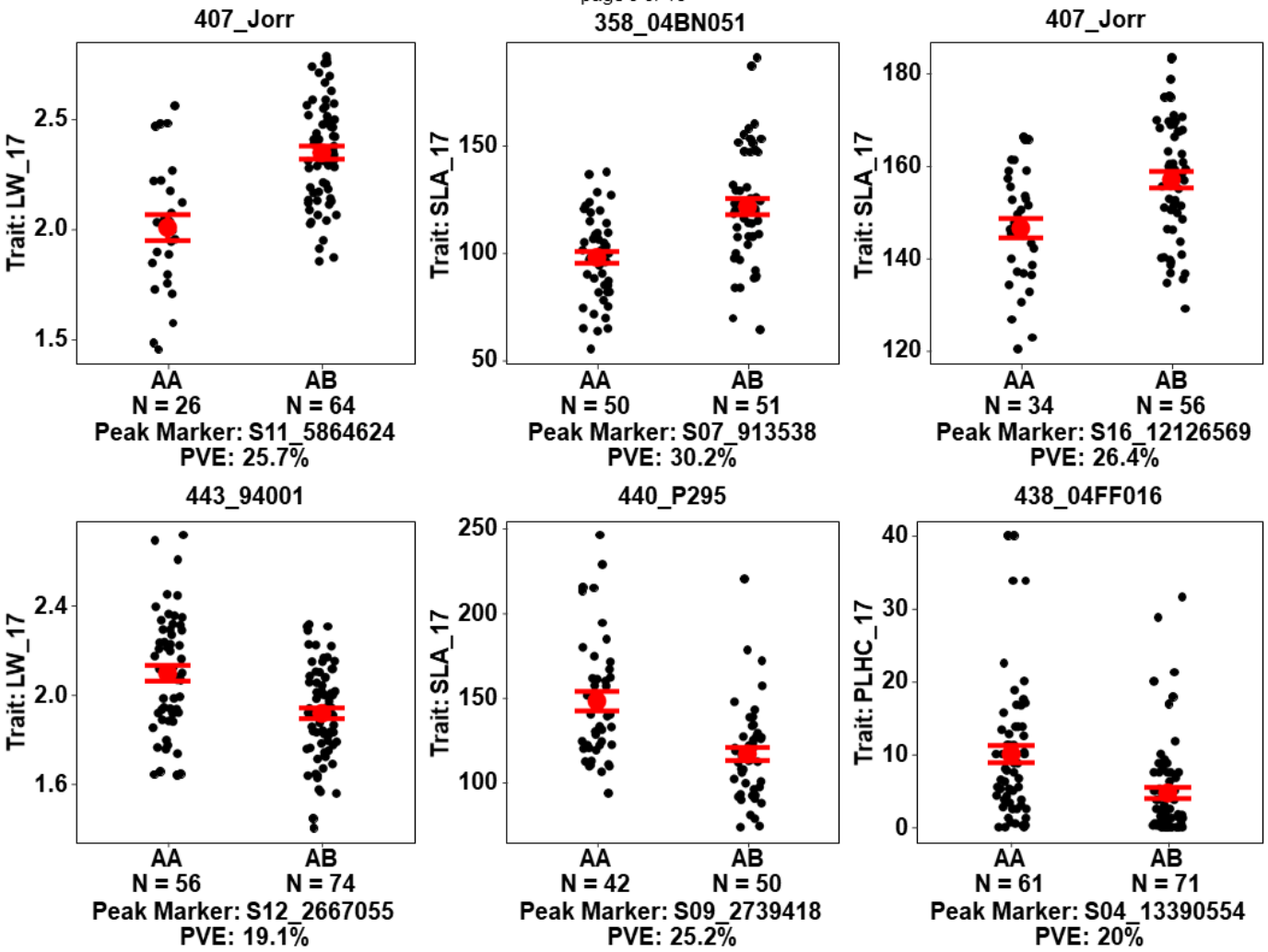


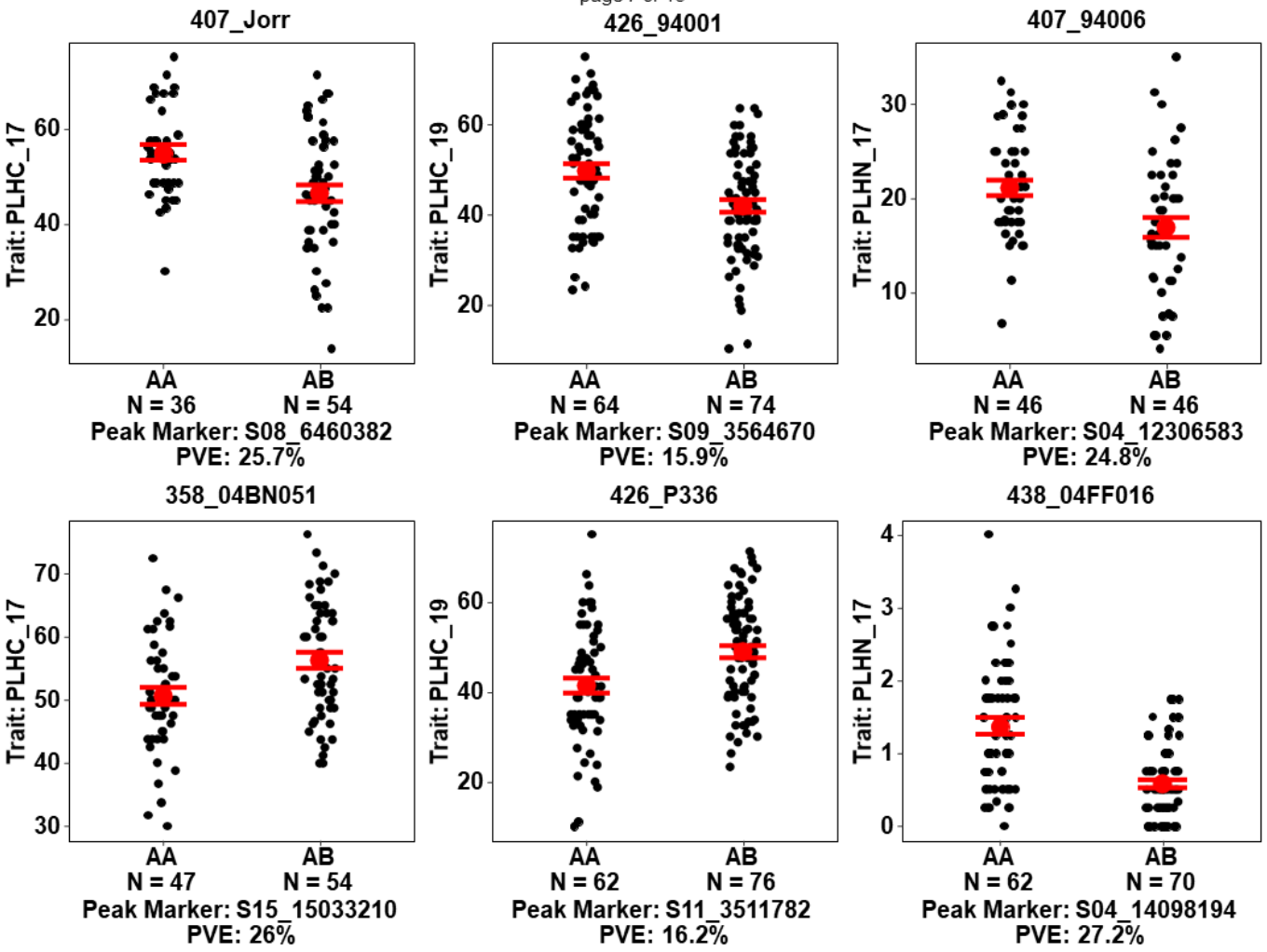

440_94001
Trait: LA_17
25
20
15
AA
N = 52
AB
N = 38
Peak Marker: S11_6463776
PVE: 26.1%
438_04FF016
Trait: LA_18
10
8
6
AA
N = 70
AB
N = 62
Peak Marker: S18_8591308
PVE: 16.3%
426_P336
Trait: LFR_17
0.70
0.65
0.60
0.55
0.50
AA
N = 80
AB
N = 58
Peak Marker: S16_13503299
PVE: 17.9%
443_94001
Trait: LA_17
30
25
20
15
10
AA
N = 56
AB
N = 74
Peak Marker: S12_2667055
PVE: 21.4%
400_P63
Trait: LFF_18
0.5
0.4
0.3
AA
N = 40
AB
N = 39
Peak Marker: S01_13205488
PVE: 29.1%
440_P295
Trait: LFR_18
7
6
5
AA
N = 45
AB
N = 47
Peak Marker: S03_1210507
PVE: 26.3%

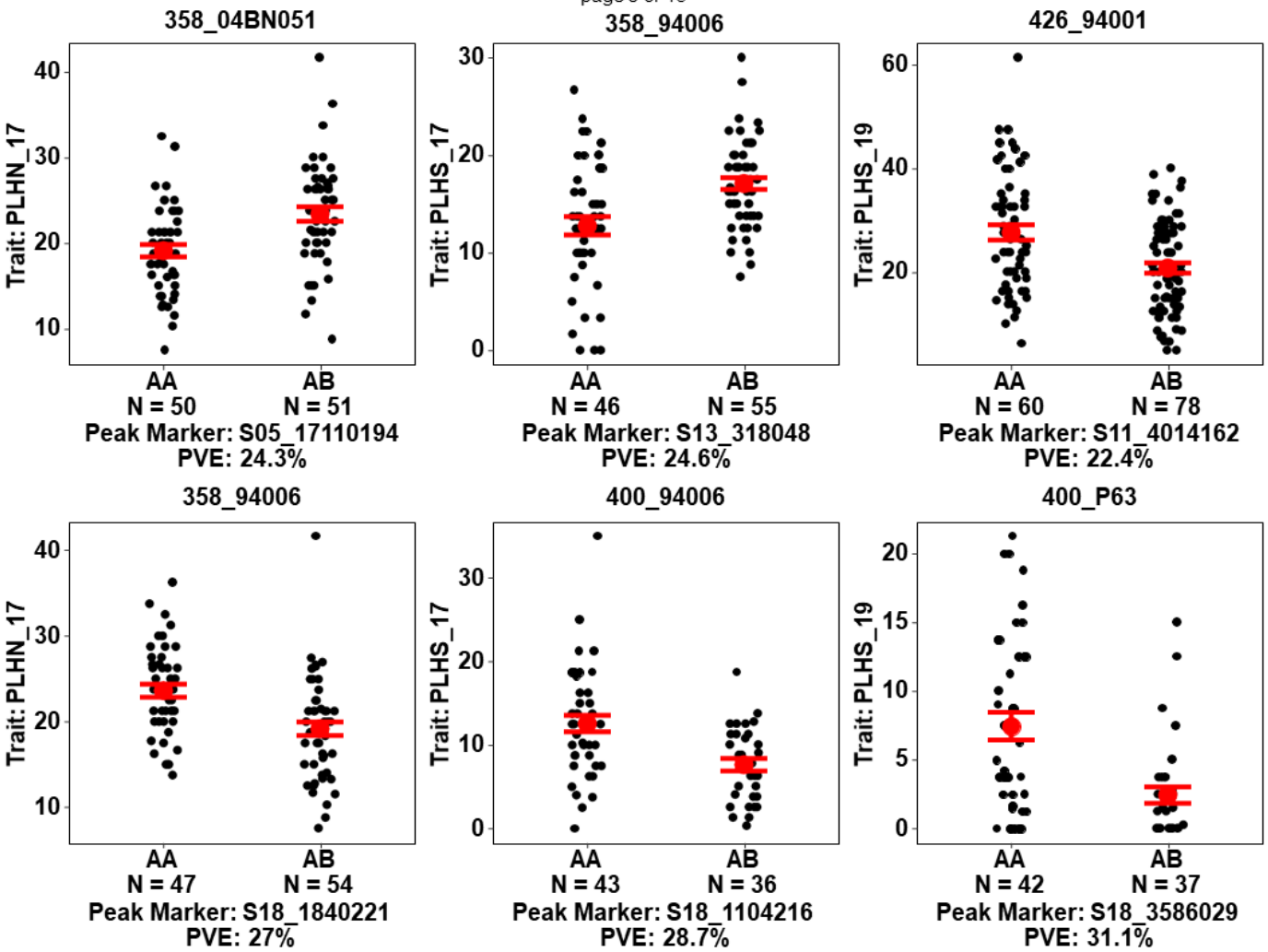

426_P336
Trait: LFR_18
AA
N = 74
AB
N = 64
Peak Marker: S10_11333223
PVE: 14.8%
438_04FF016
Trait: LL_17
AA
N = 65
AB
N = 67
Peak Marker: S11_3394026
PVE: 16.3%
358_04BN051
Trait: LL_17
AA
N = 50
AB
N = 51
Peak Marker: S12_3498325
PVE: 26.3%
440_94001
Trait: LL_17
AA
N = 49
AB
N = 41
Peak Marker: S02_9626854
PVE: 23.8%
440_94001
Trait: LL_17
AA
N = 54
AB
N = 36
Peak Marker: S11_5863198
PVE: 28.7%
443_94001
Trait: LL_17
AA
N = 57
AB
N = 73
Peak Marker: S12_2666953
PVE: 17.7%

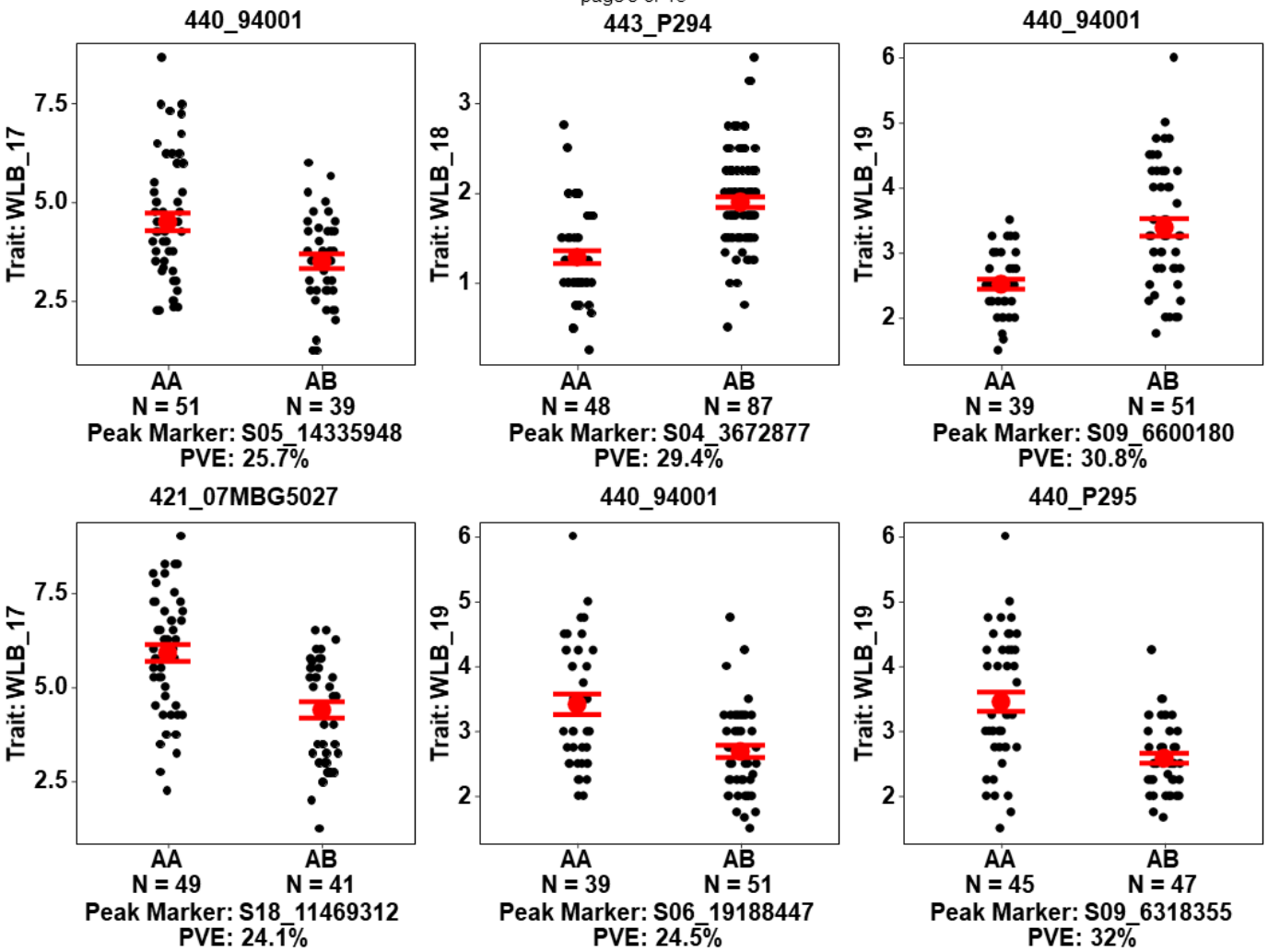

421_94001
Trait: LL_17
AA
N = 53
AB
N = 37
Peak Marker: S16_21093932
PVE: 32.6%
407_94006
Trait: LL_18
AA
N = 43
AB
N = 49
Peak Marker: S03_9498577
PVE: 23.3%
440_94001
Trait: LL_18
AA
N = 45
AB
N = 45
Peak Marker: S19_303999
PVE: 24.9%
438_04FF016
Trait: LL_17
AA
N = 66
AB
N = 66
Peak Marker: S18_7880504
PVE: 17.7%
400_P63
Trait: LL_18
AA
N = 41
AB
N = 38
Peak Marker: S17_5304192
PVE: 27%
443_94001
Trait: LL_18
AA
N = 57
AB
N = 73
Peak Marker: S19_7047151
PVE: 18.3%

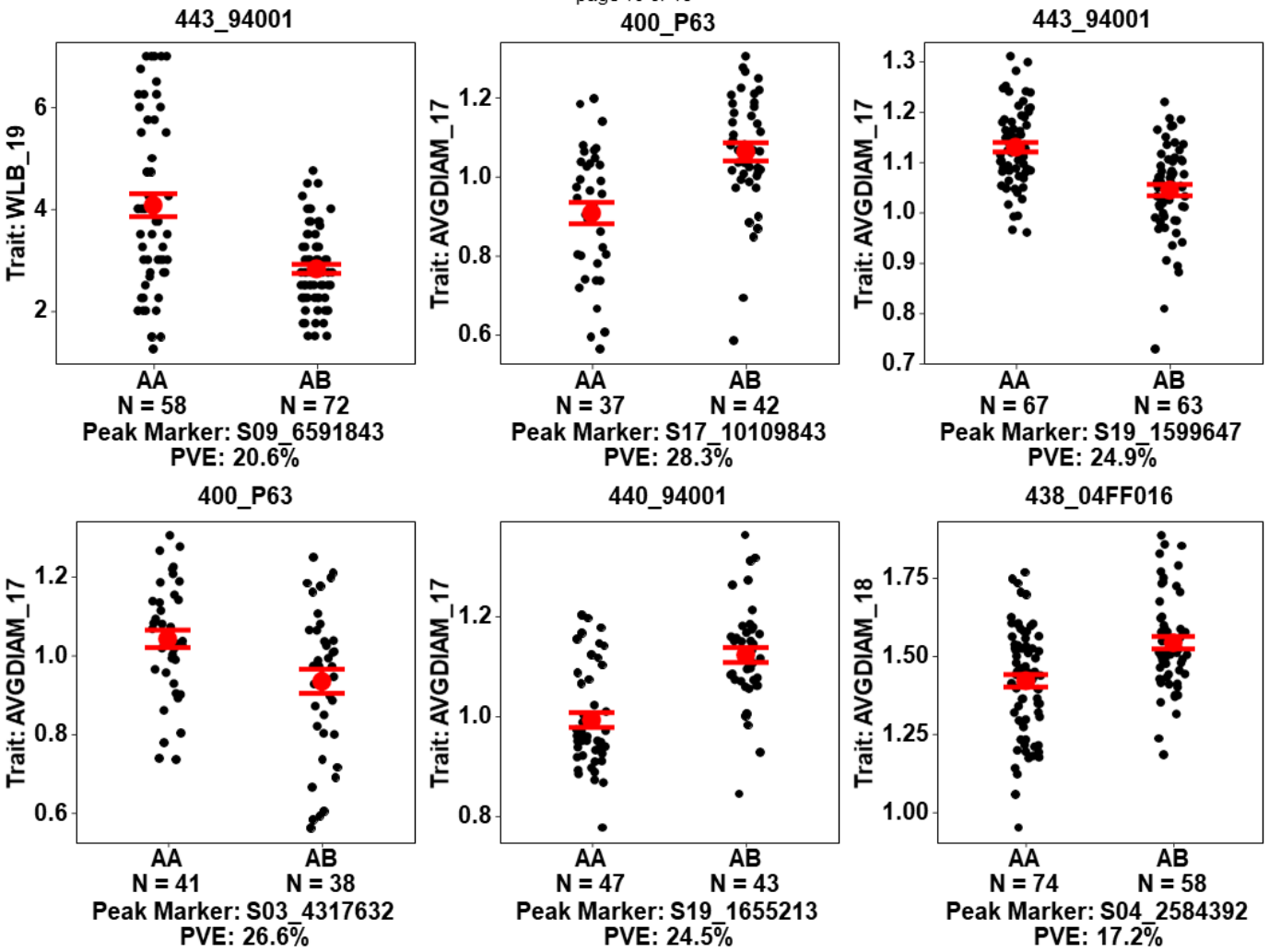

440_94001
Trait: LP_17
30
27
24
21
18
AA
N = 52
AB
N = 38
Peak Marker: S11_6463776
PVE: 29%
443_P294
Trait: LP_17
32
28
24
20
AA
N = 65
AB
N = 70
Peak Marker: S13_2049510
PVE: 18.2%
443_94001
Trait: LP_18
25
20
15
AA
N = 56
AB
N = 74
Peak Marker: S19_4025005
PVE: 18.6%
358_04BN051
Trait: LP_17
20
15
AA
N = 50
AB
N = 51
Peak Marker: S12_3498325
PVE: 28.3%
400_P63
Trait: LP_18
22.5
20.0
17.5
15.0
AA
N = 41
AB
N = 38
Peak Marker: S17_5304192
PVE: 28.3%
438_94006
Trait: LW_17
2.25
2.00
1.75
1.50
1.25
AA
N = 75
AB
N = 55
Peak Marker: S01_5462318
PVE: 16.8%

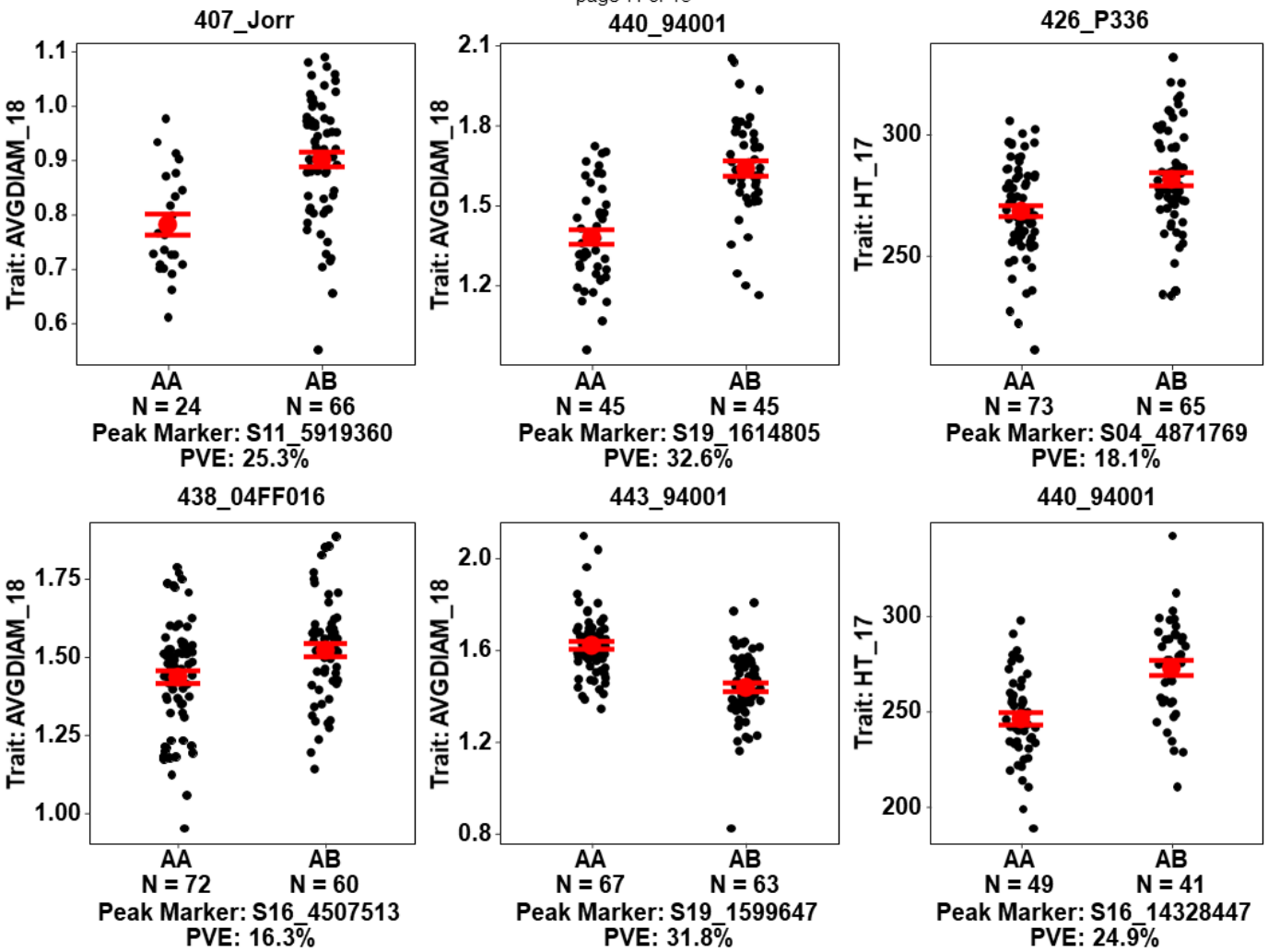

407_Jorr
Trait: LW_17
2.5
2.0
1.5
AA
N = 26
AB
N = 64
Peak Marker: S11_5864624
PVE: 25.7%
358_04BN051
Trait: SLA_17
150
100
50
AA
N = 50
AB
N = 51
Peak Marker: S07_913538
PVE: 30.2%
407_Jorr
Trait: SLA_17
180
160
140
120
AA
N = 34
AB
N = 56
Peak Marker: S16_12126569
PVE: 26.4%
443_94001
Trait: LW_17
2.4
2.0
1.6
AA
N = 56
AB
N = 74
Peak Marker: S12_2667055
PVE: 19.1%
440_P295
Trait: SLA_17
250
200
150
100
AA
N = 42
AB
N = 50
Peak Marker: S09_2739418
PVE: 25.2%
438_04FF016
Trait: PLHC_17
40
30
20
10
0
AA
N = 61
AB
N = 71
Peak Marker: S04_13390554
PVE: 20%

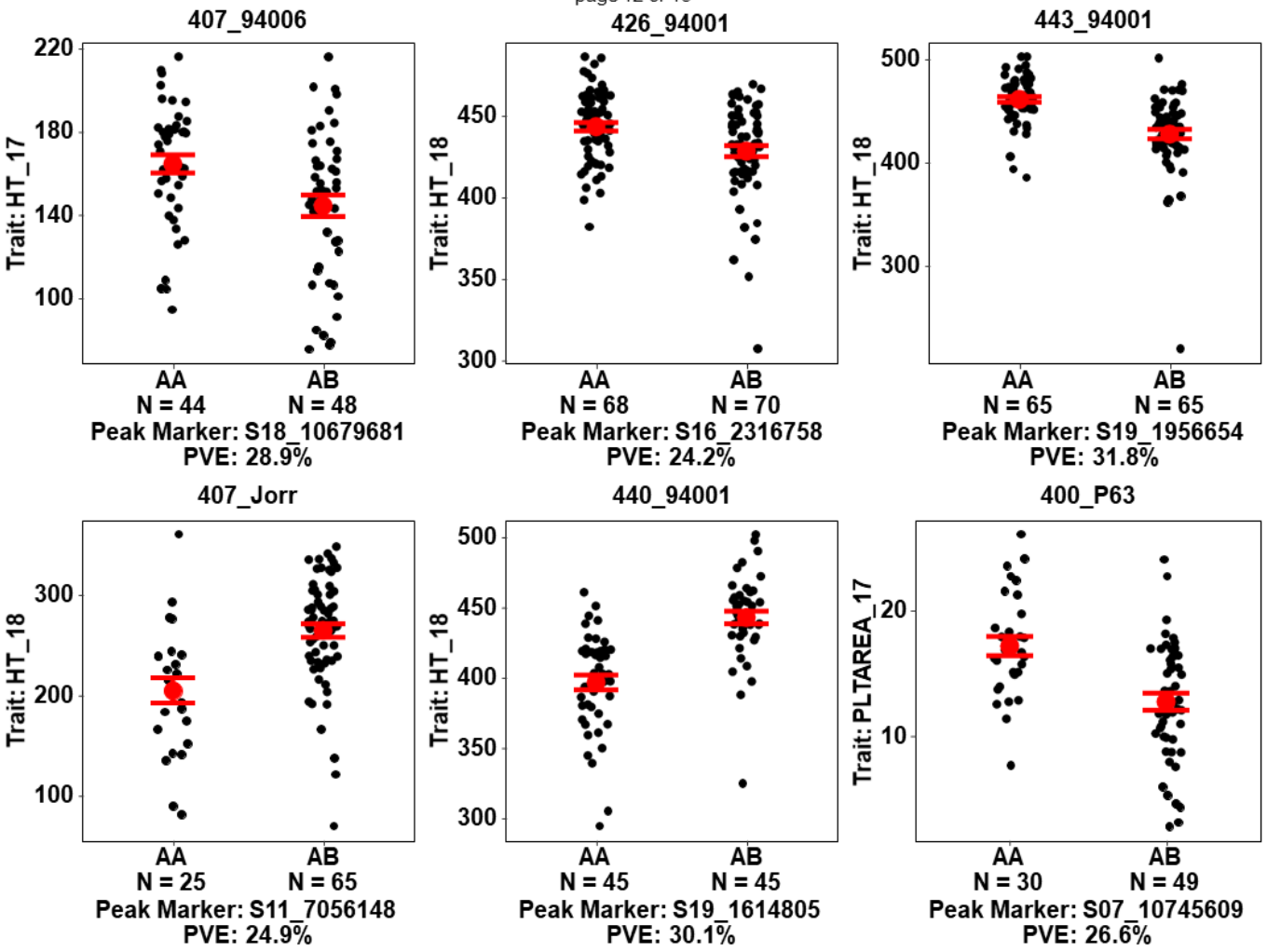

407_Jorr
Trait: PLHC_17
AA
N = 36
AB
N = 54
Peak Marker: S08_6460382
PVE: 25.7%
426_94001
Trait: PLHC_19
AA
N = 64
AB
N = 74
Peak Marker: S09_3564670
PVE: 15.9%
407_94006
Trait: PLHN_17
AA
N = 46
AB
N = 46
Peak Marker: S04_12306583
PVE: 24.8%
358_04BN051
Trait: PLHC_17
AA
N = 47
AB
N = 54
Peak Marker: S15_15033210
PVE: 26%
426_P336
Trait: PLHC_19
AA
N = 62
AB
N = 76
Peak Marker: S11_3511782
PVE: 16.2%
438_04FF016
Trait: PLHN_17
AA
N = 62
AB
N = 70
Peak Marker: S04_14098194
PVE: 27.2%

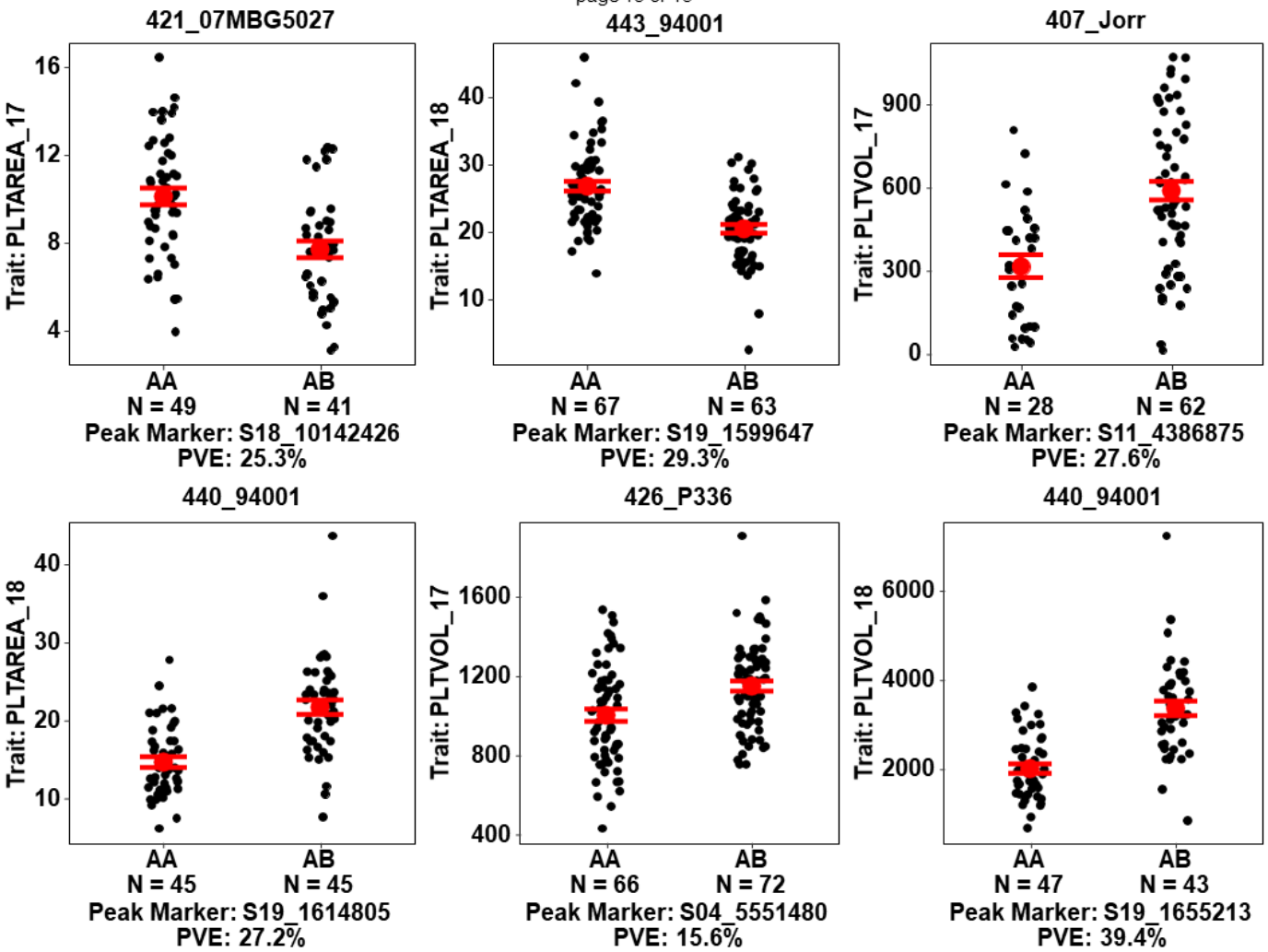

358_04BN051
Trait: PLHN_17
AA
N = 50
AB
N = 51
Peak Marker: S05_17110194
PVE: 24.3%
358_94006
Trait: PLHS_17
AA
N = 46
AB
N = 55
Peak Marker: S13_318048
PVE: 24.6%
426_94001
Trait: PLHS_19
AA
N = 60
AB
N = 78
Peak Marker: S11_4014162
PVE: 22.4%
358_94006
Trait: PLHN_17
AA
N = 47
AB
N = 54
Peak Marker: S18_1840221
PVE: 27%
400_94006
Trait: PLHS_17
AA
N = 43
AB
N = 36
Peak Marker: S18_1104216
PVE: 28.7%
400_P63
Trait: PLHS_19
AA
N = 42
AB
N = 37
Peak Marker: S18_3586029
PVE: 31.1%

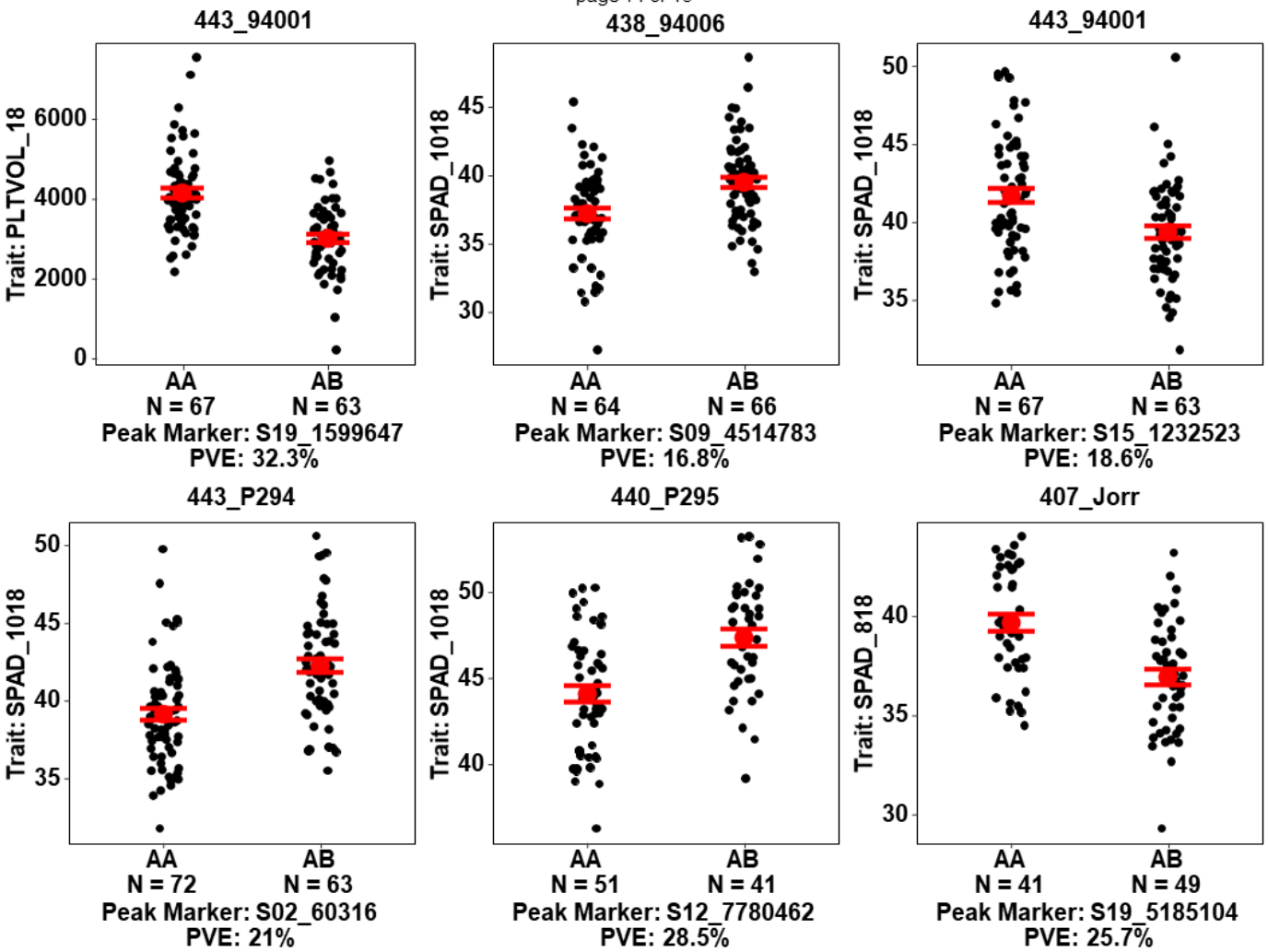

440_94001
Trait: WLB_17
7.5
5.0
2.5
AA
N = 51
AB
N = 39
Peak Marker: S05_14335948
PVE: 25.7%
443_P294
Trait: WLB_18
3
2
1
AA
N = 48
AB
N = 87
Peak Marker: S04_3672877
PVE: 29.4%
440_94001
Trait: WLB_19
6
5
4
3
2
AA
N = 39
AB
N = 51
Peak Marker: S09_6600180
PVE: 30.8%
421_07MBG5027
Trait: WLB_17
7.5
5.0
2.5
AA
N = 49
AB
N = 41
Peak Marker: S18_11469312
PVE: 24.1%
440_94001
Trait: WLB_19
6
5
4
3
2
AA
N = 39
AB
N = 51
Peak Marker: S06_19188447
PVE: 24.5%
440_P295
Trait: WLB_19
6
5
4
3
2
AA
N = 45
AB
N = 47
Peak Marker: S09_6318355
PVE: 32%

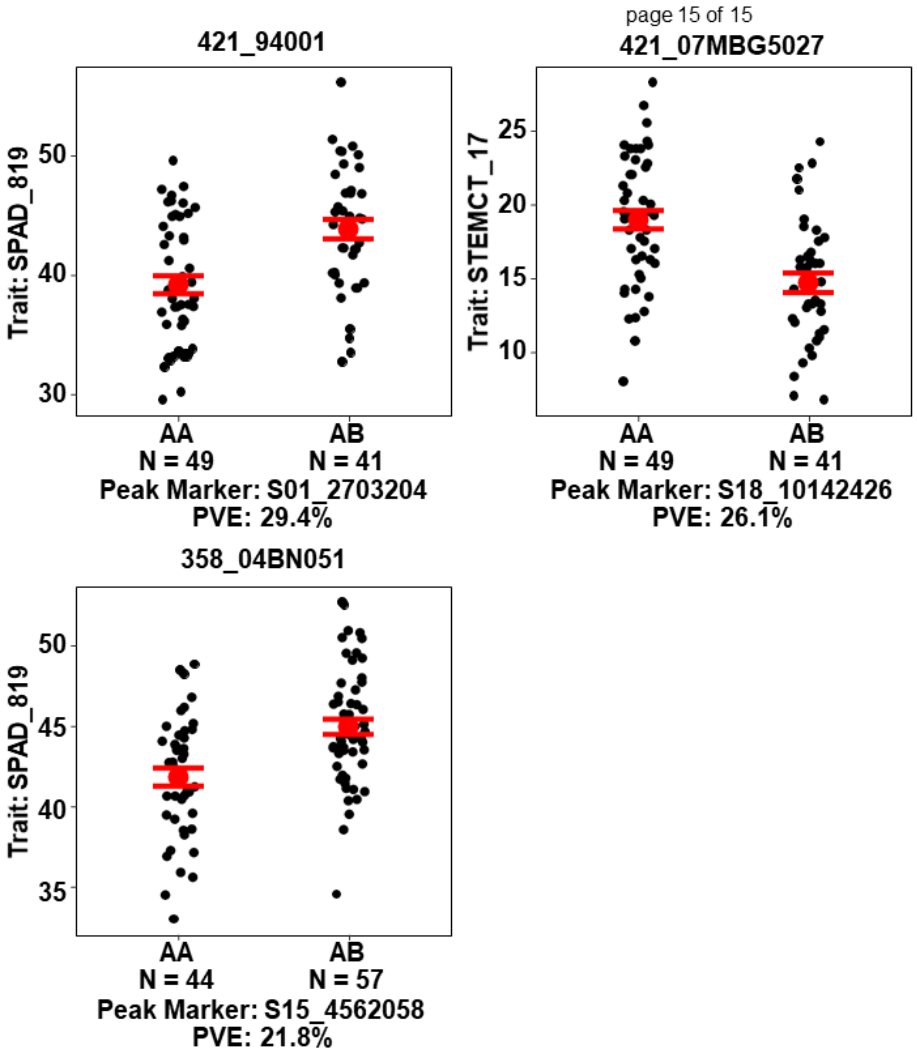

443_94001
Trait: WLB_19
AA
N = 58
AB
N = 72
Peak Marker: S09_6591843
PVE: 20.6%
400_P63
Trait: AVGDIAM_17
AA
N = 37
AB
N = 42
Peak Marker: S17_10109843
PVE: 28.3%
443_94001
Trait: AVGDIAM_17
AA
N = 67
AB
N = 63
Peak Marker: S19_1599647
PVE: 24.9%
400_P63
Trait: AVGDIAM_17
AA
N = 41
AB
N = 38
Peak Marker: S03_4317632
PVE: 26.6%
440_94001
Trait: AVGDIAM_17
AA
N = 47
AB
N = 43
Peak Marker: S19_1655213
PVE: 24.5%
438_04FF016
Trait: AVGDIAM_18
AA
N = 74
AB
N = 58
Peak Marker: S04_2584392
PVE: 17.2%


407_Jorr
Trait: AVGDIAM_18
AA
N = 24
AB
N = 66
Peak Marker: S11_5919360
PVE: 25.3%
440_94001
Trait: AVGDIAM_18
AA
N = 45
AB
N = 45
Peak Marker: S19_1614805
PVE: 32.6%
426_P336
Trait: HT_17
AA
N = 73
AB
N = 65
Peak Marker: S04_4871769
PVE: 18.1%
438_04FF016
Trait: AVGDIAM_18
AA
N = 72
AB
N = 60
Peak Marker: S16_4507513
PVE: 16.3%
443_94001
Trait: AVGDIAM_18
AA
N = 67
AB
N = 63
Peak Marker: S19_1599647
PVE: 31.8%
440_94001
Trait: HT_17
AA
N = 49
AB
N = 41
Peak Marker: S16_14328447
PVE: 24.9%


407_94006
Trait: HT_17
220
180
140
100
AA
N = 44
AB
N = 48
Peak Marker: S18_10679681
PVE: 28.9%
426_94001
Trait: HT_18
450
400
350
300
AA
N = 68
AB
N = 70
Peak Marker: S16_2316758
PVE: 24.2%
443_94001
Trait: HT_18
500
400
300
AA
N = 65
AB
N = 65
Peak Marker: S19_1956654
PVE: 31.8%
407_Jorr
Trait: HT_18
300
200
100
AA
N = 25
AB
N = 65
Peak Marker: S11_7056148
PVE: 24.9%
440_94001
Trait: HT_18
500
450
400
350
300
AA
N = 45
AB
N = 45
Peak Marker: S19_1614805
PVE: 30.1%
400_P63
Trait: PLTAREA_17
20
10
AA
N = 30
AB
N = 49
Peak Marker: S07_10745609
PVE: 26.6%


421_07MBG5027
Trait: PLTAREA_17
16
12
8
4
AA
N = 49
AB
N = 41
Peak Marker: S18_10142426
PVE: 25.3%
443_94001
Trait: PLTAREA_18
40
30
20
10
AA
N = 67
AB
N = 63
Peak Marker: S19_1599647
PVE: 29.3%
407_Jorr
Trait: PLTVOL_17
900
600
300
0
AA
N = 28
AB
N = 62
Peak Marker: S11_4386875
PVE: 27.6%
440_94001
Trait: PLTAREA_18
40
30
20
10
AA
N = 45
AB
N = 45
Peak Marker: S19_1614805
PVE: 27.2%
426_P336
Trait: PLTVOL_17
1600
1200
800
400
AA
N = 66
AB
N = 72
Peak Marker: S04_5551480
PVE: 15.6%
440_94001
Trait: PLTVOL_18
6000
4000
2000
AA
N = 47
AB
N = 43
Peak Marker: S19_1655213
PVE: 39.4%


443_94001
Trait: PLTVOL_18
6000
4000
2000
0
AA
N = 67
AB
N = 63
Peak Marker: S19_1599647
PVE: 32.3%
438_94006
Trait: SPAD_1018
45
40
35
30
AA
N = 64
AB
N = 66
Peak Marker: S09_4514783
PVE: 16.8%
443_94001
Trait: SPAD_1018
50
45
40
35
AA
N = 67
AB
N = 63
Peak Marker: S15_1232523
PVE: 18.6%
443_P294
Trait: SPAD_1018
50
45
40
35
AA
N = 72
AB
N = 63
Peak Marker: S02_60316
PVE: 21%
440_P295
Trait: SPAD_1018
50
45
40
AA
N = 51
AB
N = 41
Peak Marker: S12_7780462
PVE: 28.5%
407_Jorr
Trait: SPAD_818
40
35
30
AA
N = 41
AB
N = 49
Peak Marker: S19_5185104
PVE: 25.7%


421_94001
Trait: SPAD_819
50
40
30
AA
N = 49
AB
N = 41
Peak Marker: S01_2703204
PVE: 29.4%
421_07MBG5027
Trait: STEMCT_17
25
20
15
10
AA
N = 49
AB
N = 41
Peak Marker: S18_10142426
PVE: 26.1%
358_04BN051
Trait: SPAD_819
50
45
40
35
AA
N = 44
AB
N = 57
Peak Marker: S15_4562058
PVE: 21.8%