

DISSERTATION

AN NLR GENE LIKELY UNDERLYING *RMES1* PROVIDES GLOBAL SORGHUM RESISTANCE
BOLSTERED BY *RMES2*

Submitted by

Carl VanGessel

Department of Soil and Crop Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2023

Doctoral Committee

Advisor: Geoffrey Morris

Vamsi Nalam
Robyn Roberts
Esten Mason

Copyright by Carl VanGessel 2023

All Rights Reserved

ABSTRACT

AN NLR GENE LIKELY UNDERLYING *RMESI* PROVIDES GLOBAL SORGHUM RESISTANCE BOLSTERED BY *RMES2*

Breeding for aphid host plant resistance in sorghum has been an area of interest since the emergence of *Melanaphis sorghi* in North America a decade ago. In order to develop durable sorghum aphid resistance, breeders must be equipped with tools (trait package) and knowledge (molecular mechanisms) of host plant resistance. In this dissertation, I characterize the current state of sorghum aphid breeding and propose a genotype to phenotype map for the major source of global resistance, *Resistance to Melanaphis sorghi 1*. Relying on near-isogenic lines, I demonstrate that *RMESI* is applying selection pressure to sorghum aphid through reduction in fecundity that discriminates among aphid species. In global sorghum lines, *RMESI* is rare whereas a second resistance source, *RMES2*, is common and present in historic breeding germplasm. I mapped *RMES2* in Haitian breeding populations where it contributes fitness increases while lacking antagonistic pleiotropy and is selected for alongside *RMESI*. These results suggest breeding programs may unknowingly be deploying both sources of resistance which in combination are reducing the likelihood of *M. sorghi* biotype shifts to overcome *RMESI*. As aphid resistance may rely on phytochemical and/or induction with extended phenotypes regarding aphid populations, I used pan-genomic, transcriptomic, and metabolomic resources to describe the molecular mechanism of *RMESI*. Structural variation at the Chr06 locus underlies presence/absence variation of several nucleotide-binding leucine-rich repeat receptor (NLR) genes. Two of these candidate genes, SbPI276837.06G016400 and SbPI276837.06G016600, are representatives of two orthologous NLR groups which have genomic and transcriptomic evidence of underlying *RMESI* resistance. The *PAL* branch of the salicylic acid pathway is the primary phytohormone pathway responsible for *RMESI*-induced resistance. Finally, metabolome reorganization mirroring transcriptome changes suggest *RMESI*

is inducing multiple downstream mechanisms responsible for reducing aphid fecundity. While the causal gene underlying *RMESI* remains to be cloned and the eliciting aphid factor is unknown, this research suggests that gene-for-gene dynamics could lead to resistance-breaking biotype shifts and combining *RMESI* with additional resistance genes e.g. *RMES2*, will help achieve durability.

ACKNOWLEDGEMENTS

I would like to thank all the mentors that have helped me reach the end of my Ph.D. Drs. David Marshall and Gina Brown-Guedira first introduced me to quantitative genetics and breeding for biotic pest resistance, topics I will be researching for the remainder of my career. I'd also like to thank Dr. Stephen Pearce and the wheat lab I first joined at CSU for many great memories and lab skills I learned. And thank you to Dr. Geoff Morris for taking in an unhoused grad student and helping me develop as a scientist, showing me how to do rigorous and impactful research. Thank you to my lab mates, current and former, for your support, guidance, and patience.

Kaela, you inspire me to be a better scientist every day. Sharing this journey with you, ups and downs, has made it worthwhile. I can't wait for what is in store for us next.

To Frank, Claudia, and Gladis, thank you for always being there for me when I've needed it and humbling me as only siblings can when I don't.

Above all, I'd like to thank my first mentors who encouraged endless curiosity and have always been my greatest support. Mom and Dad, I could not ask for better role models in life and intellectual pursuits.

PREFACE

A third chapter is included but not discussed in the remainder of this dissertation. This work is from research done in Stephen Pearce's wheat genetics lab between 2019 and summer 2021. This work was published in Fall 2022 and can be found at <https://pubmed.ncbi.nlm.nih.gov/36241895/>.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
PREFACE	v
Chapter 1: Introduction	1
1.1 Importance of breeding for host plant resistance to insects in sorghum.....	1
1.2 Major aphid pests on cereals.....	1
1.3 HPR is a sustainable and equitable avenue for IPM.....	2
1.4 Historical aphid breeding for greenbug resistance.....	3
1.5 Lessons from greenbug breeding efforts.....	4
1.6 Current state of breeding for Sorghum aphid resistance.....	5
1.7 Sorghum aphid may be held at bay by at least two HPR genes sources.....	6
Chapter 1 Figures	8
Chapter 1 References	9
Chapter 2 - The globally deployed sorghum aphid resistance gene <i>RMES1</i> is vulnerable to biotype shifts but being bolstered by <i>RMES2</i>	16
2.1 Summary	16
2.2 Introduction.....	16
2.3 Material and Methods	19
2.3.1 Sorghum genotypes	19
2.3.2 Whole genome resequencing of NILs.....	19
2.3.3 Aphid Cultures	19
2.3.4 Choice assay	20
2.3.5 No-choice assay	20
2.3.6 Resequencing-based GWAS of community association panels.....	20
2.3.7 Population genomic analyses of <i>RMES2</i> in landraces and early breeding germplasm.....	21
2.3.8 Association analysis in Haitian breeding population.....	22
2.4 Results.....	22
2.4.1 Marker assisted back crossing isolates <i>RMES1</i> in near-isogenic lines.....	22
2.4.2 Sorghum aphid fecundity is lowered by <i>RMES1</i>	23
2.4.3 Sorghum aphid settling preference is not affected by <i>RMES1</i>	23
2.4.4 <i>RMES1</i> does not provide resistance to <i>Rhopalosiphum padi</i>	24
2.4.5 GWAS with resequencing finds quantitative resistance in global landraces but not <i>RMES1</i>	24
2.4.6 <i>RMES2</i> is common in global diversity lines	24
2.4.7 <i>RMES1</i> and <i>RMES2</i> provide resistance in Haitian breeding population	25
2.4.8 <i>RMES1</i> and <i>RMES2</i> are being selected for in breeding program.....	26
2.5 Discussion.....	26
Chapter 2 Tables	33
Chapter 2 Figures	34
Chapter 2 References	39
Chapter 3 - Genome to phenome characterization of <i>RMES1</i> resistance to sorghum aphid indicates NLR-based mechanism.....	46
3.1 Summary	46
3.2 Introduction.....	46
3.3 Material and Methods	48
3.3.1 NIL development.....	48
3.3.2 Aphid assays	48
3.3.3 Transcriptome sequencing.....	49
3.3.4 Sequencing data analysis.....	50
3.3.5 Structural Variation and Orthology analysis	51
3.3.6 Phytohormone and Metabolome quantification	51

3.4 Results.....	54
3.4.1 RMES1 NILs are appropriate for testing molecular mechanism.....	54
3.4.2 Resistant NILs undergo widespread transcriptomic changes after aphid infestation	54
3.4.3 Pangenomes reveal structural variation at RMES1	55
3.4.4 The salicylic acid pathway is differentially regulated along with phytohormone accumulation	57
3.4.5 Metabolite signatures are associated with resistance.....	59
3.5 Discussion.....	59
Chapter 3 Tables	64
Chapter 3 Figures	66
Chapter 3 References	76
Chapter 4 - Transcriptional signatures of wheat inflorescence development.....	84
4.1 Summary	84
4.2 Introduction.....	85
4.3 Material and Methods	87
4.3.1 Plant materials and growth conditions.....	87
4.3.2 RNA-seq library construction and sequencing	88
4.3.3 Transcription factors	89
4.3.4 Spike-dominant expression analysis.....	89
4.3.5 Principal Component Analysis (PCA), Differential Expression, and GO enrichment	90
4.3.6 Causal Structure Inference Network.....	92
4.3.7 Conversion of wheat, rice, and barley gene IDs.....	92
4.3.8 Enrichment analysis	93
4.3.9 QTL proximity and definition of homoeologous pairs	93
4.4 Results.....	93
4.4.1 Early wheat inflorescence development is defined by two major transcriptional shifts.....	93
4.4.2 Co-expression networks reveal predominant transcriptome profiles during inflorescence development	95
4.4.3 Inflorescence meristem development is associated with the down-regulation of RAV and TCP transcription factors.....	96
4.4.4 A small number of genes are transiently expressed during double ridge formation	97
4.4.5 Inflorescence transition and spike architecture genes are upregulated at W3.0.....	97
4.4.6 Inflorescence and spikelet meristem formation is associated with sequential activation of different classes of TFs	98
4.4.7 Gene regulatory networks predict high-confidence interactions between transcription factors.....	98
4.4.8 Identification of CLE/WOX genes expressed during wheat inflorescence development.....	101
4.5 Discussion.....	102
Chapter 4 Figures	106
Chapter 4 References	111
Chapter 5 – Conclusion	118
5.1 Summary	118

Chapter 1: Introduction

1.1 Importance of breeding for host plant resistance to insects in sorghum

Aphids are damaging pests on crops which have historically limited yields. There are more than 4,000 aphid species within the superfamily Aphidoidea (order Hemiptera) with many that can infest cereals (Dixon 1997; Blackman and Eastop 2000). Aphids feed on plants by piercing the epidermis with specialized mouthparts called stylets which probe for sieve elements and ingesting phloem sap (Nalam, Louis, and Shah 2019). In addition to removing photoassimilates from the plant, aphids indirectly impact crops by vectoring viral diseases and the excretion of ‘honeydew’, a by-product of nutritional imbalances in the phloem sap that can lead to mold and mechanical issues at harvest. Aphids can also modulate host responses in order to avoid defense responses and favor infestation. Aphids can reproduce asexually (parthenogenesis), have short and prolific generation times, and can produce winged morphs capable of long dispersal leading them to be among the fastest colonizers in the world (Wellings 1994; Quisenberry and Peairs 1998).

1.2 Major aphid pests on cereals

In the Great Plains, cereal production has been heavily impacted by several aphid species over the last century. Greenbug, *Schizaphis graminum* (Rondani), was first reported in 1884 and infests several grasses including wheat, barley, and sorghum (Webster and Phillips 1912). Greenbug vectors maize dwarf mosaic virus and barley yellow dwarf virus as well as predisposing plants to charcoal rot (*Macrophomina phaseolina*) (G. L. Teetes et al. 1973; Wallin and Loonan 1971; Daniels and Toler 1969). Sorghum aphid, *Melanaphis sorghi* (Zehntner) is a major pest of sorghum that can also infest sugarcane and other grasses (Armstrong et al. 2015). A sorghum aphid outbreak in 2011 was the most recent cereal pest to emerge in North America and heavily impacted sorghum production. By 2013, sorghum aphid had spread to all major growing regions of North America and yield loss could reach 50%-100% (Harris-Shultz, Armstrong, and Jacobson 2020; Michael J. Brewer et al. 2017). At the height of the sorghum aphid

outbreak in 2015, monetary losses could range between \$25 and \$175 per acre and regional impact to the Lower Rio Grande Valley alone was estimated to be \$31.6 million (Zapata et al. 2016; R. D. Bowling et al. 2016).

1.3 HPR is a sustainable and equitable avenue for IPM

Crop management for aphid pests can involve cultural practices, insecticide use, natural enemies, and host plant resistance. Reduced tillage, adjusted planting times, and removal of host-bridges are cultural practices designed to reduce pest populations and avoid peak flight windows (Harris-Shultz, Armstrong, and Jacobson 2020). These strategies can be effective when implemented but other crop production decisions may prevent farmers from doing so. Insecticides are an effective method of greenbug and sorghum aphid control and have been used in the Great Plains as seed treatments or applied foliarly (Harris-Shultz, Armstrong, and Jacobson 2020). While insecticide use is common in large-scale farm settings, it can be cost-prohibitive for small-holder farmers. Finally, natural enemies such as entomopathogenic fungi and aphidophagous insects can reduce aphid density but successful cases of their use are rare. Bird cherry oat aphid is a pest of wheat that is kept below economic thresholds thanks to natural predators like lady beetles, lacewings, hover flies, and parasitic wasps (Michaud 2008). In contrast for economically relevant pests, the release of several imported greenbug predators did not result in their permanent establishment and aphid control, likewise, entomopathogenic fungi can cause aphid population crashes when infestation levels are severe but commercial formulations have been ineffective in field environments (M.J. Brewer and Elliott 2004; James et al. 1998).

Genetic variation for resistance to insects, or HPR, is a consequence of competitive co-evolution between crops and pests that has resulted in a multitude of constitutive and induced, biochemical and physiological, or direct and indirect mechanisms to plant defense. Breeding for crop improvement under aphid pressure has focused on either resistance traits which reduce the amount of injury a plant sustains under infestation or tolerance traits which reduce damage (typically quantified in crop yield) despite successful infestation (Stout 2013). Tolerance is tractable in theory since it does not select on insect

populations, however the genetic basis is poorly understood and its likely polygenic nature makes it difficult to breed for (Peterson, Varella, and Higley 2017). Resistance has been characterized as either antibiosis which reduces fecundity or antixenosis which deters behavior resulting in infestation using no-choice and choice assays, respectively. The distinction between antibiosis and antixenosis can be blurred, such as a molecular deterrent which discourages settling behavior but also reduces life history traits when suitable alternative hosts are unavailable (Stout 2013). A more useful distinction for resistance is induced or constitutive defenses which respond to infestation or are pre-formed, respectively (Stout 2013). Host plant resistance, either antibiosis or antixenosis which reduces infestation and injury, by definition imposes selection pressure on insect populations which can result in biotype shifts to overcome genetic-based defenses. In this way, HPR and insecticides are vulnerable tools for IPM that can be lost when mismanaged or deployed in isolation. For HPR, it is critical to isolate the genetic basis and combine with additional HPR sources in order to build durable, oligo-genic resistance in gene-pyramids (MacIntosh 2019; Mundt 2018).

The use of multiple HPR sources in released varieties will not only give better crop protection, it will positively benefit the lifespan and durability of each HPR source. The current understanding of the vulnerability and durability of genetic resistance to biotic pests has largely been informed by breeding for plant pathogen resistance. Rust diseases on wheat have shown arms-race patterns of new virulent strains developing after corresponding resistance is widely deployed and subsequently overcome (Wan et al. 2007). The boom-and-bust cycle seen in greenbug breeding suggests some of the deployed HPR were contributing to a similar arms-race.

1.4 Historical aphid breeding for greenbug resistance

Greenbug has been a recurrent pest of sorghum and small grains in the Great Plains for several decades. Large-scale outbreaks occurred roughly every 5-10 years in the southern Great Plains over the second half of the 20th century (Kenneth J. Starks and Burton 1977; Harris-Shultz, Armstrong, and Jacobson 2020). When greenbug resistance breeding began in wheat in the mid 20th century, aphid

populations with differential ability to damage plants were identified as distinct *biotypes* (Puterka and Peters 1990). The first biotype which was identified as injurious to sorghum was biotype C in 1968 (Harvey and Hackerott 1969; Weng et al. 2010). Biotypes C was phenologically distinct from biotypes A and B with field populations peaking in late summer as opposed to cool season outbreaks of greenbug and the first identified to colonize sorghum and sudangrass (*S. sudanese* (Piper) Stapf.) (Harvey and Hackerott 1969). Three other important greenbug biotypes emerged since then with biotype E in 1980, biotype I in 1990, and biotype K in 1992 (Harvey et al. 1991; K. B. Porter, Peterson, and Vise 1982; Harvey, Wilde, and Kofoed 1997). The dominant greenbug biotypes in the southern great plains have shifted over the decades from C to E and E to I (Weng et al. 2010). A combination of improved HPR in sorghum varieties and cultural practices like natural enemies and insecticide has prevented large outbreaks of greenbug in the Great Plains (Michaud 2017; R. Bowling and Wilde 1996).

Greenbug biotype shifts coincided with the wide deployment of HPR sources in farmer varieties. The identification of biotypes E and I were based on their colonization of sorghum with biotype C and E HPR, respectively, when those sources of resistance were widespread. Host adaptation and geographic isolation appear to be significant drivers of greenbug biotypic diversity, however, whether biotypes are pre-adapted and present prior to selection pressure or arise after HPR is deployed via de novo genetic diversity is unknown (Weng et al. 2010). It can not be confirmed that biotype K was identified after biotype I HPR was released supporting the pre-adaptation hypothesis, and the deployment of HPR is not clearly linked to emergence of injurious biotypes (D. R. Porter et al. 1997). Despite the absence of new agricultural biotypes in the 21st century, the history of greenbug population shift and increasing sorghum acreage means that the potential for new injurious biotypes remains a threat.

1.5 Lessons from greenbug breeding efforts

Host plant resistance to aphids can be exceptionally rare and therefore it is critical to maintain HPR through proper stewardship (Harris-Shultz, Armstrong, and Jacobson 2020). HPR sources within a species can be extremely rare. After the emergence of biotype E, more than 23,000 accessions of sorghum

screened for resistance resulted in only six sources (Harvey et al. 1991). Another source of HPR is wild relatives which represent a wider diversity pool. Resistance to biotype C was first identified in a handful of sorghum lines, all of which were believed to trace from tunis grass (*S. virgatum* (Hackel) Stapf) (D. R. Porter et al. 1997). In wheat breeding for greenbug resistance, a rye (*Secale cereale* L.) translocation was the source of *Gb2* providing resistance to biotype B and C. Another source of HPR in wheat (*Gb3*, provides resistance to 6 different biotypes of greenbug) was derived from an *Aegilops tauschii* Coss. landrace (Hollenhorst and Joppa 1983). Screening large numbers of genotypes and introgression from wild relatives have been necessary for HPR identification but laborious and time consuming, impeding breeding efforts. Therefore, it is important to protect HPR sources that are already in use.

Genome-level phenotype comparisons have limited the mechanistic understanding of individual HPR sources. Aphid resistant genotypes are often described as one or a combination of antibiosis, antixenosis, and tolerance but whole-genotype comparisons prevent meaningful insights on the genotype to phenotype map. Several mapping studies identified between 8 and 9 QTLs for resistance to various biotypes of greenbug (Katsar et al. 2002a; Punnuri and Huang 2017; Nagaraj et al. 2005; Agrama et al. 2002). These are useful for dissecting the genetic architecture of resistance traits which whole genome comparisons cannot provide. Tx2783 also had a QTL for biotype C resistance on SBI-09 suggesting these could be shared by descent or the chromosome has more than one aphid resistance gene (Katsar et al. 2002b). Bloomless sorghum NILs showed nonpreference or antixenosis to greenbug at seedling stages (Weibel and Starks 1986).

1.6 Current state of breeding for Sorghum aphid resistance

Sorghum aphid was first described in 1904 in Sudan and has been known to cause heavy damage on sorghum in Africa (Balfour 1904; Vuillet and Vuillet 1914). The sorghum aphid outbreak originating in Beaumont, TX in 2013 was first identified as sugarcane aphid (*M. sacchari* (Zehntner, 1897)), however morphometric and molecular evidence have now distinguished sugarcane aphid from sorghum aphid, and identified sorghum aphid as the species in the recent outbreak (Nibouche et al. 2021; Villanueva et al.

2014). *M. sacchari* and *M. sorghii* have a host species preference and associated fitness increase for sugarcane and sorghum (Boukari et al. 2020; Paudyal, Armstrong, Harris-Shultz, et al. 2019; Nibouche et al. 2021). A survey of *Melanaphis spp.* samples found both *M. sorghii* and *M. sacchari* present in North, Latin, and South America as well as Africa and Asia, however only *M. sacchari* was present in North and Latin America prior to 2013 (Nibouche et al. 2021). By 2016, sorghum aphid had been identified in all U.S. sorghum growing regions and Mexico underlining its rapid spread in suitable environments (Peterson et al. 2018; R. D. Bowling et al. 2016). North American sorghum production has recovered from the significant impact of sorghum aphid. Sorghum production in the Rio Grande Valley saw revenues in 2014 and 2015 reduced by \$50/acre and total production in Haiti dropped from 100,000 tonnes to 30,000 between 2015 and 2018 (USDA-FAS 2020; Zapata et al. 2018). The identification of HPR to sorghum aphid and deployment in commercial lines has contributed to the rebound of sorghum production (Muleta et al. 2022).

1.7 Sorghum aphid may be held at bay by at least two HPR genes sources

At least two sources of sorghum aphid HPR have been identified and are likely in use. *RMES1* (Wang et al. 2013; Muleta et al. 2022) and *RMES2* (putatively *SbWRKY86*) (Poosapati et al. 2022) are the only QTLs identified which provide sorghum aphid resistance. *RMES1* is likely the main source of resistance to sorghum aphid in North America as it is the only loci with a publicly available marker (Muleta et al. 2022). The Haitian variety Papepichon was released in 2017, while Tx2783 has been common in US breeding programs, both of which contain *RMES1* (Muleta et al. 2022). Tx2783 and the hybrid DKS 37-07 are resistant to sorghum aphid which were relatively common in breeders and farmers fields due to having greenbug biotypes C and E resistance as well (Limaje et al. 2018; Peterson et al. 1984; Szczepaniec et al. 2018). *RMES1* is a simply inherited major resistance gene on SBI-06 (Wang et al. 2013; Muleta et al. 2022). The resistance allele was traced to Ethiopian landraces, including accession PI257599, but was disseminated in US programs via the conversion line SC110, and later Tx2783. Tx2783 has a high degree of tolerance and antibiosis to sorghum aphid (Paudyal, Armstrong, Giles, et al.

2019; Armstrong et al. 2015). This resistance may be distinct from greenbug HPR (derived from Capbam) considering SC110 is also in the pedigree of Tx2783 (Muleta et al. 2022; Harris-Shultz, Armstrong, and Jacobson 2020).

The putative *SbWRKY86*, or conservatively discussed as *RMES2*, expression is associated with aphid resistance but the extent to which beneficial alleles are used in breeding programs to increase HPR is unclear (Poosapati et al. 2022). This source of HPR was identified in germplasm consisting of the Sorghum Association Panels and Sorghum Bioenergy Panels and may be widespread in sorghum global germplasm. This transcription factor regulated carbon-nitrogen metabolism, callose deposition, and other defense mechanisms associated with antibiosis when heterologously expressed in Arabidopsis (Poosapati et al. 2022).

Chapter 1 Figures

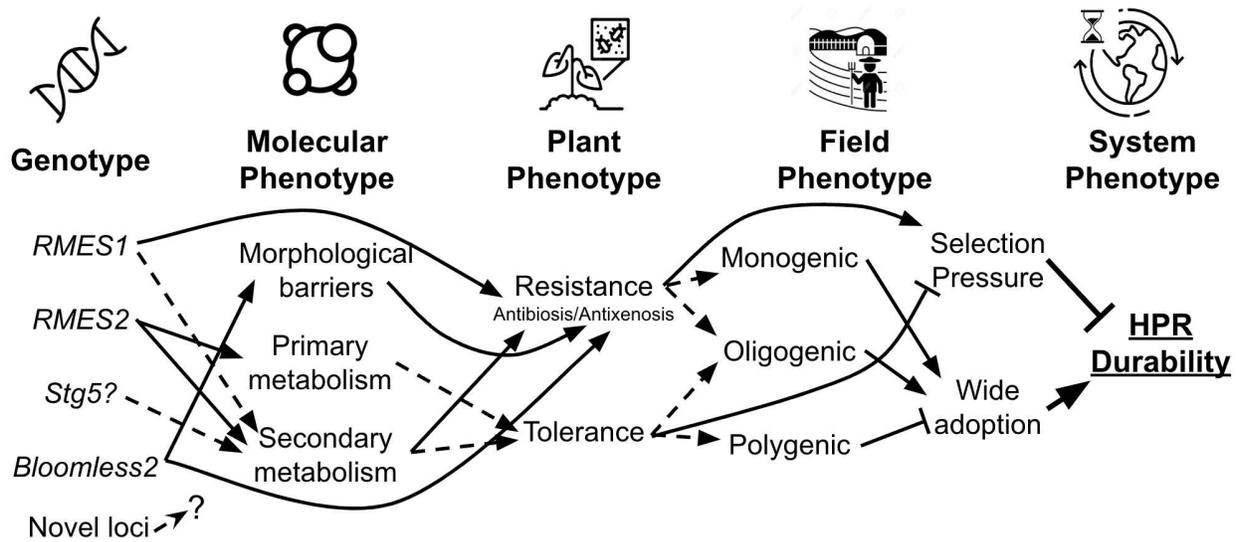


Figure 1.1 - Hypothetical genotype to phenotype map for sorghum aphid resistance

Chapter 1 References

- Adrianna Szczepanec. 2018. “Interactive Effects of Crop Variety, Insecticide Seed Treatment, and Planting Date on Population Dynamics of Sugarcane Aphid (*Melanaphis Sacchari*) and Their Predators in Late-Colonized Sorghum.” *Crop Protection* 109 (July): 72–79.
<https://doi.org/10.1016/j.cropro.2018.03.002>.
- Agrama, H., G. Widle, J. Reese, L. Campbell, and M. Tuinstra. 2002. “Genetic Mapping of QTLs Associated with Greenbug Resistance and Tolerance in Sorghum Bicolor.” *Theoretical and Applied Genetics* 104 (8): 1373–78. <https://doi.org/10.1007/s00122-002-0923-3>.
- Andrews, David J., Paula J. Bramel-Cox, and Gerald E. Wilde. 1993. “New Sources of Resistance to Greenbug, Biotype I, in Sorghum.” *Crop Science* 33 (1): crops1993.0011183X003300010035x.
<https://doi.org/10.2135/cropsci1993.0011183X003300010035x>.
- Armstrong, Scott J., William L. Rooney, Gary C. Peterson, Raul T. Villeneuve, Michael J. Brewer, and Danielle Sekula-Ortiz. 2015. “Sugarcane Aphid (Hemiptera: Aphididae): Host Range and Sorghum Resistance Including Cross-Resistance From Greenbug Sources.” *Journal of Economic Entomology* 108 (2): 576–82. <https://doi.org/10.1093/jee/tou065>.
- Balfour, Andrew. 1904. Report of the Wellcome Research Laboratories at the Gordon Memorial College Khartoum, By... Andrew Balfour. Departm. of Education, Sudan Government.
- Blackman, R. L., and V. F. Eastop. 2000. “Aphids on the World’s Crops: An Identification and Information Guide.” *Aphids on the World’s Crops: An Identification and Information Guide.*, no. Ed. 2. <https://www.cabdirect.org/cabdirect/abstract/20001106823>.
- Boukari, Wardatou, Chunyan Wei, Lihua Tang, Martha Hincapie, Moramay Naranjo, Gregg Nuessly, Julien Beuzelin, Sushma Sood, and Philippe Rott. 2020. “Lack of Transmission of Sugarcane Yellow Leaf Virus in Florida from Columbus Grass and Sugarcane to Sugarcane with Aphids or Mites.” *PLOS ONE* 15 (3): e0230066. <https://doi.org/10.1371/journal.pone.0230066>.

- Bowling, R., and G. Wilde. 1996. "Mechanisms of Resistance in Three Sorghum Cultivars Resistant to Green Bug (Homoptera: Aphididae) Biotype I." *Journal of Economic Entomology* 89 (2): 558–61. <https://doi.org/10.1093/jee/89.2.558>.
- Bowling, Robert D., Michael J. Brewer, David L. Kerns, John Gordy, Nick Seiter, Norman E. Elliott, G. David Buntin, et al. 2016. "Sugarcane Aphid (Hemiptera: Aphididae): A New Pest on Sorghum in North America." *Journal of Integrated Pest Management* 7 (1): 12. <https://doi.org/10.1093/jipm/pmw011>.
- Brewer, Michael J., John W. Gordy, David L. Kerns, James B. Woolley, William L. Rooney, and Robert D. Bowling. 2017. "Sugarcane Aphid Population Growth, Plant Injury, and Natural Enemies on Selected Grain Sorghum Hybrids in Texas and Louisiana." *Journal of Economic Entomology* 110 (5): 2109–18. <https://doi.org/10.1093/jee/tox204>.
- Brewer, M.J., and N.C. Elliott. 2004. "Biological Control of Cereal Aphids in North America and Mediating Effects of Host Plant and Habitat Manipulations." *Annual Review of Entomology* 49 (1): 219–42. <https://doi.org/10.1146/annurev.ento.49.061802.123149>.
- Daniels, Norris E., and Robert W. Toler. 1969. "Transmission of Maize Dwarf Mosaic by the Greenbug, *Schizaphis Graminum*." *The Plant Disease Reporter* 53: 59–61.
- Dixon, A. F., and G. Aphid Ecology. "An Optimization Approach." (1998).
- Harris-Shultz, Karen, Scott Armstrong, and Alana Jacobson. "Invasive cereal aphids of North America: Biotypes, genetic variation, management, and lessons learned." *Trends Entomol* 15 (2019): 99–122.
- Harvey, T. L., and H. L. Hackerott. 1969. "Recognition of a Greenbug Biotype Injurious to Sorghum1." *Journal of Economic Entomology* 62 (4): 776–79. <https://doi.org/10.1093/jee/62.4.776>.
- Harvey, T. L., K. D. Kofoid, T. J. Martin, and P. E. Sloderbeck. 1991. "A New Greenbug Virulent to E-Biotype Resistant Sorghum." *Crop Science* 31 (6): crops1991.0011183X003100060062x. <https://doi.org/10.2135/cropsci1991.0011183X003100060062x>.

- Harvey, T. L., G. E. Wilde, and K. D. Kofoid. 1997. "Designation of a New Greenbug, Biotype K, Injurious to Resistant Sorghum." *Crop Science* 37 (3): cropsci1997.0011183X003700030047x. <https://doi.org/10.2135/cropsci1997.0011183X003700030047x>.
- Hollenhorst, Margaret M., and Leonard R. Joppa. 1983. "Chromosomal Location of Genes for Resistance to Greenbug in 'Largo' and 'Amigo' Wheats1." *Crop Science* 23 (1): cropsci1983.0011183X002300010026x. <https://doi.org/10.2135/cropsci1983.0011183X002300010026x>.
- James, R. R., B. A. Croft, B. T. Shaffer, and B. Lighthart. 1998. "Impact of Temperature and Humidity on Host-Pathogen Interactions Between *Beauveria Bassiana* and a Coccinellid." *Environmental Entomology* 27 (6): 1506–13. <https://doi.org/10.1093/ee/27.6.1506>.
- Katsar, Catherine S., Andrew H. Paterson, George L. Teetes, and Gary C. Peterson. 2002a. "Molecular Analysis of Sorghum Resistance to the Greenbug (Homoptera: Aphididae)." *Journal of Economic Entomology* 95 (2): 448–57. <https://doi.org/10.1603/0022-0493-95.2.448>.
- Limaje, Ankur, Chad Hayes, J. Scott Armstrong, Wyatt Hoback, Ali Zarrabi, Sulochana Paudyal, and John Burke. 2018. "Antibiosis and Tolerance Discovered in USDA-ARS Sorghums Resistant to the Sugarcane Aphid (Hemiptera: Aphididae)1." *Journal of Entomological Science* 53 (2): 230–41. <https://doi.org/10.18474/JES17-70.1>.
- MacIntosh, Gustavo C. 2019. "Gene Pyramids and the Balancing Act of Keeping Pests at Bay." *Journal of Experimental Botany* 70 (18): 4591–93. <https://doi.org/10.1093/jxb/erz216>.
- Michaud, J. P. 2017. "IPM Case Studies: Sorghum." *Aphids as Crop Pests*, no. Ed. 2: 557–68.
- Michaud, J.P. 2008. "Wheat Insects: Bird Cherry-Oat Aphid, *Rhopalosiphum Padi* (Hemiptera: Sternorrhyncha: Aphididae)." KSU Entomology. <https://entomology.k-state.edu/extension/insect-information/crop-pests/wheat/bird-cherry.html>.
- Muleta, Kebede T., Terry Felderhoff, Noah Winans, Rachel Walstead, Jean Rigaud Charles, J. Scott Armstrong, Sujana Mamidi et al. "The recent evolutionary rescue of a staple crop depended on over half a century of global germplasm exchange." *Science advances* 8, no. 6 (2022): eabj4633.

- Mundt, Christopher C. 2018. "Pyramiding for Resistance Durability: Theory and Practice." *Phytopathology*® 108 (7): 792–802. <https://doi.org/10.1094/PHYTO-12-17-0426-RVW>.
- Nagaraj, Nandi, John C. Reese, Mitchell R. Tuinstra, C. Michael Smith, Paul St. Amand, M. B. Kirkham, Kenneth D. Kofoed, Leslie R. Campbell, and Gerald. E. Wilde. 2005. "Molecular Mapping of Sorghum Genes Expressing Tolerance to Damage by Greenbug (Homoptera: Aphididae)." *Journal of Economic Entomology* 98 (2): 595–602. <https://doi.org/10.1093/jee/98.2.595>.
- Nalam, Vamsi, Joe Louis, and Jyoti Shah. 2019. "Plant Defense against Aphids, the Pest Extraordinaire." *Plant Science* 279 (February): 96–107. <https://doi.org/10.1016/j.plantsci.2018.04.027>.
- Nibouche, Samuel, Laurent Costet, Raul F. Medina, Jocelyn R. Holt, Joëlle Sadeyen, Anne-Sophie Zoogones, Paul Brown, and Roger L. Blackman. 2021. "Morphometric and Molecular Discrimination of the Sugarcane Aphid, *Melanaphis Sacchari*, (Zehntner, 1897) and the Sorghum Aphid *Melanaphis Sorghi* (Theobald, 1904)." *PLOS ONE* 16 (3): e0241881. <https://doi.org/10.1371/journal.pone.0241881>.
- Paudyal, Sulochana, John Scott Armstrong, Kristopher L Giles, Mark E Payton, George P Opit, and Ankur Limaje. 2019. "Categories of Resistance to Sugarcane Aphid (Hemiptera: Aphididae) Among Sorghum Genotypes." *Journal of Economic Entomology* 112 (4): 1932–40. <https://doi.org/10.1093/jee/toz077>.
- Paudyal, Sulochana, John Scott Armstrong, Karen Harris-Shultz, Hongliang Wang, Kristopher Giles, Philippe Rott, and Mark Payton. 2019. "Evidence of Host Plant Specialization among the U.S. Sugarcane Aphid (Hemiptera: Aphididae) Genotypes." *Trends in Entomology*. <https://agritrop.cirad.fr/594367/>.
- Peterson, G. C., J. S. Armstrong, B. B. Pendleton, M. Stelter, and M. J. Brewer. 2018. "Registration of RTx3410 through RTx3428 Sorghum Germplasm Resistant to Sugarcane Aphid [*Melanaphis Sacchari* (Zehntner)]." *Journal of Plant Registrations* 12 (3): 391–98. <https://doi.org/10.3198/jpr2018.02.0007crg>.

- Peterson, G. C., J. W. Johnson, G. L. Teetes, and D. T. Rosenow. 1984. "Registration of Tx2783 Greenbug Resistant Sorghum Germplasm Line." *Crop Science* 24 (2): 390–390.
- Robert K. D. Peterson, Andrea C. Varella, and Leon G. Higley, "Tolerance: The Forgotten Child of Plant Resistance," *PeerJ* 5 (October 16, 2017): e3934, <https://doi.org/10.7717/peerj.3934>.
- Poosapati, Sowmya, Elly Poretsky, Keini Dressano, Miguel Ruiz, Armando Vazquez, Evan Sandoval, Adelaida Estrada-Cardenas, et al. 2022. "A Sorghum Genome-Wide Association Study (GWAS) Identifies a WRKY Transcription Factor as a Candidate Gene Underlying Sugarcane Aphid (*Melanaphis Sacchari*) Resistance." *Planta* 255 (2): 37. <https://doi.org/sz>.
- Porter, David R., John D. Burd, Kevin A. Shufran, James A. Webster, and George L. Teetes. 1997. "Greenbug (Homoptera: Aphididae) Biotypes: Selected by Resistant Cultivars or Preadapted Opportunists?" *Journal of Economic Entomology* 90 (5): 1055–65. <https://doi.org/10.1093/jee/90.5.1055>.
- Porter, K. B., G. L. Peterson, and O. Vise. 1982. "A New Greenbug Biotype1." *Crop Science* 22 (4): [cropsci1982.0011183X002200040035x](https://doi.org/10.2135/cropsci1982.0011183X002200040035x). <https://doi.org/10.2135/cropsci1982.0011183X002200040035x>.
- Punnuri, Somashekhar, and Yinghua Huang. 2017. "Identification and Confirmation of Greenbug Resistance Loci in an Advanced Mapping Population of Sorghum." *The Journal of Agricultural Science* 155 (10): 1610–22. <https://doi.org/10.1017/S0021859617000685>.
- Puterka, G. J., and D. C. Peters. "Sexual reproduction and inheritance of virulence in the greenbug, *Schizaphis graminum*(Rondani)." (1990).
- Morrison, W. P., F. B. Peairs, S. S. Quisenberry, and F. B. Peairs. "Response Model for an Introduced Pest—The Russian Wheat Aphid." *Thomas Say publications in entomology. Entomological Society of America, Lanham* (1998): 1-11.
- Starks, K. J., and D. J. Schuster. 1976. "Greenbug: 1 Effects of Continuous Culturing on Resistant Sorghum 2." *Environmental Entomology* 5 (4): 720–23. <https://doi.org/10.1093/ee/5.4.720>.

- Starks, Kenneth J., and Robert L. Burton. 1977. "Preventing Greenbug [*Schizaphis Graminum*] Outbreaks [Small Grains and Sorghum]." Leaflet - U.S. Dept. of Agriculture (USA). No. 309.
- Starks, Kenneth J., Robert L. Burton, and O. G. Merkle. 1983. "Greenbugs (Homoptera: Aphididae) Plant Resistance in Small Grains and Sorghum to Biotype E1." *Journal of Economic Entomology* 76 (4): 877–80. <https://doi.org/10.1093/jee/76.4.877>.
- Stout, Michael J. 2013. "Reevaluating the Conceptual Framework for Applied Research on Host-Plant Resistance." *Insect Science* 20 (3): 263–72. <https://doi.org/10.1111/1744-7917.12011>.
- Teetes, G. L., D. T. Rosenow, R. A. Frederiksen, and J. W. Johnson. 1973. "Predisposing Influence of Greenbugs on Charcoal Rot of Sorghum." Tex Agric Exp Stn Prog Rep Consol Prog Rep.
- Teetes, George L., Curtis A. Schaefer, and Jerry W. Johnson. 1974. "Resistance in Sorghums to the Greenbug; Laboratory Determination of Mechanisms of Resistance¹²." *Journal of Economic Entomology* 67 (3): 393–96. <https://doi.org/10.1093/jee/67.3.393>.
- USDA-FAS. 2020. "Grain and Feed Annual Report: Haiti (HA2020-0001)." USDA-FAS.
- Villanueva, R. T., and D. Sekula. "A new pest of sorghum: The sugarcane aphid." In *Proceedings of the 20th Annual Rio Grande Valley Cotton & Grain Pre-Plant Conference, Edcouch, TX, USA*, vol. 17. 2014.
- Vuillet, Jean François, and André Vuillet. 1914. *Les Pucerons Du Sorgho Au Soudan Français: Par Jean Vuillet et André Vuillet*. E. Larose.
- Wallin, J. R., and D. V. Loonan. 1971. "Low-Level Jet Winds, Aphid Vectors, Local Weather, and Barley Yellow Dwarf Virus Outbreaks." *Phytopathology*.
- Wan, A. M., X. M. Chen, Z. H. He, A. M. Wan, X. M. Chen, and Z. H. He. 2007. "Wheat Stripe Rust in China." *Australian Journal of Agricultural Research* 58 (6): 605–19. <https://doi.org/10.1071/AR06142>.
- Wang, Faming, Songmin Zhao, Yonghua Han, Yutao Shao, Zhenying Dong, Yang Gao, Kunpu Zhang, et al. 2013. "Efficient and Fine Mapping of RMES1 Conferring Resistance to Sorghum Aphid

- Melanaphis Sacchari.” *Molecular Breeding* 31 (4): 777–84. <https://doi.org/10.1007/s11032-012-9832-6>.
- Webster, F. M., and W. J. Phillips. 1912. *The Spring Grain-Aphis or “Green Bug.”* Washington, D.C: U.S. Dept. of Agriculture, Bureau of Entomology. <https://doi.org/10.5962/bhl.title.109852>.
- Weibel, D. E., and K. J. Starks. 1986. “Greenbug Nonpreference for Bloomless Sorghum1.” *Crop Science* 26 (6): [cropsci1986.0011183X002600060014x](https://doi.org/10.2135/cropsci1986.0011183X002600060014x).
<https://doi.org/10.2135/cropsci1986.0011183X002600060014x>.
- Wellings, P. W. 1994. “How Variable Are Rates of Colonisation?” *European Journal of Entomology* 91 (1): 121–25.
- Weng, Yiqun, Azhaguvel Perumal, John D. Burd, and Jackie C. Rudd. 2010. “Biotypic Diversity in Greenbug (Hemiptera: Aphididae): Microsatellite-Based Regional Divergence and Host-Adapted Differentiation.” *Journal of Economic Entomology* 103 (4): 1454–63.
<https://doi.org/10.1603/EC09291>.
- Zapata, Samuel D., Rebekka Dudensing, Danielle Sekula, Gabriela Esparza-Díaz, and Raul Villanueva. 2018. “Economic Impact of the Sugarcane Aphid Outbreak in South Texas.” *Journal of Agricultural and Applied Economics* 50 (1): 104–28. <https://doi.org/10.1017/aae.2017.24>.
- Zapata, S. D., Villanueva, R., Sekula, D., Esparza-Diaz, G., Duke, K., & Mutaleb, M. (2016). The economic impact of the sugarcane aphid on sorghum production.
<https://doi.org/10.22004/ag.econ.229992>.

Chapter 2 - The globally deployed sorghum aphid resistance gene *RMES1* is vulnerable to biotype shifts but being bolstered by *RMES2*

2.1 Summary

Durability of host plant resistance (HPR) to insect pests is critical for sustainable agriculture. Natural variation exists for aphid HPR in sorghum (*Sorghum bicolor*) but the genetic architecture and phenotype has not been clarified for most sources. To assess the current progress of breeding for sorghum aphid (*Melanaphis sorghi*) resistance we characterized the phenotype of *Resistance to Melanaphis sorghi 1* (*RMES1*) and contributing HPR architecture in two breeding populations, which are selected under strong aphid infestation in Haiti. We developed near-isogenic lines segregating for *RMES1* with 83.6% genomic similarity. *RMES1* reduces sorghum aphid fecundity but not *Rhopalosiphum padi* fecundity, suggesting a discriminant HPR response typical of gene-for-gene interactions. Using whole-genome resequencing of a global association panel, we found *RMES2* resistant alleles were more frequent than *RMES1* resistant alleles in landraces and historic US breeding germplasm. *RMES2* contributes early and mid-season aphid resistance in an segregating population of F₂'s, however *RMES1* was only significant with mid-season fitness. In a fixed population with high aphid resistance, *RMES1* and *RMES2* were selected for demonstrating their value for breeders. Globally, therefore, a vulnerable HPR source (*RMES1*) is likely bolstered by a second common source of resistance in breeding programs (*RMES2*) which may be staving off a biotype shift.

2.2 Introduction

Plant breeding indirectly affects insect populations by applying selection pressure via deployed host plant resistance (HPR) which deters infestation. It is important for breeders to consider what HPR is being deployed in order to reduce the likelihood of population shifts in agronomically important pests or their emergence. Insect populations have regularly overcome HPR with genetic or geographical shifts into open niches. For example, fall armyworm (*Spodoptera frugiperda*) which is native to the Americas has expanded as far as southeast Asia through long-distance dissemination while agronomically important

biotypes of Russian wheat aphid (*Diuraphis noxia*) and greenbug (*Schizaphis graminum*) have shifted in the Great Plains in the 20th century (Nagoshi et al. 2019; Harris-Shultz, Armstrong, and Jacobson 2020). The *Nr* and *Ag1* resistance genes were overcome by resistance breaking aphid biotypes of *Pemphigus bursarius* and *Amphorophora agathonica*, respectively, in Europe (Keep 1989; Arendt, Ester, and Schijndel 1999). Climate change is expected to exacerbate this problem as climate-change-induced hybridization and habitable range expansion drive genetic diversity (Arce-Valdés and Sánchez-Guillén 2022).

Aphids are economically significant pests which remove photoassimilates and vector viruses. Plants have multiple layers of defense against aphids including morphological barriers and chemical compositions which can prevent feeding and infestation by deterring aphid behavior (antixenosis) and/or reducing fecundity (antibiosis) (Nalam, Louis, and Shah 2019). In contrast to resistance and HPR, tolerance allows the plant to maintain fitness under moderate infestation (Painter 1951). The molecular mechanisms of relatively few aphid HPR sources are well understood. In one example of monogenic constitutive antixenosis, sorghum aphid (*Melanaphis sorghi*, Theobald 1904) feeding preference was affected in choice-assays by a bloomless gene knockout which lacked cuticular wax while reproduction in no-choice assays was not (Cardona et al. 2022). Two aphid resistance genes have been cloned, *Mi-1* and *Vat*, which encode nucleotide binding leucine-rich repeat (NLR) receptors (Dogimont et al. 2014; Nombela, Williamson, and Muñiz 2003). Translational evidence from plant-pathogen dynamics where NLR-based resistance suggests that a gene-for-gene arms race could occur in plant-aphid systems, since resistance breaking biotypes of *Mi-1* and *Vat* have been identified (Dogimont et al. 2010; Kaloshian 2004). This is one hypothesis for the boom and bust cycles of greenbug biotype-specific HPR that were seen in cereals of the Great Plains.

Sorghum (*Sorghum bicolor* L. [Moench]) is among the world's most important cereals and a staple crop for small-holder farmers in semi-arid regions (Rakshit et al. 2014). Sorghum aphid emerged as a major pest in North America a decade ago, believed to be due to a range expansion rather than biotype shift (Nibouche et al. 2021; Armstrong et al. 2015). Within five years of its introduction, the sorghum

aphid was found in all production regions (Harris-Shultz, Armstrong, and Jacobson 2020). One major HPR source to sorghum aphid is the globally-deployed *Resistance to Melanaphis sorghi 1 (RMES1)* on chromosome 6 (Chr06) found in African landraces where *M. sorghi* was first described (Muleta et al. 2022; Vuillet 1914). The evolution of such HPR sources likely occurs at longer time-scales than the breakdown of HPR due to selection pressures in modern agriculture and must therefore be deployed judiciously as a finite source of natural variation exists. A second sorghum aphid HPR QTL which we will refer to as *RMES2* is co-located with a *WRKY* transcription factor whose functional allele may act as a regulatory hub for induced defenses (Poosapati et al. 2022). The prevalence and utility of this second sorghum aphid HPR source in breeding programs has not been established.

The durability of HPR to emergent aphid populations is also expected to vary between tolerance and resistance since tolerance is not expected to apply the same selection pressure as resistance, but study systems to test are not tractable (Peterson, Varella, and Higley 2017). The source of greenbug biotype C resistance in sorghum was considered tolerance derived from *S. virgatum* but was overcome by biotype E shortly after wide deployment (Hackerott, Harvey, and Ross 1969). A biotype C and E resistant grain sorghum breeding line, RTx2783, was identified as having tolerance and antibiosis-resistance to sorghum aphid with moderate-high infestation levels but low damage (Armstrong et al. 2015). RTx2783 inherited the resistant *RMES1* allele from SC110 and was donor for aphid resistance in many cultivars grown on the Great Plains (Muleta et al. 2022). Hypothesis testing on whether a *RMES1* provides tolerance or HPR to sorghum aphid, as well as whether it provides resistance to both species (broad resistance), would provide insight on its durability and the likelihood of biotype shifts but have not been tested.

Sorghum aphid and sorghum is an agronomically relevant system to study HPR and pest population dynamics. There are competing hypotheses of what crop protection phenotype is conferred by *RMES1*, resistance, tolerance, and/or broad resistance. The contribution of *RMES2* to global breeding is unknown and molecular breeding tools have not been developed (Poosapati et al. 2022). It is also unclear whether *RMES1* and *RMES2* act additively or are epistatic. Here, we used NILs to test the hypotheses that *RMES1* (1) provides antibiosis-based resistance as opposed to a tolerance mechanism, (2) resistance is

specific discriminant Aphididae species, and (3) is likely bolstered by a secondary HPR source found in global germplasm.

2.3 Material and Methods

2.3.1 *Sorghum* genotypes

RMESI near-isogenic lines (NILs) were developed with a donor parent IRAT204 (*M. sorghi* resistant, *RMESI* donor) and recurrent backcrossing to RTx430 (*M. sorghi* susceptible). Single plant selections were made at the BC_xF₂ using a KASP marker for *RMESI* (Sbv3.1_06_02892438R, Muleta et al. 2022) and homozygous plants were backcrossed to RTx430. Population development was done at Kansas State University. BC₂F₃'s (NIL+, NIL-) and parental genotypes were used for aphid bioassays and whole genome resequencing.

2.3.2 Whole genome resequencing of NILs

Genomic DNA was collected from leaf tissue of BC₂F₃ NIL+, NIL-, IRAT204, and RTx430. Four plants of each genotype were grown, DNA extracted with Zymo Plant DNA Isolation Kits, and sequenced individually. Samples were sequenced at the Genomics Shared Resource Core at the University of Colorado Anschutz Medical Campus. Raw reads were trimmed using trimmomatic v0.39 and mapped to the RTx430v2 reference with BWA v0.7.17-r1188 (H. Li and Durbin 2009). Duplicate reads were identified using Picard v2.26 (McKenna et al. 2010). Finally, variants were called using GATK v4.2.5.0.

2.3.3 Aphid Cultures

M. sorghi used in this study were received from Dr. Scott Armstrong at the USDA-ARS Stillwater, Oklahoma. Aphids were reared on Tx7000 seedlings under laboratory conditions as described in Nalam et al. 2021. Seedlings were grown in 4.5 inch pots with potting soil and top layer of greens grade to reduce damping off. Colonies were housed in cages covered with organdy cloth. *Rhopalosiphum padi* cultures were obtained from Dr. Vamsi Nalam at CSU and maintained on Tx7000 seedlings similar to *M. sorghi* cultures.

2.3.4 Choice assay

A choice assay was done to determine aphid settling preference at the seedling stage. A pairwise comparison was done with NIL+ and NIL-. Seedlings of each genotype were grown approx. 2 inches apart in 1 gallon pots using potting soil and a top layer of greens grade. Seedlings were thinned to one plant of each genotype per pot. At 3-4 weeks of age, twenty 3-4 day old apterous *M. sorghi* aphids were placed in the center of a paper bridge between the seedlings. A clear plastic cylinder was placed over the plants to prevent aphids from leaving the pot with an organdy cloth covering for ventilation. The number of aphids on each plant were counted at 6 hours post infestation (6-hpi), 12-hpi, 24-hpi, and 48-hpi. The first replicate (n=9) started 1/6/22 and the second replicate (n=7) started 1/31/22. Statistical analyses were done in R to detect differences in *M. sorghi* preferences between genotypes (R Core Team 2021). A Student's *t*-test was used to compare genotypes at each time point.

2.3.5 No-choice assay

No-choice assays were used to compare aphid fecundity on various genotypes. A single seedling was grown in 6 inch pots using potting soil and a top layer of greens grade. At 3-4 weeks of age, three 3-4 day old apterous *M. sorghi* aphids were placed at the base of the seedlings with a camel hair brush. A clear plastic cylinder was placed over the plant to prevent aphids from leaving the pot with an organdy cloth covering for ventilation. The number of aphids on each plant were counted daily for a week at ~12pm. A no-choice assay with *R. padi* was used to determine broad resistance. NIL+ and NIL- lines were infested with three 4-5 day old apterous aphids and counted daily for one week and at two weeks after infestation.

2.3.6 Resequencing-based GWAS of community association panels

Whole genome resequencing for 665 sorghum genotypes from the Sorghum Association Panel (SAP) and Bioenergy Association Panel (BAP) were used for association analyses. Raw reads for the SAP were retrieved from the European Nucleotide Archive (RJE50066) (Boatwright et al. 2022). For the BAP, accessions were obtained through the USDA-ARS Germplasm Resources Information Network

and from Steven Kresovich and grown to seedlings. Tissue was collected and DNA extracted by Todd Mockler's lab at the Donald Danforth Plant Science Center. DNA was submitted to HudsonAlpha Genome Sequencing Center and The Department of Energy Joint Genome Institute for sequencing. Raw reads for both association panels were mapped to BTx623 v5.1 (sorghumbase.org/Sorghum_bicolorv5) reference using BWA-mem (H. Li and Durbin 2009). Samtools (H. Li et al. 2009) was used to select properly paired reads and sort, Picard (<http://broadinstitute.github.io/picard>) was used to remove duplicate reads, Samtools was used to remove low quality reads ("-Q 30"), and VarScan (<https://sourceforge.net/projects/varscan/>) used for variant calling. Variants were filtered using bcftools v1.15.1 commands 'F_MISSING < 0.9' and 'MAF > 0.01' (Danecek et al. 2021). Variants were imputed with BEAGLE v5.2 using default parameters (Browning, Zhou, and Browning 2018). A random subset of 500,000 variants was used to estimate population structure using TASSEL 5.0 CLI - PrincipalComponentsPlugin function. A general linear model was used to determine associations with the -FixedEffectLMPlugin using the first three PCs and normalized sorghum aphid phenotypes (Bradbury et al. 2007; Poosapati et al. 2022). A second association analysis was performed with the highest associated variant at *RMES2* (S09_61521444) included as a fixed-effect covariate. Manhattan plots were generated in base R (v4.2.2).

The most recent reference genome version (BTx623v5, Rice et al. 2024) is used throughout this paper and coordinates refer to the v5 coordinate system unless otherwise noted. For example S09_61433682 refers to the variant at 61,433,682 bp on Chr09 in the BTx623v5 genome while S09_57630053.v3 represents a variant at 57,630,053 in BTx623v3 with a different coordinate system.

2.3.7 Population genomic analyses of RMES2 in landraces and early breeding germplasm

Whole genome resequencing for 647 sorghum genotypes was used to determine allelic distributions for *RMES1* and *RMES2*. Genotypes partially overlapped with the SAP and BAP members, and genomic data was generated in the same method as described above. Longitude, latitude, and germplasm origin

metadata was manually curated. Geographic distributions and pie charts were generated using *r/ggplot2* (v3.4.1).

2.3.8 Association analysis in Haitian breeding population

A diverse Haitian breeding population of F₂'s was developed and described in Rice et al, (2023). Plants were phenotyped for fitness (alive/dead) at flowering initiation (~6-7 leaf) and booting (~8-9 leaf) growth stages. Sorghum is grown year round in Haiti's tropical climate where sorghum aphid is continuously present. No chemical prevention was used in field studies presented here.

Tissue was collected for genotyping approximately one month after planting and genotyped using Diverse Array Technology (Jaccoud et al. 2001). Sequencing was mapped to the BTx623 v3.0.1 reference genome (McCormick et al. 2018). Data was processed in R package *dartR*, and markers which were monomorphic, had <50% call rate, or <90% repeatability were removed. Data was converted to VCF format using a custom R script and imputation was done using Beagle 5.4 (Browning, Zhou, and Browning 2018). Data was converted to numeric representation of reference alleles (0,1,2). After filtering, there were 1,172 individuals with 8,195 markers.

Quantitative genetic analysis was done using *AsremlR* (Butler et al. 2017). Phenotype data was treated as binary and modeled as a linear model with the binomial logit function. A detailed description of the models used is given in Rice et al, (2023). A custom script was used to run individual marker associations genomewide with the first two principal components and *K* to estimate population structure, computed with *rTassel* (Monier et al. 2022).

2.4 Results

2.4.1 Marker assisted back crossing isolates *RMESI* in near-isogenic lines

Isolating the genetic basis of *RMESI* is necessary to characterize the molecular mechanism (tolerance, resistance, broad resistance) and phenotype (selection pressure) (Figure 2.1a). To develop near-isogenic lines, we backcrossed *RMESI*-donor IRAT204 to the susceptible RTx430. We then

confirmed that marker-assisted backcrossing successfully reduced the donor genome complement in NILs using whole-genome sequencing. There was 71.7% and 24.8% of the NIL+ genome fixed for the recurrent and donor parent genomes, respectively, and the remaining 3.4% was segregating (Figure 2.1b). The NIL- genome was 80.7% and 14.2% fixed for the recurrent and donor parent genomes, respectively, and 5.1% segregated (Figure 2.1b). Comparing between NIL+ and NIL- sibling lines, 83.6% of the genome was isogenic and the remaining 16.3% was segregating, including the majority of Chr06 (Figure 2.1c). In the region of *RMESI* on Chr06, both NIL lines were fixed however there is a breakpoint in NIL+ lines between 2,051,868 and 2,169,917 (RTx430v2 coordinates). The *RMESI* region, mapped in BTx623, corresponds to 2.9 – 3.1-Mb of the RTx430v2 genome.

2.4.2 *Sorghum* aphid fecundity is lowered by *RMESI*

To determine whether antibiosis is a component of *RMESI* and placing selection pressure on aphid populations (Figure 2.1a), we assessed aphid reproduction in a no-choice assay. Differences in population growth would indicate the *RMESI* mechanism is retarding infestation and placing selection pressure on aphids. The number of aphids was lower on NIL+ than NIL- at 7 days post infestation (7-dpi) ($p < 0.001$; Figure 2.2b). A moderately significant difference ($p > 0.05$) was first seen at 3-dpi and highly significant ($p > 0.001$) at 6-dpi. Overall, the population increased by 3.5 aphids per day on NIL+ and 7.1 aphids per day on NIL- over one week.

2.4.3 *Sorghum* aphid settling preference is not affected by *RMESI*

To determine whether *RMESI* affects the behavior and settling preference of sorghum aphid, we conducted a choice assay. Under the hypothesis that *RMESI* HPR includes constitutively present mechanisms deterring infestation, we would expect to see aphid feeding choice differences in a choice assay. We found that aphid settling was not significantly different between NILs at any time point in the first 48 hours of infestation (Figure 2.2c). This indicates that sorghum aphid feeding choice is not significantly influenced by *RMESI* as well as a lack of constitutively expressed epidermal or volatile features that deter aphid settling.

2.4.4 *RMESI* does not provide resistance to *Rhopalosiphum padi*

The presence of *RMESI* in the commonly used breeding line RTx2783 with greenbug resistance led to the hypothesis that the locus provides broad resistance for other Aphididae species. We tested NILs with *R. padi* and found there was not a significant effect on reproduction over 7 and 14 day infestation (Figure 2.2d). Aphid reproduction was lower for *R. padi* than *M. sorghi* in previous no-choice assay, with an increase of 0.8 aphids per day on NIL+ and 0.7 aphids per day on NIL- over one week. Therefore, *RMESI* is discriminant and not providing general Aphididae resistance.

2.4.5 GWAS with resequencing finds quantitative resistance in global landraces but not *RMESI*

It was previously shown that the *RMESI* SNP with the highest fixation signature in a Haitian breeding program (S6_2995581, v3.1 reference) was rare in the global diversity panel. Having used GBS data and a breeding population to identify this *RMESI* associated SNP, it remains possible that a SNP in higher LD with *RMESI* exists and is at detectable frequency in association panels. We retested the rare *RMESI* hypothesis by combining recently published landrace and improved lines (SAP and BAP) phenotypes with whole-genome resequencing data and the *S. bicolor* v5.1 reference and performed genome-wide association analyses. No association peak existed at *RMESI* on Chr06 (Figure 2.3a). We included the major source of resistance, S09_61521444, as a fixed-effect in our GLM model and confirmed the Chr06 locus was not significant after controlling for potentially confounding variation (Figure 2.3b). Several loci on chromosomes 1, 2, 3, and 10 were apparent in one or both analyses and are candidates for additional HPR sources (Table 2.1).

2.4.6 *RMES2* is common in global diversity lines

RMES2 was previously reported as a source of HPR in two association panels and the candidate gene *SbWRKY86* was proposed as the causal locus due to the strongest associations falling within the gene model (Poonsapati et al. 2022). In order to test hypotheses on *RMES2* using resequencing data which contains ~200 times more variants than previous GBS datasets, we generated new genome-wide associations and reanalyzed Chr09. We found the peak association ($p < 10e-14$) at S09_61521444 (Figure

2.3a) for the resistant associated reference allele (T) and present in 213 of 665 genotypes (32%). A second strong association at *RMES2* was found 1.1Mb from the peak association at S09_62669680 ($p < 10e-14$) (Figure 2.3c). The previously reported peak association (S09_57630053.v3 / S09_61433682) located in the promoter of *SbWRKY86* (Sobic.009G238200) was 88.9 kb from the peak association at S09_61521444 (Figure 2.3c) (Poosapati et al. 2022). This previously reported variant remained significant ($p < 10e-10$) in the current analysis.

The *RMESI* allele identified via selection signatures in resistant breeding lines was previously shown to be globally rare (Muleta et al. 2022). The presence of the resistant *RMES2* allele (S09_61521444) in 32% of SAP and BAP genotypes suggested it was globally common relative to *RMESI*. We used resequencing data for genotypes with known landrace or breeding origin to test this hypothesis. Of 492 sorghum landraces, the resistant *RMESI* (S06_3096975) present in just 13 (2.6%) while *RMES2* allele was present in 78 (15.8%) of lines (Figure 2.3d,e). *RMES2* did not show a clear geographic cline and was present in African and Indian lines. The resistant *RMES2* allele was present at a very high frequency in historic US breeding lines (35.8%) and sorghum conversion lines (67.3%) than landraces, whereas the resistant *RMESI* allele was absent in both historic sets.

2.4.7 *RMESI* and *RMES2* provide resistance in Haitian breeding population

RMESI was mapped in biparental populations for aphid resistance and through fixations signatures in an inbred fixed-resistance breeding population, whereas *RMES2* has been mapped in diversity lines including landraces for aphid resistance (Muleta et al. 2022; B. Wang et al. 2021; F. Wang et al. 2013; Poosapati et al. 2022). We tested the hypothesis that both loci are present and increase fitness under heavy aphid infestation in a highly-recombinant population of F₂s (segregating population). The two most significant QTL for mid-season survival were found at *RMES2* (S09_61988551, $p < 10e-11$) and *RMESI* (S06_2170466, $p < 10e-11$). There was not a significant interaction between the two loci indicating no epistasis detectable in this population. The third most significant mid-season survival QTL was on Chr08 at S08_45925614 ($p < 10e-10$).

To determine whether *RMES1* and *RMES2* contribute resistance at earlier developmental stages, we looked at associations with survival at an early-season timepoint. Another *RMES2* variant (S09_59794306, $p < 10e-12$) associated with resistance at early time points, however *RMES1* was not significant (Figure 2.4a). A QTL on Chr01 (S01_1230714, $p < 10e-11$) was within 1 kb from the cyanogenic glucoside biosynthetic cluster (*CYP79A1*, *CYP71E1*, *UGT85B1*). Other early-season fitness QTL included Chr01 (S01_73926645, $p < 10e-13$), Chr03 (S03_59797138, $p < 10e-12$), and Chr08 (S08_11501481, $p < 10e-13$) (Table 2.1) indicates HPR is polygenic in architecture at earlier growth stages and distinct, aside from *RMES2*, from oligogenic resistance at later growth stages.

2.4.8 *RMES1* and *RMES2* are being selected for in breeding program

Evidence of *RMES1* and *RMES2* underlying fitness in a population segregating for resistance led to the hypothesis that an aphid resistant population will be fixed for resistance alleles at both loci (if antagonistic pleiotropy or linkage drag is not present). The fixation scan previously published on the resistant Haitian population mapped *RMES1* and could be reanalysed with new information about *RMES2* (Muleta et al. 2022). We found that variants at *RMES2* were among loci with strongest signatures of selection in this inbred population fixed for resistance to aphids. Outliers of fixation signatures corresponded to staygreen (*Stg3a*, S02_58906500, $p < 10e-17$) and cyanogenic (*Dhr2*, S08_11535994, $p < 10e-16$) QTL as well as *RMES1* (S06_3096975, $p < 10e-17$) and *RMES2* (S09_58038553, $p < 10e-10$). While *RMES2* variants were significant, the most significant marker on Chr09 was at S09_49792740 ($p < 10e-14$, Figure 2.5). This indicates that *RMES2* is beneficial for aphid resistance and suggests it does not co-segregate with negative phenotypes that would be selected against for Haitian target population environments, supporting its broader use in breeding programs.

2.5 Discussion

Research on aphid resistance and breeding lags behind other crop-pest systems where quantitative variation and evolutionary impact of HPR is better understood. Breeding focused research on HPR often

identifies genetically dissimilar germplasm differing for traits such as antibiosis or tolerance preventing mechanisms from being resolved at the genetic level. It is important to Mendelize traits via NILs, induced mutation, or gene-editing in order to test gene-specific hypotheses which untangle the many levels of HPR and understand individual mechanisms. The majority isogenic background and single resistance locus segregating between the parents of our NILs allowed us to define *RMESI* antibiosis-resistance for sorghum aphid. Future investigations of the molecular mechanism will benefit from continued population development of these NILs.

The common breeding line RTx2783 contains the resistant allele of *RMESI* and showed tolerance and resistance to *M. sorghi* and *S. graminum* leading to the hypothesis that *RMESI* is pleiotropic for these traits (Armstrong et al. 2015). *RMESI* + NILs had no effect in the *R. padi* no-choice assay (Figure 2.2d) showing that the molecular mechanism (activation or mode of action) of resistance is not shared by all *RMESI* – Aphididae interactions, however there remains the hypothesis that it is effective to *S. graminum* or a limited number of cereal aphid species. Resistance to *S. graminum* has not been mapped to the *RMESI* region further supporting RTx2783 resistance to greenbug is from non-*RMESI* sources (Harris-Shultz, Armstrong, and Jacobson 2020). *RMESI* is the only HPR source with breeder-friendly marker technology and is widely effective in public and private sorghum breeding programs (Muleta et al. 2022). This HPR should be managed by breeders and IPM strategies to avoid being overcome by a new biotype of sorghum aphid. New biotypes of greenbug emerged on sorghum varieties in 1979 and 1990 which had previously had genetic resistance (Harris-Shultz, Armstrong, and Jacobson 2020). Similarly, Russian wheat aphid (*Diuraphis noxia*) resistance locus *Dn4* in wheat was effective until the emergence of the RWA2 biotype in 2003 (Haley et al. 2004). To avoid a similar loss of *RMESI* resistance to novel sorghum aphid biotypes, breeders must proactively develop cultivars with multiple and mechanistically diverse sources of HPR.

Tradeoffs in plant defense strategies such as constitutive and induced mechanisms have impacts on resource allocation, development, and ultimately productivity (Kempel et al. 2011, Monson et al. 2022). In addition, some induced resistance mechanisms are at risk of loss to resistance-breaking biotypes

and pathotypes. Candidate genes proposed to underlie *RMESI* include Sobic.006G016900, Sobic.006G017200, Sobic.006G017400, and Sobic.006G017500 but remain to be tested. Under the constitutive hypothesis, *SbCASI* (Sobic.006G016900) is involved in detoxification of HCN produced by the cyanogenic glucoside dhurrin (Gleadow et al. 2021). Cyanogenesis is involved in antibiosis resistance to *Spodoptera frugiperda* in sorghum, however involvement in aphid resistance has not been tested (Gruss et al. 2022). Under the induced hypothesis, nucleotide-binding leucine-rich repeat (NLR) receptors (Sobic.006G017200, Sobic.006G017400, Sobic.006G017500), which recognize molecular patterns of infestation and activate host defenses, are predicted in the region of *RMESI* in the BTx623 (susceptible) and RTx2783 (resistant) genomes (Muleta et al. 2022; B. Wang et al. 2021). Five sorghum genotypes (RTx430, RTx2783, RTx436, BTx623, Rio) contained between 319 and 363 predicted NLR genes (B. Wang et al. 2021). This is a strong candidate class for causal genes due to numerous reports of NLR genes driving resistance in aphids and other pest systems (Snoeck, Guayazán-Palacios, and Steinbrenner 2022; Dogimont et al. 2014). This class of resistance mechanisms are expected to be less durable if the selection pressure is strong and the herbivore associated molecular pattern (HAMP) can withstand mutations to evade the host receptor. Such ‘gene-for-gene’ dynamics can lead to boom-and-bust cycles similar to those seen in cereal-biotic pest systems (Dogimont et al. 2010; Mundt 2018).

As future studies establish the relative durability of antibiosis, antixenosis, and tolerance mechanisms, knowledge of *RMESI*-antibiosis will inform how to combine and utilize all HPR available. Differences in durability would be expected depending on the target of detection (under induced hypothesis, fitness importance of herbivore associated molecular pattern or effector) as well as mode of action (fitness importance of an effective site for a toxic compound). RTx2783 appears to harbor additional tolerance and HPR sources as it was reported to retain growth despite moderate aphid infestation (Limaje et al. 2018). Alternatively, epistatic interactions from *RMESI* may contribute additional tolerance mechanisms in different backgrounds. Regardless of the epistatic tolerance hypothesis, a significant reduction in fecundity of aphids on *RMESI* NILs demonstrates a selection pressure on *M. sorghi* and increasing the likelihood of a biotype shift. This HPR source should not be

solely relied upon for crop protection as there is precedent for monogenic HPR breakdown in cereal-aphid systems (Harris-Shultz, Armstrong, and Jacobson 2020).

Frequency of a resistance allele may reflect evolutionary selection pressures on plant defense and fitness tradeoffs – for example, *RPS5* variation for *Pseudomonas syringae* resistance in Arabidopsis is maintained by balancing selection (Tian et al. 2002). The resistant *RMES1* allele identified by GBS data in a fixation scan is rare in global landraces (Muleta et al. 2022). We retested this *RMES1*-rarity hypothesis using dense genotyping of SAP and BAP lines and found that it was not contributing resistance in these diversity lines, even when *RMES2* was included in the model (Figure 2.3b). QTLs on chromosomes 1, 2, 3, and 10 were detected when controlling for *RMES2* and are potential sources of quantitative minor-effect resistance to sorghum aphid. The Chr01 QTL (S01_1536310) was 367 kb from the dhurrin biosynthetic gene cluster which includes CYP79A1, CYP71E1, and UGT85B1 (Sobic.001G012300, Sobic.001G012200, Sobic.001G012400) (Gleadow et al. 2021). Interestingly, this QTL is also 25 kb and 46 kb from two paralogs of CYP71E1 (Sobic.001G018300, Sobic.001G018600). There were no strong candidate genes for aphid HPR QTLs on chromosomes 2, 3, and 10, however QTL for physiological traits (sucrose content, photosynthetic rate) were associated nearby and may provide tolerance through growth rate (<https://aussorgm.org.au/sorghum-qt1-atlas/>).

The two strongest associations with GBS SNPs were inside the gene model of *SbWRKY86*, a transcription factor responsive to aphid infestation (Poosapati et al. 2022; Kiani and Szczepaniec 2018). Two regions 88 kb and 1.24 Mb from *SbWRKY86* had a higher association than the original SNP in our analysis (Figure 2.3c). One methodological hypothesis for discrepancies is differences between the previous panel of 697 genotypes and our subset of 665 genotypes with available resequencing. Another hypothesis is that genotypic variation at *SbWRKY86* is not causal for *RMES2* HPR but *trans*-regulating elements modulate this transcription factor (P. Li et al. 2015; Atamian, Eulgem, and Kaloshian 2012). Finally, it is possible that *SbWRKY86* is not involved in *RMES2* resistance and one or both of the QTL (S09_61521444, S09_62669680) are in higher LD with the causal gene. This QTL is in a gene-dense

region of Chr09 with 278 genes annotated between 61–63 Mb, demonstrating the need for fine-mapping or gene editing to confirm or exclude the hypothesis that *SbWRKY86* underlies *RMES2* resistance.

We determined that *RMES2* is globally common in sorghum landraces and breeding lines, highlighting its availability for sorghum improvement. While reanalysis of HPR origin highlighted the rarity of *RMESI*, it is notable that previous geographic distribution results indicated *RMESI* was confined to East Africa (Muleta et al. 2022), however resequencing data suggests that the same variant was present in several lines outside East Africa (Figure 2.3d). One explanation is that records of the naming and origin of accessions may be incomplete, however it is possible the previous report was under-powered for geographic representation or the imputation of the SNP in the GBS data set was inaccurate. Our peak *RMES2* variant was at higher frequency in global germplasm as well as being dispersed across the African continent and India (Figure 2.3e). The resistant allele's high frequency in breeding and conversion germplasm indicates *RMES2* is widely available for breeding in diverse and adapted backgrounds, and suggests that breeding programs may be unknowingly selecting it during phenotypic selection under sorghum aphid infestation.

With *RMES2* resistance observed in diversity panels and at high frequency in early breeding germplasm, it may already be providing resistance in current breeding programs where it would be selected for sorghum aphid resistance. The Chibas breeding program at the University of Quisqueya in Haiti breeds dual purpose (grain and forage) tropical sorghum varieties in high aphid pressure environments (Muleta et al. 2022). The historic recombination in the highly-intercrossed founders of the F₂ segregating HBP for survival under aphid pressure allowed high-power to determine genetic associations with fitness (Rice et al. 2024; Cockram and Mackay 2018). *RMESI* and *RMES2* were the most significant QTL for mid-season fitness, however *RMESI* was not associated with resistance at the earlier stage (Figure 2.4a) suggesting either different stressors determining survival at early-season than mid-season or that *RMESI* resistance may be developmentally regulated akin to adult plant resistance (APR). However, antibiosis-resistance was seen in the 3-4 leaf stage (Figure 2.2b) and therefore molecular plant-aphid interactions may be more complex in the field than controlled settings in the

greenhouse. Another possibility is that there were abiotic (nutrient, environmental) or biotic factors other than *M. sorghi* pressure determining fitness at early different from late developmental stages.

Phenotyping of the BAP and SAP was initially done on 9 week-old plants which approximately corresponds to flowering stage (Poosapati et al. 2022) and therefore agrees with the observation that *RMES2* is providing mid-season fitness. *RMES2* was among several significant associations at the earlier flowering initiation (Figure 2.4a) indicating quantitative resistance determines fitness at earlier life stages. There are no known aphid resistance QTL co-located with early-season survival, however other biotic resistance QTL have been reported (<https://aussorgm.org.au/sorghum-qt1-atlas/>). In summary, *RMES1* and *RMES2* have distinct field and phenology dynamics but are not the exhaustive list of sorghum aphid HPR.

The cyanogenic glucoside dhurrin is known to provide resistance to herbivores and has been proposed as a candidate mechanism for aphid resistance (Gruss et al. 2022). However, it has also been pointed out that stylet-feeding of aphids may not be expected to cause the tissue disruption necessary to bring dhurrin (epidermal cell) and dhurrinase (mesophyll cells) in contact for cyanide release (Thayer and Conn 1981). Interestingly, our results show the dhurrin bioactivation *Dhr2* loci on Chr08 (Sobic.008G080400) (Cicek and Esen 1998) is ~1.9 Mb from the peak loci for early-survival (S08_11501481, Figure 2.4a). Similarly, dhurrin biosynthesis *TCD1* on Chr01 (Sobic.001G012200, Sobic.001G012300, Sobic.001G012400) (Blomstedt et al. 2016) is within a kilobase of a QTL on Chr01 for early-survival (S01_1253672) (Figure 2.5a). This region also closely coincides with the QTL S01_1536310 identified in the association panels when *RMES2* was accounted for (Figure 2.3b), providing further evidence for variation at the biosynthetic gene cluster on Chr01 to contribute resistance to aphid. These findings suggest further investigations are warranted of dhurrin as a mechanism of resistance against sorghum aphid.

In a different HBP population fixed for aphid resistance, a selection sweep was previously used to map *RMES1*, however there were no other known *M. sorghi* resistance QTL known at that time (Muleta et al. 2022). A common HPR source for early and late resistance would be expected to have positive selection, if not as significant as globally rare alleles like *RMES1*. Many variants on Chr09 spanning the

centromere show selection signatures, including at *RMES1*. Approximately 1.1-Mb from *RMES2* is the dwarfing allele *dwl* (Sobic.009G229800), which was selected for in the conversion of temperate sorghums, and is in a genomic region of low diversity in global landraces (Morris et al. 2013; Hilley et al. 2016). This and other unknown QTLs spanning Chr09 were likely under selection in the active breeding program for other important agronomic traits.

Host plant resistance sources to insects are valuable agricultural resources that need to be understood in order to unlock its full potential and durability. *RMES1* is at risk of being overcome by new biotypes of sorghum aphid and potentially in gene-for-gene dynamics, however the globally-common *RMES2* is likely playing a supporting role to increase overall resistance and the durability of *RMES1* in breeding programs.

Chapter 2 Tables

Table 2.1 - Significant associations in quantitative genomic analyses

Marker	Chr	Pos	<i>p</i> - value	Analysis
S09_61521444	9	61521444	2.64E-15	GWAS, SAP+BAP
S09_62669680	9	62669680	2.71E-15	GWAS, SAP+BAP
S10_50970852	10	50970852	4.13E-10	GWAS, SAP+BAP, RMES2 fixed
S03_59780623	3	59780623	3.58E-09	GWAS, SAP+BAP, RMES2 fixed
S02_41893124	2	41893124	9.24E-09	GWAS, SAP+BAP, RMES2 fixed
S01_1536310	1	1536310	2.43E-08	GWAS, SAP+BAP, RMES2 fixed
S09_60932782	9	60932782	8.76E-13	Segregating HBP, mid-season survival
S06_2170466	6	2170466	8.33E-12	Segregating HBP, mid-season survival
S08_45925614	8	45925614	6.00E-13	Segregating HBP, mid-season survival
S08_11501481	8	11501481	2.46E-14	Segregating HBP, early-season survival
S01_73908160	1	73908160	5.00E-16	Segregating HBP, early-season survival
S03_59797138	3	59797138	1.53E-13	Segregating HBP, early-season survival
S09_58935337	9	58935337	5.54E-13	Segregating HBP, early-season survival
S01_1253672	1	1253672	1.01E-12	Segregating HBP, early-season survival

Chapter 2 Figures

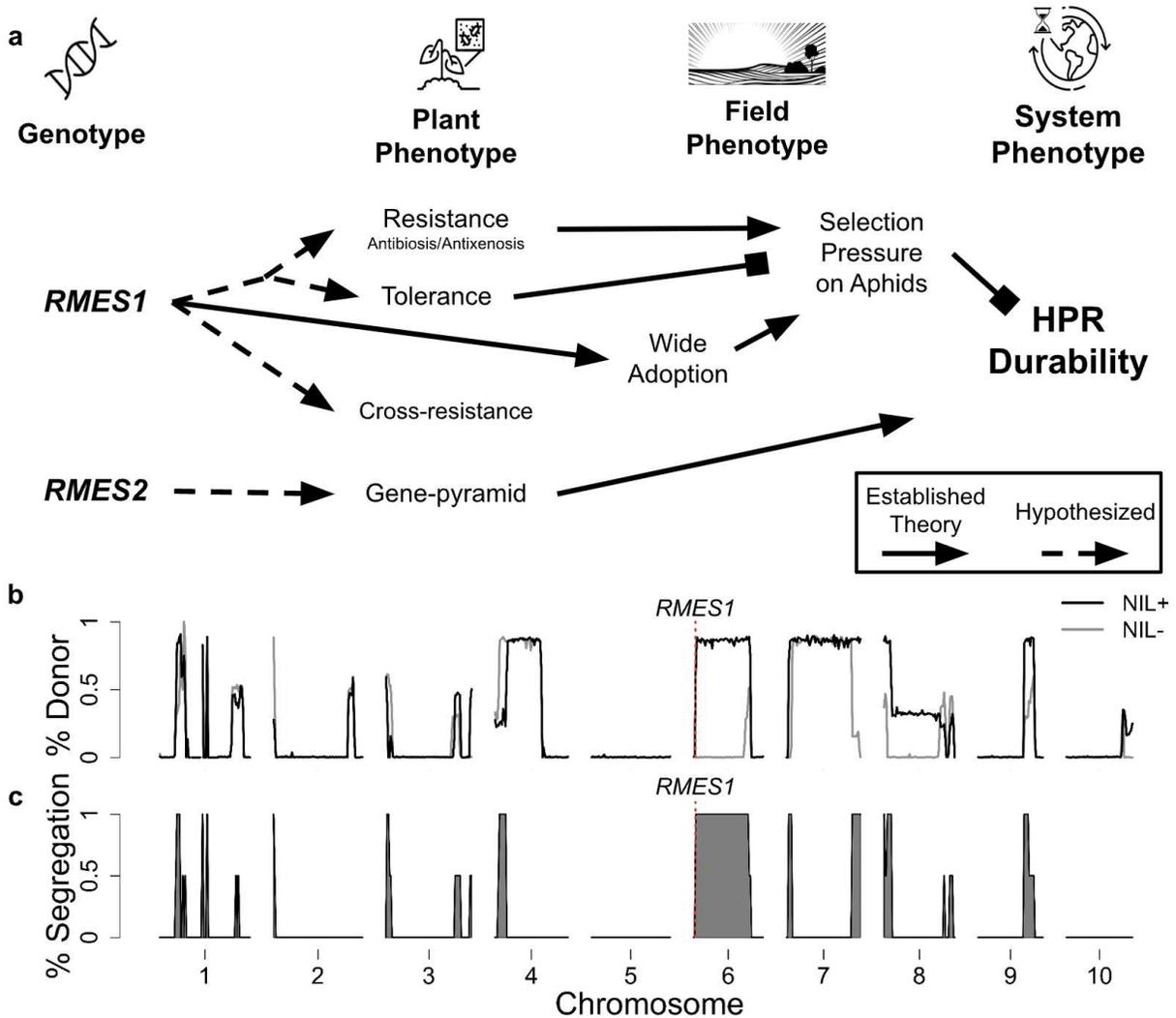


Figure 2.1 - Genotype to phenotype hypotheses for *RMES1* can be tested with NILs
a) The genotype to phenotype map for *RMES1* and HPR durability containing several hypothesized relationships. b) Donor (IRAT204, 100%) genome contribution in BC₂F₃ NILs relative to recurrent genome (RTx430, 0%). c) Genomic regions segregating between NILs. Red dotted line indicates *RMES1* position.

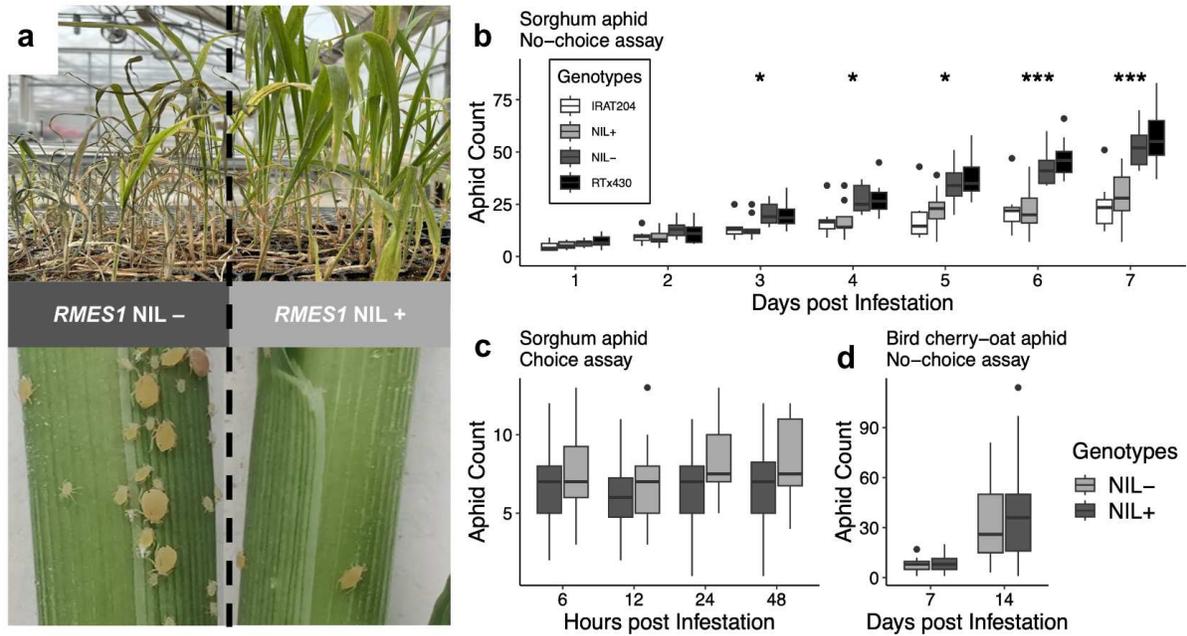


Figure 2.2 - *RMES1* NILs define antibiosis sorghum aphid resistance

a) Aphid infestation on *RMES1* NILs. b) No-choice assay (n=9) aphid counts over 7 day infestation. c) Choice assay (n=16) aphid counts over 48 hour infestation. d) No-choice assay (n=18) of *Rhopalosiphum padi* at 7-dpi and 14-dpi. Significant pairwise comparison between NILs (student's t-test) shown with asterisks (p<0.05 *, p<0.01 **, p<0.001 ***)

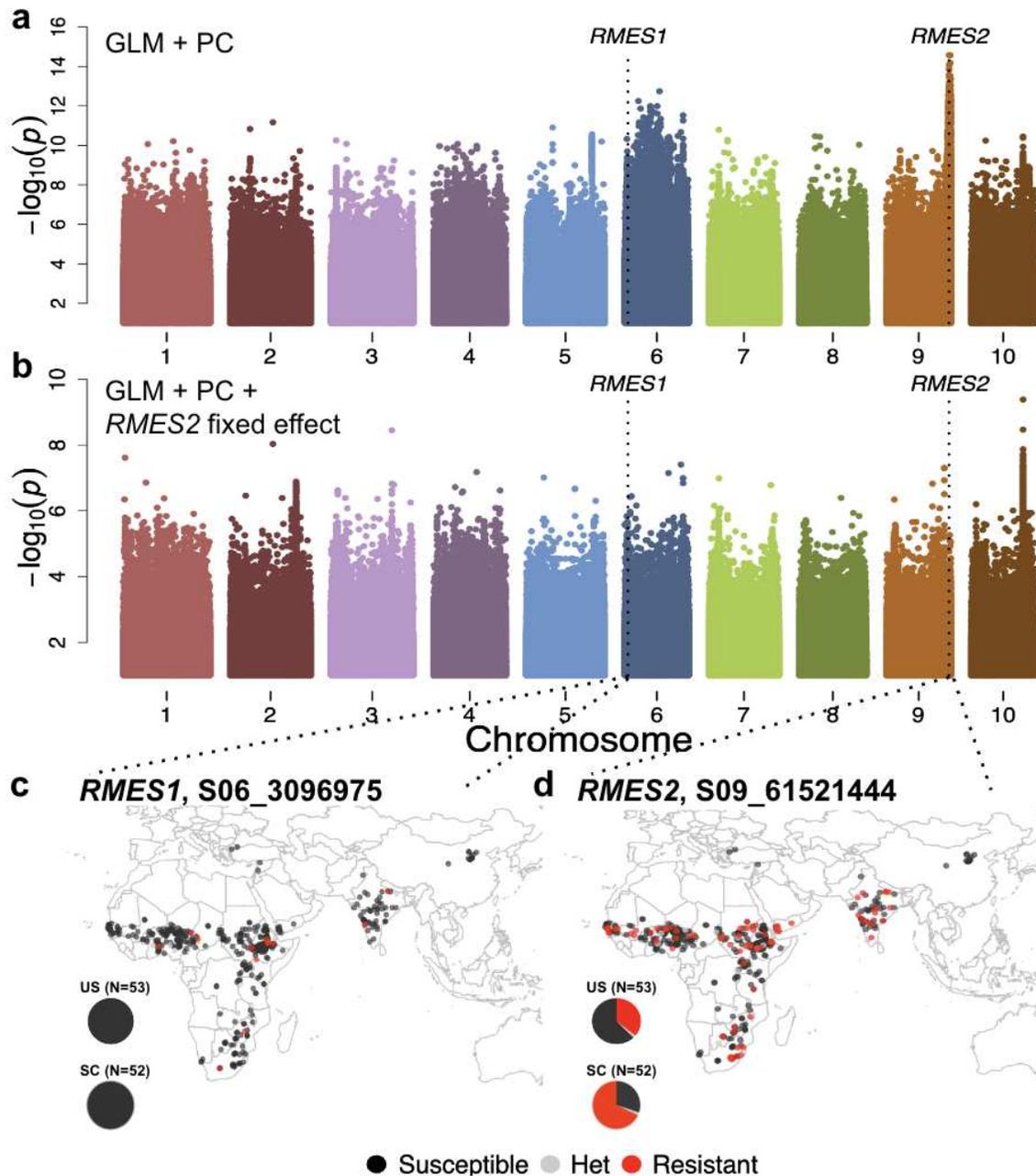


Figure 2.3 - Genome-wide associations in two global association panels with *M. sorghi* resistance show *RMES1* is globally rare.

a) Association of resequencing variants in BAP and SAP determined with a general linear model that included principal components 1-3 of global population structure as fixed effects. b) Association of resequencing variants determined with a general linear model that included principal components 1-3 and the peak *RMES2* SNP (S09_61521444) from the GLM as fixed effects. c) Distribution of *RMES1* associated SNP previously identified in Haitian fixation scan (Muleta et al. 2022). d) Distribution of *RMES2* associated SNP identified in the present study. Available resequencing data was used to estimate the frequency of alleles in known historic American breeding germplasm (US) and lines from the sorghum conversion program (SC).

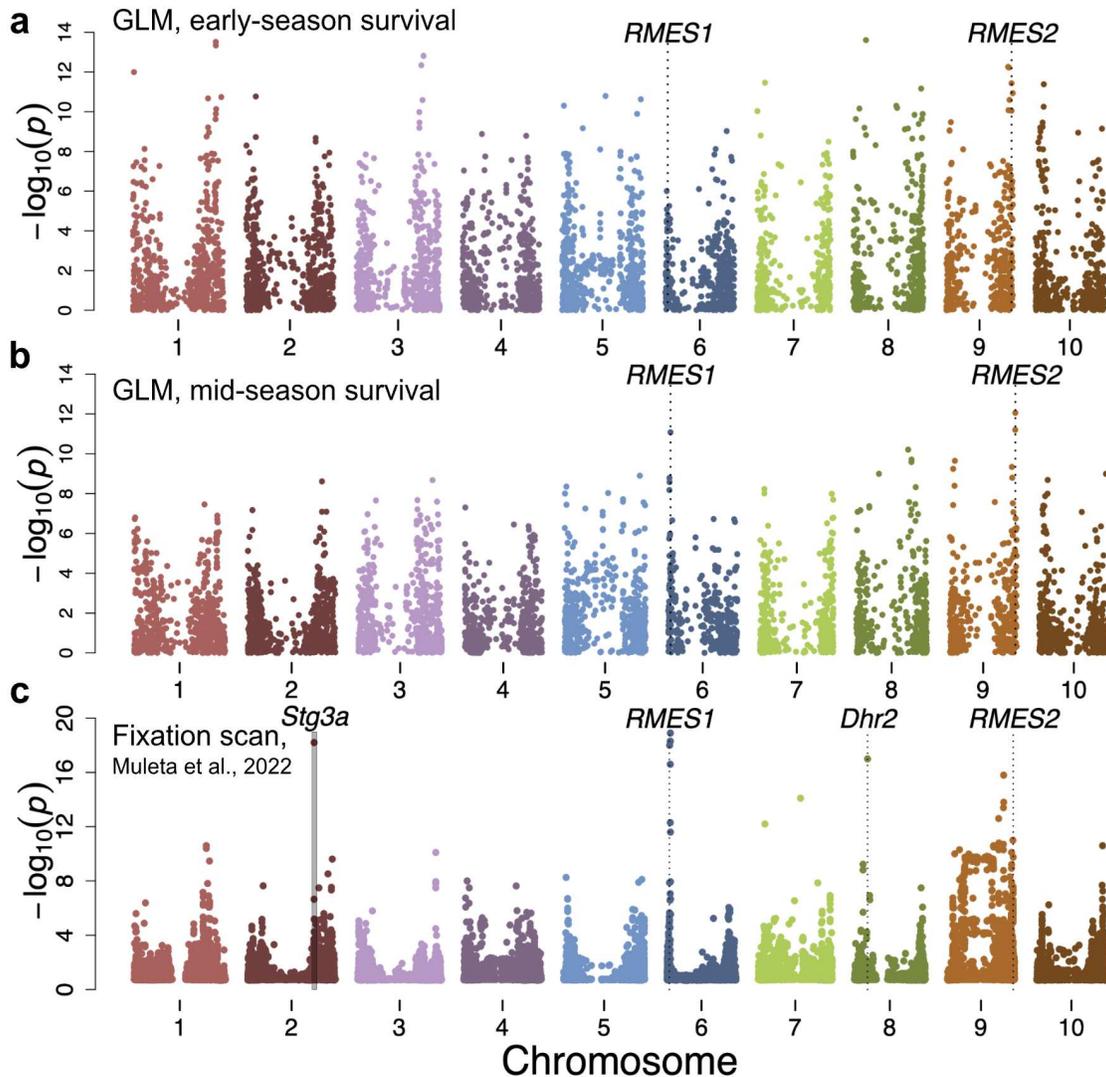


Figure 2.4 - *RMES1* and *RMES2* provide resistance in Haitian breeding programs.

Associations with fitness under heavy aphid infestation is shown for an early-generation breeding population in which survival (aphid resistance) is segregating. a) Early-season (flowering initiation) survival associations. b) Mid-season (booting) survival associations. The two highest associations are indicated in red and black dotted lines for *RMES1* and *RMES2*, respectively. c) Fixation scores (F_{st}) determined for the resistant population compared with the global diversity panel. Data published in Muleta et al. (2022). Notable signatures are indicated with shaded areas or dashed lines and the corresponding candidate region (*Stg3a*, *RMES1*, *Dhr2*, *RMES2*).

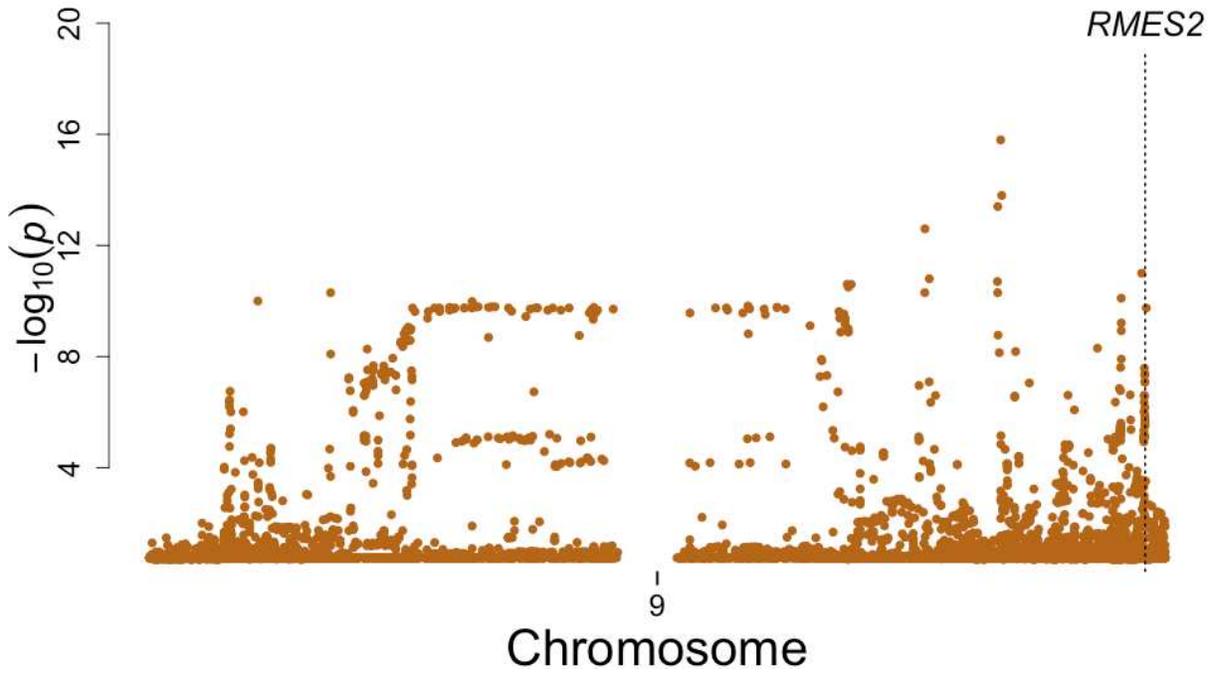


Figure 2.5 - Large regions of Chr09 are being selected in breeding population with high aphid selection pressure. Fixation scores (Fst) on Chr09 determined for an aphid resistant HBP compared with the global diversity panel. Data published in Muleta et al. (2022).

Chapter 2 References

- Arce-Valdés, Luis R, and Rosa A Sánchez-Guillén. 2022. “The Evolutionary Outcomes of Climate-Change-Induced Hybridization in Insect Populations.” *Current Opinion in Insect Science* 54 (December): 100966. <https://doi.org/10.1016/j.cois.2022.100966>.
- Van der Arendt, A. J. M., A. Ester, and J. T. Van Schijndel. "Developing an aphid resistant butterhead lettuce" Dynamite". In *Eucarpia Leafy Vegetables' 99, Olomouc (Czech Republic), 8-11 Jun 1999*. Palacky University, 1999.
- Armstrong, Scott J., William L. Rooney, Gary C. Peterson, Raul T. Villeneuve, Michael J. Brewer, and Danielle Sekula-Ortiz. 2015. “Sugarcane Aphid (Hemiptera: Aphididae): Host Range and Sorghum Resistance Including Cross-Resistance From Greenbug Sources.” *Journal of Economic Entomology* 108 (2): 576–82. <https://doi.org/10.1093/jee/tou065>.
- Atamian, Hagop S., Thomas Eulgem, and Isgouhi Kaloshian. 2012. “SIWRKY70 Is Required for Mi-1-Mediated Resistance to Aphids and Nematodes in Tomato.” *Planta* 235 (2): 299–309. <https://doi.org/10.1007/s00425-011-1509-6>.
- Blomstedt, Cecilia K., Natalie H. O’Donnell, Nanna Bjarnholt, Alan D. Neale, John D. Hamill, Birger Lindberg Møller, and Roslyn M. Gleadow. 2016. “Metabolic Consequences of Knocking out UGT85B1 , the Gene Encoding the Glucosyltransferase Required for Synthesis of Dhurrin in Sorghum Bicolor (L. Moench).” *Plant and Cell Physiology* 57 (2): 373–86. <https://doi.org/10.1093/pcp/pcv153>.
- Boatwright, J. Lucas, Sirjan Sapkota, Hongyu Jin, James C. Schnable, Zachary Brenton, Richard Boyles, and Stephen Kresovich. 2022. “Sorghum Association Panel Whole-Genome Sequencing Establishes Cornerstone Resource for Dissecting Genomic Diversity.” *The Plant Journal* 111 (3): 888–904. <https://doi.org/10.1111/tpj.15853>.

- Bradbury, Peter J., Zhiwu Zhang, Dallas E. Kroon, Terry M. Casstevens, Yogesh Ramdoss, and Edward S. Buckler. 2007. "TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples." *Bioinformatics* 23 (19): 2633–35. <https://doi.org/10.1093/bioinformatics/btm308>.
- Browning, Brian L., Ying Zhou, and Sharon R. Browning. 2018. "A One-Penny Imputed Genome from Next-Generation Reference Panels." *The American Journal of Human Genetics* 103 (3): 338–48. <https://doi.org/10.1016/j.ajhg.2018.07.015>.
- Butler, D. G., B. R. Cullis, A. R. Gilmour, B. J. Gogel, and R. Thompson. 2017. "ASReml-R Reference Manual Version 4."
- Cardona, Juan Betancurt, Sajjan Grover, Lucas Busta, Scott E. Sattler, and Joe Louis. 2022. "Sorghum Cuticular Waxes Influence Host Plant Selection by Aphids." *Planta* 257 (1): 22. <https://doi.org/10.1007/s00425-022-04046-3>.
- Muzaffer Cicek and Asim Esen, "Structure and Expression of a Dhurrinase (β -Glucosidase) from Sorghum," *Plant Physiology* 116, no. 4 (April 1998): 1469–78, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC35055/>.
- Cockram, James, and Ian Mackay. 2018. "Genetic Mapping Populations for Conducting High-Resolution Trait Mapping in Plants." In *Plant Genetics and Molecular Biology*, edited by Rajeev K. Varshney, Manish K. Pandey, and Annapurna Chitikineni, 109–38. *Advances in Biochemical Engineering/Biotechnology*. Cham: Springer International Publishing. https://doi.org/10.1007/10_2017_48.
- Danecek, Petr, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10 (2): giab008. <https://doi.org/10.1093/gigascience/giab008>.
- Dogimont, Catherine, Abdelhafid Bendahmane, Véronique Chovelon, and Nathalie Boissot. 2010. "Host Plant Resistance to Aphids in Cultivated Crops: Genetic and Molecular Bases, and Interactions with Aphid Populations." *Comptes Rendus Biologies, Les pucerons : modèles biologiques et ravageurs des cultures*, 333 (6): 566–73. <https://doi.org/10.1016/j.crv.2010.04.003>.

- Dogimont, Catherine, Veronique Chovelon, Jerome Pauquet, Adnane Boualem, and Abdelhafid Bendahmane. 2014. "The Vat Locus Encodes for a CC-NBS-LRR Protein That Confers Resistance to *Aphis Gossypii* Infestation and *A. Gossypii*-Mediated Virus Resistance." *The Plant Journal* 80 (6): 993–1004. <https://doi.org/10.1111/tpj.12690>.
- Gleadow, Roslyn M., Brian A. McKinley, Cecilia K. Blomstedt, Austin C. Lamb, Birger Lindberg Møller, and John E. Mullet. 2021. "Regulation of Dhurrin Pathway Gene Expression during Sorghum Bicolor Development." *Planta* 254 (6): 119. <https://doi.org/10.1007/s00425-021-03774-2>.
- Gruss, Shelby M., Manoj Ghaste, Joshua R. Widhalm, and Mitchell R. Tuinstra. 2022. "Seedling Growth and Fall Armyworm Feeding Preference Influenced by Dhurrin Production in Sorghum." *Theoretical and Applied Genetics*, January. <https://doi.org/10.1007/s00122-021-04017-4>.
- Hackerott, H. L., T. L. Harvey, and W. M. Ross. 1969. "Greenbug Resistance in Sorghums1." *Crop Science* 9 (5): crops1969.0011183X000900050046x. <https://doi.org/10.2135/cropsci1969.0011183X000900050046x>.
- Haley, Scott D., Frank B. Peairs, Cynthia B. Walker, Jeffrey B. Rudolph, and Terri L. Randolph. 2004. "Occurrence of a New Russian Wheat Aphid Biotype in Colorado." *Crop Science* 44 (5): 1589–92. <https://doi.org/10.2135/cropsci2004.1589>.
- Harris-Shultz, Karen, Scott Armstrong, and Alana Jacobson. "Invasive cereal aphids of North America: Biotypes, genetic variation, management, and lessons learned." *Trends Entomol* 15 (2019): 99–122.
- Hilley, Josie, Sandra Truong, Sara Olson, Daryl Morishige, and John Mullet. 2016. "Identification of Dw1, a Regulator of Sorghum Stem Internode Length." *PLOS ONE* 11 (3): e0151271. <https://doi.org/10.1371/journal.pone.0151271>.
- Jaccoud, Damian, Kaiman Peng, David Feinstein, and Andrzej Kilian. 2001. "Diversity Arrays: A Solid State Technology for Sequence Information Independent Genotyping." *Nucleic Acids Research* 29 (4): e25.

- Kaloshian, Isgouhi. 2004. "GENE-FOR-GENE DISEASE RESISTANCE: BRIDGING INSECT PEST AND PATHOGEN DEFENSE." *Journal of Chemical Ecology* 30 (12): 2419–38.
<https://doi.org/10.1007/s10886-004-7943-1>.
- Keep, Elizabeth. 1989. "Breeding Red Raspberry for Resistance to Diseases and Pests." *In Plant Breeding Reviews*, 245–321. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118061039.ch7>.
- Anne Kempel et al., "Tradeoffs Associated with Constitutive and Induced Plant Resistance against Herbivory," *Proceedings of the National Academy of Sciences* 108, no. 14 (April 5, 2011): 5685–89, <https://doi.org/10.1073/pnas.1016508108>.
- Kiani, Mahnaz, and Adrianna Szczepaniec. 2018. "Effects of Sugarcane Aphid Herbivory on Transcriptional Responses of Resistant and Susceptible Sorghum." *BMC Genomics* 19 (1): 774.
<https://doi.org/10.1186/s12864-018-5095-x>.
- Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* (Oxford, England) 25 (14): 1754–60.
<https://doi.org/10.1093/bioinformatics/btp324>.
- Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79.
<https://doi.org/10.1093/bioinformatics/btp352>.
- Li, Peiling, Aiping Song, Chunyan Gao, Jiafu Jiang, Sumei Chen, Weimin Fang, Fei Zhang, and Fadi Chen. 2015. "The Over-Expression of a Chrysanthemum WRKY Transcription Factor Enhances Aphid Resistance." *Plant Physiology and Biochemistry* 95 (October): 26–34.
<https://doi.org/10.1016/j.plaphy.2015.07.002>.
- Limaje, Ankur, Chad Hayes, J. Scott Armstrong, Wyatt Hoback, Ali Zarrabi, Sulochana Paudyal, and John Burke. 2018. "Antibiosis and Tolerance Discovered in USDA-ARS Sorghums Resistant to the Sugarcane Aphid (Hemiptera: Aphididae)1." *Journal of Entomological Science* 53 (2): 230–41. <https://doi.org/10.18474/JES17-70.1>.

- McCormick, Ryan F., Sandra K. Truong, Avinash Sreedasyam, Jerry Jenkins, Shengqiang Shu, David Sims, Megan Kennedy, et al. 2018. “The Sorghum Bicolor Reference Genome: Improved Assembly, Gene Annotations, a Transcriptome Atlas, and Signatures of Genome Organization.” *The Plant Journal* 93 (2): 338–54. <https://doi.org/10.1111/tpj.13781>.
- McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303. <https://doi.org/10.1101/gr.107524.110>.
- Monier, Brandon, Terry M. Casstevens, Peter J. Bradbury, and Edward S. Buckler. 2022. “RTASSEL: An R Interface to TASSEL for Analyzing Genomic Diversity.” *Journal of Open Source Software* 7 (76): 4530. <https://doi.org/10.21105/joss.04530>.
- Russell K. Monson et al., “Coordinated Resource Allocation to Plant Growth–Defense Tradeoffs,” *New Phytologist* 233, no. 3 (2022): 1051–66, <https://doi.org/10.1111/nph.17773>.
- Morris, Geoffrey P., Punna Ramu, Santosh P. Deshpande, C. Thomas Hash, Trushar Shah, Hari D. Upadhyaya, Oscar Riera-Lizarazu, et al. 2013. “Population Genomic and Genome-Wide Association Studies of Agroclimatic Traits in Sorghum.” *Proceedings of the National Academy of Sciences* 110 (2): 453–58. <https://doi.org/10.1073/pnas.1215985110>.
- Muleta, Kebede T., Terry Felderhoff, Noah Winans, Rachel Walstead, Jean Rigaud Charles, J. Scott Armstrong, Sujana Mamidi et al. "The recent evolutionary rescue of a staple crop depended on over half a century of global germplasm exchange." *Science advances* 8, no. 6 (2022): eabj4633.
- Mundt, Christopher C. 2018. “Pyramiding for Resistance Durability: Theory and Practice.” *Phytopathology*® 108 (7): 792–802. <https://doi.org/10.1094/PHYTO-12-17-0426-RVW>.
- Nagoshi, Rodney N., Isabel Dhanani, R. Asokan, H. M. Mahadevaswamy, Chicknayakanahalli M. Kalleshwaraswamy, Sharanabasappa, and Robert L. Meagher. 2019. “Genetic Characterization of Fall Armyworm Infesting South Africa and India Indicate Recent Introduction from a Common Source Population.” *PLOS ONE* 14 (5): e0217755. <https://doi.org/10.1371/journal.pone.0217755>.

- Nalam, Vamsi, Travis Isaacs, Sarah Moh, Jessica Kansman, Deborah Finke, Tessa Albrecht, and Punya Nachappa. 2021. “Diurnal Feeding as a Potential Mechanism of Osmoregulation in Aphids.” *Insect Science* 28 (2): 521–32. <https://doi.org/10.1111/1744-7917.12787>.
- Nalam, Vamsi, Joe Louis, and Jyoti Shah. 2019. “Plant Defense against Aphids, the Pest Extraordinaire.” *Plant Science* 279 (February): 96–107. <https://doi.org/10.1016/j.plantsci.2018.04.027>.
- Nibouche, Samuel, Laurent Costet, Raul F. Medina, Jocelyn R. Holt, Joëlle Sadeyen, Anne-Sophie Zoogones, Paul Brown, and Roger L. Blackman. 2021. “Morphometric and Molecular Discrimination of the Sugarcane Aphid, *Melanaphis Sacchari*, (Zehntner, 1897) and the Sorghum Aphid *Melanaphis Sorghi* (Theobald, 1904).” *PLOS ONE* 16 (3): e0241881. <https://doi.org/10.1371/journal.pone.0241881>.
- Nombela, Gloria, Valerie M. Williamson, and Mariano Muñoz. 2003. “The Root-Knot Nematode Resistance Gene Mi-1.2 of Tomato Is Responsible for Resistance against the Whitefly *Bemisia Tabaci*.” *Molecular Plant-Microbe Interactions: MPMI* 16 (7): 645–49. <https://doi.org/10.1094/MPMI.2003.16.7.645>.
- Painter, R. H. 1951. “Insect Resistance in Crop Plants.” *Insect Resistance in Crop Plants*. <https://www.cabdirect.org/cabdirect/abstract/19521603398>.
- Peterson, Robert K. D., Andrea C. Varella, and Leon G. Higley. 2017. “Tolerance: The Forgotten Child of Plant Resistance.” *PeerJ* 5 (October): e3934. <https://doi.org/10.7717/peerj.3934>.
- Poosapati, Sowmya, Elly Poretsky, Keini Dressano, Miguel Ruiz, Armando Vazquez, Evan Sandoval, Adelaida Estrada-Cardenas, et al. 2022. “A Sorghum Genome-Wide Association Study (GWAS) Identifies a WRKY Transcription Factor as a Candidate Gene Underlying Sugarcane Aphid (*Melanaphis Sacchari*) Resistance.” *Planta* 255 (2): 37. <https://doi.org/sz>.
- Rice, Brian, Carl VanGessel, Gael Pressoir, and Geoffrey Morris. 2024. “Flywheel Genomics.” TBD.
- Sujay Rakshit et al., “Changes in Area, Yield Gains, and Yield Stability of Sorghum in Major Sorghum-Producing Countries, 1970 to 2009,” *Crop Science* 54, no. 4 (2014): 1571–84, <https://doi.org/10.2135/cropsci2012.12.0697>.

- Snoeck, Simon, Natalia Guayazán-Palacios, and Adam D Steinbrenner. 2022. “Molecular Tug-of-War: Plant Immune Recognition of Herbivory.” *The Plant Cell*, January, koac009.
<https://doi.org/10.1093/plcell/koac009>.
- Susan S. Thayer and Eric E. Conn, “Subcellular Localization of Dhurrin β -Glucosidase and Hydroxynitrile Lyase in the Mesophyll Cells of Sorghum Leaf Blades 1,” *Plant Physiology* 67, no. 4 (April 1981): 617–22, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC425743/>.
- Tian et al., “Signature of Balancing Selection in Arabidopsis,” *Proceedings of the National Academy of Sciences* 99, no. 17 (August 20, 2002): 11525–30, <https://doi.org/10.1073/pnas.172203599>.
- Vuillet, A. 1914. “The millet aphids in the French Soudan.” *The millet aphids in the French Soudan.*, 563–64.
- Wang, Bo, Yiping Jiao, Kapeel Chougule, Andrew Olson, Jian Huang, Victor Llaca, Kevin Fengler, et al. 2021. “Pan-Genome Analysis in Sorghum Highlights the Extent of Genomic Variation and Sugarcane Aphid Resistance Genes.” *bioRxiv*. <https://doi.org/10.1101/2021.01.03.424980>.
- Wang, Faming, Songmin Zhao, Yonghua Han, Yutao Shao, Zhenying Dong, Yang Gao, Kunpu Zhang, et al. 2013. “Efficient and Fine Mapping of RMES1 Conferring Resistance to Sorghum Aphid *Melanaphis Sacchari*.” *Molecular Breeding* 31 (4): 777–84. <https://doi.org/10.1007/s11032-012-9832-6>.

Chapter 3 - Genome to phenome characterization of *RMESI* resistance to sorghum aphid indicates NLR-based mechanism

3.1 Summary

Mechanisms of host plant resistance (HPR) to aphids vary and can determine their utility in plant breeding for durable resistance. The globally-deployed sorghum aphid HPR locus *RMESI* reduces *M. sorghi* reproduction and classical R-gene mediated resistance and cyanogenesis have been proposed as causal mechanisms. Structural variation at the Chr06 locus includes a large insertion harboring copy number variation of two candidate nucleotide-binding leucine-rich repeat (NLR) genes whereas the β -cyanoalanine synthase enzyme *CAS* at *RMESI* is structurally conserved and constitutively expressed. Large transcriptomic changes were induced by *RMESI* including the upregulation of the salicylic acid pathway and resistance marker genes e.g. *PRI*, *PAD4*, and *LOX2*. The accumulation of salicylic acid, indole-3-acrylic acid, and IA-aspartic acid were significantly increased by *RMESI*. Diverse classes of defense metabolites upregulate in response to infestation, however the cyanogenic pathway and metabolite abundance was not altered, supporting one of several NLRs at *RMESI* as a master switch for broad aphid resistance mechanisms. Taken together, our findings suggest that *RMESI* is acting as an R-gene inducing multiple aphid resistance mechanisms.

3.2 Introduction

Plants must deal with diverse insect pests by presenting constitutive defenses or inducing host plant resistance (HPR). Constitutive HPR mechanisms are expressed regardless of the presence of a pest and examples include physical barriers and metabolic deterrents. In *Sorghum bicolor*, long-chain fatty acid content of epicuticular waxes influences the settling preference of sorghum aphid (*Melanaphis sorghi*) (Cardona et al. 2022). In maize, benzoxazinoids confer resistance to chewing and phloem sucking insects, with natural variation existing for a DIMBOA-Glc methyltransferase enzyme responsible for variation in constitutive DIMBOA-Glc concentration and corn leaf aphid (*Rhopalosiphum maidis*) resistance (Meihls et al. 2013). Induced HPR involves the expression of defense mechanisms in response

to an infestation (Chen 2008). In the Phaseolinae subtribe, oral secretions from armyworm (*Spodoptera exigua*) contain a host-derived protein fragment that induces defenses. A common R-gene mechanism of insect-induced defenses involves nucleotide-binding site leucine-rich repeat receptors (NLRs) which recognize pattern- or herbivore-associated molecular patterns (PAMPs, HAMPs) and initiate signaling cascades (Snoeck, Guayazán-Palacios, and Steinbrenner 2022). The first two aphid-resistance genes cloned, *Vat* and *Mi-1.2*, were NLRs which induced signaling and defense mechanisms (Dogimont et al. 2014; Q. Li et al. 2006). Early signaling (oxidative bursts, Ca²⁺ flux) give rise to transcriptional modulation of phytohormone pathways (salicylic acid, jasmonic acid) and expression of defense and antinutritive mechanisms (Zebelo and Maffei 2015).

Sorghum aphid emerged in 2013 in the Americas as a major pest of sorghum that quickly spread to all growing regions. The *Resistance to Melanaphis sorghi 1* locus (*RMESI*) on chromosome 6 (Chr06) was first reported between 2.67 Mb and 2.80 Mb in the reference genome BTx623v3 which lacks the resistant allele (F. Wang et al. 2013). *RMESI* was subsequently mapped using selection signatures at 2.99 Mb in a Haitian breeding population under high sorghum aphid infestation (Muleta et al. 2022). The publicly available KASP marker for *RMESI* has facilitated widespread deployment and improved management of sorghum aphid, however the selection pressure placed on aphid populations through antibiosis resistance highlights the potential for a biotype shift to occur.

The molecular mechanism for *RMESI*-based resistance is poorly understood and competing hypotheses involving cyanogenic toxicity or NLR-induced defenses have been proposed. The cyanogenic glucoside dhurrin is an antifeedant to chewing insects and the detoxification gene β -cyanoalanine synthase (*CAS*, Sobic.006G016900) (Gleadow et al. 2021; Gruss et al. 2022) is located adjacent to the *RMESI* locus (Muleta et al. 2022). Located within the original mapping interval are three predicted NLR genes, Sobic.006G017200, Sobic.006G017400, and Sobic.006G017500 (F. Wang et al. 2013). Hypotheses on *CAS* and NLR loci as resistance mechanisms would lead to different predictions on molecular phenotypes at the transcriptome and metabolome level. Polymorphisms in the CRE or CDS of *CAS* would predict limited expression changes or accumulation of dhurrin and/or hydrogen cyanide,

whereas variation for R-genes would predict larger scale modulation of gene expression, phytohormones, and the metabolome. Investigations of sorghum aphid resistance molecular mechanisms have not used material that isolates *RMESI*, such as NILs or mutants.

Structural variation, such as copy number variation (CNV), underlying important agronomic traits can be investigated with the help of pan-genomes (Song et al. 2020). For instance, in *Brassica napus*, only 57% of NLRs were shared by all 50 reference genomes and were more variable in clusters in relation to singletons (Dolatabadian et al. 2020), while in sorghum, presence absence variation and large InDels correspond to genes selected during domestication such as a 2 kb deletion in *Shattering1* (Tao et al. 2021). The breeding line RTx2783 contains the resistant *RMESI* allele and was reported to contain several structural variations >3 kb on Chr06, leading to the hypothesis that the causal gene is absent in BTx623. Here we use the *RMESI* near-isogenic lines (NILs) and the recently developed PI27683 reference genome to characterize genomic, transcriptomic, phytohormone, and metabolomic variation underlying sorghum aphid resistance.

3.3 Material and Methods

3.3.1 NIL development

RMESI near-isogenic lines (NILs) were developed with a donor parent IRAT204 (*M. sorghi* resistant, donor) and recurrent backcrossing to RTx430 (*M. sorghi* susceptible). Single plant selections were made at the F₂ using the KASP marker for *RMESI* (Sbv3.1_06_02892438R, Muleta et al. 2022) and homozygous +/+ plants were backcrossed. Population development was done at Kansas State University. *RMESI* homozygous BC₃F₄'s (NIL+, NIL-) were used for no-choice assay, RNA-sequencing, phytohormone quantification, and metabolomic analysis.

3.3.2 Aphid assays

M. sorghi were received from Dr. Scott Armstrong at the USDA-ARS Stillwater, Oklahoma. Aphids were reared on Tx7000 seedlings under laboratory conditions as described in Nalam et al. 2021.

Seedlings were grown in 4.5 inch pots with potting soil and top layer of greens grade to reduce damping off. Colonies were grown in a 46 × 46 × 76 cm cage (BioQuip Products Inc., Rancho Dominguez, CA). No-choice assays were used to compare aphid fecundity on NILs. A single seedling was grown in 6 inch pots using potting soil and a top layer of greens grade. At 3-4 weeks of age, three 4-5 day old apterous *M. sorghi* aphids were placed at the base of the seedlings with a camel hair brush. A clear plastic cylinder was placed over the plant to prevent aphids from leaving the pot with an organdy cloth covering for ventilation. The number of aphids on each plant were counted at the same time of day for a week.

Molecular hypotheses were tested using a *M. sorghi* infestation time course collected for RNA-sequencing, genotyping, and HPLC. A 2 × 3 factorial design was used with NIL+ and NIL- plants infested for 24 (24 hours post infestation, 24-hpi) and 48 hours (48-hpi) as well as an uninfested (control) sample collected at the same time as the 48-hpi sample. A 50 ml falcon tube with the conical end cut off and a hole with organdy cloth on the cap was placed over the third true leaf of two plants and infested with twenty adult apterous aphids. Cotton balls were used to cover the open end of the tube to ensure maximum response from aphid feeding. Control samples were handled the same way but were not infested. Samples were collected at 12pm after 24 and 48 hours by pooling both leaves. There were 2-4 replicates collected. The bottom (basal) inch of tissue from both leaves were flash frozen for sequencing. The remainder of the sample was flash frozen for HPLC analysis.

3.3.3 Transcriptome sequencing

RNA was extracted using Zymo Quick-RNA minipreps (Thermo Scientific) and treated with DNase using Invitrogen Turbo DNA-free kit (Thermo Scientific). RNA quality was checked using a NanoDrop 2000 (Thermo Scientific) and ~2 µg was submitted on dry-ice to Novogene Corporation Inc. (2921 Stockton Blvd, Sacramento, CA, US, 95817). Samples were sequenced on an Illumina NovaSeq 6000 Sequencing System with 150 bp paired-end reads with ~6 Gb of data generated per sample.

3.3.4 Sequencing data analysis

Trimmed and quality filtered reads generated by Novogene were aligned to the PI276837 and RTx430v2 reference genomes for transcriptome analysis and variant calling, respectively (Deschamps et al. 2018). Reference genomes were indexed and mapped to with STAR v2.7.10 using 2-pass mode (Dobin et al. 2013). For variant calling, BAM files were processed using the following in GATK v 4.2.5.0 unless otherwise noted: Duplicates were marked and read groups were added using Picard, reads were split using SplitNCigarReads, variants were called individually using HaplotypeCaller, gVCFs were combined using CombineGVCFs, and jointcalled VCF files were produced with GenotypeGVCFs (McKenna et al. 2010). Finally, VCFs were filtered using VariantFiltration ($QUAL > 30$, $SQR > 3$, $FS > 60$, $MQ < 40$). VCFs were analyzed in base R with variants summarized over 0.5-Mb windows and segregating markers were used to estimate segregation percentage between NILs.

For transcriptome analysis, transcript abundance was quantified using FeatureCounts v2.0.1 (Liao, Smyth, and Shi 2014). Differential gene expression was determined using DESeq2 v1.38.1 in R v4.2.2 (Love, Huber, and Anders 2014; R Core Team 2021). Principal component analysis was performed with R/prcomp and plotted with ggplot2 v3.4.1 (Wickham 2016). Genes were determined to be significantly differentially expressed with a Benjamini-Hochberg adjusted p -value < 0.05 and fold change greater than 1.5 ($L2FC > 0.58$).

Analysis of potential causal gene and pathway candidates were determined *a priori* through literature search and SorghumBase orthology finder (Gladman et al. 2022). Genes involved in aphid defense and/or phytohormone signaling in Arabidopsis or other species were searched on SorghumBase and all orthologous genes in BTx623v5 were considered as candidate genes (Table 1). Lipxygenase and jasmonate ZIM-domain gene families were characterized in sorghum previously and Sobic IDs were already available (Shrestha and Huang 2022; Shrestha, Pant, and Huang 2021). Transcriptome and genomic analysis was done in PI276837 in order to identify *RMESI* locus differences as it contains the resistant allele. In order to convert Sobic IDs to PI276837, orthology groups among the pangenomes were determined using OrthoFinder as described in Rice, Spiekerman, Lovell et al, (2024). In the absence of

rigorous phylogenetic analysis, all orthologous genes (homologs within the sorghum pangenome) were retained as potential candidates (Table 3.1). For analysis of potential causal genes, all 35 PI276837 genes between 2.5–3.5 Mb were considered. PI276837 genes are named as such “SbPI276837.06G016100” where Sb = *S. bicolor*, PI276837 = genotype accession, 06G = chromosome number, 016100 = gene number on chromosome where the example is the 161st annotated gene on chromosome 6.

3.3.5 Structural Variation and Orthology analysis

The one megabase region centered on 3 Mb from BTx623v5 and PI276837 were aligned using MUMMER4 and default parameters (Marçais et al. 2018). Output was converted into a text file for plotting in *r/ggplot2* (Wickham 2016). *RMES1* mapped positions reported in BTx623v3 were converted to v5 by blasting the 100-bp flanking sequence of markers. Structural variation annotation was done using Synteny and Rearrangement Identifier (SyRI) (Goel et al. 2019). Variants were filtered for highly diverged regions (HDR) and size > 100 bp.

Orthology of sorghum genes was determined using Orthofinder as described in (Emms and Kelly 2019; Rice et al. 2024). Briefly, protein sequences of 32 sorghum pangenome members and *Maize*, *Setaria*, *Brachypodium*, and *Panicum* were used for phylogenetic comparison to identify orthologous groups (OGs) of genes. Copy number variation was identified as genes belonging to the same OG and shared in both BTx623v5 and PI276837 but at different copy numbers as defined here Springer et al. 2009.

3.3.6 Phytohormone and Metabolome quantification

As reported by Analytical Resources Core, Bioanalysis and Omic (RRID: SCR_021758) - The phytohormone analysis was conducted following reference (Almeida Trapp et al. 2014) with modifications described below. Frozen samples were lyophilized. The dried samples were added with stainless steel balls and homogenized. The homogenate was added with cold 80% methanol in water, 20 μ L of phytohormone internal standards (SA-d4, ABA-d6, JA-d5, IAA-d5 in 50% methanol), and 10 μ L of internal standards for untargeted analysis (L-alanine-13C3, L-phenylalanine-13C6, fumaric acid-13C4, L-

tryptophan-13C11, and indole-3-acetic-acid-13C6 in 50% methanol). The mixture was vigorously mixed, followed by sonication, and another mixing. Then the mixture was centrifuged at 15,000 g and 4°C for 15 min. Supernatants were recovered, of which 0.1 mL was saved for untargeted analysis and 0.8 mL was transferred to a new vial. To the remaining pellets, 80% methanol was added, and the sample was mixed for 15 min. The supernatant after centrifugation was combined with first aliquot of 0.8 mL, dried down, and resuspended in 50% methanol for phytohormone analysis. A small aliquot (20 µL) was taken from each study sample and pooled to generate a quality control (QC) sample. Sample extracts and QCs were stored at -80°C until analysis. The authentic standards used in this assay included jasmonic acid-d5 (JA-D5), jasmonic acid (JA), salicylic acid-D4 (SA-D4), salicylic acid (SA), indole-3-acrylic acid (IACA), indole-3-carboxylic acid (ICA) were obtained from Sigma-Aldrich (St. Louis, MO, USA). Indole-2,4,5,6,7-d5-3-acetic acid (IAA-D5) were obtained from CDN Isotopes (Canada). 12-oxo-phytodienoic acid (OPDA) was obtained from Cayman (Ann Arbor, MI). IA-aspartic acid (IA-Asp) was obtained from Toronto Research Chemicals (Canada). Abscisic acid-D6 (ABA-D6) was obtained from Olchemim (Czech Republic). L-alanine-13C3, L-phenylalanine-13C6, fumaric acid-13C4, L-tryptophan-13C11, and indole-3-acetic-acid-13C6 were obtained from Cambridge Isotope Laboratories (MA, USA).

UPLC-MS/MS analysis was performed on a Waters ACQUITY Classic UPLC coupled to a Waters Xevo TQ-S triple quadrupole mass spectrometer. Chromatographic separations were carried out on a Waters ACQUITY HSS T3 column (2 x 50 mm, 1.7 µM). Mobile phases were (A) water with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid. The LC gradient was as follows: time = 0 min, 1% B; time = 0.65 min, 1% B; time = 2.85 min, 99% B; time = 3.5 min, 99% B; time = 3.55 min, 1% B; time = 5 min, 1% B. Flow rate was 0.5 mL/min and injection volume was 3 µL. Samples were held at 6°C in the autosampler, and the column was operated at 45°C. Mass detector was operated in ESI+ and ESI- mode. The capillary voltage set to 0.7 kV. Inter-channel delay was set to 3 msec. Source temperature was 150°C and desolvation gas (nitrogen) temperature 450°C. Desolvation gas flow was 1000 L/h, cone gas flow was 150 L/h, and collision gas (argon) flow was 0.15 mL/min. Nebulizer pressure (nitrogen) was set to 7 Bar. The MS acquisition functions were scheduled by retention times. Auto dwell feature was set for

each function and dwell time was calculated by Masslynx software (Waters) to achieve 12 points-across-peak as the minimum data points per peak. The retention time, MRM transitions, cone and collision energy of each compound were described in spreadsheet “transitions”.

All Raw data files were imported into the Skyline open source software package (MacLean et al. 2010). Each target analyte was visually inspected for retention time and peak area integration. Peak areas were extracted for target compounds detected in biological samples and normalized to the peak area of the appropriate internal standard or surrogate in each sample. Absolute quantitation (ng/g) was calculated using the linear regression equation generated for each compound from the calibration curve.

Sample run order was fully randomized, with a pooled QC sample injected approximately every 7 injections. The process (extraction) blank was also injected last. One microliter of each sample was injected onto a Waters Acquity UPLC system. Separation was achieved using a Waters ACQUITY UPLC Premier T3 1.7 μ m Column (2.1 x 100 mm), using a gradient from solvent A (0.1% formic acid in water) to solvent B (0.1% formic acid in acetonitrile) and a flow rate of 0.5 mL/min. The column and samples were held at 45 °C and 6 °C, respectively. The column eluent was infused into a Waters Xevo G2-XS Q-TOF-MS with an electrospray source in negative ionization sensitivity mode, with MSE data independent MS/MS acquisition. The following parameters were used for MS1 scan: 50-1200 m/z mass range with 0.1 seconds per scan, collision energy 6 V. MSE acquisition occurred at a scan rate of 0.1 seconds, mass range 50-1200 m/z, and collision energy was ramped from 15 to 30 V. Calibration was performed using sodium formate with 1 ppm mass accuracy. The capillary voltage was held at 700 V in positive mode or 2200 V in negative mode. The source temperature was held at 150 °C and the nitrogen desolvation temperature at 450 °C with a flow rate of 1000 L/hr. Lockspray reference mass was used to correct for drift, with 40 seconds interval between scans, 0.1 seconds/scan and signal averaged over 3 scans. LeuEnk was used for mass correction, with reference masses of either positive 556.2771 m/z or negative 554.2615 m/z.

RAMClustR version 1.2.4 in R version 4.2.2 (2022-10-31) was used to normalize, filter, and group features into spectra. MSFinder (Tsugawa 2016) was used for spectral matching, formula inference,

and tentative structure assignment. Results were imported into the RAMClustR object. A total score was calculated based on the product scores from the findmain function and the MSfinder formula and structure scores. A total of 14130 annotation hypotheses were tested for 4306 compounds. Spectra matches took precedence over computational inference based annotations. The following database(s) were assigned as 'priority': chebi, coconut. The database priority.factor was set to 0.9 to decrease scores for compounds which failed to match priority database(s). The inchikey priority.factor was set to 0.9 to decrease scores for compounds with non-matching inchikey(s). The highest total score was selected for each compound, considering all hypotheses.

3.4 Results

3.4.1 *RMESI* NILs are appropriate for testing molecular mechanism

In order to test hypotheses on *RMESI* mechanism, BC3 near-isogenic lines (NIL+, NIL-) were derived from IRAT204 (*RMESI* donor parent) and RTx430 (susceptible recurrent parent). RNA-sequencing of NILs mapped to RTx430v2 was used to determine the genotype. We found that 94.7% of the genome was isogenic in our NILs with regions segregating on chromosomes 1, 2, 3, 6, 7, and 8 (Figure 3.2a). NIL+ contained an introgression of 5.16 Mb on Chr06 from 2,126,567 to 7,281,855 (in RTx430v2 coordinates). The syntenic region on PI276837 corresponds with 2,061,383 to 7,246,956 which encompasses the *RMESI* QTL at approximately 2.75–3.25 Mb. Phenotyping of the NILs confirmed they retained the antibiosis-resistance trait (Figure 3.2b,c). A significant difference in aphid population was observed at 5 days post infestation ($p < 0.01$) and 7 days ($p < 2e-4$), with a rate of 1.44 nymphs per day on NIL- where as NIL+ had a rate of 0.73 nymphs per day.

3.4.2 Resistant NILs undergo widespread transcriptomic changes after aphid infestation

Several annotated NLR genes near *RMESI* in BTx623 have been highlighted as candidate causal genes (Wang et al. 2013). The hypothesis that *RMESI*-resistance is induced by an NLR gene, as opposed to constitutive mechanisms, would predict the transcriptome of infested NILs to be significantly altered in

NIL+ relative to NIL-. To test for global induced responses, we sequenced the transcriptomes of uninfested (control), 24 hours post infestation (24-hpi) and 48-hpi NILs. The NIL+ underwent global transcriptional changes relative to the NIL- evident in principal component analysis (Figure 3.3a). Principal component 1 (PC1) explained 29% of variance (PVE) and captured transcriptional differences in infested NILs at both time points relative to uninfested samples. Differences along PC1 were more pronounced in the NIL+ than NIL-. PC2 (PVE = 17%) captured genotype differences. PC3 (PVE = 10%) distinguished between 24 hpi and 48 hpi timepoints with control samples intermediate (Figure 3.4b). There were 312 and 555 genes significantly up-regulated and down-regulated, respectively, in NIL+ relative to NIL- (adjusted $p < 0.05$, fold change > 1.5) representing constitutive transcriptional differences. There were 1,807 and 1,950 genes up-regulated in NIL+ at 24-hpi and 48-hpi relative to uninfested controls, respectively, whereas 1,933 and 2,101 genes were down-regulated (Figure 3.4c). Only 238 and 66 genes were upregulated in NIL- at 24- and 48-hpi, respectively, and 255 and 24 genes were downregulated.

3.4.3 Pangenomes reveal structural variation at *RMES1*

RMES1 was mapped between 2.7–3.1Mb on Chr06 in BTx623v3, which lacks the resistant allele (F. Wang et al. 2013; Muleta et al. 2022). In order to generate hypotheses on what specific genomic variant(s) underlie resistance, we compared the recently developed reference genome (PI276837v1) which possesses the *RMES1* resistant allele (Muleta et al. 2022) to the susceptible BTx623v5, referred to hereafter as BTx623. We aligned a 1 Mb region centering on *RMES1* (Chr06:2500000..3500000) of PI276837 to BTx623 and identified several structural variants absent in BTx623 (Figure 3.4a). There were nine highly diverged regions (HDRs) > 10 kb near *RMES1* including two which correspond to insertions (2,929,232 – 2,978,048 and 2,989,720 – 3,176,946) (Figure 3.4b). Several smaller SVs (< 3 kb) also existed in the region.

Next, we evaluated candidate causal loci using genomic and transcriptomic evidence. There were 35 genes predicted in PI276837 between 2.5–3.5 Mb on Chr06 (Figure 3.4c). Twenty-one genes overlap

with HDRs. Candidate loci across the region based on hypothesized function include CAS (cysteine synthase C1, SbPI276837.06G016100) and eight tandem NLR loci (SbPI276837.06G016400 – SbPI276837.06G017100). *CAS1* and 7 of 8 NLR loci were located in HDRs, with the NLR genes in two HDRs corresponding to the large insertion. SbPI276837.06G016400 was in a conserved region. We used orthology groups (OG) to determine the relationship between the three BTx623 annotated NLRs and eight PI276837 NLR loci. In OG0020033 (hereafter, NLR-A), Sobic.006G017200, SbPI276837.06G016400 and SbPI276837.06G016900 were homologous and represent a CNV. In OG0000032 (hereafter NLR-B), Sobic.006G017400 and Sobic.006G017500 were homologous to the remaining six NLR loci and represent a second CNV. NLR-A did not have any homologs outside of the *RMES1* region of Chr06 and was not orthologous to NLRs in other grasses. Genes belonging to NLR-B were more diverse with homologs on chromosome 7 and 10, as well as in *Brachypodium*, *Panicum*, *Setaria*, and *Zea*.

Transcriptome sequencing and the *RMES1+* reference genome allowed hypotheses on CRE-based mechanisms to be tested. Only 28 of 35 annotated genes were expressed in one or more samples (Figure 3.4d). Two NLR-B genes (SbPI276837.06G016500 and SbPI276837.06G016700) were not expressed in either genotype. Two genes (SbPI276837.06G015300, SbPI276837.06G018000) were not expressed in resistant genotypes, whereas four genes (SbPI276837.06G016800, SbPI276837.06G017000, SbPI276837.06G0171000, SbPI276837.06G017700) were not expressed in susceptible genotypes. The most strongly expressed gene was *CAS1* (SbPI276837.06G016100). Both genes in NLR-A (SbPI276837.06G016400 and SbPI276837.06G016900) were expressed in all samples.

In comparing infested transcriptomes to controls, only *NIL+* had differentially expressed genes (Figure 3.4e). Two genes were upregulated at both timepoints (SbPI276837.06G018000, SbPI276837.06G018400) whereas SbPI276837.06G017400 was downregulated at 24-hpi. To determine genetic differences in expression (constitutive), we compared control *NIL+* to control *NIL-* and found three genes which were upregulated in resistant *NILs* relative to susceptible *NILs* (SbPI276837.06G016400, SbPI276837.06G016600, and SbPI276837.06G017700). Low expression of

several NLR-B genes led to apparent expression only in resistant NILs but did not result in significant genotype differences (Figure 3.4e).

3.4.4 The salicylic acid pathway is differentially regulated along with phytohormone accumulation

Phytohormones are a major component of induced HPR signaling. Expression of biosynthesis and signaling genes are indicative of a plant response to infestation, typically to mount a defense in ‘resistant’ plants. We examined homologs of the salicylic acid pathway in NIL transcriptomes and found strong upregulation in response to infestation in an *RMES1*-dependent manner (Figure 3.5). The chorismate mutase homologs were not differentially expressed, however the PAL and ICS biosynthesis pathways were significantly altered. Among six *PAL* homologs, only SbPI276837.06G113360 was significantly upregulated at both 24- and 48-hpi in NIL+. A homolog of *AIM*, SbPI276837.04G112200, was also significantly upregulated at both timepoints. In contrast, the single ICS homolog present in sorghum was downregulated in NIL+ at both timepoints and NIL- at 24-hpi. Several other pathway homologs (*EDS5*, *EPS1*) had reduced expression in either genotype. The final step in the ICS pathway involves *EPS1*, which has three homologs in sorghum. SbPI276837.07G011200 expression is reduced after infestation but was only significant in NIL+ at 48-hpi. SbPI276837.07G012800 belongs to an orthogroup which is only present in 15 of the 32 member pangenome and was significantly upregulated in NIL+ at both timepoints. We looked for evidence of salicylic acid accumulation in known receptor and signaling genes downstream of phytohormone biosynthesis. We found two PR1 homologs, SbPI276837.01G304400 and SbPI276837.10G018500, were significantly less expressed in NIL+ prior to infestation, but upregulated in NIL+ in response to aphid infestation. SbPI276837.10G018500 was significantly downregulated in NIL- at 24-hpi.

We next examined other phytohormone signaling pathways for response to infestation. The jasmonic acid (JA) pathway response was stronger in NIL+ plants than NIL- (Figure 3.6). Three members of the 13-LOX gene family and *AOC* were significantly upregulated in NIL+ lines. Other members of the JA biosynthesis pathway like *AOS* and *OPR7* were expressed higher in NIL+ plants but were not

significantly differentially expressed after infestation. Finally, the *COII* gene involved in JA signalling was upregulated due to infestation in NIL+ plants. While none were statistically significant between control samples, genes involved in JA biosynthesis were lower expressed in NIL+ plants than NIL-.

Jasmonate ZIM domain (JAZ) family, involved in JA downstream signaling and crosstalk, are largely unresponsive to aphid infestation regardless of genotype (Figure 3.6). Only *JAZ14* was upregulated significantly at 24-hpi in NIL+. Notably, several members of this family were expressed lower in NIL+ plants constitutively, including *JAZ3*, *JAZ4*, *JAZ6*, *JAZ8*, and *JAZ9*. Finally, we examined genes involved in ABA biosynthesis and found positive regulation in NIL+ lines after infestation but not NIL-. One homolog of *AAO4*, *NCED9*, *NAC1*, and *NAC2* were upregulated significantly at both timepoints or 48-hpi in NIL+.

Host-plant resistance mechanisms proposed or known to negatively affect aphid infestation include death acids produced by 9-LOX, WRKY86, PAD4, and the dhurrin pathway. We found that *LOX5* and *LOX9* were downregulated in NIL+ and NIL-, respectively, at 24-hpi (Figure 3.7). *LOX2* was upregulated significantly in NIL+ lines at both timepoints. *WRKY86* was not significantly differentially expressed, however its expression was reduced at 24-hpi and elevated at 48-hpi in both genotypes. Finally, *PAD4* was strongly upregulated at 24-hpi in NIL+ while its constitutive expression was significantly lower in NIL+. The dhurrin pathway was generally downregulated after infestation (Figure 3.7). Two members of the biosynthetic gene cluster on Chr01, *CYP79A1* and *CYP71E1*, and the transporter enzyme *MATE2* were downregulated after infestation.

Phytohormone quantification was used to confirm that salicylic acid and other signaling molecules were responding to aphid infestation treatment and *RMES1* genotype. Salicylic acid abundance was significantly different ($p < 0.05$) between genotype and treatment, and increased over the time course in both genotypes. Salicylic acid fold increase was 11.9 and 12.9 in NIL+ relative to the control whereas NIL- had a 1.7 and 4.8 fold increase. Jasmonic acid, OPDA, and IAA accumulation responded to the treatment only ($p < 0.01$). Jasmonic acid and OPDA had similar trends of less hormone abundance at 24-hpi relative to control and 48-hpi. Differences in ICA and IA-asp were significant ($p < 0.01$) for both

genotype and phenotype, both increasing over the time course but always more abundant in NIL+ than NIL-. ABA expression was not significantly affected by treatment or genotype.

3.4.5 Metabolite signatures are associated with resistance

Evidence of induced transcriptional and phytohormone changes depending on *RMESI* would predict metabolite responses to infestation. In order to generate hypotheses on metabolites involved in the *RMESI*-mechanism, we examined the metabolome of infested NILs. There were 4,306 metabolites detected in our time course, however only 202 were significantly altered (fold change > 2, p -value < 0.01) after infestation in at least one comparison. There were more metabolites differentially expressed in NIL+ than NIL- at both timepoints (Figure 3.9). There were slightly more down-regulated metabolites at 24-hpi in NIL+, however there were more up-regulated metabolites at 48-hpi. Whereas the number of differentially expressed metabolites increased between 24-hpi and 48-hpi for NIL+, it decreased for NIL-.

Several metabolites had large fold change due to infestation in NIL+ and are candidates for aphid-resistance mechanisms. Six analytes including a putative phenolic glycoside, hydrolyzable tannin, and pentose phosphate were among the most strongly responsive metabolites upregulated at both timepoints in NIL+ but not NIL- (Figure 3.10). A distinct pentose phosphate was highly upregulated at 48-hpi in both genotypes. A putative wax monoester and a O-glycosyl compound was downregulated in NIL+ lines throughout the time course.

3.5 Discussion

Biotic resistance traits can be overcome by evolving pest populations and must therefore be characterized and managed, however plant pathogen systems have been much better elucidated than insects. In cereal and aphid systems, few HPR mechanisms have been defined and open questions remain as to the effectiveness and durability of individual mechanisms (Mou et al. 2023; Harris-Shultz, Armstrong, and Jacobson 2020). Here, we show *RMESI* induces dramatic changes in the transcriptome and metabolome of sorghum and propose a set of candidate NLR loci underlying the QTL. In response to

selection pressure since the emergence of *M. sorghi* in 2013, the *RMESI* locus was rapidly swept to fixation in the Haitian sorghum breeding program and has since become widely used (Muleta et al. 2022). The biotype shifts of several aphid species (*S. graminum*, *D. noxia*) on cereals in the Great Plains motivates research on the molecular mechanism of *RMESI* to establish its durability and how best to deploy it. The antifeedant properties of dhurrin to *S. frugiperda* and genomic variation for the detoxification enzyme CAS1 (Figure 3.4a) support its inclusion as an *RMESI* candidate (Gruss et al. 2022). However, the lack of constitutive or induced transcriptional difference (Figure 3.4), or dhurrin abundance, do not support dhurrin as a mechanism. Likewise, dhurrin does not have a known signaling role that would result in large transcriptomic and metabolomic reorganization as are seen in NIL+ (Figure 3.3, 3.9). *CAS1* is highly expressed in developing sorghum leaves where its function detoxifying released hydrogen cyanide would be expected to decrease in order to provide resistance causing auto-toxicity as well (Gleadow et al. 2021). Interestingly, *RMESI* downregulates biosynthetic and transport steps of the dhurrin pathway (Figure 3.7) and may reflect redirection of primary and secondary metabolism facilitating plant defense, as seen in Arabidopsis and rice responses of SnRK1 and TOR to biotic pests (Margalha, Confraria, and Baena-González 2019; De Vleeschauwer et al. 2018).

Large genomic variation may be missed when a single, susceptible, reference genome is used to analyze HPR molecular mechanisms. Structural variation reported in *RMESI* genotype RTx2783 was similar to PI276837 with a large insertion that encompasses several NLR genes (B. Wang et al. 2021). Both NLR-A and NLR-B orthogroups appear to contain copy number variation in PI276837 with two and five homologs, respectively, compared to one and two homologs present in BTx623 (Muleta et al. 2022). Plants have extensive genomic variation for NLR gene which often cluster and are found in regions of structural rearrangements (Van de Weyer et al. 2019; B. Wang et al. 2021; Dolatabadian et al. 2020). The presence of NLR gene SbPI276837.06G016400 in a region largely shared by both reference genomes and its constitutive difference in expression suggests this member of the NLR-A orthogroup may provide resistance through higher expression of the receptor. The apparent expression of the other NLR-A orthogroup member, SbPI276837.06G016900, in susceptible NILs is unexpected as it lies within the

insertion and is absent in BTx623v5 and other *rmesI*- accessions of the pangenome resource. This discrepancy could be due to sequence homology between the two homologs confounding read-mapping (Robert and Watson 2015). The four expressed NLR-B orthogroup homologs in PI276837 are also candidates for *RMESI*, however transcription of SbPI276837.06G016800 and SbPI276837.06G017100 is nearly undetectable and unlikely as candidates (Figure 3.4d). SbPI276837.06G016800 and SbPI276837.06G017000 are only expressed in NIL+ similar to SbPI276837.06G016600, however the latter is expressed more strongly and a more likely candidate for *RMESI*. The *RMESI* region is likely to have undergone more than one structural rearrangement, a common feature of subtelomeric regions (N. Li et al. 2023). The BTx623 organization of one NLR-A gene and two NLR-B genes in tandem appears to be duplicated in PI276837 as well as a second duplication of one of the NLR-B loci. These two orthologous groups at *RMESI* are equally likely candidates, with transcriptomic evidence elevating SbPI276837.06G016400 and SbPI276837.06G016600 as candidate genes, however positional cloning or gene-editing will be required to conclusively determine the causal loci underlying host plant resistance. It should be noted that expression differences of four other loci near *RMESI* were observed and can not be ruled out as candidate loci without further investigations, however unlikely.

Aphid induced changes to phytohormone pathways have been observed in maize, wheat, and sorghum (Batyrshtina et al. 2020; Tzin et al. 2015; Huang, Shrestha, and Huang 2022) but their role in *RMESI* HPR has not been demonstrated. Comparisons between resistant RTx2783 and susceptible BTx623 showed JA, ABA, IAA, and ICA abundance, as well as pathway marker gene expression, was higher in the resistant genotype at either 1 or 3 day post infestation, supporting their role in antibiosis-resistance (Huang, Shrestha, and Huang 2022; Limaje et al. 2018). We found *RMESI* led to ~2.5 fold increase in JA at 48-hpi but was not significantly different between genotypes along with ABA, IAA, and ICA. JA and ABA biosynthesis genes (*LOX*, *AOC*, *NCED9*) and marker genes (*COII*, *NAC2*) were induced in NIL+ plants only, suggesting these pathways may be modulated by *RMESI* over longer periods of infestation or they have additional functions that do not increase JA and ABA abundance.

SA is a major pathway for aphid resistance in R-gene pathways (Q. Li et al. 2006) and our study suggests it is the primary signaling component of *RMESI* (Figures 3.5, 3.6, 3.8). SA was reported to increase dramatically after infestation but remained upregulated only in RTx2783 at 6 days post infestation. The exogenous application of SA induced antibiosis resistance in BTx623 to a higher degree than other phytohormones, demonstrating its role in inducing aphid defense (Huang, Shrestha, and Huang 2022). The induction of the PAL biosynthesis pathway and *PRI* marker genes in NIL+ agreed with the rapid increase in SA abundance at 24-hpi and its central role in the defense induction (Figure 3.6, 3.8). SA is an important signaling mechanism for *Mi-1.2* mediated resistance in tomato to *M. euphoribae* as well as proposed in *RAG1*-resistance to soybean aphid (Q. Li et al. 2006; Studham and MacIntosh 2013). We propose that *RMESI* is similarly activating defenses through NLR-induced SA signaling.

The PAL pathway is likely a major SA biosynthesis pathway in cereals like rice and barley, as opposed to the ICS pathway which is the most important in *Arabidopsis* (Qin et al. 2019; Duan et al. 2014). We found orthologs of both pathways in PI276837 but induction of *PAL* and *AIM* coincided with down regulation of *ICS* suggesting that *ICS* activity is antagonistic for SA accumulation or signaling (Figure 3.5b). Of the three homologs with orthology to *AtEPS1*, an acyltransferase which catalyzes the last step in the ICS pathway, two have drastically different responses to infestation and may represent diverged function or more complex SA biosynthesis mechanisms which are *PBS3/EPS1* independent (Torrens-Spence et al. 2019). Downstream of phytohormone accumulation, metabolite responses included several metabolite groups known to play roles in aphid defense. Compounds predicted to be glucosinolates, alkaloids, and cardenolides were induced by *RMESI* and have known roles in aphid defense (Züst and Agrawal 2016). Some glucosinolates can elicit callose deposition which can also contribute to defense (Clay et al. 2009). Tannins are another compound linked to aphid resistance (Grayer et al. 1992) and a candidate hydrolyzable tannin is strongly upregulated only in resistant lines (Figure 3.9). The identity of these and other compounds detected in the *RMESI* NIL metabolome will be further investigated, as well as correlated with gene expression to build stronger connections between genotype and endophenotypes relevant to resistance (Figure 3.1).

NLR genes have been proposed as *S. graminum* resistance mechanisms in sorghum (Zhang, Huang, and Huang 2022) and to underlie *RMESI* resistance (Wang et al. 2021) which imply global changes to molecular phenotypes characteristic of R-gene resistance (Wang, Song, and Chai 2023). Under the R-gene hypothesis supported here (Figure 3.3, 3.4, 3.5, 3.8, and 3.9), an undetermined HAMP is leading to effector triggered immunity through a suite of mechanisms (Snoeck, Guayazán-Palacios, and Steinbrenner 2022). The lack of transcriptional or metabolic responses in NIL- plants could be due to undetected feeding by the aphid where no other induction mechanisms are triggered. Alternatively, the aphid may be successfully manipulating host plant responses to maintain feeding site quality (Yates and Michel 2018). *RMESI* appears to act as a defense master switch that induces responses in an otherwise susceptible genomic background (Figure 3.2). R-genes were first associated with hypersensitive response (HR) in pathosystems where localized cell-death was observed (Morel and Dangl 1997). However, like *Mi-1.2* which does not induce HR in tomato to potato aphid *Macrosiphum euphorbiae* (Martinez de Ilarduya, Xie, and Kaloshian 2003), *RMESI* also does not induce HR. In the breeding population where *RMESI* was mapped through selection signatures, many unidentified QTL were also selected for suggesting natural variation for downstream defenses could exist which increase the *RMESI* phenotype (Muleta et al. 2022). *WRKY86* was not significantly regulated by *RMESI*. In susceptible genotype BTx623, this putative aphid resistance gene was downregulated immediately after infestation (2-hpi and 4-hpi) and only moderately upregulated ($p < 0.05$, L2FC ~ 0.6) at 48-hpi and therefore may still contribute to resistance at later stages of infestation (Kiani and Szczepaniec 2018). The large transcriptome and metabolome reorganization suggest multiple defense and metabolic pathways respond to infestation as opposed to an individual mechanism like cyanogenesis which would involve few genes and metabolites. Our proposed genotype to phenotype map should be rigorously tested to confirm or propose alternative hypotheses. However, this R-gene mechanism highlights the importance of mindful breeding for aphid resistance in order to bolster *RMESI* with *RMES2* and additional quantitative sources of resistance.

Chapter 3 Tables

Table 3.1, *a priori* candidate loci for signaling and defense pathways.

Gene	ID	Gene family/acronym
SbJAZ1 (JAZ)	SbP1276837.01G234400	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ2 (JAZ)	SbP1276837.01G234500	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ3 (JAZ)	SbP1276837.01G234600	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ4 (JAZ)	SbP1276837.01G234700	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ5 (JAZ)	SbP1276837.01G234900	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ6 (JAZ)	SbP1276837.01G254300	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ7 (JAZ)	SbP1276837.01G312700	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ8 (JAZ)	SbP1276837.01G442400	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ9 (JAZ)	SbP1276837.01G442500	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ10 (JAZ)	SbP1276837.01G442600	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ11 (JAZ)	SbP1276837.02G033000	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ11 (JAZ)	SbP1276837.02G053100	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ12 (JAZ)	SbP1276837.02G164200	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ13 (JAZ)	SbP1276837.02G179700	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ14 (JAZ)	SbP1276837.02G326700	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ15 (JAZ)	SbP1276837.03G372400	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ16 (JAZ)	SbP1276837.06G050000	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ17 (JAZ)	SbP1276837.06G224500	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ18 (JAZ)	SbP1276837.K050300	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbJAZ18 (JAZ)	SbP1276837.07G111400	Jasmonate-Zinc-finger Inflorescence Meristem domain
SbABA2 (ABA)	SbP1276837.01G038800	ABSCISIC ACID DEFICIENT2
SbAAO4-1 (ABA)	SbP1276837.01G056800	ALDEHYDE OXIDASE 4
SbAAO4-2 (ABA)	SbP1276837.01G056900	ALDEHYDE OXIDASE 4
SbAAO4-3 (ABA)	SbP1276837.01G057000	ALDEHYDE OXIDASE 4
SbNCED9-1 (ABA)	SbP1276837.01G141900	9-cis-epoxycarotenoid dioxygenase
SbNCED9-2 (ABA)	SbP1276837.02G034600	9-cis-epoxycarotenoid dioxygenase
SbNAC1 (ABA)	SbP1276837.01G473000	NAM-ATAF1.2-CUC2
SbNAC2 (ABA)	SbP1276837.05G016100	NAM-ATAF1.2-CUC2
SbAMI (JAA)	SbP1276837.02G330600	amidase 1
SbLOX5 (13-LOX)	SbP1276837.06G083300	13-Lipoxygenase
SbLOX9 (13-LOX)	SbP1276837.04G072200	13-Lipoxygenase
SbLOX2 (13-LOX)	SbP1276837.01G443200	13-Lipoxygenase
SbWRKY86 (aphid_defense)	SbP1276837.09G210900	WRKY transcription factor
SbPAD4 (putative) (aphid_defense)	SbP1276837.07G168900	Phytoalexin-deficient
CYP79A1 (Dhurrin-bio)	SbP1276837.01G011000	Cytochrome P450
CYP71E1 (Dhurrin-bio)	SbP1276837.01G010900	Cytochrome P450
CYP71E1 (Dhurrin-bio)	SbP1276837.01G016600	Cytochrome P450
CYP71E1 (Dhurrin-bio)	SbP1276837.01G016900	Cytochrome P450
UGT85B1 (Dhurrin-bio)	SbP1276837.01G011100	UDP-Glc p-hydroxymandelonitrile glycosyltransferase
DHR1 (Dhurrin-catabolism)	SbP1276837.08G071800	dhurrinase (β -glucosidase)
DHR2 (Dhurrin-catabolism)	SbP1276837.08G072000	dhurrinase (β -glucosidase)
DHR-like3 (Dhurrin-catabolism)	SbP1276837.08G072300	dhurrinase (β -glucosidase)
DHR-like3 (Dhurrin-catabolism)	SbP1276837.08G072500	dhurrinase (β -glucosidase)
HNL (Dhurrin-catabolism)	SbP1276837.04G305900	α -hydroxynitrile lyase
CAS C1 (Dhurrin-detox)	SbP1276837.06G016100	β -cyanolanine synthase
NIT4B2 (Dhurrin-recyc)	SbP1276837.04G201200	Nitrilase
NIT4A (Dhurrin-recyc)	SbP1276837.04G201300	Nitrilase
SbMATE2 (Dhurrin-transport)	SbP1276837.01G011300	Multidrug and toxic compound extrusion (MATE) family
SbCGTR1 (Dhurrin-transport)	SbP1276837.01G122100	nitrate/peptide family (NPF) transporter

Table 3.1, continued

Gene	ID	Gene family/acronym
CM	SbPI276837.03G273700	Chorismate mutase
CM	SbPI276837.07G119300	Chorismate mutase
PAL	SbPI276837.04G196800	Phenylalanine ammonia lyase
PAL	SbPI276837.06G133500	Phenylalanine ammonia lyase
PAL	SbPI276837.04G196900	Phenylalanine ammonia lyase
PAL	SbPI276837.06G133600	Phenylalanine ammonia lyase
PAL	SbPI276837.04G197100	Phenylalanine ammonia lyase
PAL	SbPI276837.04G197300	Phenylalanine ammonia lyase
AIM1	SbPI276837.04G112200	Abnormal inflorescence meristem1
ICS	SbPI276837.02G150600	Isochorismate synthase
EDSS	SbPI276837.04G018500	ENHANCED DISEASE SUSCEPTIBILITY 5
EPS1	SbPI276837.07G010900	ENHANCED PSEUDOMONAS SUSCEPTIBILITY 1
EPS1	SbPI276837.07G011200	ENHANCED PSEUDOMONAS SUSCEPTIBILITY 1
EPS1	SbPI276837.07G012800	ENHANCED PSEUDOMONAS SUSCEPTIBILITY 1
NPR1	SbPI276837.03G030400	NONEXPRESSER OF PR GENES 1
NPR3	SbPI276837.01G130600	NONEXPRESSER OF PR GENES 3
NPR4	SbPI276837.03G276500	NONEXPRESSER OF PR GENES 4
TGA6	SbPI276837.01G355800	TGA transcription factor
TGA6	SbPI276837.02G372300	TGA transcription factor
TGA6	SbPI276837.03G296600	TGA transcription factor
MPK	SbPI276837.01G289000	Mitogen-activated protein kinases
PR1	SbPI276837.01G304400	Pathogenesis-related protein 1
PR1	SbPI276837.02G022000	Pathogenesis-related protein 1
PR1	SbPI276837.10G018500	Pathogenesis-related protein 1
SbLOX1 (9-LOX)	SbPI276837.01G115400	9-Lipoxygenase
SbLOX3 (9-LOX)	SbPI276837.03G349400	9-Lipoxygenase
SbLOX4 (9-LOX)	SbPI276837.03G349600	9-Lipoxygenase
SbLOXm (9-LOX)	SbPI276837.01G115300	9-Lipoxygenase
SbLOXm (9-LOX)	SbPI276837.K049500	9-Lipoxygenase
SbLOXo (9-LOX)	SbPI276837.01G115200	9-Lipoxygenase
SbAOS (JA)	SbPI276837.01G071000	Allene oxide synthase
SbPLA2-a-1 (JA)	SbPI276837.01G110400	Phospholipase D
SbAOC3 (JA)	SbPI276837.01G301200	Allene oxide cyclase
SbOPR7 (JA)	SbPI276837.07G128800	Oxo-phytodienoate reductase
SbCOI1 (JA)	SbPI276837.03G324700	CORONATINE INSENSITIVE1

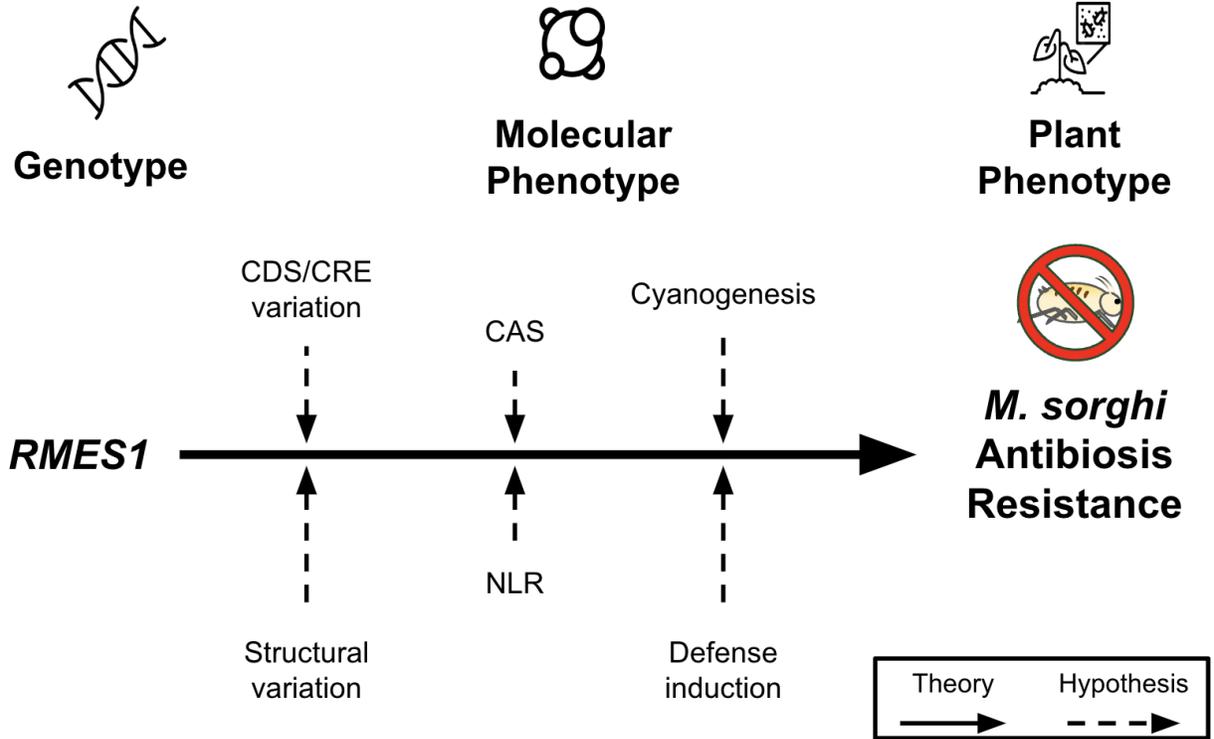


Figure 3.1 - Hypotheses on the genotype to phenotype (G2P) map for *RMES1*. Two primary hypotheses are shown and the molecular phenotype each would predict. CDS - coding sequence, CRE - *cis*-regulatory variation, CAS - β -cyanoalanine synthase, NLR - nucleotide-binding leucine-rich repeat.

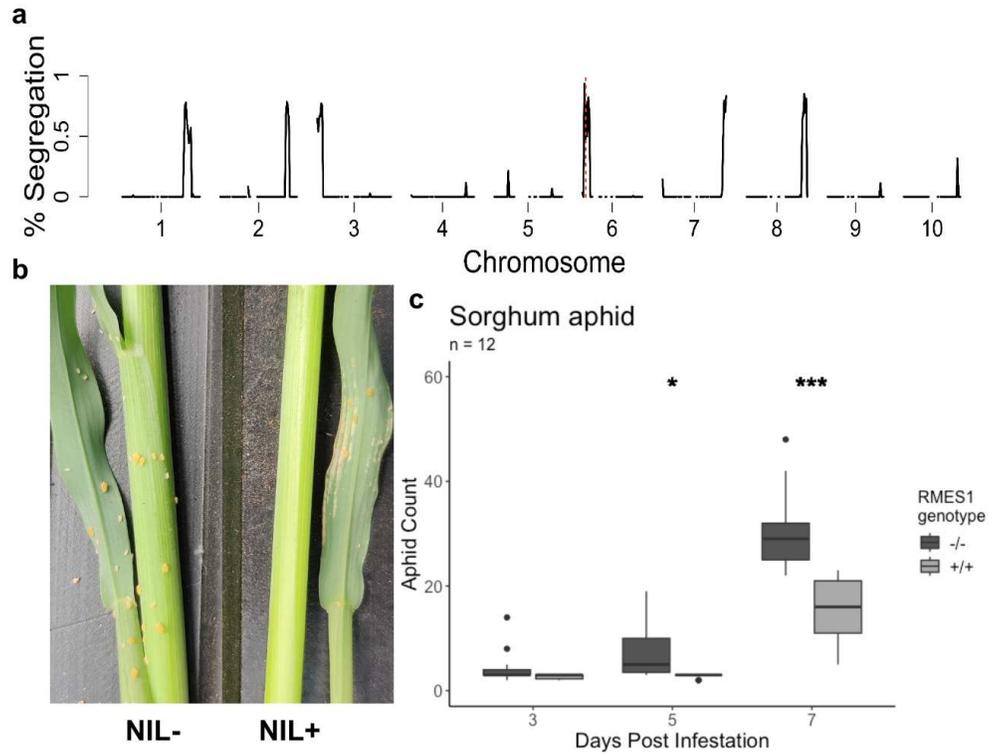


Figure 3.2 - *RMES1* BC3 NIL genome and phenotype. a) Genotype by RNA-sequencing show genomic regions segregating between NIL genotypes. Red dashed line indicates *RMES1*. b) Sorghum aphid infestation on BC₃ NILs. c) No-choice assay of NILs over 7 day infestation. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

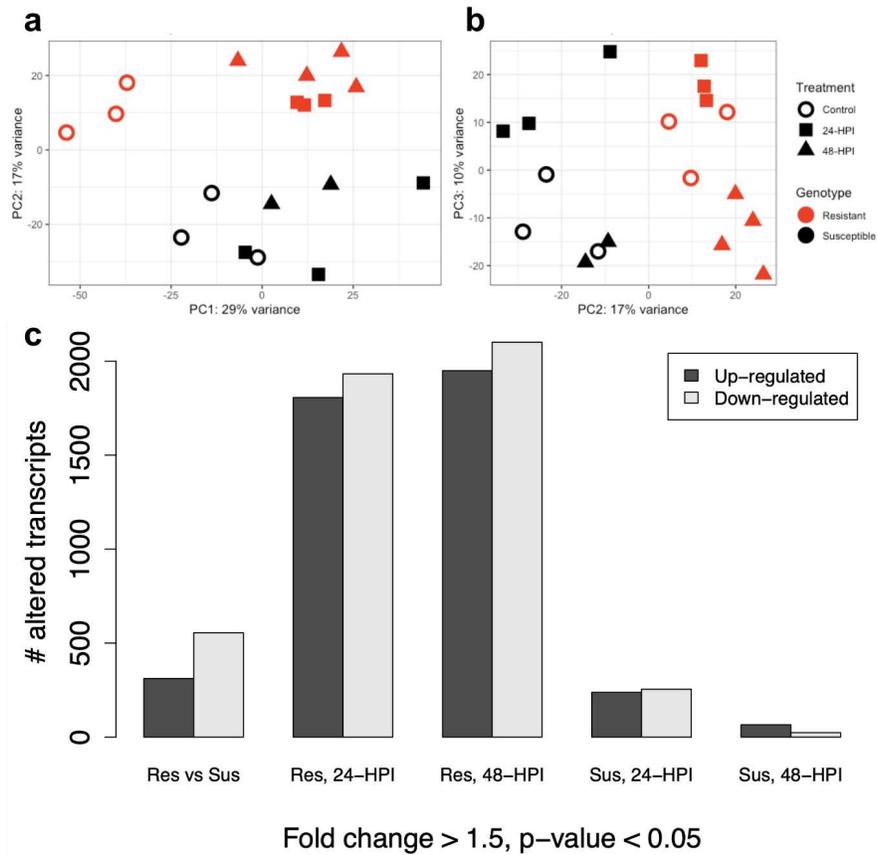


Figure 3.3 - Resistant NILs remodel their transcriptome in response to infestation. a) Principal component (PC) 1 and PC2 of global gene expression for resistant (NIL+) and susceptible (NIL-) genotypes under control (uninfested), 24-HPI, and 48-HPI treatments. b) PC2 and PC3. c) Number of significantly differentially expressed genes. Res=resistant, Sus=susceptible, hpi=Hours Post Infestation.

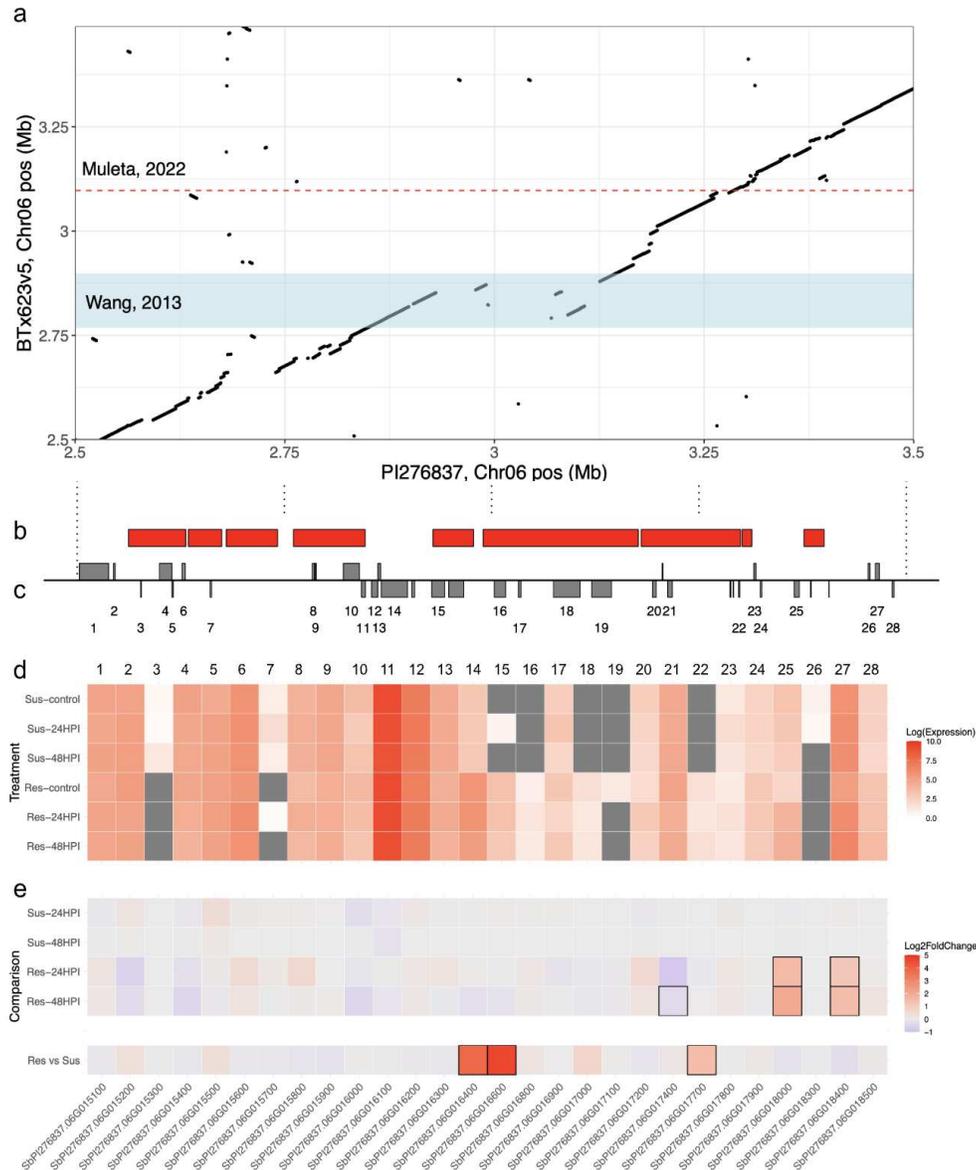


Figure 3.4 - Candidate genes on insertion present in resistant reference genomes PI276837. a) Chr06 RMES1 region alignment of PI276837 with BTx623. Previously reported QTL for *RMES1* in BTx623 converted to v5, red dashed line - S06_3096975 fixation-associated SNP, blue region - Chr06:2768472-2898239 linkage mapping (F. Wang et al. 2013; Muleta et al. 2022). b) Highly diverged regions (HDRs) between reference genomes > 10 kb. c) Annotated gene locations shown by grey boxes, forward coding strand above line, reverse coding strand below line. Numbers below corresponding genes indicate expressed transcripts. d) Heatmap of expression of genes near *RMES1* in all treatments. Normalized expression from DESeq2 was log transformed. Grey boxes indicate few to no mapped reads. Res=resistant, Sus=susceptible, hpi=Hours Post Infestation. e) Heatmap of log₂ fold changes of genes near *RMES1*. Significant (fold change > 1.5, $p < 0.05$) differentially expressed genes are indicated with black boxes around corresponding cells. Top four rows correspond to genotype-treatment compared with uninfested controls within genotype. Bottom row indicates uninfested controls compared between genotype.

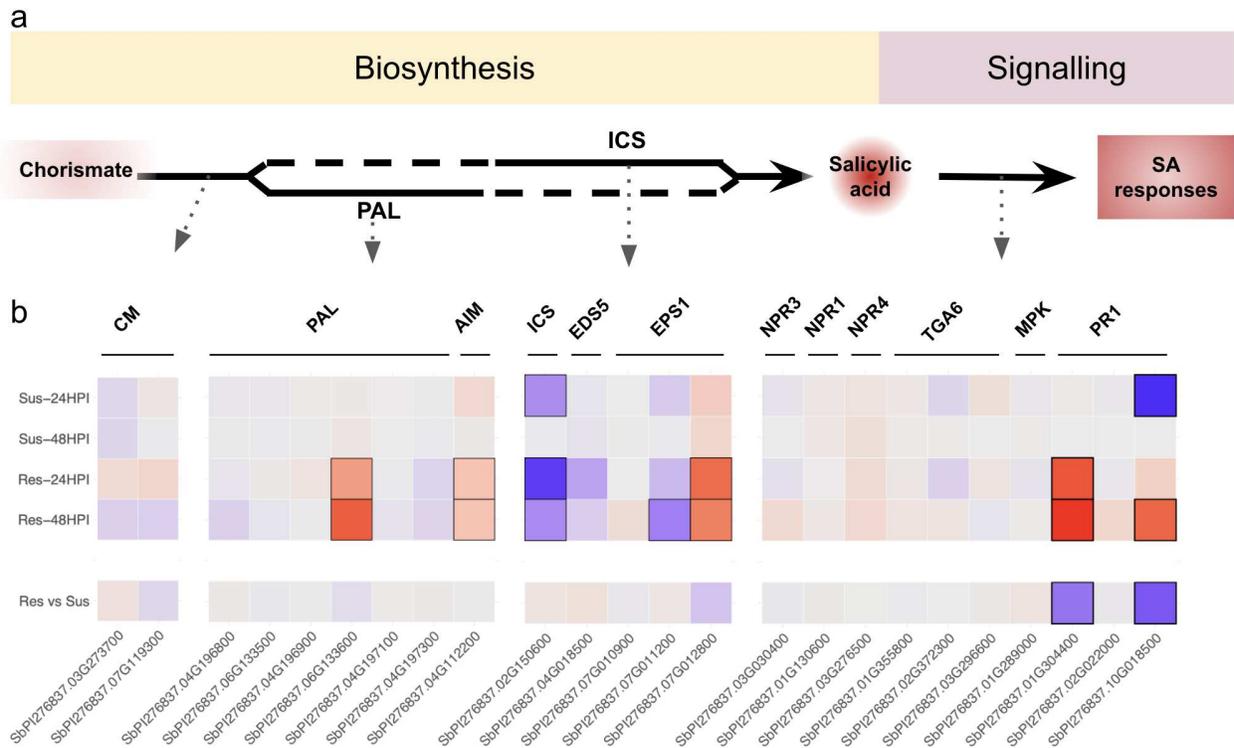


Figure 3.5 - Marker genes for the salicylic acid pathway are up-regulated by *RMES1* in response to infestation. a) Conceptual figure of salicylic acid biosynthesis and signaling pathway. b) Heatmap of \log_2 fold changes of homologs of pathway. Res=resistant, Sus=susceptible, hpi=Hours Post Infestation. Significant (fold change > 1.5 , $p < 0.05$) differentially expressed genes are indicated with black boxes around corresponding cells. Top four rows correspond to genotype-treatment compared with uninfested controls. Bottom row indicates uninfested controls compared with one another.

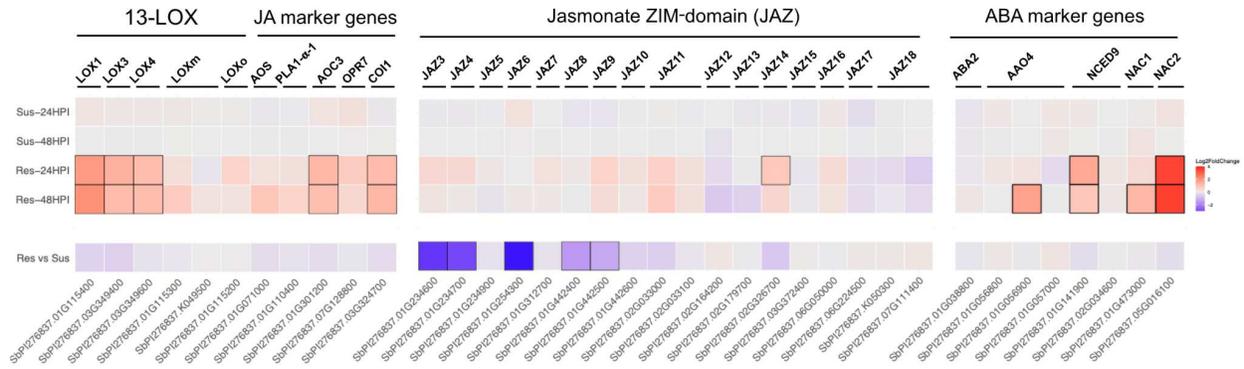


Figure 3.6 - Induced expression of JA, JAZ, and ABA pathways by RMES1 response to infestation. Heatmap of log₂ fold changes of genes in pathways of interest. Res=resistant, Sus=susceptible, hpi=Hours Post Infestation. Significant (fold change > 1.5, $p < 0.05$) differentially expressed genes are indicated with black boxes around corresponding cells. Top four rows correspond to genotype-treatment compared with uninfested controls. Bottom row indicates uninfested controls compared with one another.

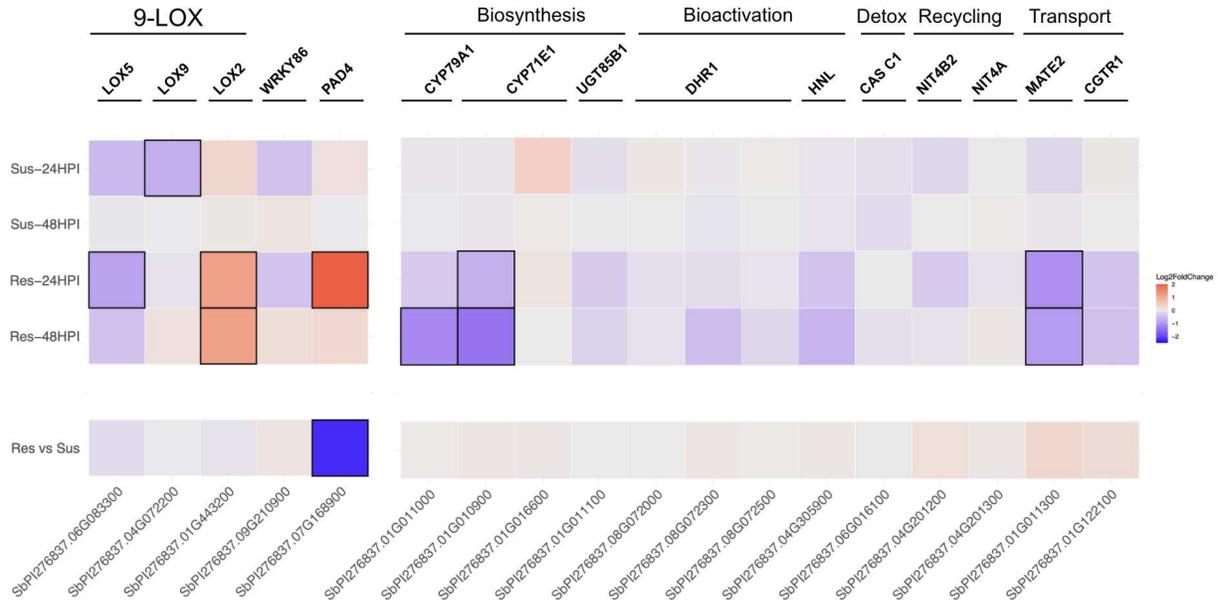


Figure 3.7 - Candidate defense mechanisms are differentially expressed by *RMES1* in response to infestation. Heatmap of log₂ fold changes of genes of interest. Res=resistant, Sus=susceptible, hpi=Hours Post Infestation. Significant (fold change > 1.5, $p < 0.05$) differentially expressed genes are indicated with black boxes around corresponding cells. Top four rows correspond to genotype-treatment compared with uninfested controls. Bottom row indicates uninfested controls compared with one another.

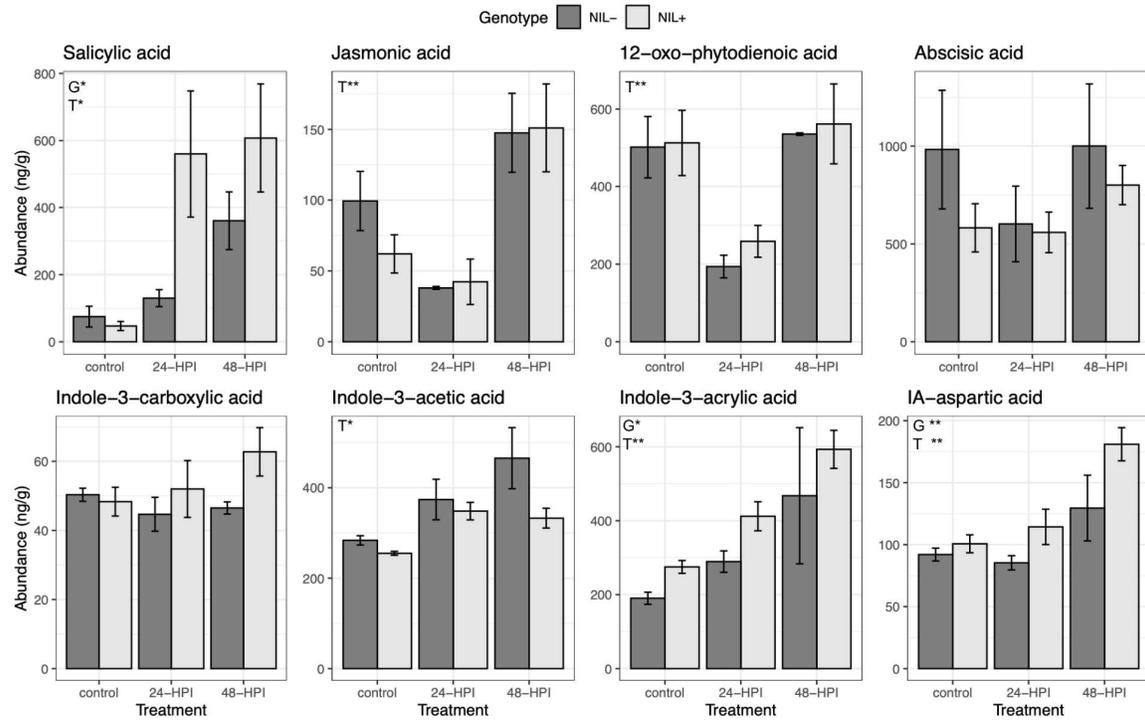


Figure 3.8 - Salicylic acid, IAA, and IA-aspartic acid phytohormones respond in an *RMESI*-dependent manner. Abundance of phytohormones in uninfested (control) and infested (24-hpi, 48-hpi) samples. G (genotype) and T (treatment) indicate significance in 2-way ANOVA (* = $p < 0.05$, ** = $p < 0.01$). Mean \pm standard error, n = 2-4. Res=resistant, Sus=susceptible, hpi=Hours Post Infestation.

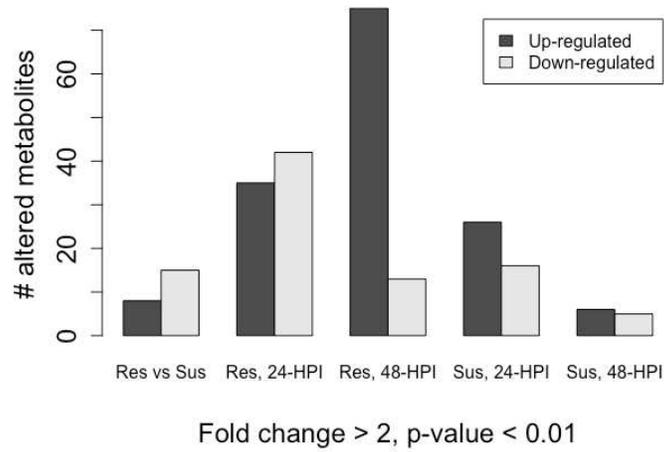


Figure 3.9 - Accumulation of metabolites in response to aphid infestation is *RMESI*-dependant. Number of significantly differentially expressed metabolites. Res=resistant, Sus=susceptible, hpi=Hours Post Infestation.

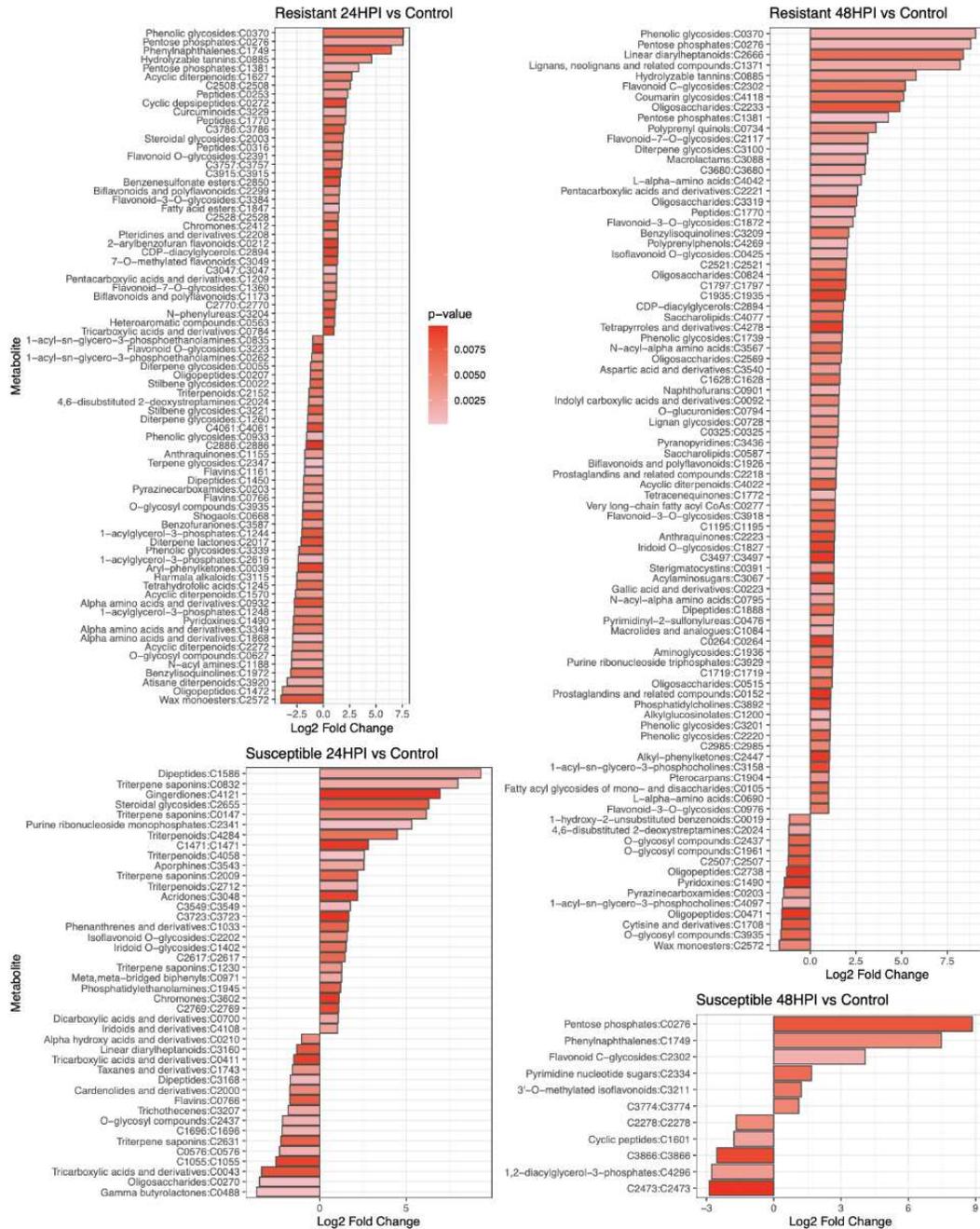


Figure 3.10 - *RMESI*-dependent differential expression of metabolites indicates diversified suite of defenses. Log2 fold change of significantly differentially expressed metabolites determined by Student's *t*-test between infested treatment groups and control groups ($L2FC > 2$, $p < 0.01$). Plotted bars are colored by *p*-value. Putative parental class of each metabolite is listed, however some analytes were unable to be annotated.

Chapter 3 References

- Almeida Trapp, Marília, Gezimar D. De Souza, Edson Rodrigues-Filho, William Boland, and Axel Mithöfer. 2014. “Validated Method for Phytohormone Quantification in Plants.” *Frontiers in Plant Science* 5. <https://www.frontiersin.org/articles/10.3389/fpls.2014.00417>.
- Batyrshina, Zhaniya S., Beery Yaakov, Reut Shavit, Anuradha Singh, and Vered Tzin. 2020. “Comparative Transcriptomic and Metabolic Analysis of Wild and Domesticated Wheat Genotypes Reveals Differences in Chemical and Physical Defense Responses against Aphids.” *BMC Plant Biology* 20 (1): 19. <https://doi.org/10.1186/s12870-019-2214-z>.
- Cardona, Juan Betancurt, Sajjan Grover, Lucas Busta, Scott E. Sattler, and Joe Louis. 2022. “Sorghum Cuticular Waxes Influence Host Plant Selection by Aphids.” *Planta* 257 (1): 22. <https://doi.org/10.1007/s00425-022-04046-3>.
- Chen, Ming-Shun. 2008. “Inducible Direct Plant Defense against Insect Herbivores: A Review.” *Insect Science* 15 (2): 101–14. <https://doi.org/10.1111/j.1744-7917.2008.00190.x>.
- Clay, Nicole K., Adewale M. Adio, Carine Denoux, Georg Jander, and Frederick M. Ausubel. 2009. “Glucosinolate Metabolites Required for an Arabidopsis Innate Immune Response.” *Science* 323 (5910): 95–101. <https://doi.org/10.1126/science.1164627>.
- Deschamps, Stéphane, Yun Zhang, Victor Llaca, Liang Ye, Abhijit Sanyal, Matthew King, Gregory May, and Haining Lin. 2018. “A Chromosome-Scale Assembly of the Sorghum Genome Using Nanopore Sequencing and Optical Mapping.” *Nature Communications* 9 (1): 4844. <https://doi.org/10.1038/s41467-018-07271-1>.
- David De Vleeschauwer et al., “Target of Rapamycin Signaling Orchestrates Growth–Defense Trade-Offs in Plants,” *New Phytologist* 217, no. 1 (2018): 305–19, <https://doi.org/10.1111/nph.14785>.
- Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics* 29 (1): 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.

- Dogimont, Catherine, Veronique Chovelon, Jerome Pauquet, Adnane Boualem, and Abdelhafid Bendahmane. 2014. “The Vat Locus Encodes for a CC-NBS-LRR Protein That Confers Resistance to *Aphis Gossypii* Infestation and *A. Gossypii*-Mediated Virus Resistance.” *The Plant Journal* 80 (6): 993–1004. <https://doi.org/10.1111/tpj.12690>.
- Dolatabadian, Aria, Philipp E. Bayer, Soodeh Tirnaz, Bhavna Hurgobin, David Edwards, and Jacqueline Batley. 2020. “Characterization of Disease Resistance Genes in the Brassica Napus Pangenome Reveals Significant Structural Variation.” *Plant Biotechnology Journal* 18 (4): 969–82. <https://doi.org/10.1111/pbi.13262>.
- Duan, Liu, Hongbo Liu, Xianghua Li, Jinghua Xiao, and Shiping Wang. 2014. “Multiple Phytohormones and Phytoalexins Are Involved in Disease Resistance to *Magnaporthe Oryzae* Invaded from Roots in Rice.” *Physiologia Plantarum* 152 (3): 486–500. <https://doi.org/10.1111/ppl.12192>.
- Emms, David M., and Steven Kelly. 2019. “OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics.” *Genome Biology* 20 (1): 238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Gladman, Nicholas, Andrew Olson, Sharon Wei, Kapeel Chougule, Zhenyuan Lu, Marcela Tello-Ruiz, Ivar Meijs, et al. 2022. “SorghumBase: A Web-Based Portal for Sorghum Genetic Information and Community Advancement.” *Planta* 255 (2): 35. <https://doi.org/10.1007/s00425-022-03821-6>.
- Gleadow, Roslyn M., Brian A. McKinley, Cecilia K. Blomstedt, Austin C. Lamb, Birger Lindberg Møller, and John E. Mullet. 2021. “Regulation of Dhurrin Pathway Gene Expression during Sorghum Bicolor Development.” *Planta* 254 (6): 119. <https://doi.org/10.1007/s00425-021-03774-2>.
- Goel, Manish, Hequan Sun, Wen-Biao Jiao, and Korbinian Schneeberger. 2019. “SyRI: Finding Genomic Rearrangements and Local Sequence Differences from Whole-Genome Assemblies.” *Genome Biology* 20 (1): 277. <https://doi.org/10.1186/s13059-019-1911-0>.
- Grayer, Renée J., Frances M. Kimmins, David E. Padgham, Jeffrey B. Harborne, and D. V. Ranga Rao. 1992. “Condensed Tannin Levels and Resistance of Groundnuts (*Arachis Hypogaea*) against

- Aphis Craccivora.” *Phytochemistry, The International Journal of Plant Biochemistry*, 31 (11): 3795–3800. [https://doi.org/10.1016/S0031-9422\(00\)97530-7](https://doi.org/10.1016/S0031-9422(00)97530-7).
- Gruss, Shelby M., Manoj Ghaste, Joshua R. Widhalm, and Mitchell R. Tuinstra. 2022. “Seedling Growth and Fall Armyworm Feeding Preference Influenced by Dhurrin Production in Sorghum.” *Theoretical and Applied Genetics*, January. <https://doi.org/10.1007/s00122-021-04017-4>.
- Huang, Jian, Kumar Shrestha, and Yinghua Huang. 2022. “Revealing Differential Expression of Phytohormones in Sorghum in Response to Aphid Attack Using the Metabolomics Approach.” *International Journal of Molecular Sciences* 23 (22): 13782. <https://doi.org/10.3390/ijms232213782>.
- Kiani, Mahnaz, and Adrianna Szczepaniec. 2018. “Effects of Sugarcane Aphid Herbivory on Transcriptional Responses of Resistant and Susceptible Sorghum.” *BMC Genomics* 19 (1): 774. <https://doi.org/10.1186/s12864-018-5095-x>.
- Li, Ning, Qiang He, Juan Wang, Baike Wang, Jiantao Zhao, Shaoyong Huang, Tao Yang, et al. 2023. “Super-Pangenome Analyses Highlight Genomic Diversity and Structural Variation across Wild and Cultivated Tomato Species.” *Nature Genetics* 55 (5): 852–60. <https://doi.org/10.1038/s41588-023-01340-y>.
- Li, Qi, Qi-Guang Xie, Jennifer Smith-Becker, Duroy A. Navarre, and Isgouhi Kaloshian. 2006. “Mi-1-Mediated Aphid Resistance Involves Salicylic Acid and Mitogen-Activated Protein Kinase Signaling Cascades.” *Molecular Plant-Microbe Interactions: MPMI* 19 (6): 655–64. <https://doi.org/10.1094/MPMI-19-0655>.
- Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* (Oxford, England) 30 (7): 923–30. <https://doi.org/10.1093/bioinformatics/btt656>.
- Limaje, Ankur, Chad Hayes, J. Scott Armstrong, Wyatt Hoback, Ali Zarrabi, Sulochana Paudyal, and John Burke. 2018. “Antibiosis and Tolerance Discovered in USDA-ARS Sorghums Resistant to

- the Sugarcane Aphid (Hemiptera: Aphididae)1.” *Journal of Entomological Science* 53 (2): 230–41. <https://doi.org/10.18474/JES17-70.1>.
- Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- MacLean, Brendan, Daniela M. Tomazela, Nicholas Shulman, Matthew Chambers, Gregory L. Finney, Barbara Frewen, Randall Kern, David L. Tabb, Daniel C. Liebler, and Michael J. MacCoss. 2010. “Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments.” *Bioinformatics* (Oxford, England) 26 (7): 966–68. <https://doi.org/10.1093/bioinformatics/btq054>.
- Marçais, Guillaume, Arthur L. Delcher, Adam M. Phillippy, Rachel Coston, Steven L. Salzberg, and Aleksey Zimin. 2018. “MUMmer4: A Fast and Versatile Genome Alignment System.” *PLOS Computational Biology* 14 (1): e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>.
- Leonor Margalha, Ana Confraria, and Elena Baena-González, “SnRK1 and TOR: Modulating Growth–Defense Trade-Offs in Plant Stress Responses,” *Journal of Experimental Botany* 70, no. 8 (April 15, 2019): 2261–74, <https://doi.org/10.1093/jxb/erz066>.
- Martinez de Ilarduya, Oscar, QiGuang Xie, and Isgouhi Kaloshian. 2003. “Aphid-Induced Defense Responses in Mi-1-Mediated Compatible and Incompatible Tomato Interactions.” *Molecular Plant-Microbe Interactions: MPMI* 16 (8): 699–708. <https://doi.org/10.1094/MPMI.2003.16.8.699>.
- McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303. <https://doi.org/10.1101/gr.107524.110>.
- Meihls, Lisa N., Vinzenz Handrick, Gaetan Glauser, Hugues Barbier, Harleen Kaur, Meena M. Haribal, Alexander E. Lipka, et al. 2013. “Natural Variation in Maize Aphid Resistance Is Associated with

- 2,4-Dihydroxy-7-Methoxy-1,4-Benzoxazin-3-One Glucoside Methyltransferase Activity[C][W].” *The Plant Cell* 25 (6): 2341–55. <https://doi.org/10.1105/tpc.113.112409>.
- Morel and Dangl, “The Hypersensitive Response and the Induction of Cell Death in Plants,” *Cell Death & Differentiation* 4, no. 8 (December 1997): 671–83, <https://doi.org/10.1038/sj.cdd.4400309>.
- Mou et al., “Monocot Crop–Aphid Interactions: Plant Resilience and Aphid Adaptation,” *Current Opinion in Insect Science* 57 (June 1, 2023): 101038, <https://doi.org/10.1016/j.cois.2023.101038>.
- Muleta, Kebede T., Terry Felderhoff, Noah Winans, Rachel Walstead, Jean Rigaud Charles, J. Scott Armstrong, Sujana Mamidi et al. "The recent evolutionary rescue of a staple crop depended on over half a century of global germplasm exchange." *Science advances* 8, no. 6 (2022): eabj4633.
- Nalam, Vamsi, Travis Isaacs, Sarah Moh, Jessica Kansman, Deborah Finke, Tessa Albrecht, and Punya Nachappa. 2021. “Diurnal Feeding as a Potential Mechanism of Osmoregulation in Aphids.” *Insect Science* 28 (2): 521–32. <https://doi.org/10.1111/1744-7917.12787>.
- Qin, Yuan, Anna Maria Torp, Gaëtan Glauser, Carsten Pedersen, Søren K. Rasmussen, and Hans Thordal-Christensen. 2019. “Barley Isochorismate Synthase Mutant Is Phylloquinone-Deficient, but Has Normal Basal Salicylic Acid Level.” *Plant Signaling & Behavior* 14 (11): 1671122. <https://doi.org/10.1080/15592324.2019.1671122>.
- R Core Team. 2021. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rice, Brian, Spiekerman John, John T. Lovell, and Geoffrey Morris. 2024. “Sorghum PanGenome Reveals Cool Science.” TBD.
- Robert, Christelle, and Mick Watson. 2015. “Errors in RNA-Seq Quantification Affect Genes of Relevance to Human Disease.” *Genome Biology* 16 (1): 177. <https://doi.org/10.1186/s13059-015-0734-x>.
- Shrestha, Kumar, and Yinghua Huang. 2022. “Genome-Wide Characterization of the Sorghum JAZ Gene Family and Their Responses to Phytohormone Treatments and Aphid Infestation.” *Scientific Reports* 12 (1): 3238. <https://doi.org/10.1038/s41598-022-07181-9>.

- Shrestha, Kumar, Shankar Pant, and Yinghua Huang. 2021. “Genome-Wide Identification and Classification of Lipxygenase Gene Family and Their Roles in Sorghum-Aphid Interaction.” *Plant Molecular Biology* 105 (4): 527–41. <https://doi.org/10.1007/s11103-020-01107-7>.
- Snoeck, Simon, Natalia Guayazán-Palacios, and Adam D Steinbrenner. 2022. “Molecular Tug-of-War: Plant Immune Recognition of Herbivory.” *The Plant Cell*, January, koac009. <https://doi.org/10.1093/plcell/koac009>.
- Jia-Ming Song et al., “Eight High-Quality Genomes Reveal Pan-Genome Architecture and Ecotype Differentiation of Brassica Napus,” *Nature Plants* 6, no. 1 (January 2020): 34–45, <https://doi.org/10.1038/s41477-019-0577-7>.
- Springer, Nathan M., Kai Ying, Yan Fu, Tieming Ji, Cheng-Ting Yeh, Yi Jia, Wei Wu, et al. 2009. “Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content.” *PLOS Genetics* 5 (11): e1000734. <https://doi.org/10.1371/journal.pgen.1000734>.
- Studham, Matthew E., and Gustavo C. MacIntosh. 2013. “Multiple Phytohormone Signals Control the Transcriptional Response to Soybean Aphid Infestation in Susceptible and Resistant Soybean Plants.” *Molecular Plant-Microbe Interactions: MPMI* 26 (1): 116–29. <https://doi.org/10.1094/MPMI-05-12-0124-FI>.
- Torrens-Spence, Michael P., Anastassia Bobokalonova, Valentina Carballo, Christopher M. Glinkerman, Tomáš Pluskal, Amber Shen, and Jing-Ke Weng. 2019. “PBS3 and EPS1 Complete Salicylic Acid Biosynthesis from Isochorismate in Arabidopsis.” *Molecular Plant* 12 (12): 1577–86. <https://doi.org/10.1016/j.molp.2019.11.005>.
- Yongfu Tao et al., “Extensive Variation within the Pan-Genome of Cultivated and Wild Sorghum,” *Nature Plants* 7, no. 6 (June 2021): 766–73, <https://doi.org/10.1038/s41477-021-00925-x>.
- Tzin, Vered, Noe Fernandez-Pozo, Annett Richter, Eric A. Schmelz, Matthias Schoettner, Martin Schäfer, Kevin R. Ahern, et al. 2015. “Dynamic Maize Responses to Aphid Feeding Are Revealed by a

- Time Series of Transcriptomic and Metabolomic Assays.” *Plant Physiology* 169 (3): 1727–43.
<https://doi.org/10.1104/pp.15.01039>.
- Van de Weyer, Anna-Lena, Freddy Monteiro, Oliver J. Furzer, Marc T. Nishimura, Volkan Cevik, Kamil Witek, Jonathan D. G. Jones, Jeffery L. Dangl, Detlef Weigel, and Felix Bemm. 2019. “A Species-Wide Inventory of NLR Genes and Alleles in *Arabidopsis Thaliana*.” *Cell* 178 (5): 1260-1272.e14. <https://doi.org/10.1016/j.cell.2019.07.038>.
- Wang, Bo, Yinping Jiao, Kapeel Chougule, Andrew Olson, Jian Huang, Victor Llaca, Kevin Fengler, et al. 2021. “Pan-Genome Analysis in Sorghum Highlights the Extent of Genomic Variation and Sugarcane Aphid Resistance Genes.” *bioRxiv*. <https://doi.org/10.1101/2021.01.03.424980>.
- Jizong Wang, Wen Song, and Jijie Chai, “Structure, Biochemical Function, and Signaling Mechanism of Plant NLRs,” *Molecular Plant* 16, no. 1 (January 2, 2023): 75–95,
<https://doi.org/10.1016/j.molp.2022.11.011>.
- Wang, Faming, Songmin Zhao, Yonghua Han, Yutao Shao, Zhenying Dong, Yang Gao, Kunpu Zhang, et al. 2013. “Efficient and Fine Mapping of RMES1 Conferring Resistance to Sorghum Aphid *Melanaphis Sacchari*.” *Molecular Breeding* 31 (4): 777–84. <https://doi.org/10.1007/s11032-012-9832-6>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
<https://ggplot2.tidyverse.org>.
- Yates, Ashley D, and Andy Michel. 2018. “Mechanisms of Aphid Adaptation to Host Plant Resistance.” *Current Opinion in Insect Science* 26 (April): 41–49. <https://doi.org/10.1016/j.cois.2018.01.003>.
- Zebelo, Simon A., and Massimo E. Maffei. 2015. “Role of Early Signalling Events in Plant–Insect Interactions.” *Journal of Experimental Botany* 66 (2): 435–48. <https://doi.org/10.1093/jxb/eru480>.
- Hengyou Zhang, Jian Huang, and Yinghua Huang, “Identification and Characterization of Plant Resistance Genes (R Genes) in Sorghum and Their Involvement in Plant Defense against Aphids,” *Plant Growth Regulation* 96, no. 3 (April 1, 2022): 443–61,
<https://doi.org/10.1007/s10725-022-00797-x>.

Züst, Tobias, and Anurag A. Agrawal. 2016. "Mechanisms and Evolution of Plant Resistance to Aphids."
Nature Plants 2 (1): 1–9. <https://doi.org/10.1038/nplants.2015.206>.

Chapter 4 - Transcriptional signatures of wheat inflorescence development¹

4.1 Summary

In order to maintain global food security, it will be necessary to increase yields of the cereal crops that provide most of the calories and protein for the world's population, which includes common wheat (*Triticum aestivum* L.). An important factor contributing to wheat yield is the number of grain-holding spikelets which form on the spike during inflorescence development. Characterizing the gene regulatory networks controlling the timing and rate of inflorescence development will facilitate the selection of natural and induced gene variants that contribute to increased spikelet number and yield.

In the current study, co-expression and gene regulatory networks were assembled from a temporal wheat spike transcriptome dataset, revealing the dynamic expression profiles associated with the progression from vegetative meristem to terminal spikelet formation. Consensus co-expression networks revealed enrichment of several transcription factor families at specific developmental stages including the sequential activation of different classes of MIKC-MADS box genes. This gene regulatory network highlighted interactions among a small number of regulatory hub genes active during terminal spikelet formation. Finally, the CLAVATA and WUSCHEL gene families were investigated, revealing potential roles for TaCLE13, TaWOX2, and TaWOX7 in wheat meristem development. The hypotheses generated from these datasets and networks further our understanding of wheat inflorescence development.

¹ This chapter was reproduced verbatim from “VanGessel, et al. Transcriptional signatures of wheat inflorescence development. *Scientific Reports* (2023)”. The text benefitted from writing and editing contributions from contributing authors and reviewers selected by the publisher. The ordering of the materials in this dissertation are consistent with the content available online but have been renumbered to reflect incorporation into this dissertation.

4.2 Introduction

The world population is expected to exceed nine billion people by 2050, signaling that further increases in grain production will be required to ensure food security (Ray et al. 2013). Because there remain few opportunities to expand arable land area, increasing the yield of major cereal crops through genetic improvement will be critical to meet this goal. In common wheat (*Triticum aestivum* L.) characterizing the genetic pathways regulating grain size and grain number will facilitate the rational combination of superior alleles in wheat breeding programs to help drive continued yield improvements (Brinton and Uauy 2019).

Grain number in wheat is determined to a large extent by inflorescence architecture. By integrating photoperiod and temperature cues, the vegetative shoot apical meristem (SAM) transitions to the reproductive inflorescence meristem (IM), during which the developing spike passes through the characteristic double ridge (DR) stage, forming a lower leaf ridge and an upper spikelet ridge (Waddington, Cartwright, and Wall 1983). The lower leaf ridge is repressed by the MIKC-MADS box transcription factors (TFs) *VRN1*, *FUL2* and *FUL3* (Li et al. 2019), whereas the upper ridges develop glumes, lemmas, and floret primordia. As the IM elongates, spikelet meristems are added at the growing apex, while basal spikelets continue to develop. Wheat spikes are determinate structures and the addition of lateral spikelets ends when the terminal spikelet is formed. Therefore, spikelet number is determined by the timing and rate of meristem development preceding terminal spikelet formation. Each spikelet has the potential to form between three and six grains (Bonnett 1966) and spikelet number is correlated with grain number and yield (Rawson 1970; Cao et al. 2020; Würschum et al. 2018).

Shoot meristems are organized around the organizing center and stem cell maintenance is governed by the conserved CLAVATA-WUSCHEL negative feedback loop (Somssich et al. 2016). In Arabidopsis, the homeodomain TF *WUS* induces *CLV3*, which encodes a secreted peptide that forms receptor complexes repressing *WUS* (Fletcher 2018). Manipulation of this pathway confers variation in locule number in tomato (*Solanum lycopersicum*) and kernel row number in maize (*Zea mays*)

(Rodriguez-Leal et al. 2017; Chen et al. 2021). The wheat genome contains 104 *CLAVATA3/EMBRYO SURROUNDING REGION (CLE)* peptides (Li et al. 2019) and 44 WUSCHEL RELATED HOMEODOMAIN (WOX) TFs (Li et al. 2020), but the specific ones regulating inflorescence meristem development in wheat are yet to be identified.

Inflorescence development is controlled by a complex regulatory network involving multiple classes of transcription factors (TFs) which orchestrate rapid and dynamic changes in gene expression. The Type II MIKC MADS-box TFs play critical roles in flower development across the angiosperms and can be divided into A, B, C, D and E-classes that interact mainly as tetrameric complexes in a spatially regulated manner to direct sepal (A- and E-), petal (A-, B-, E-), stamen (B-, C-, E-), and carpel development (C- and E-class genes) (Honma and Goto 2001; Theißen 2001). This family expanded during cereal evolution and the hexaploid wheat genome contains 201 MIKC MADS-box genes, classified into 15 phylogenetic subclades (Schilling et al. 2020).

The SHORT VEGETATIVE PHASE (SVP) subclade members *SVPI*, *VRT2*, and *SVP3* promote the transition from the vegetative SAM to the IM, along with the AP1/SQUA subclade genes *VRN1*, *FUL2* and *FUL3* (Li et al. 2019; Li et al. 2020). Subsequently, AP1/SQUA genes suppress the expression of SVP genes, which may be required to promote interactions between AP1/SQUA proteins and the E-class MIKC-MADS proteins SEPELLATA1 (SEP1) and SEP3, which are predominantly expressed in floral organogenesis during early reproductive growth (Li et al. 2021). The natural *VRT2^{pol}* allele from *Triticum polonicum* exhibits ectopic expression and is associated with elongated glumes and increased grain length (Adamski et al. 2021). *VRT2*-overexpression lines show reduced transcript levels of B-class (*PI* and *AP3*) and C-class (*AG1* and *AG2*) MIKC-MADS box genes, although the role of these latter subclades in wheat inflorescence development remains to be characterized (Li et al. 2021).

Although much has been learned about wheat inflorescence development from positional cloning, reverse genetics, and comparative genetic approaches, we lack a full understanding of the regulatory networks controlling meristem determinacy and developmental transitions. Only a fraction of the

hundreds of QTL for thousand kernel weight, kernel number per spike, and spikelet number have been cloned and validated to date, indicating that a large proportion of quantitative variation in these traits remains uncharacterized (Cao et al. 2020).

Transcriptomics provides a complementary approach to characterize the regulatory networks underlying inflorescence development that is empowered by an expanding set of wheat genomic resources (IWGSC 2018; Walkowiak et al. 2020). Co-expression and gene regulatory networks (GRNs) are powerful tools to interpret temporal correlation and causal relationships between genes, and to help identify critical hub genes that coordinate development (Rao and Dixon 2019; van den Broeck et al. 2020). Previous transcriptomic studies in wheat inflorescence tissues described the differential expression profiles of thousands of genes during vegetative and floral meristem development, including the stage-specific expression of different TFs and hormone biosynthesis and signaling genes (Feng et al. 2017; Li et al. 2018). A population-associative transcriptomic approach was used to identify regulators of wheat spike architecture, including *CEN2*, *TaPAP2/SEP1-6*, and *TaVRS1/HOX1*, which were validated in functional studies (Wang et al. 2017).

In the current study, a series of co-expression and gene regulatory networks were assembled to characterize the predominant transcriptional profiles associated with the progression of wheat inflorescence development, revealing two consecutive regulatory shifts at the DR and TS stages. Core regulatory candidate genes were identified including both known TFs and novel candidates with potential roles in regulating spike architecture.

4.3 Material and Methods

4.3.1 Plant materials and growth conditions

All experiments were performed in the tetraploid *Triticum turgidum* L. subsp. *durum* (Desf.) var. Kronos (genomes AABB). Kronos has a spring growth habit conferred by a *VRN-A1* allele containing a deletion in intron 1 and carries the *Ppd-A1a* allele that confers reduced sensitivity to photoperiod

(Wilhelm et al. 2008; Fu et al. 2005). Plants were grown in controlled conditions in PGR15 growth chambers (Conviroon, Manitoba, Canada) under a long day photoperiod (16 h light/8 h dark) at 23 °C day/17 °C night temperatures and a light intensity of $\sim 260 \mu\text{M m}^{-2} \text{s}^{-1}$. Developing apical meristems were harvested under a dissecting microscope using a sterile scalpel and placed immediately in liquid nitrogen. All samples were harvested within a one-hour period approximately 4 h after the lights were switched on (± 30 min) to account for possible differences in circadian regulation of gene expression. Approximately 20 apices were combined for each biological replicate of samples harvested at stages W1.0 (shoot apical meristem, SAM) and W2.0 (early double ridge, EDR) and approximately 12 apices for samples harvested at stages W3.0 (double ridge, DR), W3.25 (lemma primordia, LP) and W3.5 (terminal spikelet, TS) (Waddington, Cartwright, and Wall 1983). Four biological replicates were harvested at each timepoint.

4.3.2 RNA-seq library construction and sequencing

Tissues were ground into a fine powder in liquid nitrogen and total RNA was extracted using the Spectrum™ Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO). Sequencing libraries were produced using the TruSeq RNA Sample Preparation kit v2 (Illumina, San Diego, CA), according to the manufacturer's instructions. Library quality was determined using a high-sensitivity DNA chip run on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Libraries were barcoded to allow multiplexing and all samples were sequenced using the 100 bp single read module across two lanes of a HiSeq3000 sequencer at the UC Davis Genome Center.

4.3.3 RNA-seq data processing

'Kronos' RNA-seq reads were trimmed and checked for quality Phred scores above 30 using Fastp v0.20.1 (Chen et al. 2018). Trimmed reads were aligned to the IWGSC RefSeq v1.0 genome assembly consisting of A and B chromosome pseudomolecules and unanchored (U) scaffolds not assigned to any chromosome (ABU) using STAR 2.7.5 aligner (outFilterMismatchNoverReadLmax = 0.04, alignIntronMax = 10,000) (Dobin et al. 2013; IWGSC 2018). Only uniquely mapped reads were retained for expression analysis. Transcript levels were quantified by featureCounts using 190,391 gene

models from the ABU IWGSC RefSeq v1.1 annotations (Ramirez-Gonzalez et al. 2018; Liao, Smyth, and Shi 2014) and converted to Transcripts Per Million (TPM) values using a custom python script available from https://github.com/cvanges/spike_development/ (Supplementary data 1).

Raw RNA-seq reads for ‘Kenong9204’ and ‘Chinese Spring’ inflorescence development datasets were obtained from BioProjects PRJNA325489 and PRJNA383677 (Feng et al. 2017; Lie et al. 2018). RNA-seq reads were processed with Fastp as described above and aligned to the hexaploid ABDU RefSeq v1.0 genome assembly using the same methods and parameters. Transcript quantification and TPM were determined as above using the full ABDU IWGSC RefSeq v1.1 annotations. RNA-seq reads and raw count data for each sample is available from NCBI Gene Expression Omnibus under the accession GSE193126 (<https://www.ncbi.nlm.nih.gov/geo/>).

4.3.3 *Transcription factors*

There were 3,838 ABU gene models annotated as transcription factors that were grouped into 65 TF families per IWGSC v1.1 annotations (Ramirez-Gonzalez et al. 2018). The following families were consolidated: “AP2” and “APETALA2”, “bHLH” and “HRT-like”, “MADS” and “MADS1”, “NFYB” and “NF-YB”, “NFYC” and “NF-YC”, and “SBP” and “SPL”, as well as “MADS2” and “MIKC”, which were consolidated into “MIKC-MADS”. After consolidation, there were 59 TF families. A previous study described the annotation of 201 MIKC-MADS box genes placed into 15 subclades (Schilling et al. 2020). There were 30 MIKC transcription factors on the A and B genomes absent from the IWGSC TF list, which were added to this family. Investigations of the *CLE* and *WOX* gene families were based on the naming reported in Li et al., 2019b and Li et al. 2020b, with the addition of *TaWUSb* (*TraesCS2B02G775400LC*) to the *WOX* family, which was absent from these studies. In total, 3,861 TFs were included in this study (Supplementary data 2).

4.3.4 *Spike-dominant expression analysis*

Expression data (TPMs) for two developmental studies were obtained from the Grassroots Data Repository (https://opendata.earlham.ac.uk/wheat/under_license/toronto/Ramirez-Gonzalez_etal_2018-

06025-Transcriptome-Landscape/expvip/RefSeq_1.0/ByTranscript/) (Choulet et al. 2014; Ramirez-Gonzalez et al. 2018). The first dataset, in ‘Chinese Spring’, included samples from five tissue types at three timepoints (mean of two biological replicates) for 15 total tissue/stages (Choulet et al. 2014). A second dataset from the variety ‘Azhurnaya’ comprised 209 unreplicated samples grouped into 22 “intermediate tissue” groups of various sizes (Ramirez-Gonzalez et al. 2018). Twelve samples overlapping with ‘Kronos’ spike samples were removed (tissue groups “coleoptile”, “stem axis”, and “shoot apical meristem”). For early spike tissue specificity analyses, the mean TPM expression of 15 ‘Chinese Spring’ tissues (n = 2) or the mean of 22 ‘Azhurnaya’ tissues (n ranging from 3 - 30) were compared to the ‘Kronos’ sampling stage with the highest mean expression (n = 4). Comparisons were made using the Tau (τ) tissue specificity metric where $\tau = 0$ indicates ubiquitous expression and $\tau = 1$ indicates tissue specific expression (Yanai et al. 2005; Kryuchkova-Mostacci and Robinson-Rechavi 2017). A custom R script was used to calculate tissue specificity and is available at github.com/cvanges/spike_development. Genes which were expressed predominantly in ‘Kronos’ inflorescence tissues ($\tau > 0.9$) were defined ‘spike-dominant’ whereas genes only expressed in ‘Kronos’ inflorescence tissues ($\tau = 1$) were defined ‘spike-specific’ (Supplementary data 3).

4.3.5 Principal Component Analysis (PCA), Differential Expression, and GO enrichment

PCA was performed in R using `prcomp` in the `r/stats` package v2.6.2 including all replications for each time point. PCA plots were generated with `ggplot2` v3.3.2. Whole transcriptome PCA used read counts from all expressed gene models (n = 82,019) and TF PCA used expression of 2,874 expressed TFs. Randomized PCA distribution used independent random subsampling of 2,874 expressed genes without replacement. Principle component percent variation explained and eigenvalues from `prcomp` were used for comparisons between whole transcriptome PCA and TF-only PCA.

Pairwise differential expression was determined using both `EdgeR` v3.24.3 and `DESeq2` v1.22.2 for robustness (Robinson et al. 2010; Anders and Huber 2010). Pairwise comparisons between consecutive timepoints were done using raw read counts for four biological replicates at each stage.

Benjamin-Hochberg FDR adjusted P-values ≤ 0.01 was used as a stringent DE cut-off for both tools. Only genes DE using both tools were classified as pairwise DEGs (Supplementary data 5). Differential expression of ‘Chinese Spring’ and ‘Kenong9204’ inflorescence development datasets was also determined with raw read counts and EdgeR and DESeq2 using the same method as for the ‘Kronos’ dataset. Adjustments to DE tests were made to compare all four timepoints (6 pairwise comparisons) with two biological replicates in ‘Chinese Spring’ as well as the six timepoints (15 pairwise comparisons) with two biological replicates in the ‘Kenong9204’ datasets. For network analyses, a second DE test was included which reinforced longitudinal DE determination, an impulse model (ImpulseDE2, <https://github.com/YosefLab/ImpulseDE2>) was used for ‘Kronos’ data (Fischer et al. 2018; Spies et al. 2019). Raw counts were used with default parameters and genes with Benjamin-Hochberg FDR adjusted P-values ≤ 0.05 considered differentially expressed. Functional annotation to generate GO terms for each high-confidence and low-confidence gene in the IWGSC RefSeq v1.1 genome was performed as described previously (Pearce et al. 2016).

4.3.6 Standard and consensus WGCNA network construction

Genes identified using pairwise differential expression (EdgeR and DESeq2) and ImpulseDE2 (22,566 genes total) were used for co-expression analyses. A standard co-expression network was built using the R package WGCNA v1.66 with the parameters: power = 20, networkType = signed, minimum module size = 30, and mergecutheight = 0.25 (Langfelder and Horvath 2008) (Supplementary data 7). Parallel coordinate plots were produced in R by normalizing raw read counts and visualized with ggparacoord (scale = ‘globalminmax’) in GGally (version 1.5.0).

A consensus network was built using methods described in Shahan *et al.* (2018). In brief, 1,000 WGCNA runs were performed with 80% of genes randomly subsampled without replacement and random parameters for power (1, 2, 4, 8, 12, 16, 20), minModSize (40, 60, 90, 120, 150, 180, 210), and mergeCutHeight (0.15, 0.2, 0.25, 0.3). The final consensus network was built using an adjacency matrix – $adj = \text{number of times gene } i \text{ is clustered with gene } j / \text{number of times gene } i \text{ is subsampled with gene } j$ –

with parameters $\text{power} = 6$ and $\text{minModuleSize} = 30$ (Supplementary data 7). The consensus100 network was built by filtering the adjacency network for $\text{adj} = 1$ prior to network construction. Along with module assignments, we used the WGCNA package to find the connectivity of each gene with co-clustered genes (`intramodularConnectivity.fromExpr()`) and summarized module expression patterns (`moduleEigengenes()`). Python and R scripts for creating the adjacency matrix and consensus network are available at https://github.com/cvanges/spike_development. The Bioconductor package GeneOverlap was used to determine the overlap of module assignments between consensus and standard networks (<http://shenlab-sinai.github.io/shenlab-sinai/>) (Shen 2021).

4.3.6 Causal Structure Inference Network

Expression data (TPM) for 970 transcription factors retained in the consensus100 network was used to build a gene regulatory network using the Causal Structure Inference algorithm (Penfold and Wild 2011). Network construction used CSI in Cyverse with default parameters.

4.3.7 Conversion of wheat, rice, and barley gene IDs

Genes associated with wheat and rice spikelet number described in Wang et al., 2017b were identified from a previous set of annotated wheat gene models (<ftp://ftp.ensemblgenomes.org/pub/plants/release-28/>). To identify the corresponding IWGSC RefSeq v1.1 gene ID, each gene model coding sequence was extracted and used as a query in BLASTn searches against the IWGSCv1.1 ABU genome. Homologous gene pairs with $> 99\%$ identity to each query were considered spikelet number associated genes. Two previous studies reported genes DE during *H. vulgare* inflorescence development using IBSC_v2 annotations (Digel, Pankin, and von Korff 2015; Liu et al. 2020). Each barley gene model coding sequence was extracted and used as a query in BLASTn searches against the IWGSCv1.1 ABU wheat genome. Genes with percent identity $> 90\%$ were retained and considered orthologs of barley DEGs (HvDE).

4.3.8 Enrichment analysis

Enrichment and depletion of genes among modules or DEG lists was determined using the cumulative distribution function of the hypergeometric distribution (systems.crump.ucla.edu/hypergeometric/).

4.3.9 QTL proximity and definition of homoeologous pairs

Using a previously published meta-analysis of yield component QTL studies, we searched the IWGSCv1.1 genome for expressed genes in our timecourse within 500 kbp of 428 loci associated with yield component traits (kernel number per spike, thousand kernel weight, spikelet number) (Cao et al. 2020). Homoeologous gene pairs reported from Ramírez-González et al., (2018) were used to determine co-expressed homoeologs.

4.4 Results

4.4.1 Early wheat inflorescence development is defined by two major transcriptional shifts

To characterize the wheat transcriptome during inflorescence development, RNA was sequenced from tetraploid durum wheat meristem tissue at five developmental stages; vegetative meristem (W1), double ridge (W2), glume primordium (W3.0), lemma primordium (W3.25), and terminal spikelet (W3.5) (Figure 4.1A) (Waddington, Cartwright, and Wall 1983). An average of 28.9 M reads per sample (79.6% of all reads) mapped uniquely to the A, B and U genomes of the IWGSC RefSeqv1.0 assembly. Of the 190,391 gene models on these chromosomes, 82,019 (43.1%) were expressed (> zero TPM) and 45,243 (23.8%) had a mean expression greater than one TPM in at least one timepoint (Supplementary data 1). Of the 3,861 gene models annotated as TFs (2.0% of gene models), 2,874 (74.5%) were expressed (> zero TPM) and 1,703 (44.1%) had a mean expression greater than one TPM in at least one timepoint (Supplementary data 2).

Comparison of the inflorescence development transcriptome with two whole-plant wheat development transcriptome datasets (Choulet et al. 2014; Ramirez-Gonzalez et al. 2018) revealed 3,682

genes with spike-dominant expression profiles ($\tau > 0.9$, where zero means constitutive expression and one indicates tissue-specific expression) (Supplementary data 3). These genes were most strongly enriched for gene ontology (GO) terms relating to histone assembly and chromosome organization (Supplementary data 4), but also included 286 genes (7.8%) encoding TFs, including both *LEAFY* homoeologs, 15 GROWTH REGULATING FACTOR (GRF) TFs (of 20 expressed during the time course), seven SHI RELATED SEQUENCE (SRS) TFs (out of ten), 20 TCP TFs (out of 49) and ten WOX TFs (out of 28, Supplementary data 3). Despite their known roles in regulating inflorescence development, only two out of 130 MIKC-MADS box and six out of 41 SPL TFs exhibited spike-dominant expression profiles, suggesting they play more diverse roles across plant development. There were 86 spike-specific genes with zero expression in all other stages of development ($\tau = 1$) (Supplementary data 3).

Principal component analysis (PCA) using the whole transcriptome grouped the four biological replicates of each growth stage closely together and revealed that the majority of the transcriptional changes in this time course occur between the vegetative meristem and double ridge formation (Figure 4.1B). These changes are described by PC1, which accounted for 71.8 percent variation explained (PVE). The transition from W1 to W2 was associated with 6,828 DEGs, 58.6% of which were downregulated (Figure 4.1C, Supplementary data 5) and most significantly enriched for GO terms relating to “cell wall organization”, and lignin and hemicellulose metabolic processes (Supplementary data 6). Surprisingly, the 2,828 (41.4%) DEGs upregulated between W1 and W2 were most significantly enriched for GO terms relating to photosynthesis despite the transition from leaf to floral meristem development (Supplementary data 5).

The transition from W2 to W3.0 was associated with 7,531 DEGs (57.6% downregulated, Supplementary data 5, 6). The 3,191 DEGs upregulated between these timepoints were most significantly enriched for “meristem maintenance” and “flower development” GO terms (Supplementary data 6), suggesting that a number of genes triggering floral meristem formation are first activated at this stage.

By contrast, the transcriptomic changes from W3.0 to terminal spikelet formation (Figure 4.1A) were distributed across PC2, which accounts for just 7.4 PVE (Figure 4.1B) and were associated with 12.3-fold fewer DEGs than during the transition from vegetative meristem to stage W3.0 (Figure 4.1C). Just 535 DEGs were found between W3.0 and W3.25 (55.3% upregulated) and 628 DEGs between W3.25 and W3.5 (48.6% upregulated) (Supplementary data 5). Genes upregulated across these three timepoints were most significantly enriched for “floral organ identity” (Supplementary data 6). There are fewer developmental changes between W3.25 and W3.5, relative to changes between W1 and W3.0, which may be due in part to basal and apical spikelets being at similar developmental stages between the latter timepoints (Backhaus et al. 2022).

Of the 11,669 DEGs in at least one of the four consecutive pairwise comparisons, 899 (7.7%) encoded a TF, a 2.2-fold enrichment (hypergeometric $P = 2.22 \text{ e-}62$). This enrichment was strongest after DR through terminal spikelet formation (5.2-fold enrichment, $P = 8.73 \text{ e-}73$) where TFs accounted for 19.8% and 20.5% of all DEGs in pairwise comparisons (Figure 4.1C). A PCA using only TF expression resulted in the same spatial arrangement of biological samples as in the whole-transcriptome PCA but with improved resolution between stages (Figure 4.1B), and explained a greater proportion of variation for PC2 than when including the whole transcriptome (Figure S4.1).

Taken together, these analyses show that less than half of the wheat transcriptome but nearly three-quarters of TFs are expressed during inflorescence development, including a set of genes which are spatially and temporally restricted to early inflorescence tissues. Terminal spikelet formation is associated with comparatively less transcriptional variation relative to stages preceding W3.25 and the strong enrichment in TFs suggests they play critical roles during this stage.

4.4.2 Co-expression networks reveal predominant transcriptome profiles during inflorescence development

Co-expression networks were assembled to identify highly correlated modules of genes that define the major transcriptional profiles during early inflorescence development. All networks were

assembled using a set of 22,566 genes that were differentially expressed in at least one of the ten possible pairwise combinations between timepoints (Figure 4.1D) and that were also defined as significantly differentially expressed using ImpulseDE2, a package used to analyze longitudinal transcriptomic datasets (Supplementary data 5).

A consensus network constructed with repeated subsampling and randomized parameters with WGCNA (see Materials and Methods) assembled these genes into 21 modules with a mean connectivity score of 0.485 (Figure 4.2A, Supplementary data 7). A standard WGCNA network was also constructed using ‘best practices’ parameters but with no repeated subsampling and randomization and had a connectivity score of 0.327 which skewed to zero (Figure 4.2A). In both networks, the majority of genes clustered into modules 1 and 2, which contained many of the same genes (Jaccard index > 0.86, Figure S4.2). However, other modules exhibited dissimilar expression profiles between networks (Jaccard index < 0.5), indicating the consensus network clustered genes into a greater number of modules with distinct expression profiles not captured in the standard network. Based on the improved correlation of co-clustered genes within modules and the detection of distinct regulatory profiles, the consensus network was used in all subsequent analyses.

4.4.3 Inflorescence meristem development is associated with the down-regulation of RAV and TCP transcription factors

Module 1 was the largest in the network and grouped 10,102 genes defined by high transcript levels in the vegetative meristem and early meristem transition followed by down-regulation after DR and as the spike develops (Figure 4.2B). Several TF families were enriched in this module, including 101 basic Helix-Loop-Helix (bHLH) TFs, 47 MYB TFs and eight of the nine differentially expressed RELATED TO ABI3 AND VP1 (RAV) TFs included in the network (Figure 4.2F). Twenty-six of the 33 total TCP TFs clustered in this module, nine of which were also spike-dominant expressed (Figure 4.2F). Although at the whole family level MIKC-MADS TFs are significantly under-represented in module 1 (Figure 4.2F, hypergeometric $P = 8.6 \text{ e-}4$), all six SVP genes (*SVPI*, *VRT2* and *SVP3*) cluster in this

module, consistent with their specific role regulating early stages of inflorescence development. In addition, both AGL12 subclade genes, and three of the six FLC subclade genes clustered in this module (Figure 4.2G).

4.4.4 A small number of genes are transiently expressed during double ridge formation

Genes which showed a peak at the double ridge stage (W2) followed by a decline in later stages were clustered in modules 11 (131 genes), 15 (104 genes), 20 (44 genes) and 21 (42 genes). These clusters share broadly similar expression profiles (Figure 4.2C) and were enriched for genes with spike-dominant expression profiles (between 2.1 and 3.0-fold enrichment). Genes in modules 15 and 20 were significantly enriched for development functional terms including “shoot system development” and “carpel development” (Supplementary data 8) including three *TERMINAL FLOWER1*-like genes *CENTRORADIALIS2* (*CEN2*), *CEN4*, and *CEN-5A* (Supplementary data 7). All three modules were enriched for the functional term “response to auxin” and included several auxin-responsive factors (ARF), indole-3 acetic acid (IAA), and SAUR-like protein family members, indicating that auxin signaling may promote double ridge formation.

4.4.5 Inflorescence transition and spike architecture genes are upregulated at W3.0

Modules 6 (267 genes), 8 (211 genes), and 10 (144 genes) share broadly similar profiles defined by maximum expression at stage W3.0 and subsequent downregulation (Figure 4.2D). Each of these modules was significantly enriched (between 2.3 and 5.3-fold) for spike-dominant genes, indicating they likely play highly specific roles restricted to developing meristems and inflorescence initiation. Module 6 included 18 genes previously associated with variation in spikelet number and five orthologs of rice genes with roles in panicle development, including the ERF TF *WHEAT FRIZZY PANICLE* (*WFZP*) and *KAN2*, a MYB TF which functions in establishing lateral organ polarity in *Arabidopsis* (Emery et al. 2003; Shaw et al. 2019).

4.4.6 Inflorescence and spikelet meristem formation is associated with sequential activation of different classes of TFs

The 8,971 genes in module 2 were defined by the inverse transcriptional profile to module 1, with low expression in the vegetative meristem followed by sustained upregulation from the double ridge stage onwards (Figure 4.2E). Transcription factors were under-represented in this module, and only the B3 family (42 of 77 B3 TFs assembled in the co-expression network) was significantly enriched (Figure 4.2F). There were 18 MIKC-MADS box TFs which were upregulated early in the transition to the inflorescence meristem including all genes in the AP1/SQUA subclade (with the exception of *VRN-A1*) and six of the thirteen genes in the SEP1 subclade (Figure 4.2G). Several genes with characterized roles in inflorescence development clustered in this module, including *FLOWERING LOCUS T2 (FT-A2)*, *Q*, and *RAMOSA2 (TaRA-B2)* (Supplementary data 7) (Shaw et al. 2019; Debernardi et al. 2020).

The 708 genes clustered in module 3 exhibited a similar transcriptional profile to module 2, with a delayed upregulation and stronger peak at the terminal spikelet stage (Figure 4.2E). These genes are significantly enriched for developmental functional terms including “specification of floral organ identity”, suggesting they include floral patterning and developmental genes that regulate spikelet meristem formation (Supplementary data 8). This module was significantly enriched for both spikelet-dominant expressed genes (106 genes, $P < 0.001$) and for TFs (86 genes, 12.1%, $P < 0.001$), consistent with pairwise DE analysis between stages W3.0 and W3.5 (Figure 4.1C). These included four members of the SRS TF family, four YABBY TFs, and the HD-zip TFs *Grain Number Increase 1 (GN1)* and *HOX2* (Supplementary data 7). All members of the MIKC-MADS subclades PI, AGL6 and SEP3 were clustered in module 3, as well as two of the three AP3 subclade genes, four of the five AG/STK subclade genes and five SEP1 subclade genes (Figure 4.2G).

4.4.7 Gene regulatory networks predict high-confidence interactions between transcription factors

To identify the most robust co-expression patterns, the consensus adjacency matrix used for previous co-expression analyses was filtered for genes which co-clustered with at least one gene every

time they were co-sampled in 1,000 networks assembled with variable, randomized parameters. The 18,174 genes that met this criterion were assembled into a consensus100 network consisting of 924 modules with a median size of three (Supplementary data 7).

Module 9 of this network comprised 167 genes (including 32 TFs) which were most highly expressed at the terminal spike stage (Figure S4.3) and significantly enriched for the GO terms “specification of floral organ identity” and “flower development” (Supplementary data 9), suggesting it may represent a core regulatory network for wheat spikelet and/or floret development. The genes with the highest connectivity (Kw, a measure of each gene’s intramodular co-expression) in this module are *SEPI-A2* and *SEPI-B2*, which may be related with the intermediate position of the *SEP* genes between the meristem identity SQUAMOSA MADS-BOX genes and the anther and carpel development MADS-box genes. This module also groups *WAPO-A1*, that influences spikelet number and stamen identity (Kuzay et al. 2022) and a gene encoding an F-box protein that is a component of an SCF ubiquitin ligase that may be targeted by *TBI* (Supplementary data 7) (Dong et al. 2017).

To predict interactions between TFs during inflorescence development, a *de novo* Causal Structure Inference (CSI) network was constructed using all 970 TFs from the consensus100 network. This gene regulatory network consisted of 704 genes (nodes) with 5,604 predicted interactions (edges) with interaction strength (edge weight) > 0.001 (Supplementary data 10). To prioritize the most important regulatory candidate genes, the network was screened for interactions with an edge weight ≥ 0.03 , leaving 88 genes with 177 interactions. The majority of these genes were from consensus modules 1 (37 genes, 42.0%) and 3 (36 genes, 40.9%), with 27 of the latter genes clustered in consensus100 module 9 (Figure 4.3).

Most predicted interactions were between genes in the same consensus module, with the majority occurring within module 3 and involving MIKC-MADS box TFs, suggesting a closely coordinated network during spikelet meristem and terminal spikelet formation (Figure 4.3). Among the genes with the highest betweenness centrality, a measure of each gene’s importance in the overall network, were *AGL6-*

AI and *AGL6-B1* which were predicted to interact with 31 other TFs in the network, including 13 MIKC-MADS genes such as *PI-1*, *SEP3-1*, *AP3-1*, *SEP1-1* and *AG1* (Figure 4.3). Interaction strengths implicated a role for *AG-D1* as a regulatory hub with strong incoming interactions from other MIKC-MADS-box genes from the SEP1, SEP3, AG, PI, and AP3 subclades, as well as outgoing interactions with genes such as the LOFSEP MIKC-MADS box TF *SEP1-1* (Figure 4.3). The BES1 TF *BES1/BZR1 HOMOLOG 2-like* had high betweenness centrality and was predicted to have outgoing interactions with MIKC-MADS, Trihelix and HD-ZIP TFs (Figure 4.3).

Cross-module interactions included 16 outgoing edges from module 3 to module 1, including six outgoing interactions to a PCF-type TCP TF (Figure 4.3). Although only four TFs from module 5 were assembled in the network, they included *SEP1-A3* and a C2H2 TF with ten incoming interactions from module 3 including *AGL6-B1*, *BES1/BZR1 HOMOLOG 2-like* and *AG-D1* (Figure 4.3).

4.4.8 Integrating transcriptomics to prioritize candidate genes underlying natural variation

The consensus network includes 4,637 high confidence homoeologous gene pairs, the majority of which (3,636, 78.4 %) clustered either in the same module, or in modules with highly similar expression profiles (Supplementary data 7). We hypothesized that homoeologous genes clustering in different modules may have divergent expression profiles resulting from natural variation in one homoeolog. Of these 1,001 divergently expressed gene pairs, 221 encoded TFs, including *VRN1* (where the dominant *VRN-A1* spring allele is expressed at an earlier stage of inflorescence development compared to the wild-type *VRN-B1* allele), *RHT1* (where the *Rht-B1b* semi-dwarfing allele is more highly expressed in the vegetative meristem than *RHT-A1*), and *TEOSINTE BRANCHED 1 (TB1)*, where *TB-B1* expression is maintained at higher levels than *TB-A1* during terminal spikelet formation, Figure 4.4A).

Each of the three genes from Figure 4.4A lies within 250 kb of a QTL for either grain number or grain size (Supplementary data 7), so we hypothesized that other differentially expressed homoeologs located close to a yield-component QTL might point to natural variation for yield traits in wheat. For

example, *UPBEAT-A1* is upregulated at the double ridge stage to a much greater degree than *UPBEAT-B1* (Figure 4.4B), is close to a QTL for TKW, and encodes an ortholog of a bHLH TF that regulates cell proliferation in Arabidopsis (Tsukagoshi, Busch, and Benfey 2010). Similarly, *TRYPTOPHAN AMINOTRANSFERASE RELATED-A1 (TAR-A1)* is also upregulated at the double ridge stage compared to *TAR-B1* (Figure 4.4B) and is proximal to a QTL for grain yield (Supplementary data 7). These genes encode enzymes in the IAA biosynthesis pathway and their overexpression has previously been shown to modify inflorescence development in wheat (Shao et al. 2017). Co-expression networks and observations from meta-analysis are available for developing hypotheses on inflorescence development (Supplementary data 7).

4.4.8 Identification of *CLE/WOX* genes expressed during wheat inflorescence development

To identify members of the conserved CLAVATA-WUSCHEL pathway that may regulate stem cell maintenance in wheat spike meristems, the expression profiles of genes encoding WOX TFs and *CLE* peptides were analyzed. Of 29 WOX TFs, 28 were expressed during early inflorescence development and 11 were both significantly differentially expressed during the time course and exhibited a spike-specific expression profile (Figure 4.5A). Two orthologs of *OsWOX4* were co-expressed in module 1 with rapid down-regulation before transition to the inflorescence meristem, suggesting they may play a role in vegetative meristem maintenance but not in inflorescence development. Seven *WOX* genes clustered in module 2, characterized by rising expression during inflorescence development, including the orthologs of *AtWUS (TaWUSa and b)*. The homoeologues *TaWOX2a* and *2b* are both associated with variation in spikelet number and are clustered into separate co-expression modules (Supplementary data 7). Of the 64 *CLE* genes, 35 were expressed during inflorescence development and just nine were differentially expressed across the time course (Figure 4.5B). Three wheat genes orthologous to *OsFON2/4* (putatively *TaCLV3*, *TraesCS2A02G329300* and *TraesCS2B02G353000*) exhibit spike-dominant DR-peaking expression profiles.

4.5 Discussion

Temporal transcriptomic datasets can help to characterize the regulatory networks controlling the development of complex organs such as the wheat inflorescence. One strategy to reduce spurious co-clustering of genes is to assemble a consensus co-expression network using a matrix of co-clustering frequencies from multiple independent networks, each assembled with randomized parameters and gene selection (Monti et al. 2003; Wu et al. 2002; Shahan et al. 2018). Co-expression networks have been successfully applied to unravel gene function in yeast (*Saccharomyces cerevisiae*), floral and fruit developmental pathways in strawberry (*Fragaria vesca*), and regulatory networks underlying leaf development in maize (*Zea mays*) (Wu et al. 2002; Shahan et al. 2018; Miculan et al. 2021). In the current study, this approach generated a consensus network with a larger number of modules with improved intramodular connectivity compared to a standard WGCNA network (Figure 4.2A). A further refinement to screen for genes co-clustering in every network assembly that they were both included revealed a consensus100 module 9 of 167 genes that likely contribute to spikelet meristem and terminal spikelet formation (Figure S4.3), indicating that consensus networks can help improve the accuracy of co-expression predictions and module assignment.

Beyond co-expression profiles, context-specific gene regulatory networks provide information on the centrality of each gene (a measure of its importance to the flow of information through a network), as well as the strength and directionality of interactions between individual genes (Penfold and Wild 2011). This network predicts that the MIKC-MADS box TF *AGL6* is a critical gene in inflorescence development regulatory networks, and functions together with MIKC-MADS TFs from the PI and SEP subclades (Figure 4.3). This is consistent with its role in rice, where *AGL6* functions as a cofactor with A, B, C, and D class proteins during floral development, as well as in wheat, where it interacts with ABCDE proteins, likely as a bridge in complex protein-protein interactions to regulate whorl development (Su et al. 2019; Li et al. 2011; Kong et al. 2022). This network also revealed novel candidate genes for future characterization studies. For example, the *BES1* TF *BES1/BZRI HOMOLOG 2-like* is predicted to interact

with several TFs, including two HD-ZIP TFs with homology to *HvVRS1*, suggesting a role for brassinosteroid signaling in wheat inflorescence development.

During the inflorescence development time course in tetraploid Kronos presented here, 43.1% of genes were expressed in at least one timepoint, comparable to the 40.2% and 42.5% of genes expressed in similar inflorescence development time courses in the hexaploid wheat genotypes ‘Chinese Spring’ and ‘Kenong 9204’ when these reads were reanalyzed using the same mapping parameters and reference genome (Feng et al. 2017; Li et al. 2018). Of these genes, 3,682 exhibited spike-dominant expression profiles ($\tau > 0.9$). Among these genes were seven of ten SRS TFs, including the wheat ortholog of *six-rowed spike 2* (*HvVRS2*) that modulates hormone activity in the developing barley spike (Yousef et al. 2016). Its expression profile in wheat, coupled with its association with spikelet number in an earlier study (Li et al. 2018), suggests it plays a conserved role in wheat inflorescence development. It would also be interesting to characterize the function of four other SRS TFs that exhibit spike-specific expression profiles peaking towards terminal spikelet formation (Supplementary data 7). Ten of fifteen GRF TFs were expressed predominantly in spike tissues, including *TaGRF4* which improves regeneration efficiency in tissue culture when co-expressed with GIF cofactors (Debernardi et al. 2020). The broadly similar, spike-specific expression profiles of genes in this family suggest other members may also contribute to meristem differentiation and inflorescence development (Supplementary data 7).

A subset of *WOX* TFs and *CLE* peptides exhibited dynamic and spike-dominant expression profiles across the time course, consistent with the differential regulation of *OsWUS*, *OsWOX3*, *OsWOX4*, and *OsWOX12* during panicle development in rice (Cheng et al. 2014). The overexpression of *TaWOX5* (named *TaWOX9* in the current study) enhances wheat transformation and callus regeneration efficiency (Wang et al. 2022). Several other *WOX* TFs are co-clustered with this gene and exhibit similar expression profiles in the wheat inflorescence (Figure 4.5), suggesting they may also be candidates to enhance regeneration efficiency (Figure 4.5). Among *CLE* peptides, *TaCLV3* was negatively associated with

spikelet number in a set of Chinese wheat landraces (Wang et al. 2017), consistent with its proposed role as a negative regulator of SAM size and activity in rice and maize (Chu et al. 2006; Hu et al. 2021).

Analyses of principal components and co-expression profiles indicate that the transition from the vegetative meristem to the double ridge stage is associated with major reprogramming of the wheat transcriptome (Figure 4.1), consistent with an earlier study (Li et al. 2018). Several TF families were enriched in module 1, characterized by high expression in the vegetative meristem before rapid downregulation after the double ridge stage, including eight of the nine RAV TFs in the consensus network. In Arabidopsis, the RAV genes *TEMPRANILLO1* (*TEM1*) and *TEM2* repress *FT* to prevent precocious flowering (Hu et al. 2021; Castillejo and Pelaz 2008). In rice, the *TEM* orthologs *OsRAV8* and *OsRAV9* bind the promoters of *OsMADS14* and *Hd3a* to suppress the floral transition, indicating this function is conserved in monocots (Osnato et al. 2020). The rapid downregulation of the wheat orthologs of these genes before double ridge formation, as well as homologs of *OsRAV11* and *OsRAV12* that act in reproductive patterning in rice (Osnato et al. 2020), suggests this family may act as local repressors of meristem identity genes in the developing wheat spike.

There were also 26 TCP TFs clustered in module 1, including *TaTCP-A9* and *TaTCP-B9*, negative regulators of spikelet number and grain size in durum wheat (Zhao et al. 2018). It is likely that other members of the TCP TF family also play roles as negative regulators of grain development. For example, *TaTCP-A17* and *-B17* are both downregulated during inflorescence development, are within 250 kbp of QTL for grain size, and are orthologous to genes associated with spikelet number variation in rice (Supplementary data 7). Eight TCP TFs clustered in different modules and were most highly expressed during spikelet meristem formation, including *TEOSINTE BRANCHED 1*, which integrates photoperiod signals to regulate spike architecture in a dosage-dependent manner (Dixon et al. 2018), and a paralogous copy on chromosome 5B, *BRANCHED AND INDETERMINATE SPIKE*, that regulates spike architecture in barley (Shang et al. 2020). Four other uncharacterized TCP TFs with homology to *RETARDED*

PALEA1 exhibit spike-dominant expression profiles and would be promising candidates to characterize their role in inflorescence development in wheat (Supplementary data 7).

Although association and linkage mapping studies in wheat have described hundreds of QTL for agronomic traits, relatively few causative genes have been cloned and validated (Cao et al. 2020). Transcriptomic data can help prioritize candidate gene selection within a mapping interval based on spatial or temporal expression profiles (Yang et al. 2021). Furthermore, changes in transcription may indicate the presence of dominant or semi-dominant gain-of-function variants in cis-regulatory elements or of structural variation that confer changes in phenotype through modified expression profiles. Because of the functional redundancy of the polyploid wheat genome, such variants underlie the majority of cloned genes to date (Gaurav et al. 2022), including domestication alleles of *PPD1*, *VRN1* and *RHT1*, which clustered in different co-expression modules to their wild-type homoeologous allele (Figure 4.4). Such divergent expression profiles, especially for those genes in close proximity to QTL for traits relating to grain number and grain size, might be strong candidates for allele mining to explore the extent of natural variation in wheat germplasm collections, and to engineer novel variation by targeted editing of cis-regulatory regions (Swinnen, Goossens and Pauwels 2016).

Chapter 4 Figures

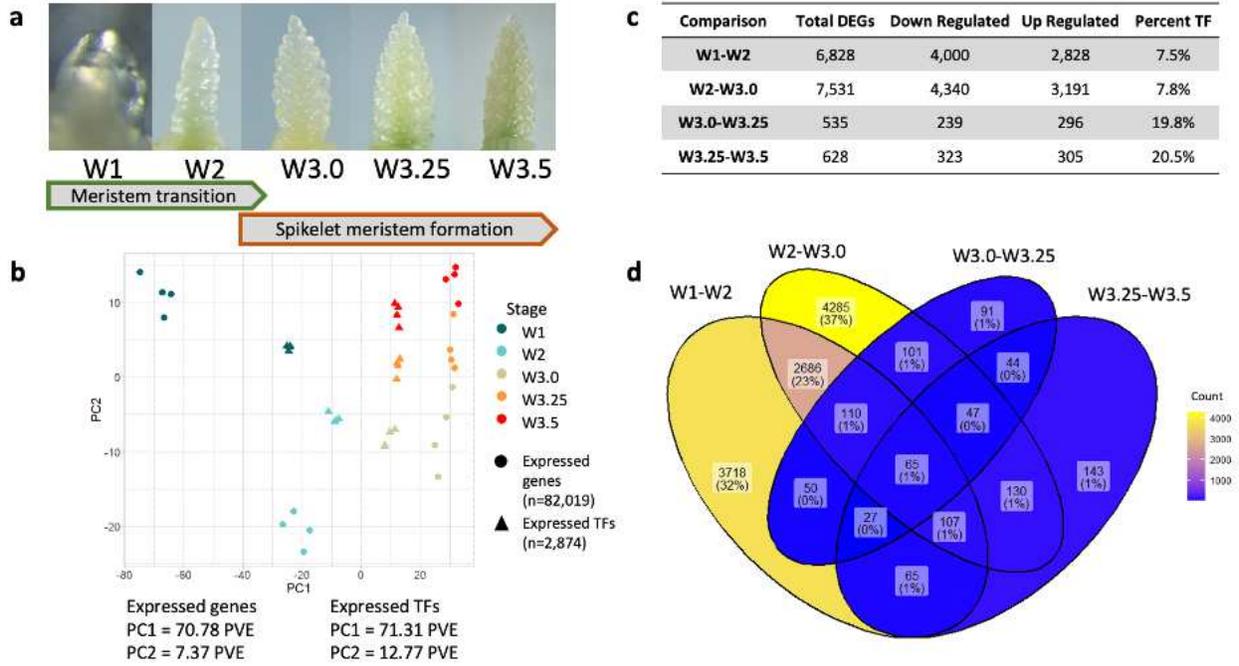


Figure 4.1: The early wheat inflorescence development transcriptome. (A) Sampling stages of Kronos apical meristems according to the Waddington development scale³; W1.0 – vegetative meristem, W2 – double ridge, W3.0 – glume primordium, W3.25 – lemma primordium, W3.5 – terminal spikelet. (B) Whole transcriptome and transcription factor expression principal component analysis of samples, PC1 plotted on the x-axis and PC2 plotted on the y-axis. PVE = Percent Variance Explained. (C) Differentially expressed genes (DEGs) in sequential pairwise comparisons (W1 – W2, W2 – W3.0, W3.0 – W3.25, W3.25 – W3.5). The total number of genes, the number up- and down-regulated and the proportion encoding transcription factors (TF) are described. (D) Venn diagram of DEGs in each consecutive pairwise comparison from (C). Each category is shaded according to the number of sequential DEGs shared among the four comparisons.

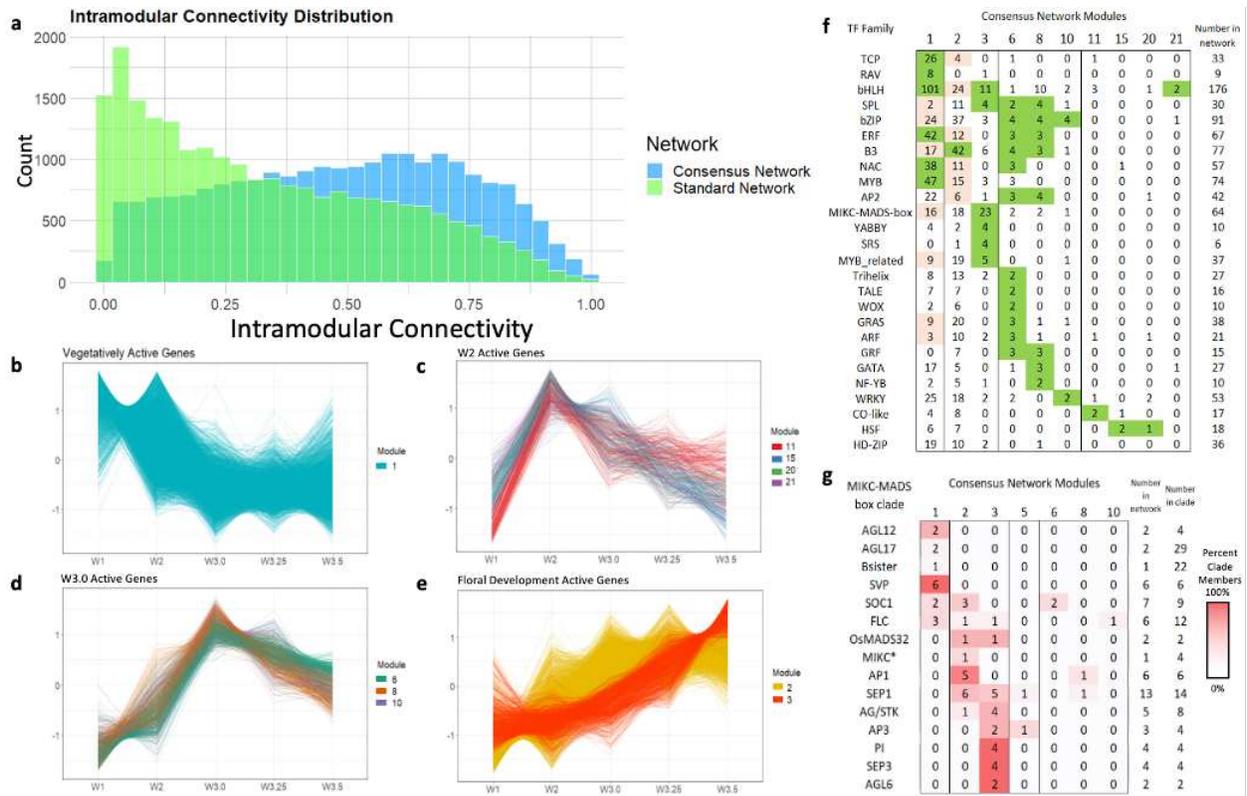


Figure 4.2: Co-expression networks showing the dominant transcriptional profiles during wheat inflorescence development. (A) Histogram of intramodular connectivity scores for 22,566 genes clustered in consensus (blue) or standard (green) network. (B - E) Expression profiles during inflorescence development of discussed modules in the consensus network. Lines represent scaled time course expression of each gene in the module. Modules with similar expression profiles are grouped together for comparison. (F) Number of TF family members clustered in each discussed consensus module. Modules enriched (green) or depleted (pink) for TF families are highlighted ($P < 0.01$). (G) Number of MIKC-MADS box clade members clustered in each consensus modules. Co-expressed MIKC-MADS box groups are shaded relative to the total number of genes in the clade.

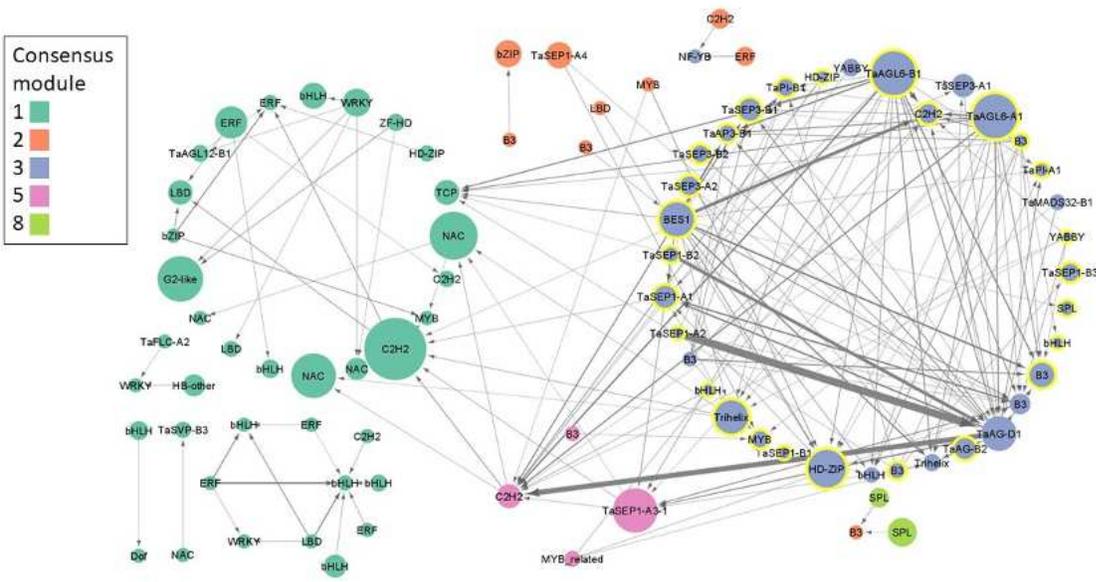


Figure 4.3: Causal structural inference prediction of interacting transcription factors, filtered for edge weight ≥ 0.03 . Nodes (genes) are colored by their consensus network modules, and consensus100 module 9 genes are highlighted with a yellow border. Node diameter is scaled to betweenness centrality to indicate its importance within the network. Directional interactions are indicated by arrows and width is scaled to predicted interaction strength.

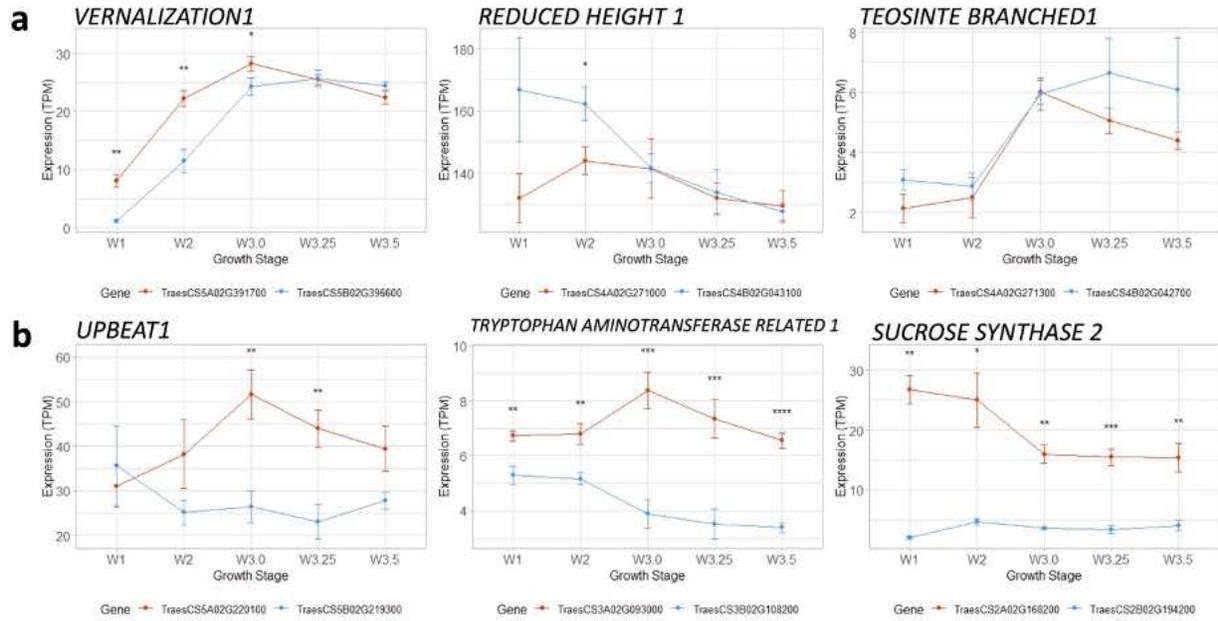


Figure 4.4: Divergent expression of homoeologous gene pairs during inflorescence development. Expression profiles of (A) Characterized domestication and adaptation alleles and (B) Genes close to QTL for spike architecture or grain size. Expression values are in TPM \pm standard error. A-genome homoeologs are in orange, B-genome homoeologs are in blue. Paired t-tests were used to indicate differences between homoeolog expression at each time point, P values < 0.05 (*), 0.01 (**), 0.001 (***), 0.0001 (****).

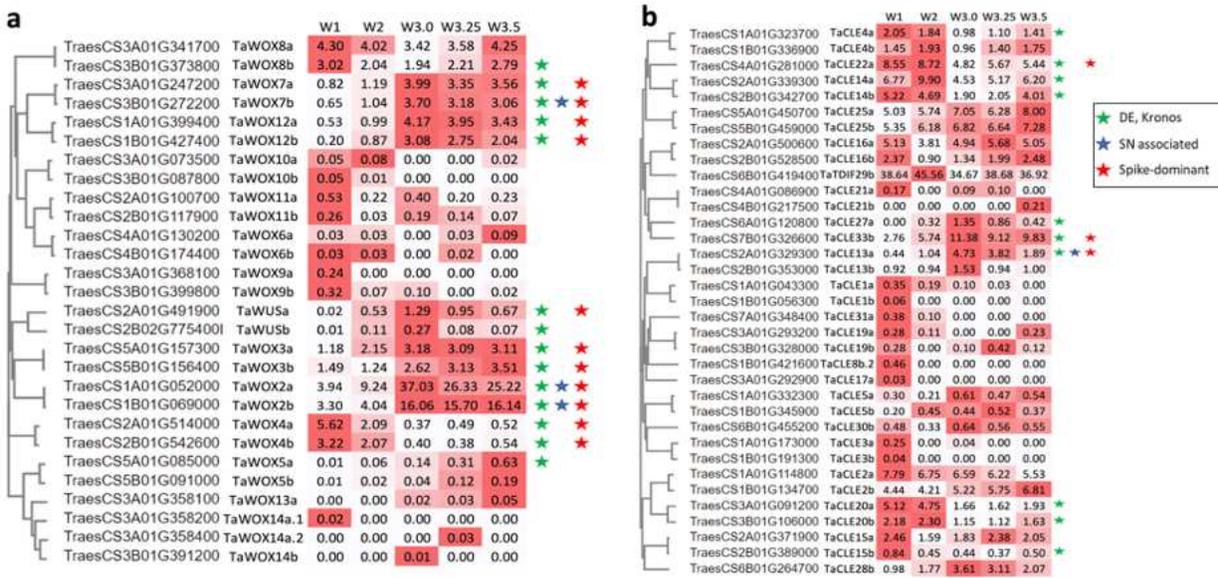


Figure 4.5: Expression profiles of WOX TFs (A) and CLE peptides (B) during wheat inflorescence development. Stars indicate additional evidence of a possible role in spike regulation (green = differential expression in ‘Kronos’ inflorescence, blue = associated with variation in spikelet number, red = spike-dominant expression profile). Heatmaps show expression (TPM) relative to each gene’s minimum and maximum expression. Only genes with TPM ≥ 0.05 are shown.

Chapter 4 References

- Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8, e66428 (2013).
- Brinton, J. & Uauy, C. A reductionist approach to dissecting grain weight and yield in wheat. *J Integr Plant Biol* 61, 337–358 (2019).
- Waddington, S. R., Cartwright, P. M. & Wall, P. C. A quantitative scale of spike initial and pistil development in barley and wheat. *Annals of Botany* Vol 51, 119–130 (1983).
- Li, C. *et al.* Wheat *VRN1*, *FUL2* and *FUL3* play critical and redundant roles in spikelet development and spike determinacy. *Development* 146, dev175398 (2019).
- Bonnett, O. T. Inflorescences of maize, wheat, rye, barley, and oats: their initiation and development. *University of Illinois (Urbana-Champaign) College of Agriculture, Agricultural Research Center Bulletin* 721, (1966).
- Rawson, H. M. Spikelet number, its control and relation to yield per ear in wheat. *Aust J Biol Sci* 23, 1–15 (1970).
- Cao, S., Xu, D., Hanif, M., Xia, X. & He, Z. Genetic architecture underpinning yield component traits in wheat. *Theoretical and Applied Genetics* 133, 1811–1823 (2020).
- Würschum, T., Leiser, W. L., Langer, S. M., Tucker, M. R. & Longin, C. F. H. Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. *Theoretical and Applied Genetics* 131, 2071–2084 (2018).
- Somssich, M., Je, B. il, Simon, R. & Jackson, D. CLAVATA-WUSCHEL signaling in the shoot meristem. *Development* 143, 3238–3248 (2016).
- Fletcher, J. C. The CLV-WUS stem cell signaling pathway: a roadmap to crop yield optimization. *Plants* 7, 87 (2018).
- Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E. & Lippman, Z. B. Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171, 470–480 (2017).

- Chen, Z. *et al.* Structural variation at the maize *WUSCHEL1* locus alters stem cell organization in inflorescences. *Nat Comms* 12, 1–12 (2021).
- Li, Z. *et al.* Identification and functional analysis of the CLAVATA3/EMBRYO SURROUNDING REGION (CLE) gene family in wheat. *Int J Mol Sci* 20, 4319 (2019).
- Li, Z. *et al.* Identification of the WUSCHEL-Related Homeobox (WOX) gene family, and interaction and functional analysis of *TaWOX9* and *TaWUS* in wheat. *Int J Mol Sci* 21, 1581 (2020).
- Honma, T. & Goto, K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529 (2001).
- Theißen, G. Development of floral organ identity: Stories from the MADS house. *Current Opinion in Plant Biology* 4, 75–85 (2001).
- Schilling, S., Kennedy, A., Pan, S., Jermiin, L. S. & Melzer, R. Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytologist* 225, 511–529 (2020).
- Li, K. *et al.* Interactions between SQUAMOSA and SHORT VEGETATIVE PHASE MADS-box proteins regulate meristem transitions during wheat spike development. *The Plant Cell* 33, 3621–3644 (2021).
- Adamski, N. M. *et al.* Ectopic expression of *Triticum polonicum VRT-A2* underlies elongated glumes and grains in hexaploid wheat in a dosage-dependent manner. *The Plant Cell* 33, 2296–2319 (2021).
- The International Wheat Genome Sequencing Consortium (IWGSC) *et al.* Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, eaar7191 (2018).
- Walkowiak, S. *et al.* Multiple wheat genomes reveal global variation in modern breeding. *Nature* 588, 277–283 (2020).
- Rao, X. & Dixon, R. A. Co-expression networks for plant biology: why and how. *Acta Biochim Biophys Sin (Shanghai)* 51, 981–988 (2019).
- van den Broeck, L., Gordon, M., Inzé, D., Williams, C. & Sozzani, R. Gene regulatory network inference: Connecting plant biology and mathematical modeling. *Frontiers in Genetics* 11, 457 (2020).

- Feng, N. *et al.* Transcriptome profiling of wheat inflorescence development from spikelet initiation to floral patterning identified stage-specific regulatory genes. *Plant Physiology* 174, 1779–1794 (2017).
- Li, Y. *et al.* A genome-wide view of transcriptome dynamics during early spike development in bread wheat. *Scientific Reports* 8, 1–16 (2018).
- Wang, Y. *et al.* Transcriptome association identifies regulators of wheat spike architecture. *Plant Physiology* 175, 746–757 (2017).
- Choulet, F. *et al.* Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345, eaar1249721 (2014).
- Ramírez-González, R. H. *et al.* The transcriptional landscape of polyploid wheat. *Science* 361, eaar6089 (2018).
- Backhaus, A. E. *et al.* High expression of the MADS-box gene *VRT2* increases the number of rudimentary basal spikelets in wheat. *Plant Physiology* 189, 1536–1552 (2022).
- Emery, J. F. *et al.* Radial patterning of *Arabidopsis* shoots by Class III HD-ZIP and KANADI genes. *Current Biology* 13, 1768–1774 (2003).
- Du, D. *et al.* FRIZZY PANICLE defines a regulatory hub for simultaneously controlling spikelet formation and awn elongation in bread wheat. *New Phytologist* 231, 814–833 (2021).
- Shaw, L. M. *et al.* *FLOWERING LOCUS T2* regulates spike development and fertility in temperate cereals. *Journal of Experimental Botany* 70, 193–204 (2019).
- Debernardi, J. M., Greenwood, J. R., Jean Finnegan, E., Jernstedt, J. & Dubcovsky, J. APETALA 2-like genes *AP2L2* and *Q* specify lemma identity and axillary floral meristem development in wheat. *The Plant Journal* 101, 171–187 (2020).
- Kuzay, S. *et al.* *WAPO-A1* is the causal gene of the 7AL QTL for spikelet number per spike in wheat. *PLOS Genetics* 18, e1009747 (2022).

- Dong, Z. *et al.* Ideal crop plant architecture is mediated by *tassels replace upper ears1*, a BTB/POZ ankyrin repeat gene directly targeted by TEOSINTE BRANCHED1. *Proc Natl Acad Sci U S A* 114, E8656–E8664 (2017).
- Tsukagoshi, H., Busch, W. & Benfey, P. N. Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. *Cell* 143, 606–616 (2010).
- Shao, A. *et al.* The auxin biosynthetic *TRYPTOPHAN AMINOTRANSFERASE RELATED TaTAR2.1-3A* increases grain yield of wheat. *Plant Physiology* 174, 2274–2288 (2017).
- Monti, S. *et al.* Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning* 52, 91–118 (2003).
- Wu, L. F. *et al.* Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. *Nat Genet* 31, 255–265 (2002).
- Shahan, R. *et al.* Consensus coexpression network analysis identifies key regulators of flower and fruit development in wild strawberry. *Plant Physiology* 178, 202–216 (2018).
- Miculan, M. *et al.* A forward genetics approach integrating genome-wide association study and expression quantitative trait locus mapping to dissect leaf development in maize (*Zea mays*). *The Plant Journal* 107, 1056–1071 (2021).
- Penfold, C. A. & Wild, D. L. How to infer gene networks from expression profiles, revisited. *Interface Focus* 1, 857–870 (2011).
- Su, Y. *et al.* Wheat AGAMOUS like 6 transcription factors function in stamen development by regulating the expression of *TaAPETALA3*. *Development* 146, (2019).
- Li, H. *et al.* Rice *MADS6* interacts with the floral homeotic genes *SUPERWOMANI*, *MADS3*, *MADS58*, *MADS13*, and *DROOPING LEAF* in specifying floral organ identities and meristem fate. *The Plant Cell* 23, 2536–2552 (2011).
- Kong, X. *et al.* The wheat *AGL6*-like MADS-box gene is a master regulator for floral organ identity and a target for spikelet meristem development manipulation. *Plant Biotechnology Journal* 20, 75–88 (2022).

- Youssef, H. M. *et al.* *VRS2* regulates hormone-mediated inflorescence patterning in barley. *Nat Genet* 49, 157–161 (2016).
- Debernardi, J. M. *et al.* A GRF–GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nature Biotechnology* 38, 1274–1279 (2020).
- Cheng, S., Huang, Y., Zhu, N. & Zhao, Y. The rice *WUSCHEL*-related homeobox genes are involved in reproductive organ development, hormone signaling and abiotic stress response. *Gene* 549, 266–274 (2014).
- Wang, K. *et al.* The gene *TaWOX5* overcomes genotype dependency in wheat genetic transformation. *Nature Plants* 8, 110–117 (2022).
- Chu, H. *et al.* The *FLORAL ORGAN NUMBER4* gene encoding a putative ortholog of Arabidopsis *CLAVATA3* regulates apical meristem size in rice. *Plant Physiology* 142, 1039–1052 (2006).
- Bommert, P., Je, B. il, Goldshmidt, A. & Jackson, D. The maize $G\alpha$ gene *COMPACT PLANT2* functions in *CLAVATA* signalling to control shoot meristem size. *Nature* 502, 555–558 (2013).
- Hu, H. *et al.* *TEM1* combinatorially binds to *FLOWERING LOCUS T* and recruits a Polycomb factor to repress the floral transition in *Arabidopsis*. *Proc Natl Acad Sci U S A* 118, e2103895118 (2021).
- Castillejo, C. & Pelaz, S. The balance between *CONSTANS* and *TEMPRANILLO* activities determines *FT* expression to trigger flowering. *Current Biology* 18, 1338–1343 (2008).
- Osnato, M., Matias-Hernandez, L., Aguilar-Jaramillo, A. E., Kater, M. M. & Pelaza, S. Genes of the *RAV* family control heading date and carpel development in rice. *Plant Physiology* 183, 1663–1680 (2020).
- Zhao, J. *et al.* Genome-Wide Identification and Expression Profiling of the *TCP* Family Genes in Spike and Grain Development of Wheat (*Triticum aestivum* L.). *Frontiers in Plant Science* 9, (2018).
- Dixon, L. E. *et al.* *TEOSINTE BRANCHED1* regulates inflorescence architecture and development in bread wheat (*Triticum aestivum*). *The Plant Cell* 30, 563–581 (2018).
- Shang, Y. *et al.* A *CYC/TB1*-type *TCP* transcription factor controls spikelet meristem identity in barley. *Journal of Experimental Botany* 71, 7118–7131 (2020).

- Yang, Y. *et al.* Large-scale integration of meta-QTL and genome-wide association study discovers the genomic regions and candidate genes for yield and yield-related traits in bread wheat. *Theoretical and Applied Genetics* 134, 3083–3109 (2021).
- Gaurav, K. *et al.* Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement. *Nature Biotechnology* 40, 422–431 (2022).
- Swinnen, G., Goossens, A. & Pauwels, L. Lessons from domestication: Targeting cis-regulatory elements for crop improvement. *Trends in Plant Science* 21, 506–515 (2016).
- Uauy, C., Wulff, B. B. H. & Dubcovsky, J. Combining traditional mutagenesis with new high-throughput sequencing and genome editing to reveal hidden variation in polyploid wheat. *Annual Review of Genetics* 51, 435–54 (2017).
- Wilhelm, E. P., Turner, A. S. & Laurie, D. A. Photoperiod insensitive *Ppd-A1a* mutations in tetraploid wheat (*Triticum durum* Desf.). *Theoretical and Applied Genetics* 118, 285–294 (2008).
- Fu, D. *et al.* Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Mol Genet Genomics* 273, 54–65 (2005).
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, 884–890 (2018).
- Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013).
- Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930 (2014).
- Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21, 650–659 (2005).
- Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Briefings in Bioinformatics* 18, 205–214 (2017).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010).

- Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biology* 11, R106 (2010).
- Fischer, D. S., Theis, F. J. & Yosef, N. Impulse model-based differential expression analysis of time course sequencing data. *Nucleic Acids Research* 46, e119 (2018).
- Spies, D., Renz, P. F., Beyer, T. A. & Ciaudo, C. Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Briefings in Bioinformatics* 20, 228–298 (2019).
- Pearce, S., Kippes, N., Chen, A., Debernardi, J. M. & Dubcovsky, J. RNA-seq studies using wheat *PHYTOCHROME B* and *PHYTOCHROME C* mutants reveal shared and specific functions in the regulation of flowering and shade-avoidance pathways. *BMC Plant Biology* 16, 141 (2016).
- Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 1–13 (2008).
- Shen, L. GeneOverlap: Test and visualize gene overlaps. Preprint at <http://shenlab-sinai.github.io/shenlab-sinai/> (2021).
- Digel, B., Pankin, A. & von Korff, M. Global transcriptome profiling of developing leaf and shoot apices reveals distinct genetic and environmental control of floral transition and inflorescence development in barley. *Plant Cell* 27, 2318–2334 (2015).
- Liu, H. *et al.* Transcriptome profiling reveals phase-specific gene expression in the developing barley inflorescence. *Crop Journal* 8, 71–86 (2020).

Chapter 5 – Conclusion

5.1 Summary

Breeding for insect resistance in has historically been reactionary and in response to emerged pests after they have become economically relevant. This dissertation provides new information about sorghum aphid resistance in *Sorghum bicolor* which has informed pre-breeding and breeding efforts which I hope will proactively reduce the likelihood of biotype shifts.

The reliance on *RMES1* as the primary source of host-plant resistance in breeding programs necessitated the need to determine if selection pressure was significant and what other source of HPR may exist or are in use. I showed *RMES1* is placing selection pressure on *M. sorghi* but not all aphid species and was likely being bolstered by *RMES2* in breeding programs. These findings will guide marker development for the intentional use and deployment of both sources to reduce the likelihood of a biotype shift as expected by antibiosis. To further characterize the threat of *RMES1*-resistance being overcome, we tested hypotheses on the molecular mechanism of *RMES1*. Genomic, transcriptomic, and metabolomic evidence elevated the NLR hypothesis over the HCN hypothesis which has implications for R-gene deployment and IPM strategies.