

ECOLOGY, MECHANISMS, AND STRUCTURE OF SELFISH GENETIC ELEMENTS IN
MAIZE

by

MEGHAN J. BRADY

(Under the Direction of R. Kelly Dawe)

ABSTRACT

Selfish genetic elements are sections of DNA that get transmitted to the next generation more than expected by chance, referred to as preferential transmission. Very large haplotype based selfish genetic elements can be referred to as chromosomal drive haplotypes. Maize has at least three well characterized chromosomal drive haplotypes (Abnormal chromosome 10, K10L2, B chromosome) and ample resources making it the ideal system to study them. The maize genome is littered with large heterochromatic tandem repeats called knobs that are classified by their repeat type: knob180 and TR-1. Ab10 is a large structural variant of normal chromosome 10, with several knobs of both classes, which exhibits preferential transmission in female meiosis. Ab10 encodes two proteins, KINDR and TRKIN, which pull knobs, including those on Ab10, towards the poles of the meiotic tetrad resulting in their overrepresentation in the egg cell. K10L2 is also a structural variant of normal chromosome 10, but only encodes TRKIN and exhibits much more subtle preferential transmission than Ab10. The B chromosome exhibits preferential transmission through the male. Here we assess the structure of Abnormal chromosome 10 (Ab10) and identify a classical maize marker gene, *striate leaves 2*, in a

previously unidentified structural variant on Ab10. Then we explored the function of TRKIN on Ab10, finding that it appears to be deleterious. Finally, we surveyed over 10,000 maize lines for Ab10, K10L2, and the B chromosome and explored their individual relationships to the genome and the abiotic environment. We found that Ab10 distribution is influenced by unlinked genetic modifiers while K10L2 and B chromosomes distribution is influenced by both genetic modifiers and the environment. The work we presented here greatly increases our understanding of selfish genetic element behavior as well as maize genome evolution.

INDEX WORDS: Selfish genetic elements, Abnormal Chromosome 10, B Chromosome, K10L2, Maize

ECOLOGY, MECHANISMS, AND STRUCTURE OF SELFISH GENETIC ELEMENTS IN
MAIZE

by

MEGHAN J. BRADY

BS, Gettysburg College, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2025

© 2025

Meghan J. Brady

All Rights Reserved

ECOLOGY, MECHANISMS, AND STRUCTURE OF SELFISH GENETIC ELEMENTS IN
MAIZE

by

MEGHAN J. BRADY

Major Professor:	R. Kelly Dawe
Committee:	Kelly Dyer
	Brad Nelms
	Douda Bensasson
	John Burke

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2025

DEDICATION

For those who believed in me when I did not believe in myself.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Kelly Dawe, for his kind mentorship. Under his tutelage I have grown as a scientist and as a person, for which I will always be grateful. I also thank the entire Dawe lab, particularly Jonathan Gent, Rebecca Piri, Yibing Zeng, and Mingyu Wang, who made my science better and my days enjoyable. I would also like to thank my committee, Kelly Dyer, Brad Nelms, Douda Bensasson, and John Burke. Their expertise and generosity with their time improved my science immeasurably.

Finally, I would like to thank my friends and family for supporting me over the last several years. Particularly, my parents Ann and Michael Brady who have always believed I was capable of anything. It has been a long journey, and I would not have made it without them.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW	1
The Genome as an Ecosystem	1
Selfish Genetic Elements - Mechanism	2
Selfish Genetic Elements - Structure	3
Selfish Genetic Elements - Synthetic.....	4
Maize as a Model System	4
Abnormal Chromosome 10	5
K10L2	6
Maize B Chromosome	7
Dissertation Summary	7
References	9

2	THE MAIZE STRIATE LEAVES 2 GENE ENCODES A CONSERVED DUF3732 DOMAIN AND IS HOMOLOGOUS TO THE RICE YSS1 GENE	12
	Abstract	13
	Introduction	13
	Results	15
	Discussion	20
	Methods	21
	Author Contributions	26
	References	27
3	ANTAGONISTIC KINESIN-14S WITHIN A SINGLE CHROMOSOMAL DRIVE HAPLOTYPE	38
	Abstract	39
	Introduction	39
	Results	43
	Discussion	54
	Methods	56
	Author Contributions	72
	References	73
4	GENETIC AND ENVIRONMENTAL INFLUENCES ON THE GEOGRAPHIC DISTRIBUTION OF THREE CHROMOSOMAL DRIVE HAPLOTYPES IN MAIZE	90
	Abstract	91

Introduction.....	91
Results.....	96
Discussion.....	103
Methods.....	105
Author Contributions	116
References.....	117
5 CONCLUSIONS.....	127
Future Directions	128
References.....	131
APPENDICES	
A SUPPLEMENTARY TABLES AND FIGURES – CHAPTER 2.....	132
B SUPPLEMENTARY TABLES AND FIGURES – CHAPTER 3.....	142
C SUPPLEMENTARY TABLES AND FIGURES – CHAPTER 4.....	166

LIST OF TABLES

	Page
Table 2.1: Name of genes involved in Ab10 duplication and locations	31
Table 2.2: Summary of leaf margin striping phenotypes in crosses with Mu alleles	32
Table 3.1: Ab10 <i>Trkin</i> Modeling	77
Table 3.2: Ab10 <i>Trkin</i> Model Parameters.....	78

LIST OF FIGURES

	Page
Figure 2.1: Orthologs between the N10 and Ab10 assembly	33
Figure 2.2: Duplicated region on Ab10	35
Figure 2.3: Differential expression of duplicated genes between sr2 and wildtype	36
Figure 2.4: Phenotypes of sr2 mutants.....	37
Figure 3.1: Diagram of Maize Chromosome 10 Haplotypes	79
Figure 3.2: Sequence Comparison of Trkin Bearing Region of Ab10 and K10L2	80
Figure 3.3: Comparison of trkin and Mutants.....	81
Figure 3.4: Comparison of Transposable Elements Composition Between All Trkin genes	83
Figure 3.5: Gene Ortholog Comparisons Among Chromosome 10 Haplotypes	84
Figure 3.6: TRKIN Immunofluorescence in Various trkin Genotype Male Meiocytes	85
Figure 3.7: FISH for Neocentromere Activity in Various trkin Genotype Male Meiocytes	86
Figure 3.8: Effect of Trkin on the meiotic drive of Ab10 when paired with K10L2.....	87
Figure 3.9: Ab10 Drive and Plant Fitness of Trkin in Ab10 Heterozygotes	88
Figure 3.10: How long can Ab10 Trkin(+) persist in a population being invaded by trkin(-)	89
Figure 4.1: Detection of CDHs in Genotype By Sequencing Data.....	122
Figure 4.2: Detection of CDHs in Experimental Samples	123
Figure 4.3: Relationship of CDHs to the Environment and Genetic Loci	125

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

The Genome as an Ecosystem:

Mendel's first law provided an essential framework that gave the nascent field of genetics guard rails that have facilitated its development. In the 1920s the first violations of this quintessential rule of genetics were characterized in *Drosophila obscura* (Gershenson 1928), and *Mus musculus* (Burt and Trivers 2008). In 1942 Rhoades found another such violation and called the responsible region abnormal chromosome 10 (Rhoades 1941), perfectly encapsulating the way that these rule breaking loci have been and continue to be thought of today. Mendel's first law defines normality, and all violations are oddities; quirks of genetics that are interesting toy systems but of minimal broad importance and appropriate to disregard in most contexts. As research has continued an ever-growing list of rule breaking loci, now broadly referred to as drive systems, have been discovered (Burt and Trivers 2008; Burga, Ben-David, and Kruglyak 2020; Saupe and Johannesson 2022; Torgasheva et al. 2019; Lampson and Black 2017; Finseth, Nelson, and Fishman 2021; Ahmad and Martins 2019; Dawe 2022; Herrick 1999; Lopez Hernandez and Zanders 2018; Zanders et al. 2014). The preponderance of drive systems, transcending taxonomic categories, gently challenges the dogma established by Mendel's first law. Drive systems, which have always been considered abnormal phenomena, are clearly not. They have all exploited similar vulnerabilities in the vertical transmission of DNA that exist in most inheritance systems. They are all playing a similar role in the genome, occupying similar

genomic niches if you will. Viewed through this lens, rule breaking drive systems are not aberrant toy systems that can be disregarded, but important players in a complex genomic ecosystem that are necessarily rare.

Mendel's first law set the stage for the genome to be considered a harmonious monolith uniformly and incrementally being selected to produce the most fit organism (Agren and Clark 2018). The story that has emerged since indicates that the genome is much more akin to an ecosystem with umpteen interacting players jockeying for a finite number of positions in the next generation. Selection acts based on both fitness at the genomic level and fitness at the organismal level. Rather than a harmonious monolith, each genome is a community of discrete players with a unique ecology. The analogy of selfish genetic elements to community ecology has been proposed and developed to various extents with regards to transposable elements (Brookfield 2005; Kidwell and Lisch 1997; Stitzer et al. 2021). This analogy is also useful in understanding drive systems.

Selfish Genetic Elements-Mechanism:

Selfish genetic elements (SGE) increase their own transmission to the next generation despite conferring no selectively advantageous fitness effects. SGEs have been implicated in evolutionary processes such as extinction, speciation, recombination, and genome size evolution (Agren and Clark 2018). Transposable elements are likely the most recognized SGE, accounting for large portions of eukaryotic genomes (Almojil et al. 2021). There is a huge amount of variety within transposable elements; with hundreds of families each having their own unique behaviors and mechanisms (Venner, Feschotte, and Biémont 2009). However, transposable elements represent only a small portion of the diversify within the broader class of SGEs (Werren 2011).

Examples include heritable organelles and microbes, the classical example being Wolbachia in arthropods (Kaur et al. 2021), as well as biased gene converters, such as homing endonucleases found in bacteria (Stoddard 2011). Additionally, there is a large class of SGE broadly referred to as drivers or drive systems. These typically create the appearance of biased transmission in meiosis though they do not always directly affect meiosis. Previous classifications have broken drive systems into meiotic drivers and post segregation distorters based on their mode of action. Meiotic drivers gain their advantage in meiosis specifically, while post segregation distorters gain their advantage outside of meiosis (Werren 2011). Classical examples of true meiotic drive are abnormal chromosome 10 (Ab10) in *Zea mays* (Dawe 2022) and centromere drive in monkey flower (Finseth, Nelson, and Fishman 2021). While one of the most frequently discussed examples of true post segregation distortion is the *wtf* locus in fission yeast (Lopez Hernandez and Zanders 2018). However given that it is often difficult to determine the mode of action, the broader term drive system is often more useful (James et al. 2023).

Selfish Genetic Elements- Structure:

In addition to grouping SGE's by mechanism SGEs can also be grouped by structure. Broadly, they can be grouped into mobile elements and haplotype-based elements. Mobile elements include transposable elements, but also toxin antidote drive systems (which can also be transposable) in fungi and nematode (Lai and Vogan 2023; Burga, Ben-David, and Kruglyak 2020). Haplotype based elements are large non-recombining regions often containing multiple elements necessary for drive. Examples include maize Ab10 (Dawe 2022), the maize B chromosome (Birchler and Yang 2021), *Drosophila* sex ratio distorters (Courret et al. 2019), and

the mouse t haplotype (Herrmann and Bauer 2012). The largest haplotype based drivers are also often referred to as chromosomal drive haplotypes.

Selfish Genetic Elements- Synthetic:

Synthetic SGEs are artificially created genetic elements that spread a desired trait through a population (Bier 2022). Their potential was recognized as early as the 1940s with the potential use of chromosomal translocations (Serebrovsky 1940). Since that time there has been large amounts of interest specifically in the control of pests such as mosquitos and mice (Bier 2022). A variety of methods have been targeted including endosymbionts, homing endonucleases, CRISPR/CAS9 based methods and many others (Wang et al. 2021). While synthetic genetic elements hold great promise, they also carry significant risks. Synthetic SGEs would need to be released into the environment and this would necessitate government regulation (James et al. 2023). Unfortunately, SGEs may not respect borders, making it important to demonstrate that the spread of any synthetic gene drive system could be predicted or controlled to avoid accidental international incidents. However, this modeling is complicated by a lack of research about the way that gene drives interact with their environment (Lindholm et al. 2016).

Maize as a Model System:

Maize is among the most important and productive food crops in the world. It is also an exceptional model system for genetic research broadly. Maize can be loosely broken into two groups: modern maize and maize landraces. Modern maize lines (inbreds) are improved, inbred, and used in the production of hybrid lines for commercial consumption (Anderson and Brown 1952). Maize landraces have a distinct identity and historical origin, but lack formal crop

improvement. They also are commonly genetically diverse, locally adapted, and associated with traditional farming (Villa et al. 2005). The economic importance and plethora of resources make maize an ideal system for genetic studies.

Maize has a moderately large genome at between 2.3 and 2.5 Gb that is comprised of 85% transposable elements (TE) (Schnable et al. 2009; Llaca, Campbell, and Deschamps 2011). The maize genome also contains large blocks of repetitive heterochromatic DNA referred to as knobs. Knobs fall into two categories, TR-1 or knob180, depending on the repeat element they are comprised of. TR-1 knobs are comprised of a 350 bp repeat while knob180 is comprised of a 180 bp repeat. These repeats are the primary driver behind the remarkable intraspecific genome size variation within maize (Grandbastien and Casacuberta 2012)

Abnormal Chromosome 10

Ab10 is an ~80 Mb addition to normal chromosome 10 consisting of three TR-1 knobs and one knob180 knob as well as unique euchromatic regions. The knob180 knob on the Ab10 haplotype is of unknown size, as it was not fully assembled (Liu et al. 2020). However, cytologically it is the largest knob in the maize genome (Liu et al. 2020; Dawe 2009). The Ab10 haplotype shares homology with the N10 haplotype in a 4.4MB and 8.3 Mb inversion (Liu et al. 2020). These inversions as well as the knobs prevent recombination between Ab10 and N10 keeping the haplotype together.

The Ab10 haplotype acts as a true meiotic driver by converting knobs throughout the genome into neocentromeres that rapidly move towards the spindle poles during meiosis I and II (Rhoades and Vilkomerson 1942). This occurs because the Ab10 haplotype encodes the kinesin driver protein (KINDR), a minus end directed motor. During meiosis KINDR associates with

knob180 repeats and drags them along microtubules towards the spindle poles ahead of centromeres. This mechanism results in the preferential transmission of knobs and Ab10 itself when heterozygous (Dawe et al. 2018). Additionally, Ab10 encodes a second minus end directed kinesin, TRKIN, which associates with TR-1 knobs and creates neocentromeres. While Ab10 is only present in a small portion of the population (Kanizay, Pyhajarvi, et al. 2013) it is the driving force behind the exceptional proliferation of knobs in the maize genome. While Ab10 is not generally present in commercially produced maize today, it is a vital part of the evolutionary story that produced it (Buckler et al. 1999).

K10L2

K10L2 is a structurally and functionally unique variant of chromosome 10 that expresses TRKIN during meiosis and induces neocentromere activation at TR-1 repeats (Kanizay, Albert, et al. 2013) (Figure 1). While K10L2 exhibits relatively weak meiotic drive (1-2%), this effect is reproducible across seasons and statistically significant (Kanizay, Albert, et al. 2013).

Furthermore, its presence has been documented in at least 12 distinct maize landrace populations, suggesting a potential role within the Ab10 system (Kanizay, Albert, et al. 2013). A drive rate of 1-2% should be sufficient for K10L2 to rapidly propagate within a population, provided it does not impose fitness costs (Hartl 1970). Additionally, K10L2 effectively competes against Ab10, nearly eliminating Ab10 drive when the two are paired (Kanizay, Albert, et al. 2013). It has been proposed that both K10L2's drive and its suppressive influence on Ab10 are facilitated by the TRKIN/TR-1 system (Swentowsky *et al.* 2020).

Maize B Chromosome

The maize B chromosome is a ~150MB supernumerary chromosome composed primarily of repetitive elements like TE's and a B specific repeat element (Marques, Klemme, and Houben 2018). The B chromosome is a chromosomal drive haplotype that accumulates via nondisjunction at the second pollen mitosis and through preferential fertilization of the egg by sperm containing the B chromosome (Roman 1947; Carlson 1978). Nondisjunction is independent of centromere function and is controlled by the heterochromatic region surrounding the centromere and a second region at the distal tip of the chromosome (Su et al. 2018). The B chromosome is most often found in fewer than 8 copies (Birchler and Yang 2021), though under laboratory conditions it has been accumulated up to 30 (Randolph 1941). It has been noted to have no fitness effects below 15 copies with increasingly deleterious fitness effects at higher copy numbers. However, B chromosome copy number has sometimes been found to have a positive relationship with altitude in populations with fewer than 15 copies (Rosato et al. 1998). This may reflect the fact that maize genomes are generally smaller at high altitudes (Bilinski et al. 2018), and B chromosomes are known to be more prevalent in smaller genomes (Rosato et al. 1998).

Dissertation Summary

The subsequent body of this dissertation explores the structure, mechanism, and ecology of selfish genetic elements in the maize genome. Chapter II contains a published manuscript detailing the discovery of a novel structural rearrangement on Ab10 and the subsequent discovery of a classical maize marker gene. We utilized comparative genomics and traditional genetic mutants to identify the *striate leaves 2* gene as homologous to the *yss1* gene in rice.

Chapter III contains a manuscript submitted for publication describing our efforts to determine the function of the *Trkin* gene on Ab10 and K10L2. We generated PacBio HiFi assemblies for both Ab10 and K10L2 and compared them to understand the evolution of *Trkin*. We then created *Trkin* mutants and conducted a broad range of drive and fitness assays. Finally, in Chapter IV we developed a novel method to detect large structural variants in genotype by sequencing data and screened >10,000 maize and teosinte accessions for Ab10, K10L2, and the B chromosome. We used these data to explore the relationship between all three SGEs and the environment.

References

- Agren, J. A., and A. G. Clark. 2018. "Selfish Genetic Elements." *PLoS Genetics* 14 (11): e1007700.
- Ahmad, S. F., and C. Martins. 2019. "The Modern View of B Chromosomes Under the Impact of High Scale Omics Analyses." *Cells* 8 (2). <https://doi.org/10.3390/cells8020156>.
- Almojil, Dareen, Yann Bourgeois, Marcin Falis, Imtiyaz Hariyani, Justin Wilcox, and Stéphane Boissinot. 2021. "The Structural, Functional and Evolutionary Impact of Transposable Elements in Eukaryotes." *Genes* 12 (6): 918.
- Anderson, Edgar, and William L. Brown. 1952. "Origin of Corn Belt Maize and Its Genetic Significance." In *Heterosis - A Record of Researches Directed Toward Explaining and Utilizing the Vigor of Hybrids*, edited by J. W. Gowen, 124–48. Ames, IA: Iowa State College Press.
- Bier, Ethan. 2022. "Gene Drives Gaining Speed." *Nature Reviews. Genetics* 23 (1): 5–22.
- Bilinski, P., P. S. Albert, J. J. Berg, J. A. Birchler, M. N. Grote, A. Lorant, J. Quezada, K. Swarts, J. Yang, and J. Ross-Ibarra. 2018. "Parallel Altitudinal Clines Reveal Trends in Adaptive Evolution of Genome Size in Zea Mays." *PLoS Genetics* 14 (5): e1007162.
- Birchler, J. A., and H. Yang. 2021. "The Supernumerary B Chromosome of Maize: Drive and Genomic Conflict." *Open Biology* 11 (11): 210197.
- Brookfield, J. F. 2005. "The Ecology of the Genome - Mobile DNA Elements and Their Hosts." *Nature Reviews. Genetics* 6 (2): 128–36.
- Buckler, E. S., T. L. Phelps-Durr, C. S. Keith-Buckler, R. K. Dawe, J. F. Doebley, and T. P. Holtsford. 1999. "Meiotic Drive of Chromosomal Knobs Reshaped the Maize Genome." *Genetics* 153: 415–26.
- Burga, A., E. Ben-David, and L. Kruglyak. 2020. "Toxin-Antidote Elements Across the Tree of Life." *Annual Review of Genetics* 54: 387–415.
- Burt, Austin, and Robert Trivers. 2008. *Genes in Conflict The Biology of Selfish Genetic Elements*. Cambridge, MA and London, England: Harvard University Press.
- Carlson, Wayne R. 1978. "The B Chromosome of Corn." *Annual Review of Genetics* 16: 5–23.
- Courret, C., C. H. Chang, K. H. Wei, C. Montchamp-Moreau, and A. M. Larracuente. 2019. "Meiotic Drive Mechanisms: Lessons from Drosophila." *Proceedings. Biological Sciences / The Royal Society* 286 (1913): 20191430.
- Dawe, R. K. 2009. *Handbook on Maize: Chapter 12 Maize Centromeres and Knobs*. Vol. 2. Springer Science+Business Media LLC.
- . 2022. "The Maize Abnormal Chromosome 10 Meiotic Drive Haplotype: A Review." *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*. <https://doi.org/10.1007/s10577-022-09693-6>.
- Dawe, R. K., E. G. Lowry, J. I. Gent, M. C. Stitzer, K. W. Swentowsky, D. M. Higgins, J. Ross-Ibarra, et al. 2018. "A Kinesin-14 Motor Activates Neocentromeres to Promote Meiotic Drive in Maize." *Cell* 173 (4): 839–850 e18.
- Finseth, F. R., T. C. Nelson, and L. Fishman. 2021. "Selfish Chromosomal Drive Shapes Recent Centromeric Histone Evolution in Monkeyflowers." *PLoS Genetics* 17 (4): e1009418.
- Gershenson, S. 1928. "A New Sex-Ratio Abnormality In Drosophila Obscura." *Genetics* 13: 488–507.

- Grandbastien, Marie-Angele, and Josep M. Casacuberta. 2012. "Plant Transposable Elements." In *Topics in Current Genetics*, edited by Stefan Hohmann, 24:41–59. Topics in Current Genetics. Springer.
- Hartl, D. L. 1970. "Analysis of a General Population Genetic Model of Meiotic Drive." *Evolution; International Journal of Organic Evolution* 24 (3): 538.
- Herrick, G. 1999. "Paternal Genome Elimination in Insects." *Results and Problems in Cell Differentiation* 25: 41–72.
- Herrmann, B. G., and H. Bauer. 2012. "The Mouse *t*-Haplotype: A Selfish Chromosome – Genetics, Molecular Mechanism, and Evolution." In *Evolution of the House Mouse*, edited by Miloš Macholán, Stuart J. E. Baird, Pavel Munclinger, and Jaroslav Piálek, 297–314. Cambridge University Press.
- James, Stephanie L., David A. O’Brochta, Filippo Randazzo, and Omar S. Akbari. 2023. "A Gene Drive Is a Gene Drive: The Debate over Lumping or Splitting Definitions." *Nature Communications* 14 (1): 1749.
- Kanizay, L. B., P. S. Albert, J. A. Birchler, and R. K. Dawe. 2013. "Intragenomic Conflict between the Two Major Knob Repeats of Maize." *Genetics* 194 (1): 81–89.
- Kanizay, L. B., T. Pyhajarvi, E. G. Lowry, M. B. Hufford, D. G. Peterson, J. Ross-Ibarra, and R. K. Dawe. 2013. "Diversity and Abundance of the Abnormal Chromosome 10 Meiotic Drive Complex in Zea Mays." *Heredity* 110 (6): 570–77.
- Kaur, Rupinder, J. Dylan Shropshire, Karissa L. Cross, Brittany Leigh, Alexander J. Mansueto, Victoria Stewart, Sarah R. Bordenstein, and Seth R. Bordenstein. 2021. "Living in the Endosymbiotic World of Wolbachia: A Centennial Review." *Cell Host & Microbe* 29 (6): 879–93.
- Kidwell, Margret G., and Damon Lisch. 1997. "Transposable Elements as Sources of Variation in Animals and Plants." *Proceedings of the National Academy of Sciences of the United States of America* 94: 7704–11.
- Lai, Eric C., and Aaron A. Vogan. 2023. "Proliferation and Dissemination of Killer Meiotic Drive Loci." *Current Opinion in Genetics & Development* 82 (102100): 102100.
- Lampson, M. A., and B. E. Black. 2017. "Cellular and Molecular Mechanisms of Centromere Drive." *Cold Spring Harbor Symposia on Quantitative Biology* 82: 249–57.
- Lindholm, A. K., K. A. Dyer, R. C. Firman, L. Fishman, W. Forstmeier, L. Holman, H. Johannesson, et al. 2016. "The Ecology and Evolutionary Dynamics of Meiotic Drive." *Trends in Ecology & Evolution* 31 (4): 315–26.
- Liu, J., A. S. Seetharam, K. Chougule, S. Ou, K. W. Swentowsky, J. I. Gent, V. Llaca, et al. 2020. "Gapless Assembly of Maize Chromosomes Using Long-Read Technologies." *Genome Biology* 21 (1): 121.
- Llaca, Victor, Matthew A. Campbell, and Stéphane Deschamps. 2011. "Genome Diversity in Maize." *Journal of Botany* 2011: 1–10.
- Lopez Hernandez, J. F., and S. E. Zanders. 2018. "Veni, Vidi, Vici: The Success of Wtf Meiotic Drivers in Fission Yeast." *Yeast* 35 (7): 447–53.
- Marques, A., S. Klemme, and A. Houben. 2018. "Evolution of Plant B Chromosome Enriched Sequences." *Genes* 9 (10). <https://doi.org/10.3390/genes9100515>.
- Randolph, L. F. 1941. "Genetic Characteristics of the B Chromosomes in Maize." *Genetics* 26: 608–31.
- Rhoades, M. M. 1941. "Preferential Segregation In Maize." *Genetics* 27: 395–407.

- Rhoades, M. M., and Hulda Vilkomerson. 1942. "On the Anaphase Movement of Chromosomes." *Proceedings of the National Academy of Sciences of the United States of America* 28 (10): 433–36.
- Roman, Herschel. 1947. "Directed Fertilization In Maize." *Proceedings of the National Academy of Sciences of the United States of America* 34: 36–42.
- Rosato, M., A. M. Chiavarino, C. A. Naranjo, J. C. Hernandez, and L. Poggio. 1998. "Genome Size and Numerical Polymorphism for the B Chromosome in Races of Maize (*Zea Mays* Ssp. *Mays*, Poaceae)." *American Journal of Botany* 85 (2): 168–74.
- Saupe, S. J., and H. Johannesson. 2022. "On the Mechanistic Basis of Killer Meiotic Drive in Fungi." *Annual Review of Microbiology* 76: 305–23.
- Schnable, Patrick S., Doreen Ware, Robert S. Fulton, Joshua C. Stein, Fusheng Wei, Shiran Pasternak, Chengzhi Liang, et al. 2009. "The B73 Maize Genome: Complexity, Diversity, and Dynamics." *Science* 326 (5956): 1112–15.
- Serebrovsky, A. S. 1940. "On the Possibility of a New Method for the Control of Insect Pests." *Zoologicheskii Zhurnal* 19: 618–30.
- Stitzer, M. C., S. N. Anderson, N. M. Springer, and J. Ross-Ibarra. 2021. "The Genomic Ecosystem of Transposable Elements in Maize." *PLoS Genetics* 17 (10): e1009768.
- Stoddard, Barry L. 2011. "Homing Endonucleases: From Microbial Genetic Invaders to Reagents for Targeted DNA Modification." *Structure (London, England: 1993)* 19 (1): 7–15.
- Su, Hangdong, Yalin Liu, James A. Birchler, and Fangpu Han. 2018. "The Behavior of the Maize B Chromosome and Centromere." *Genes* 9: 476.
- Swentowsky, K. W., J. I. Gent, E. G. Lowry, V. Schubert, X. Ran, K. F. Tseng, A. E. Harkess, W. Qiu, and R. K. Dawe. 2020. "Distinct Kinesin Motors Drive Two Types of Maize Neocentromeres." *Genes & Development* 34 (17–18): 1239–51.
- Torgasheva, A. A., L. P. Malinovskaya, K. S. Zadesenets, T. V. Karamysheva, E. A. Kizilova, E. A. Akberdina, I. E. Pristiyazhnyuk, et al. 2019. "Germline-Restricted Chromosome (GRC) Is Widespread among Songbirds." *Proceedings of the National Academy of Sciences of the United States of America* 116 (24): 11845–50.
- Venner, Samuel, Cédric Feschotte, and Christian Biémont. 2009. "Dynamics of Transposable Elements: Towards a Community Ecology of the Genome." *Trends in Genetics: TIG* 25 (7): 317–23.
- Villa, Tania Carolina Camacho, Nigel Maxted, Maria Scholten, and Brian For-Lloyd. 2005. "Defining and Identifying Crop Landraces." *Plant Genetic Resources: Characterization and Utilization* 3 (3): 373–84.
- Wang, Guan-Hong, Stephanie Gamez, Robyn R. Raban, John M. Marshall, Luke Alphey, Ming Li, Jason L. Rasgon, and Omar S. Akbari. 2021. "Combating Mosquito-Borne Diseases Using Genetic Control Technologies." *Nature Communications* 12 (1): 4388.
- Werren, John H. 2011. "Selfish Genetic Elements, Genetic Conflict, and Evolutionary Innovation." *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 2 (Suppl 2): 10863–70.
- Zanders, Sarah E., Michael T. Eickbush, Jonathan S. Yu, Ji-Won Kang, Kyle R. Fowler, Gerald R. Smith, and Harmit Singh Malik. 2014. "Genome Rearrangements and Pervasive Meiotic Drive Cause Hybrid Infertility in Fission Yeast." *ELife* 3 (June): e02630.

CHAPTER 2

THE MAIZE STRIATE LEAVES 2 GENE ENCODES A CONSERVED DUF3732 DOMAIN AND IS HOMOLOGOUS TO THE RICE YSS1 GENE

Brady, Meghan J., Maya Cheam, Jonathan I. Gent, and R. Kelly Dawe (2024) The Maize Striate Leaves2 (Sr2) Gene Encodes a Conserved DUF3732 Domain and Is Homologous to the Rice Yss1 Gene. *Plant Direct* **8**(2): e567.

Reprinted here with permission of the publisher.

Abstract

Maize *striate leaves2* (*sr2*) is a mutant that causes white stripes on leaves that has been used in mapping studies for decades, though the underlying gene has not been identified. The *sr2* locus has been previously mapped to small regions of the normal chromosome 10 (N10) and a rearranged variant called Abnormal Chromosome 10 (Ab10). A comparison of assembled genomes carrying N10 and Ab10 revealed only five candidate *sr2* genes. Analysis of a stock carrying the *sr2* reference allele (*sr2-ref*) showed that one of the five genes has a transposon insertion that disrupts its protein sequence and has a severe reduction in mRNA. An independent Mutator transposon insertion in the gene (*sr2-Mu*) failed to complement the *sr2-ref* mutation, and plants homozygous for *sr2-Mu* showed white striped leaf margins. The *sr2* gene encodes a DUF3732 protein with strong homology to a rice gene with a similar mutant phenotype called *young seedling stripe1* (*yss1*). These and other published data suggest that *sr2* may have a function in plastid gene expression.

Introduction

A large number of maize nuclear genes provide products necessary for chloroplast function. Mutations in these genes result in albino, virescent, pale green, yellow, or white striped plants. In the striped class there are at least eight different maize loci - *striate leaves1*, *striate leaves2*, *striate leaves3*, *striate leaves4*, *japonica striping1*, *japonica striping2*, *iojap striping1* and *iojap striping2* (Gerald Neuffer et al., 1997). These mutants impact the efficiency of plastid transcription, translation or morphogenesis such that chloroplast function is impaired but not abolished. Striping is likely caused by the sorting out of mixed populations of functional and non-functional chloroplasts in a way that some lineages inherit no functional chloroplasts and

appear as white sectors (Birky, 1983; Coe et al., 1988). White stripes tend to be wider and more common at the margins of leaves, because cells at the margin undergo more division to expand the width of the leaf than cells in the center of the leaf (Walbot and Coe, 1979; Han et al., 1992; Park et al., 2000).

The *iojap striping1 (ijl)* gene has been described at the molecular level (Rhoades, 1943; Han et al., 1992). Chloroplasts within white stripes of *ijl* mutants are present but lack ribosomes, suggesting a failure in ribosome assembly (Shumway and Weier, 1967; Walbot and Coe, 1979; Siemenroth et al., 1980). *Ioja*p is a member of a conserved family of ribosomal silencing factor A/DUF143-containing proteins (Häuser et al., 2012; Wang et al., 2016) that function in bacterial and chloroplast ribosome biogenesis and protein synthesis (Walbot and Coe, 1979; Trösch and Willmund, 2019). More recently a mutant called *white and green striate leaves1 (wgsll)* was described (Li et al., 2023), which may be an allele of *striate leaves4* on chromosome 6. The *wgsll* gene encodes a 16S rRNA processing protein that is thought to be required for ribosome maturation (Li et al., 2023).

Striate leaves2 (sr2) was originally identified in Waseca Minnesota as a spontaneous mutant (Joachim and Burnham, 1953). Electron microscope analysis of white tissue revealed chloroplasts with unorganized internal structure and a lack of visible ribosomes, similar to *ijl* (Williams and Kermicle, 1974). The *sr2* gene is located on the long arm of chromosome 10, both on the normal form of chromosome 10 (N10) and a variant of chromosome 10 known as abnormal chromosome 10 (Ab10) (Rhoades and Dempsey, 1985). Here we combine comparative genomics with genetic and molecular analyses of two alleles of the *sr2* gene to demonstrate that it encodes a DUF3732 protein with homology to rice *Young Seedling Stripe1* (Zhou et al., 2017), a gene that is thought to modulate gene expression by plastid-encoded plastid RNA polymerase.

Results

Comparative genomics of N10 and Ab10

We started our analysis with a careful comparison of two forms of chromosome 10, the normal chromosome 10 (N10) and abnormal chromosome 10 (Ab10). The two chromosomes are syntenic except at the ends of their long arms, where, on Ab10, there is a large ~55 Mb meiotic drive haplotype (Liu et al., 2020; Dawe, 2022). The Ab10 meiotic drive haplotype causes the preferential transmission of Ab10 when crossed as a female (Dawe, 2022). Portions of the end of normal chromosome 10 (N10) are present within the Ab10 haplotype, though the order of genes is altered by inversions and rearrangements (Rhoades and Dempsey, 1985). On N10, four genes with visible mutant phenotypes called *white2* (*w2*), *opaque7* (*o7*), *luteus13* (*l13*) and *striate leaves2* (*sr2*) occur in the order *w2-o7-l13-sr2*, whereas on Ab10 the gene order is *l13-o7-w2-sr2* (Rhoades and Dempsey, 1985). More comprehensive mapping demonstrated that there are two separate inversions but that *sr2* is not included in either one (Mroczek et al., 2006). Complete genome assembly of the Ab10 haplotype (Liu et al., 2020) and subsequent whole genome alignments confirmed the two inversions but did not identify an obvious uninverted region of homology at the ends of the shared region (Liu et al., 2020; Song et al., 2022), raising concerns about the original interpretations of gene order.

To confirm the location of *sr2*, we grew and analyzed a line homozygous for the Ab10 terminal deletion line known as Ab10-Df(K). Ab10-Df(K) had been described as having the *sr2* phenotype and a reduced stature, but otherwise appearing normal, suggesting that only a small section of the Ab10 shared region, including *sr2*, was missing (Rhoades and Dempsey, 1985). We grew homozygous Ab10-Df(K) plants and confirmed that they are striated (see below). In addition, we Illumina sequenced the genomes of plants containing Ab10-Df(K) and two other

deletion chromosomes (Ab10-Df(L) and Ab10-Df(M) (Hiatt and Dawe, 2003b)) that do not have the *sr2* phenotype, and aligned the short read data to the Ab10 reference (Liu et al., 2020). The results showed that nearly all of the shared region is present in Ab10-Df(K), suggesting that *sr2* must lie at the end of the shared region of both haplotypes, presumably within a very small region or rearrangement (Figure 2.1, Figure S2.1) (Rhoades and Dempsey, 1985)).

Identification of a duplicated and inverted region in Ab10

We used OrthoFinder and BLAST to detect ortholog gene pairs within the *sr2* candidate region on both N10 (distal to *o7*, (Wang et al., 2011)) and Ab10 (between the Ab10-Df(K) and Ab10-Df(M) breakpoints). This approach revealed six homologous gene pairs (Figure 2.1, Table S2.1). Surprisingly, we found that this region is also present within the larger inversion proximal to the Ab10-Df(K) breakpoint on Ab10. The results suggest that a segment carrying the six genes was duplicated and inserted distal to the Ab10-Df(K) breakpoint (Figure 2.2, Table 2.1). The fact that it is a duplication helps to explain why it was not visible in alignments that display one-to-one homology relationships (Liu et al., 2020; Song et al., 2022). The OrthoFinder output and our own analysis of the structure and transcripts of each gene indicate that only 7 of the total 12 genes are functional (Table 2.1). In the terminal duplicated segment, there are five *sr2* candidate genes, which we will refer to in the next two sections as genes 1, 2, 4, 5, and 6. The reference names for these genes can be found in Table 2.1.

RNA-seq of plants homozygous for the sr2-ref reference allele implicates gene 1 as the most likely sr2 candidate

Given that the *sr2-ref* mutation has a similar phenotype to Ab10-Df(K), it seemed possible that the *sr2-ref* allele may also be associated with reduced expression. We performed a differential expression analysis using mRNA from leaf tissue. Because the genetic background of the *sr2-ref* allele is not known, we used the W22 inbred as the negative control. We aligned the *sr2-ref* and W22 RNA-seq data to the B73 v5 reference genome and performed differential expression analysis. Of the six duplicated genes, only gene 1 and gene 2 showed significant differential expression between *sr2-ref* and W22 (Figure 2.3). Expression of gene 2 was 44% higher in *sr2-ref* than in W22, while the expression of gene 1 was dramatically reduced to only 2% of the levels observed in the W22 inbred (Figure S2.2).

De novo assembly of the (relatively few) gene 1 (Zm00001eb434490) transcripts in *sr2-ref* suggest that there is an insertion that often creates a chimeric transcript that omits the first exon (Figure S2.3). The insertion itself appears to be a chimera of the second through fourth exons of the CASP-like protein 4A2 (Zm00001eb231550) and a DNA transposon (Zm10271_AC186904_1). The transposon Zm10271_AC186904_1 is annotated as a Robertson's mutator (Mu) element, but we can find no homology to Mu in the terminal inverted repeat sequence (TIR). If the chimeric transcript from *sr2-ref* is translated, only a portion of the encoded protein would be homologous to the wild type gene 1 protein. There is also a very small amount of full length gene 1 transcript in *sr2-ref*, but it is >99% reduced relative to W22 (Figure S2.3).

To confirm the presence of the insertion in the gene1 (Zm00001eb434490), we designed three pairs of primers based on transcript isoforms, each with a forward primer in the insertion

and a reverse primer in the gene (Table S2.2). All three primer pairs produced amplicons from *sr2-ref* DNA, but not from W22 control. We then Sanger sequenced the shortest amplicon (794 bp, forward primer matching the part of the insertion homologous to CASP like 4A2). Its sequence included not only CASP like 4A2 sequence, but also the 3' 309-bp TIR of Zm10271_AC186904_1 that had been spliced out of all detected transcripts (the 5' TIR is present in some transcripts, see Figure S2.3, isoform 5). It also revealed its precise insertion point in the first intron of gene 1 (Figure S2.4). These results suggest that the reduced expression in *sr2-ref* is because a DNA transposon carrying a truncated piece of CASP like 4A2 inserted into its first intron and disrupted both transcription and splicing.

Complementation tests using transposon-induced alleles confirm that sr2 is gene 1

To further test which of the candidate genes were *sr2*, we carried out complementation tests with Mu insertions for genes 1, 2, 4, and 5 from the UniformMu collection ((Settles et al., 2007), there are no mutants for gene 6). Each of the mutant alleles contained a Robertson's mutator element within the first exon (Table S2.3). Unfortunately the genetic background of these lines is not ideal for testing recessive alleles of *sr2*. All UniformMu lines carry an allele of the *colored1* gene known as *R1-r:standard* (Settles et al., 2007; McCarty et al., 2013). *R1-r:standard* is tightly linked to a dominant allele of *inhibitor of striate1* (*Isr1*) that inhibits the *sr2* mutant phenotype (Kermicle and Axtell, 1981; Park et al., 2000). Although it is theoretically possible that *Isr1* was recombined from *R1-r:standard* during the preparation of the UniformMu lines (McCarty et al., 2005), it seems highly unlikely, as *Isr1* (Zm00001eb429350 on chr10:141210513-141213010) is only ~20 kb from the P component of the complex *r1* locus (Zm00001eb429330 on chr10:141187279-141196584) (Walker et al., 1995; Park et al., 2000).

One copy of *Isr1* reduces the striping in a homozygous *sr2* line, while two copies nearly eliminates the *sr2* phenotype. In plants homozygous for both *Isr1* and *sr2*, leaves are thinner, and white stripes are only observed at the edges of leaf margins (Park et al., 2000).

To generate material for the complementation tests we crossed lines carrying Mu alleles of genes 1, 2, 4, and 5 to *sr2-ref*. The expectation was that if one of the genes is the *sr2* gene, the Mu allele/*sr2-ref* heterozygote for that gene would show a striated phenotype. We also self-crossed the UniformMu lines to obtain plants that were homozygous for Mu alleles of each of the four genes. Phenotypic analyses of the progeny of these crosses, along with positive and negative controls, were carried out in both the greenhouse and field. Homozygous *sr2-ref* and Ab10-Df(K) had white stripes on leaf sheaths and blades in both environments, with *sr2-ref* being more striped in the field (Figure S2.5).

The complementation data indicate that gene 1 (Zm00001eb434490) is the *sr2* gene. We observed leaf margin striping in the complementation tests for gene 1 and in plants homozygous for the gene 1 Mu allele (called *sr2-Mu* here forward), but not in lines carrying Mu alleles of the other three genes (Figure 2.4, Table 2.2). The *sr2-Mu* homozygous plants grew poorly and had thin leaves (Figure 2.4F), particularly in the field where they died before striping is normally visible. However, when they grew to maturity in the greenhouse, *sr2-Mu* homozygous plants consistently had white stripes on the edges of sheath margins (Table 2.2, Figure 2.4).

The *sr2* gene is annotated as a “BTB/POZ domain protein TNFAIP protein.” However *sr2* has no homology to either the BTB/POZ domain or TNFAIP1. Rather, *sr2* is a DUF3732 domain protein, with strong (86%) protein homology to the rice *Young Seedling Stripe1* (*yss1*) gene (LOC_Os04g59570, (Zhou et al., 2017)). The fact that *sr2* and *yss1* have similar structure

and function is consistent with the fact that *sr2* is a core gene in maize, found in all 26 NAM founder inbreds and is highly expressed in leaf tip tissue (Hufford et al., 2021) (Figure S2.2).

Discussion

Striate leaves2 (*sr2*) is a morphological marker that has been known since the 1940s and has played an important role in understanding the structure of abnormal chromosome 10. Comparative genomics and analysis of two independent alleles of *sr2* indicate that the causal gene is a DUF3732 domain-containing gene that is homologous to rice *yss1*. The phenotypes of *sr2* and *yss1* are similar, except that *yss1* stripes are only present in early leaves, which is not the case in *sr2*.

The rice YSS1 protein is localized to nucleoids (chloroplast genomes) and displays reduced expression of genes that are transcribed by plastid-encoded RNA polymerase (Zhou et al., 2017). In Arabidopsis, the homolog of *sr2* and *yss1* is AT4G33480. Biochemical data show that the AT4G33480 protein physically associates with psbA mRNA, which encodes a component of the Photosystem II reaction center in chloroplasts (McDermott et al., 2019). These data are consistent with the interpretation that SR2/YSS1/AT4G33480 functions at the level of chloroplast gene expression (Zhou et al., 2017). Our data further show that SR2 is unlikely to be absolutely required for chloroplast function (at least in maize), because homozygous Ab10-Df(K) plants lack the *sr2* gene yet still grow to maturity. Similar to *ij1*, *sr2* appears to promote faster and/or more accurate chloroplast biogenesis (Han et al., 1992).

The *sr2* phenotype is highly variable. Homozygous *sr2-ref* plants grown in the same environment can vary from having very few to very many stripes (Figure S2.5). The degree of striping is also environmentally sensitive, with full sibling *sr2-ref* plants grown in the field

having more severe leaf margin striping than those grown concurrently in the greenhouse (Figure S2.5). The powerful effects of *Isr1* in reducing the *sr2* phenotype illustrates that there are also genetic modifiers of *sr2* (Park et al., 2000). The *Isr1* gene suppresses the growth of white tissue in *sr2* and other white striped mutants such as *ij1*. Our observation that plants homozygous for *sr2-Mu* tend to have thin leaves and few stripes can be at least partially explained by the presence of *Isr1* in the UniformMu background (Figure 2.4E, F). The heavy reliance on the UniformMu resource in recent years may have inadvertently limited the identification of the multiple other *striated leaves*, *iojap striping* and *japonica striping* loci that have yet to be described at the molecular level (Gerald Neuffer et al., 1997).

Methods

Plant material and growth

Ab10-Df(K) was obtained from Marcus Rhoades and backcrossed to the W23 inbred (Hiatt and Dawe, 2003a). Ab10-Df(L) and Ab10-Df(M) were identified in our lab (Hiatt and Dawe, 2003b). Seeds carrying *sr2-ref* (X16D) and all Mu alleles were obtained from the Maize Genetics Cooperation Stock Center in Urbana, IL (Table S2.3). Most experiments were conducted in the UGA Botany greenhouses in Athens GA. We also grew plants in an outdoor field site in Athens GA in April-June 2023.

Sequencing deletion lines

We collected young leaf tissue from single plants homozygous for Ab10-Df(K), Ab10-Df(L), and Ab10-Df(M) and extracted DNA with the IBI Genomic DNA Mini Kit (Plant)

(IB47230) using the GP1 buffer and 16,000xg for all centrifuge steps. We used a Kapa HyperPrep Kit to prepare the sequencing library (KK8580) and adapters from Netflex DNA Barcodes (Nova-520996). Sequencing was performed by GENEWIZ from Azenta using a HiSeq 4000. We trimmed the resulting reads using cutadapt version 2.7 (Martin, 2011), mapped the reads to the Zm-B73_AB10-REFERENCE-NAM-1.0 reference (Liu et al., 2020) using BWAmem 0.7.17 (Li and Durbin, 2009), and removed duplicates using picard version 2.16.0 (Tools, 2018). We determined the breakpoints by filtering the primary alignments with a MAPQ $\geq$ 20, counting the number of reads over each bp using IGV tools, and visually inspecting the resulting files (Robinson et al., 2011). N10 reads shown in Figure 2.1 and Figure S2.1 are B73 ~30X Illumina short reads from the NAM project (Hufford et al., 2021). Plots were made using R v4.3.1.

Identifying and analyzing orthologs in the duplicated region of Ab10

We selected the protein sequence for the longest isoforms of the annotated genes between the *colored1* (*r1*) gene (Zm00001eb429330, B73 v5 annotation) and the ends of the long arm of chromosome 10 in the the Zm-B73-REFERENCE-NAM-5.0 (<https://www.maizegdb.org/genome/assembly/Zm-B73-REFERENCE-NAM-5.0>) and Zm-B73_AB10-REFERENCE-NAM-1.0 (https://download.maizegdb.org/Zm-B73_AB10-REFERENCE-NAM-1.0/) assemblies (Liu et al., 2020; Hufford et al., 2021). These files were used to run orthofinder v2.5.2 using default parameters (Emms and Kelly, 2019). Orthofinder identified five genes that were duplicated at the end of the shared region. BLAST v 2.2.26 (Camacho et al., 2023) was used to identify a sixth gene within the duplication. Plots were made using R v4.3.1.

We analyzed the transcripts for all six genes in the Ab10 inverted, Ab10 duplicated and N10 regions using the multiple sequence aligner and Clustal Omega in Geneious Prime v2022.0.2 (<https://www.geneious.com/>). We used the coding sequence (CDS) data published with the B73 and B73-Ab10 reference genomes (Liu et al., 2020; Hufford et al., 2021). For gene 1 within the inversion, all transcript isoforms were truncated due to an unknown insertion in exon 4 which caused a frameshift and premature stop codon in exon 5. For gene 2 within the inversion, we found that all transcript isoforms either lacked significant homology to the N10 gene or were truncated for unknown reasons. For gene 3, we used Clustal Omega to determine that the duplication homologs are missing 288bp of coding sequence and are unlikely to be functional. For gene 4, we used Clustal Omega to determine that both the Ab10 inverted and duplicated copies are similar to the N10 homolog and are likely both functional. For gene 5, we used Clustal Omega to determine that the Ab10 inversion homolog has one isoform that is severely truncated for an unknown reason and one that results in the insertion of a proline within the PaaI thioesterase domain (Marchler-Bauer et al., 2017). This may or may not disrupt function. For gene 6, we used Clustal Omega to determine that both Ab10 homologs are longer than the N10 counterpart and that both the Ab10 inversion and duplication homologs have a similar amount of homology to the N10 copy. This may or may not disrupt function in the Ab10 homologs. These data in conjunction with the results from OrthoFinder indicate that genes 1, 2, 4, 5, and 6 are candidates for *sr2* (Table 2.1).

RNA-seq of plants homozygous for sr2-ref

RNA isolation and Sequencing

We collected mature leaf tissue from homozygous *sr2-ref* and W22 individuals grown side by side in the UGA Botany greenhouse and immediately froze it in liquid nitrogen. RNA was extracted using the IBI Total RNA Mini Kit (Plant) (IB47340) and cDNA was prepared using BioRad iSCRIPT Transcription Supermix for RT-PCR (1708891). GENEWIZ from Azenta performed a library preparation with PolyA selection using an NEBNext Ultra II RNA Library Prep Kit followed by paired end 150bp illumina sequencing on a NovaSeq 6000.

Differential Expression Analysis

We trimmed reads using Trimmomatic version 0.39, and checked quality before and after using FastQC version 0.11.9 (Bioinformatics, 2014; Bolger et al., 2014). We aligned reads to the Zm-B73-REFERENCE-NAM-5.0 reference using HISAT2 version 2.1.0 using default parameters (Kim et al., 2019; Hufford et al., 2021). Using HTSeq version 0.13.5 with default parameters, we determined the number of reads mapped to each annotated feature (Anders et al., 2015). All genes with fewer than 10 reads were removed. We used DESEQ2 version 1.38.3 to perform a differential expression analysis with default parameters. Plots were made using R v4.3.1.

De novo assembly of sr2-ref transcripts

We performed a Trinity v2.10.0 de novo transcriptome assembly on pooled data for all three biological replicates of *sr2-ref* (Haas et al., 2013). BLASTv 2.2.26 was used to identify isoforms with homology to the *sr2* gene. Some of the isoforms also contained sequences with

homology to CASP-like protein 4A2 (B73 v5 annotation Zm00001eb231550) and L-aspartate oxidase (B73 v5 annotation Zm00001eb231540). The relative abundance of each isoform was determined using Kallisto within Trinity v2.8.4 (Haas et al., 2013; Bray et al., 2016). All transcript assemblies were then aligned to the B73 *sr2* reference gene (Zm00001eb434490) using Geneious Prime v2022.0.2 (<https://www.geneious.com/>) MiniMap2 with default parameters followed by minimal manual curation.

The *sr2-ref* allele insertion was confirmed via PCR using primers to CASP like 4A2 and *sr2* (Table S2.2). The shortest PCR product was Sanger sequenced by Eton Biosciences .

Genotyping

All genotyping DNA extractions were performed using a CTAB protocol (Clarke, 2009). Polymerase chain reactions were performed using Promega GoTaq Green Master Mix (M7123). The wild type locus for each gene was detected using gene specific primers (Table S2.2). The Mu allele for each gene was detected using a primer to the Mu terminal inverted repeat (5'-GCCTCYATTTCGTCGAATCCS-3') and either a forward or reverse gene specific primer. All genotyping reactions used the following temperature profile: Hold 95°C, 2min 95°C, 30(30sec 95°C, 30sec 60°C, 45sec 72°C), 5 min 72°C.

Complementation tests using Mu-induced alleles

For genes 2, 4, and 5 we produced seed that was homozygous for the Mu insertion of interest and crossed these to *sr2-ref* to produce seed that was all heterozygous for the insertion and *sr2-ref*. For gene 1, we generated segregating populations of mutant and wild type by self crossing a plant heterozygous for the Mu insertion and crossing this plant to the *sr2-ref* tester.

Plants were grown in both the greenhouse and field as detailed in Table 2.2. In the field we randomized plant location and surrounded experimental plants with buffer corn to limit any environmental effects. In all cases we grew *sr2-ref* and Ab10-Df(K) plants alongside the Mu bearing plants to confirm that the conditions were appropriate to see the phenotype.

Author Contributions

MJB and RKD designed the research. MJB, JIG, and MC performed research; MJB analyzed data. MJB and RKD wrote the paper. All authors agreed that this work may be included in this dissertation.

References

- Anders S, Pyl PT, Huber W** (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169
- Bioinformatics B** (2014) FastQC. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>,
- Birky CW Jr** (1983) Relaxed cellular controls and organelle heredity. *Science* **222**: 468–475
- Bolger AM, Lohse M, Usadel B** (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120
- Bray NL, Pimentel H, Melsted P, Pachter L** (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–527
- Camacho C, Boratyn GM, Joukov V, Vera Alvarez R, Madden TL** (2023) ElasticBLAST: accelerating sequence search via cloud computing. *BMC Bioinformatics* **24**: 117
- Clarke JD** (2009) Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harb Protoc* **2009**: db.prot5177
- Coe EH Jr, Thompson D, Walbot V** (1988) PHENOTYPES MEDIATED BY THE IOJAP GENOTYPE IN MAIZE. *Am J Bot* **75**: 634–644
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al** (2021) Twelve years of SAMtools and BCFtools. *Gigascience*. doi: 10.1093/gigascience/giab008
- Dawe RK** (2022) The maize abnormal chromosome 10 meiotic drive haplotype: a review. *Chromosome Res* **30**: 205–216
- Dawe RK, Lowry EG, Gent JI, Stitzer MC, Swentowsky KW, Higgins DM, Ross-Ibarra J, Wallace JG, Kanizay LB, Alabady M, et al** (2018) A Kinesin-14 Motor Activates Neocentromeres to Promote Meiotic Drive in Maize. *Cell* **173**: 839–850 e18
- Emms DM, Kelly S** (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**: 238
- Gerald Neuffer M, Coe EH, Wessler SR** (1997) *Mutants of Maize*. Cold Spring Harbor Laboratory Press
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al** (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**: 1494–1512
- Han CD, Coe EH Jr, Martienssen RA** (1992) Molecular cloning and characterization of iojap (ij), a pattern stripping gene of maize. *EMBO J* **11**: 4037–4046
- Häuser R, Pech M, Kijek J, Yamamoto H, Titz B, Naeve F, Tovchigrechko A, Yamamoto K,**

- Szaflarski W, Takeuchi N, et al** (2012) RsfA (YbeB) proteins are conserved ribosomal silencing factors. *PLoS Genet* **8**: e1002815
- Hiatt EN, Dawe RK** (2003a) Four Loci on Abnormal Chromosome 10 Contribute to Meiotic Drive in Maize. *Genetics* **164**: 699–709
- Hiatt EN, Dawe RK** (2003b) The meiotic drive system on maize abnormal chromosome 10 contains few essential genes. *Genetica* **117**: 67–76
- Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y, et al** (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**: 655–662
- Joachim G, Burnham CR** (1953) Inheritance of Waseca stripe. *Maize Genetics Cooperation Newsletter* **27**: 66
- Kermicle JL, Axtell JD** (1981) Modification of chlorophyll striping by the R region. *Maydica* **26**: 185–197
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL** (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**: 907–915
- Li H, Durbin R** (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760
- Li Q, Du J, Qiao Z, Pan C, He W, Zhang L, Li X, Nie Y, Li X, Pan G, et al** (2023) White and green striate leaves 1, predicted to encode a 16S rRNA processing protein, plays a critical role in the processing of chloroplast ribosomes in maize (*Zea mays* L.). *Mol Breed* **43**: 65
- Liu J, Seetharam AS, Chougule K, Ou S, Swentowsky KW, Gent JI, Llaca V, Woodhouse MR, Manchanda N, Presting GG, et al** (2020) Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol* **21**: 121
- Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al** (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* **45**: D200–D203
- Martin M** (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12
- McCarty DR, Settles AM, Suzuki M, Tan BC, Latshaw S, Porch T, Robin K, Baier J, Avigne W, Lai J, et al** (2005) Steady-state transposon mutagenesis in inbred maize. *Plant J* **44**: 52–61
- McCarty DR, Suzuki M, Hunter C, Collins J, Avigne WT, Koch KE** (2013) Genetic and molecular analyses of UniformMu transposon insertion lines. *Methods Mol Biol* **1057**: 157–166
- McDermott JJ, Watkins KP, Williams-Carrier R, Barkan A** (2019) Ribonucleoprotein

- Capture by in Vivo Expression of a Designer Pentatricopeptide Repeat Protein in Arabidopsis. *Plant Cell* **31**: 1723–1733
- Mroczek RJ, Melo JR, Luce AC, Hiatt EN, Dawe RK** (2006) The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. *Genetics* **174**: 145–154
- Park SH, Chin HG, Cho MJ, Martienssen RA, Han C d.** (2000) Inhibitor of striate conditionally suppresses cell proliferation in variegated maize. *Genes Dev* **14**: 1005–1016
- Rhoades MM** (1943) Genic Induction of an Inherited Cytoplasmic Difference. *Proc Natl Acad Sci U S A* **29**: 327–329
- Rhoades MM, Dempsey E** (1985) Structural heterogeneity of chromosome 10 in races of maize and teosinte. In M Freeling, ed, *Plant Genetics*. Alan R. Liss, New York, pp 1–18
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP** (2011) Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26
- Settles AM, Holding DR, Tan BC, Latshaw SP, Liu J, Suzuki M, Li L, O'Brien BA, Fajardo DS, Wroclawska E, et al** (2007) Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genomics* **8**: 116
- Shumway LK, Weier TE** (1967) The chloroplast structure of iojap maize. *Am J Bot* **54**: 773–780
- Siemenroth A, Börner T, Metzger U** (1980) Biochemical studies on the iojap mutant of maize. *Plant Physiol* **65**: 1108–1110
- Song B, Marco-Sola S, Moreto M, Johnson L, Buckler ES, Stitzer MC** (2022) AnchorWave: Sensitive alignment of genomes with high sequence diversity, extensive structural polymorphism, and whole-genome duplication. *Proc Natl Acad Sci U S A*. doi: 10.1073/pnas.2113075119
- Swentowsky KW, Gent JI, Lowry EG, Schubert V, Ran X, Tseng KF, Harkess AE, Qiu W, Dawe RK** (2020) Distinct kinesin motors drive two types of maize neocentromeres. *Genes Dev* **34**: 1239–1251
- Tools P** (2018) Picard. <http://broadinstitute.github.io/picard>,
- Trösch R, Willmund F** (2019) The conserved theme of ribosome hibernation: from bacteria to chloroplasts of plants. *Biol Chem* **400**: 879–893
- Udy DB, Belcher S, Williams-Carrier R, Gualberto JM, Barkan A** (2012) Effects of reduced chloroplast gene copy number on chloroplast gene expression in maize. *Plant Physiol* **160**: 1420–1431
- Walbot V, Coe EH** (1979) Nuclear gene *iojap* conditions a programmed change to ribosome-less plastids in *Zea mays*. *Proc Natl Acad Sci U S A* **76**: 2760–2764

- Walker EL, Robbins TP, Bureau, T. E., Kermicle J, Dellaporta SL** (1995) Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. *EMBO J* **14**: 2350–2363
- Wang G, Sun X, Wang G, Wang F, Gao Q, Sun X, Tang Y, Chang C, Lai J, Zhu L, et al** (2011) Opaque7 encodes an acyl-activating enzyme-like protein that affects storage protein synthesis in maize endosperm. *Genetics* **189**: 1281–1295
- Wang M, Jiang L, Da Q, Liu J, Feng D, Wang J, Wang H-B, Jin H-L** (2016) DELAYED GREENING 238, a Nuclear-Encoded Chloroplast Nucleoid Protein, Is Involved in the Regulation of Early Chloroplast Development and Plastid Gene Expression in *Arabidopsis thaliana*. *Plant Cell Physiol* **57**: 2586–2599
- Williams E, Kermicle JL** (1974) Fine structure of plastids in maize leaves carrying the striate-2 gene. *Protoplasma* **79**: 401–408
- Zhou K, Ren Y, Zhou F, Wang Y, Zhang L, Lyu J, Wang Y, Zhao S, Ma W, Zhang H, et al** (2017) Young Seedling Stripe1 encodes a chloroplast nucleoid-associated protein required for chloroplast development in rice seedlings. *Planta* **245**: 45–60

Table 2.1. Names of genes involved in Ab10 duplication in all locations.

Gene Number	B73 N10 Name	Ab10 Inversion Name ¹	Ab10 Duplication Name ²	Ortholog on Ab10
1	Zm00001eb434490	Zm00043a049649	Zm00043a050219	Duplication Only
2	Zm00001eb434500	Zm00043a049648	Zm00043a050220	Duplication Only
3	Zm00001eb434510	Zm00043a049638	Zm00043a050221 + Zm00043a050222	Inversion Only
4	Zm00001eb434520	Zm00043a049637	Zm00043a050224	Both Duplication and Inversion
5	Zm00001eb434530	Zm00043a049636	Zm00043a050225	Duplication Only
6	Zm00001eb434540	Zm00043a049635	Zm00043a050226	Duplication Only

¹ Inversion refers to genes annotated in their expected location in the known Ab10 inversion

² Duplication refers to genes annotated outside of the known Ab10 inversion

Table 2.2. Summary of leaf margin striping phenotypes in crosses with Mu alleles^{1,2}.

	Greenhouse Fall 2023					Field Summer 2023					Greenhouse Summer 2023					Greenhouse Spring 2023					Total				
Gene Number	<u>Mu</u> sr2	<u>+</u> sr2	<u>Mu</u> Mu	<u>+</u> Mu	<u>+</u> +	<u>Mu</u> sr2	<u>+</u> sr2	<u>Mu</u> Mu	<u>+</u> Mu	<u>+</u> +	<u>Mu</u> sr2	<u>+</u> sr2	<u>Mu</u> Mu	<u>+</u> Mu	<u>+</u> +	<u>Mu</u> sr2	<u>+</u> sr2	<u>Mu</u> Mu	<u>+</u> Mu	<u>+</u> +	<u>Mu</u> sr2	<u>+</u> sr2	<u>Mu</u> Mu	<u>+</u> Mu	<u>+</u> +
1	6	6	10 ³	4	6	5	4	2-Died	9	7	4	0	2	0	0	0	0	2	6	2	15	10	15	19	15
2	0	0	0	0	0	4	0	4	0	0	3	0	2	0	0	0	0	8	6	2	7	0	14	6	2
4	0	0	0	0	0	4	0	3	0	0	3	0	3	0	0	0	0	0	0	0	7	0	6	0	0
5	0	0	0	0	0	4	0	3	0	0	3	0	3	0	0	0	0	0	0	0	7	0	6	0	0

¹ Numbers indicate the number of plants that were scored.

² Grey shaded and bolded cells indicate that all plants had leaf margin stripes. White cells indicate that no plants had leaf margin stripes.

³ One additional plant did not show leaf margin striping. It suffered an infection that stunted its growth and prevented it from flowering. Lack of the phenotype may be a result of the combination of *Isr1*, stunted growth, and lack of flowering or may reflect incomplete penetrance.

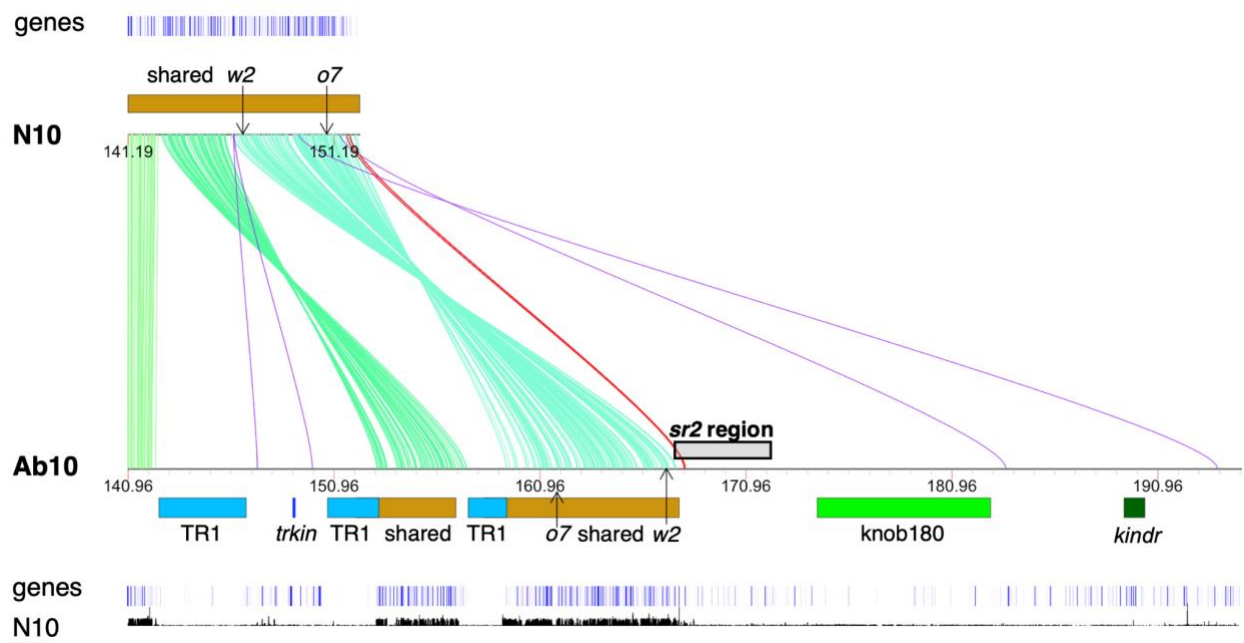


Figure 2.1. Orthologs between the N10 and Ab10 assembly. Lines indicate orthologs between N10 and Ab10 determined by OrthoFinder (Emms and Kelly, 2019). Shades of green indicate expected orthologs, red and purple indicate unexpected orthologs that could or could not be *sr2*, respectively. Location of genetic markers with known physical position, *o7* and *w2*, are shown (Wang et al., 2011; Udy et al., 2012). TR1 (light blue) and knob180 (bright green) are maize knob types. *Trkin* (dark blue) and *kindr* (dark green) are kinesin proteins responsible for the preferential transmission of Ab10 (Dawe et al., 2018; Swentowsky et al., 2020). Shared (brown) indicates the regions of known homology between Ab10 and N10. The *sr2* region is where the *sr2* gene has been mapped on Ab10, as defined by the breakpoints of Ab10-Df(K) and Ab10-Df(M) (Hiatt and Dawe, 2003b) (Figure S2.1). Annotated genes are indicated as blue vertical bars (Liu et al., 2020; Hufford et al., 2021). The plot at the bottom shows short reads from B73 (which has N10) mapped to the Ab10 reference assembly with a mapping quality greater than or equal to 20 (Hufford et al., 2021). This alignment shows the traditionally defined shared region.

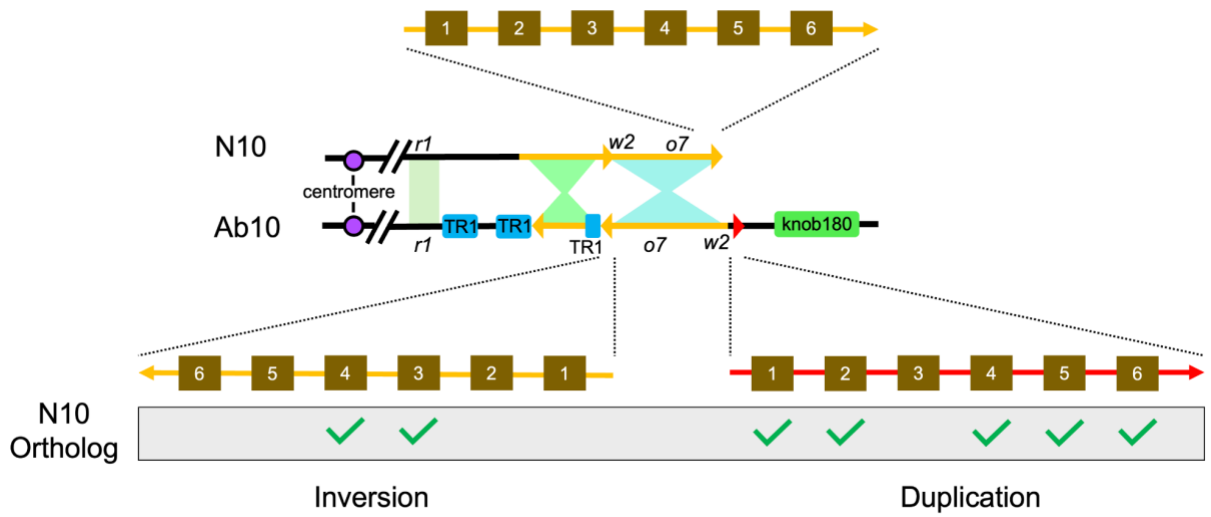


Figure 2.2. Duplicated region on Ab10. Cartoon representation of the duplicated region on Ab10 identified via OrthoFinder and BLAST (Emms and Kelly, 2019; Camacho et al., 2023). Shaded regions between N10 and Ab10 indicate regions of homology where the hourglass shape indicates an inversion. TR1 and knob180 are maize knob types and *r1* is a kernel and plant color locus marking the edge of the Ab10 haplotype. Location of genes with known physical position, *o7* and *w2*, are shown (Wang et al., 2011; Udy et al., 2012). Brown squares indicate an annotated gene (Liu et al., 2020; Hufford et al., 2021). Green checks indicate genes that appear functional.

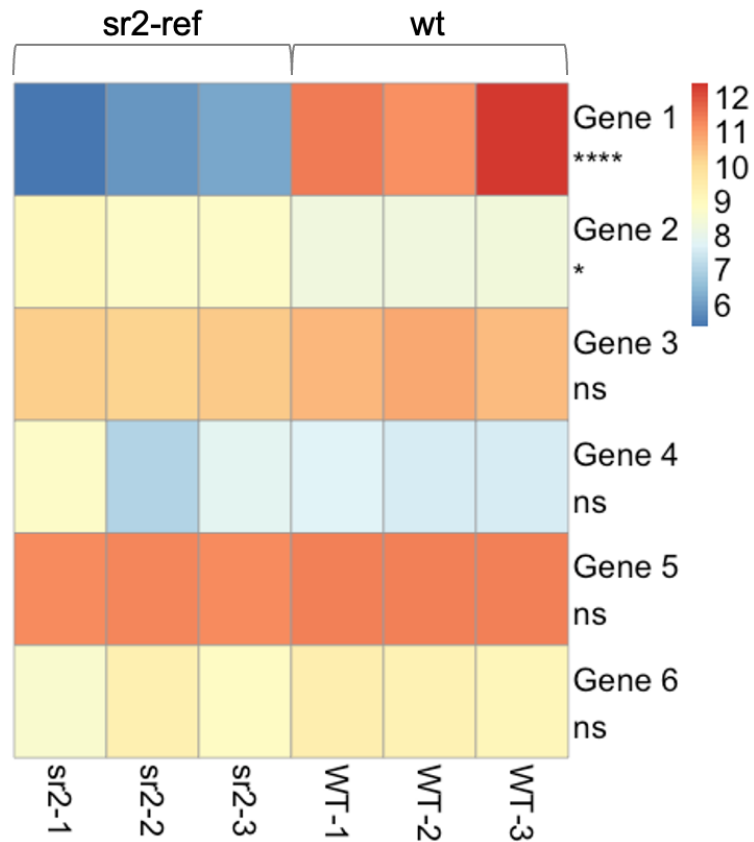


Figure 2.3. Differential expression of duplicated genes between *sr2* and wild type plants.

Gene numbers refer to those defined in Table 2.1. Color represents log₂ transformed expression value for each gene. ****= less than 0.0001, *= less than 0.05, ns = not significant.

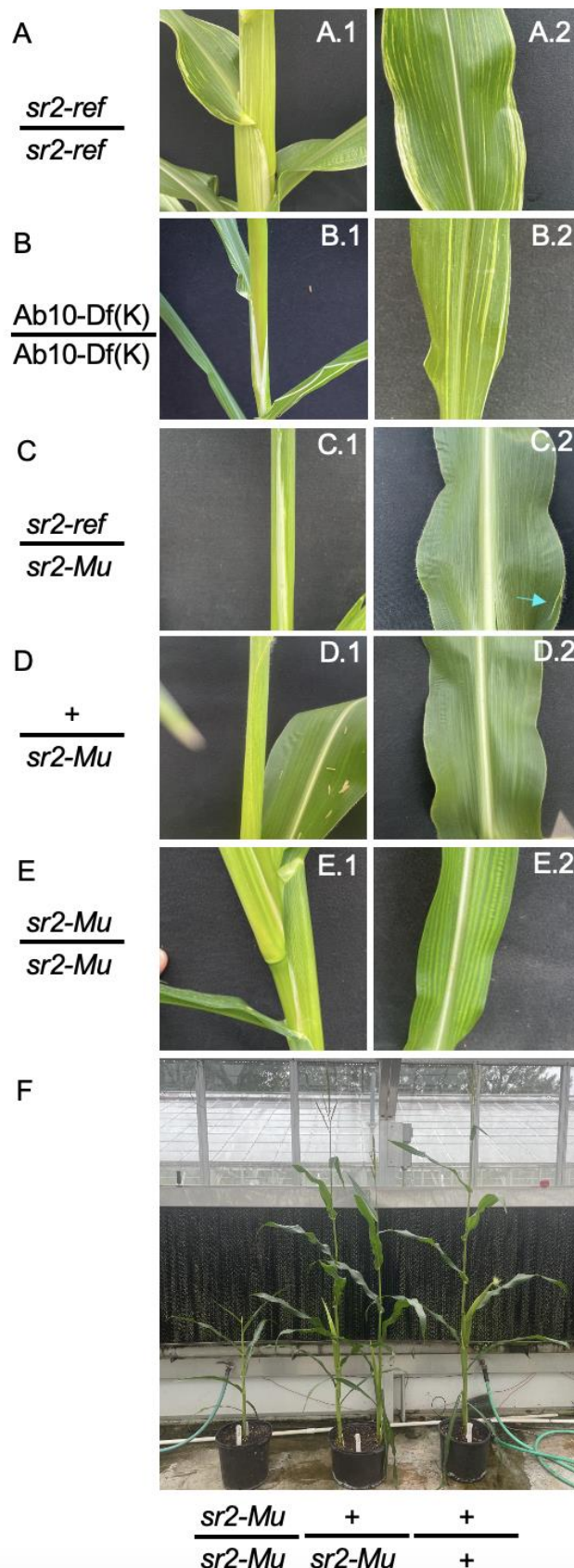


Figure 2.4. Phenotypes of *sr2*

mutants. A. Homozygous *sr2-ref* plants grown in the field. B.

Homozygous Ab10-Df(K) plant grown in the greenhouse. C. Plant heteroallelic for *sr2-ref/sr2-Mu* grown in the

greenhouse. These plants typically had few small leaf margin stripes indicated by the blue arrow. D. Plant

heterozygous for *+/sr2-ref* (where + is wild type) grown in the greenhouse.

This plant is a sibling of the plant shown in C. E. Plant homozygous for *sr2-Mu* grown in the greenhouse. F.

Sibling plants demonstrating the severe phenotype in *sr2-Mu/sr2-mu*

homozygous plants relative to siblings.

CHAPTER 3

ANTAGONISTIC KINESIN-14S WITHIN A SINGLE CHROMOSOMAL DRIVE HAPLOTYPE

Brady M. J., Gupta A., Gent J. I., Swentowsky K. W., Unckless R. L., and Dawe R. K.

Dawe, 2025 Antagonistic Kinesin-14s within a Single Chromosomal Drive Haplotype.

Submitted to *Genetics* 2/5/2025

Abstract

In maize, there are two meiotic drive systems that operate on large tandem repeat arrays called knobs that are found on chromosome arms. One meiotic drive haplotype, Abnormal chromosome 10 (Ab10), encodes two kinesin proteins that interact with two distinct tandem repeat arrays in a sequence-specific manner to confer meiotic drive. The kinesin KINDR associates with knob180 repeats while the kinesin TRKIN associates with TR-1 repeats. Prior data show that meiotic drive is conferred primarily by the KINDR/knob180 system, with the TRKIN/TR-1 system having little or no role. The second meiotic drive haplotype, K10L2, shows low levels of meiotic drive and only encodes the TRKIN/TR-1 system. Here we used long-read sequencing to assemble the K10L2 haplotype and showed that it has strong homology to an internal portion of the Ab10 haplotype. We also carried out CRISPR mutagenesis of *Trkin* to test the role of *Trkin* on Ab10 and K10L2. The data indicate that the *Trkin* gene on Ab10 does not improve drive or fitness but instead has a weak deleterious effect when paired with a normal chromosome 10. The deleterious effect is more severe when Ab10 is paired with K10L2: in this context functional *Trkin* on either chromosome nearly abolishes Ab10 drive. We modeled the effect of *Trkin* on Ab10 and found it should not persist in the population. We conclude that *Trkin* either confers an advantage to Ab10 in untested circumstances or that it is in the process of being purged from the Ab10 population.

Introduction

Selfish genetic elements (i.e. transmission ratio distorters) are structural elements of the genome that increase their own representation in the next generation despite conferring no fitness advantage (Burt and Trivers 2008). Meiotic drivers, one class of selfish genetic element, gain

their advantage by altering meiosis so that they are transmitted to more than 50% of the gametes (Lindholm *et al.* 2016). Examples of meiotic drive that operate at the level of meiosis are centromere drive, where larger centromeres are preferentially transmitted over smaller centromeres (Fishman and Kelly 2015; Lampson and Black 2017; Clark and Akera 2021; Dawe 2022), the segregation of certain B chromosomes (Fishman and Kelly 2015; Lampson and Black 2017; Clark and Akera 2021; Dawe 2022), and the maize Abnormal chromosome 10 haplotype (Ab10) (Fishman and Kelly 2015; Lampson and Black 2017; Clark and Akera 2021; Dawe 2022). There are also many other examples of drivers that exhibit preferential transmission but gain their advantage outside of meiosis (Lindholm *et al.* 2016). Selfish genetic elements are implicated in critical evolutionary processes such as extinction, speciation, recombination, and genome size evolution (Agren and Clark 2018). Ab10 is of particular interest as it has had a significant impact on shaping the evolution of maize, one of the most economically important crops (Buckler *et al.* 1999).

As much as >15% of the maize genome is composed of tandem repeat arrays (Hufford *et al.* 2021). One form of tandem repeat is referred to as knobs, and come in two different sequence classes, TR-1 and knob180. The Ab10 meiotic drive haplotype contains long arrays of both knob repeats as well as two kinesin protein-encoding genes: *Kindr* and *Trkin*. KINDR physically associates with knob180 knobs and TRKIN associates with TR-1 knobs (Figure 3.1a). Both kinesins pull their respective knobs ahead of the centromere during meiotic anaphase to cause their preferential transmission to the egg cell during female meiosis (Dawe 2022) (Figure 3.1b). Knobs throughout the genome are also preferentially transmitted when Ab10 is present. Both knob180 and TR-1 are conserved and abundant across the *Zea* genus and in *Tripsacum dactyloides* suggesting that Ab10 may have originated deep in the evolutionary history of the

grass family (Buckler *et al.* 1999; Swentowsky *et al.* 2020). The KINDR/knob180 system is primarily responsible for the preferential transmission of Ab10 while the TRKIN/TR-1 system contributes little, if at all (Kanizay *et al.* 2013a; Dawe *et al.* 2018). Nevertheless, *Trkin* is present on multiple Ab10 haplotypes in both teosinte and maize suggesting it may have been maintained via selection over the ~8700 years since their divergence (Piperno *et al.* 2009; Higgins *et al.* 2018; Swentowsky *et al.* 2020).

K10L2 is a structurally and functionally distinct variant of chromosome 10 that expresses TRKIN during meiosis and activates neocentromeres at TR-1 repeats (Kanizay *et al.* 2013a) (Figure 3.1). K10L2 demonstrates weak (1-2%) but statistically-significant meiotic drive (Kanizay *et al.* 2013a). Additionally, it has been identified in at least 12 disparate maize landrace populations suggesting it may be an important part of the Ab10 system (Kanizay *et al.* 2013a). One to two percent drive should be sufficient to cause K10L2 to rapidly spread throughout a population as long as it isn't associated with negative fitness consequences (Hartl 1970). K10L2 is also a very effective competitor against Ab10. When Ab10 is paired with K10L2, Ab10 drive is almost completely suppressed (Kanizay *et al.* 2013a). It has been speculated that both the drive of K10L2 and the suppressive effect of K10L2 on Ab10 are mediated by the TRKIN/TR-1 system (Swentowsky *et al.* 2020).

The fitness costs commonly imposed on the genome by selfish genetic elements selects for suppressors throughout the genome (Price *et al.* 2020). In the Ab10 system, K10L2 and N10 both represent disadvantaged loci. K10L2 can be thought of as both a disadvantaged locus carrying a highly effective suppressor when interacting with Ab10 and as an independent driver when interacting with N10. Both Ab10 and K10L2 have what appear to be suppressors of the KINDR/knob180 drive system. N10 carries a pseudo-*Kindr* locus that produces siRNAs that may

suppress *Kindr* expression and reduce drive (Dawe *et al.* 2018). K10L2 also acts as a suppressor of Ab10 with the likely mechanism being the TRKIN/TR-1 drive system. The evolution of suppressors by co-opting the machinery of drive has been observed before (Price *et al.* 2020). For example, the *wtf* genes in *Schizosaccharomyces pombe* represent a toxin-antidote system. There are *wtf* loci carrying only the antidote that behave as suppressors to intact *wtf* loci (Bravo Núñez, María Angélica, Lange, Jeffrey J, Zanders, Sarah E 2018). If the Ab10 drive system followed the same model, we would expect that the TRKIN/TR-1 system (i.e. a suppressor) would appear only on K10L2 or N10. How or why *Trkin* persists on Ab10 while conferring no apparent benefit in terms of drive, and likely contributing to the suppression of drive when paired with K10L2, is unclear.

Several hypotheses have been proposed to resolve the conundrum of the *TRKIN/TR-1* drive system on Ab10, all suggesting that *Trkin* improves the fitness of Ab10. The main ideas are that *Trkin* may: 1) increase Ab10 drive or 2) reduce the negative fitness effects associated with Ab10 (Swentowsky *et al.* 2020). In previous work the favored hypothesis was that *Trkin* reduces meiotic errors caused by the rapid movement of knobs during meiotic anaphase (Swentowsky *et al.* 2020). In this study, we set out to determine what effect *Trkin* has on Ab10 that may help to explain its persistence. We assembled the K10L2 haplotype and compared it to Ab10, then conducted drive and fitness assays of K10L2 and Ab10 haplotypes carrying *trkin* null alleles. Finally, we used mathematical modeling to better understand the predicted population dynamics of Ab10 haplotypes that carry *Trkin*.

Results

Assembly of K10L2 and Ab10

We began by generating a new assembly of Ab10 using PacBio HiFi. The Ab10 haplotype has been challenging to accurately assemble due to the prevalence of multiple repetitive arrays (i.e. knobs) that are notoriously difficult to assemble (Tørresen *et al.* 2019). The previous assembly of B73-Ab10 v1 was conducted with PacBio CLR data (single long reads) which have a higher error rate (Liu *et al.* 2020; Hon *et al.* 2020). To assess the quality and fidelity of the new assembly, we compared sequence homology between B73-Ab10 v1 (Liu *et al.* 2020) and the new assembly, B73-Ab10 v2. We found strong homology between the assemblies and the same relationship to N10 as previously reported (Figure S3.1, Figure S3.2b). In both assemblies the Ab10 haplotype is located at the end of the long arm of chromosome 10 as expected (Liu *et al.* 2020; Dawe 2022). The total size is unknown because of N-gaps predominantly within tandem-repeat arrays but, using the B73-Ab10 v2 assembly, we estimate the Ab10 haplotype contains about 77 Mb of sequence, with the proximal edge traditionally defined as the *colored1 (r1)* gene (a linked marker used to track Ab10 in crosses). We identified two large inverted segments homologous to N10 within the haplotype of 4.8 Mb and 9.5 Mb respectively (shared region) (Figure 3.1a, Figure S3.2b). These are slightly longer than reported in B73-Ab10 v1 assembly (Liu *et al.* 2020). There are three TR-1 knobs (assembled length=8.7 Mb collectively) and a very large knob180 knob (partially assembled length=8.5 Mb). Both the TR-1 and knob180 knobs assembled lengths are slightly lower than in the B73-Ab10 v1 assembly (Liu *et al.* 2020). Using data from terminal deletion lines of Ab10 in a different inbred background, we determined that the Ab10 knob is ~30.67 Mb long indicating it is only 28% assembled (Brady *et al.* 2024). There is also at least ~22 Mb of sequence that is unique to Ab10.

The 1.8 Mb region between the first two TR-1 knobs includes two copies of *Trkin* (Figure 3.2). The region to the right of the large knob180 knob contains an array of *Kindr* genes. Interestingly, there was a marked reduction in percent identity between the two assemblies over large tandem arrays like *Kindr* (Figure S3.1). This is likely due to the increased accuracy of PacBio HiFi reads (Hon *et al.* 2020). In fact, we identified 10 copies of *Kindr* in B73-Ab10 v2 instead of 9 as previously reported in B73-Ab10 v1 (Liu *et al.* 2020) (Figure S3.2d).

We next assembled the K10L2 haplotype. We found a distinct structure with two large TR-1 knobs (15.5 Mb collectively) and a 2.7 Mb non-shared region with a single copy of *Trkin* between them (Figure 3.1a, non-shared means a lack of homology to N10). Otherwise, we found no large inversions or other rearrangements relative to N10 (Figure S3.2a). Additionally, we found no tandemly repeated genes (i.e. *Kindr* array), which are common on Ab10 (Figure S3.2c,d) (Dawe *et al.* 2018). Sequence comparisons revealed the region between the two TR-1 knobs on K10L2 has strong homology to the *Trkin* bearing region on Ab10. However, unlike K10L2, Ab10 contains an inverted duplication with a second copy of *Trkin* (Figure 3.2, Figure S3.2e) (Swentowsky *et al.* 2020). The second copy of *Trkin* on Ab10 was previously thought to be a pseudogene and was referred to as Ab10 *pseudo-Trkin1* (Swentowsky *et al.* 2020). During this study we found that the coding sequence of *pseudo-Trkin1* was misinterpreted, and that it instead encodes a full-length open reading frame. Accordingly, we have renamed *pseudo-Trkin1* to *Trkin2* (Figure 3.3).

Genomic sequence of three Trkin genes reveals near identical intronic transposons

We annotated the K10L2 and new B73-Ab10 v2 assemblies using BRAKER v3.0.8 (Gabriel *et al.* 2024), which was not available at the time of the B73-Ab10 v1 assembly (Liu *et*

al. 2020). This allowed us to identify the full unbiased structure of each independent copy of *Trkin* on both Ab10 and K10L2. In line with the strong homology between the K10L2 haplotype and Ab10, inspection of the *Trkin* genomic sequence revealed a similar atypical structure between all three *Trkin* genes. Ab10 *Trkin1* spans 113 Kb and Ab10 *Trkin2* spans 99 Kb, while the K10L2 *Trkin* spans 89 Kb. The size differences are due to the presence of nine transposable elements in the introns of Ab10 *Trkin1* and two transposable elements in the introns of Ab10 *Trkin2* relative to K10L2 *Trkin*. The transposable elements in Ab10 *Trkin1* and Ab10 *Trkin2* are not shared suggesting duplication and divergence after separation from the K10L2 *Trkin*. Notably, Ab10 *Trkin1* and *Trkin2* carry all the transposable elements that are present in K10L2 *Trkin* (Figure 3.4). These data suggest that K10L2 *Trkin* is ancestral to the Ab10 *Trkin* genes.

Comparison of three Trkin CDS sequences reveals very few differences

Interrogation of the *Trkin* annotated coding sequence revealed that all three *Trkin* genes are remarkably similar with no significant evidence of functional divergence (Figure 3.3a). The K10L2 *Trkin* CDS contains six point mutations relative to Ab10 *Trkin1*. Five of these produce nonsynonymous amino acid substitutions (one in an unstructured region, one in the coiled coil domain, and three in the motor domain). The K10L2 *Trkin* CDS contains only four point mutations relative to Ab10 *Trkin2*, of which three cause nonsynonymous amino acid substitutions (one in an unstructured region and two in the motor domain). Ab10 *Trkin1* and Ab10 *Trkin2* differ by only two point mutations resulting in non-synonymous amino acid substitutions (one in the coiled coil domain and one in the motor domain) (Figure 3.3a). These data suggest that differing effects of *Trkin* between Ab10 and K10L2, if any exist, are not due to differences in the protein itself.

We next wondered what the relationship between the three *Trkin* genes might be. We generated a neighbor joining tree using the amino acids of the motor domain of all three *Trkin* gene as well as their most similar maize gene as an outgroup. We found that Ab10 *Trkin1* and Ab10 *Trkin2* are more similar to each other than to K10L2 *Trkin* (Figure 3.3b). This relationship suggests that the Ab10 *Trkin* genes duplicated after they diverged from K10L2 *Trkin*, in agreement with the inferences from the TE profile (Figure 3.4).

Gene orthology between three chromosome 10 haplotypes finds high agreement in the Trkin bearing region and unexpected orthologs in the Ab10 non-shared region.

We next investigated the gene orthology between all three assembled structural variants of chromosome 10 (Figure 3.5). We define the shared regions of both K10L2 and Ab10 as the regions with significant homology to N10, and the non-shared regions as the regions without significant homology to N10 (Figure S3.2, Figure 3.5). We found that there were 12 gene ortholog pairs between the Ab10 *Trkin* region and K10L2 *Trkin* region representing 44% (12/27) of annotated genes in this region on K10L2 and 66% (12/18) of the annotated genes in this region of Ab10 (Table 3.1, Table S3.2, Table S3.3, Figure 3.5). There were also unexpected gene ortholog pairs particularly between the shared region of K10L2 and the non-shared region of Ab10 (Table S3.4, Figure 3.5). Interestingly, using our new annotations, we identified 10 previously unknown gene orthologs between N10 and Ab10 in the non-shared region (Table S3.5, Figure 3.5). Among the newly identified genes are nine partial copies of a gene homologous to *nripd2/e2*, which is related to RNA dependent DNA methylation (Figure 3.5, Figure S3.3). This is of particular interest as it has been hypothesized that RNA dependent DNA

methylation may be related to the antagonistic dynamics between Ab10 and the host genome (Dawe *et al.* 2018).

Ab10 non-shared region annotations are enriched for RNA dependent DNA methylation GO terms.

We went on to perform a functional annotation of the Ab10 and K10L2 haplotypes using EnTAP (Table S3.1, Table S3.2) (Hart *et al.* 2020; Gabriel *et al.* 2024). Incorporating all gene annotations, Ab10 is significantly enriched for GO terms related to RNA dependent DNA methylation (Figure S3.4), a result that likely reflects the high copy number of *nrpd2/e2*. We also reduced all known tandemly duplicated genes to a single copy and reran the analysis. Under these circumstances, Ab10 is enriched for GO terms related to meiotic organization and microtubule based movement in agreement with our understanding of the mechanism (Figure S3.5) (Dawe 2022). Ab10 is enriched for RNA dependent DNA methylation when considering gene copy number, but not when considering only unique genes. In contrast, the K10L2 region was only significantly enriched for general reproductive processes, ATP hydrolysis, and several other miscellaneous GO terms (Figure S3.6).

Trkin expression in K10L2 and Ab10 lines

The *Trkin* copy number difference between Ab10 and K10L2 led us to wonder if they may also have expression level differences. We obtained RNA sequencing for Ab10 and K10L2 and mapped it to the B73-Ab10 v1 assembly (Liu *et al.* 2020; Swentowsky *et al.* 2020). The data revealed no consistent difference in *Trkin* expression between Ab10 bearing two copies and K10L2 bearing one copy of *Trkin* (Figure S3.7).

We also assessed the relative expression levels of *Trkin1* and *Trkin2* on Ab10. Analysis of RNA-seq data from ten tissues from a homozygous Ab10 line (Liu *et al.* 2020) indicated that the expression of *Trkin2* is ~93% lower on average than *Trkin1* ($t = 6.5$, $df = 41.4$, $p\text{-value} = 6e-08$) (Figure S3.8).

Generation of trkin knockout mutants on K10L2 and Ab10

To knock out the *trkin* gene on both K10L2 and Ab10, we designed a CRISPR construct with three guide RNAs targeting exon 3 and exon 4 of the *Trkin* gene (Figure 3.3). When we initiated the CRISPR mutagenesis, we were under the impression that Ab10 *Trkin2* was a pseudogene, and did not assay it for mutations; the primers were designed to be specific to Ab10 *Trkin1* (Table S3.6) (Swentowsky *et al.* 2020). Later, when we determined that Ab10 *Trkin2* is likely functional, we developed primers specific to Ab10 *Trkin2* and found that it is mutated in the line we were using as a positive control (Table S3.6, Figure 3.3d). We isolated the following mutations: K10L2 *Trkin*(+), K10L2 *trkin*(-), Ab10 *Trkin1*(+) *trkin2*(-), Ab10 *trkin1*(-) *Trkin2*(+), and Ab10 *trkin1*(-) *trkin2*(-) (Figure 3.3c,d,e). For K10L2, we had both a true wild type and a *trkin* mutant. For Ab10, we lacked a true wild type, so compared lines carrying either *Trkin1* or *Trkin2* alone to double mutants lacking both *trkin1* and *trkin2*.

Based on the strong correlation between *Trkin* and TR-1 neocentromere activity (Swentowsky *et al.* 2020), we expected *trkin* mutants to lack TRKIN protein and visible TR-1 neocentromeres at meiosis. In the Ab10 *trkin1*(-) *trkin2*(-) double mutant plants we could not detect TRKIN by immunostaining and observed no TR-1 neocentromeres by FISH (Figure 3.6, Figure 3.7), whereas Ab10 *Trkin1*(+) *Trkin2*(-) showed strong TRKIN staining and TR-1 neocentromeres (Figure 3.6, Figure 3.7). In the K10L2 *trkin*(-) mutant plants we could not detect

TRKIN by immunostaining, whereas K10L2 *Trkin*(+) showed strong TRKIN staining (Figure 3.6). However, we did not observe TRKIN localization or TR-1 neocentromeres in plants of the Ab10 *trkin1*(-) *Trkin2*(+) genotype, which likely reflects the fact that *Trkin2* is expressed at very low levels (Figure S3.8).

The Trkin gene is required for K10L2 to suppress meiotic drive of Ab10

Prior work had established that when Ab10 is paired with K10L2, meiotic drive is strongly suppressed (Kanizay *et al.* 2013a). We hypothesized that K10L2 *Trkin* may be responsible for this phenomenon. Using Ab10 *Trkin1*(+) *trkin2*(-) and K10L2 *Trkin*(+) as positive controls, we tested the effect of *Trkin* on Ab10 and K10L2 competition. We found that when *trkin* was completely knocked out on both Ab10 and K10L2, drive was fully restored to Ab10/N10 levels (Figure 3.8, Figure S3.9). This demonstrates that *Trkin* is necessary for K10L2 to compete with Ab10. Using reciprocal crosses, we further determined that one copy of Ab10 *Trkin1* or K10L2 *Trkin* is sufficient to fully suppress drive.

These data suggest that Ab10 encodes its own context dependent suppressor. Ab10 with active *Trkin1* should lose most of its drive whenever it encounters K10L2, variants of K10L2 that lack *Trkin*, or any other chromosome 10 with a large TR-1 knob.

Field and greenhouse experiments reveal no positive fitness effect of Trkin

Given the persistence of *Trkin* on the Ab10 haplotype, it seemed possible that it provides some benefit either through increased drive or reduced fitness effects (Buckler *et al.* 1999; Swentowsky *et al.* 2020). We tested this hypothesis by crossing our Ab10 *trkin* mutant lines as heterozygotes (*R1*-Ab10 (edited *trkin* alleles)/*r1*-N10) with pollen from *r1/r1* homozygous plants

in a large, randomized field design. Drive was measured by counting kernels carrying the dominant *R1* allele, which makes the kernels purple (*r1/r1* is colorless). We found that Ab10 *trkin1(-) trkin2(-)* had significantly higher drive than both Ab10 single *trkin* mutants with a mean difference of 0.41% (1 - 2 +) and 0.96% (1+ 2-) (Figure 3.9a). These effect sizes are quite small and right at the edge of what our experiment had power to detect. We had 51.8% power to detect a 1% change in drive and 82.8% power to detect a 1.2% change in drive. These data indicate that *Trkin* is not increasing Ab10 drive under the tested experimental conditions. Instead, *Trkin* appears to decrease drive.

It has previously been suggested that *Trkin* may improve Ab10 fitness by preventing anaphase segregation errors that might occur when centromeres and neocentromeres move in opposite directions on the spindle (Swentowsky *et al.* 2020). Such errors would be expected to cause increased numbers of aborted kernels. On the same ears used for testing drive, we found that Ab10 *Trkin1(+)* *trkin2(-)* had a significantly higher proportion of defective kernels than Ab10 *trkin1(-)* *Trkin2(+)* with a mean difference of 0.41%. However, Ab10 *trkin1(-) trkin2(-)* did not have a significantly different proportion of defective kernels than either single mutant (Figure 3.9b). We had 13% power to detect a 0.4% change and 78.2% power to detect a 0.8% change in kernel abortion. We also tested the effect of *Trkin* on the total number of kernels and found no significant differences between any genotypes (Figure 3.9). We had 80% power to detect down to a 30 kernel (~8.54%) difference. These data indicate that *Trkin1* does not reduce kernel abortion or alter total kernel count.

It is well understood that Ab10 causes severe reductions in kernel count and weight when homozygous (Higgins *et al.* 2018). We hypothesized that *trkin* may be ameliorating some of the deleterious fitness effects when Ab10 is homozygous. We created an F2 population segregating

for Ab10 *Trkin1*(+) *trkin2*(-) and Ab10 *trkin1*(-) *trkin2*(-) and conducted greenhouse fitness experiments. We found no significant effects on plant height, average kernel weight, or competitiveness between Ab10 haplotypes (Intra-Ab10 competition) with respect to *trkin* genotype (Figure S3.10). We had 80% power to detect differences of the following magnitudes: Height = 52 cm (32% change), average kernel weight = 0.07 g (48% change), intra-Ab10 competition = 21% change. Although in this small study we only could have detected large changes, the data indicate that *Trkin1* does not improve the fitness of Ab10 in the homozygous state.

The Trkin1 gene does not reduce the frequency of meiotic errors in male meiosis

To test the effects of Ab10 *Trkin* on the accuracy of male meiosis, we screened Ab10 homozygous male meiocytes under the microscope for meiotic errors. Prior data demonstrated that homozygous Ab10 plants have reduced pollen viability (Higgins *et al.* 2018). We found no differences in the meiotic errors between Ab10 *Trkin1*(+) *trkin2*(-), Ab10 *trkin1*(-) *Trkin2*(+), Ab10 *trkin1*(-) *trkin2*(-) lines or N10 lines (Figure S3.11). We had 80% power to detect down to the following differences: Tetrad Micronuceli = 5%, Tetrad Microcyte = >0%, Dyad Micronuclei = 36%, Total Meiotic Errors = 6%. These data provide further evidence that Ab10 *Trkin1* does not reduce the frequency of meiotic segregation errors that might occur when centromeres and neocentromeres move in opposite directions on the spindle (Swentowsky *et al.* 2020).

The Trkin1 gene does not affect the degree of meiotic drive at an unlinked mixed knob

Trkin is known to activate neocentromeres throughout the genome (Dawe 2022). It seemed possible that *Trkin* behaved differently with other TR-1 knobs in the genome. To test the

effect of *Trkin* on knobs elsewhere in the genome, we looked at its effect on the transmission of a large mixed knob on chromosome 4L marked by a GFP-encoding insertion that expresses in kernel endosperm (Li *et al.* 2013). We found no significant difference in segregation of the 4L knob between Ab10 with functional *Trkin1* or without functional *trkin*. We also found no difference in K10L2 *Trkin*(+) or *trkin*(-). We had 80% power to detect down to an 8% difference in segregation (Figure S3.12). Together these data indicate that *Trkin* does not have an outsized effect on knobs elsewhere in the genome, just as it has little or no effect on Ab10.

Ab10 Trkin(+) should not persist in maize populations and will quickly get replaced by *Ab10 trkin(-)*

The above evidence indicates that *Trkin* has a negative effect on Ab10 fitness. While it remains possible that two copies of *Trkin* have different effects or that *Trkin* has some benefit we were unable to detect, we wanted to examine the population dynamics of *Trkin* in the long-term using a modeling approach. We built on the prior Ab10 meiotic drive model (Hall and Dawe 2018) to include Ab10 *Trkin*(+), Ab10 *trkin*(-), K10L2, and N10, and examined Ab10 *Trkin*(+) dynamics in populations. Specifically, we asked three questions for a subset of parameters representative of the empirical system: (1) When and how often does Ab10 *Trkin*(+) outcompete Ab10 *trkin*(-) in a population, (2) Is the persistence of Ab10 *Trkin*(+) dominated by natural selection or genetic drift, and (3) How long does it take for Ab10 *trkin*(-) to eventually replace Ab10 *Trkin*(+) in a population?

We began with simulations following a deterministic model (assuming discrete non-overlapping generations, diploid organisms, and a single panmictic population of infinite size). We found that Ab10 *Trkin*(+) cannot invade a population at equilibrium with Ab10 *trkin*(-).

Additionally, we found that Ab10 *trkin*(-) can always invade a population at equilibrium with Ab10 *Trkin*(+). Thus, unless the Ab10 *Trkin*(+) allele has some hidden or context-dependent benefit, it should not invade or segregate in a population assuming a deterministic model.

Next, we considered the strength of selection against Ab10 *Trkin*(+) reasoning that if selection is weak enough, genetic drift might dominate over natural selection in small populations. If so, genetic drift might explain the persistence of Ab10 *Trkin*(+). We calculated the selection coefficient against Ab10 *Trkin*(+) compared to Ab10 *trkin*(-) for various values of reduction in drive due to *Trkin*. Selection predominates drift if $2*Ne*s > 1$, where s is the selection coefficient and Ne is the effective population size (Hartl and Clark 2007). So, we calculated $2*Ne*s$ for a range of reductions of drive and effective population sizes. There are almost no combinations of parameters where selection against Ab10 *Trkin*(+) would be dominated by genetic drift ($2*Ne*s < 1$). In fact, the population size would need to be less than 100 and the reduction in drive close to zero for genetic drift dynamics to dominate: neither of which are realistic. Therefore we concluded that selection against Ab10 *Trkin*(+) is strong enough that drift cannot explain its persistence.

Though genetic drift is unlikely to prevent Ab10 *trkin*(-) from overtaking Ab10 *Trkin*(+) in a population, drift may influence how long the process takes. Given that we know both Ab10 *trkin*(-) and *Trkin*(+) segregated in wild ancestors, this suggests both have persisted for at least 8700 generations (Piperno *et al.* 2009; Swentowsky *et al.* 2020). Therefore, we assessed whether, given estimated parameters, the Ab10 *trkin*(-) might still be in the process of replacing Ab10 *Trkin*(+). Thus, we extended our deterministic model to a stochastic model (choosing genotypes from a multinomial distribution to simulate genetic drift). We asked how long it takes for Ab10 *trkin*(-) to replace Ab10 *Trkin*(+) when Ab10 *Trkin*(+) starts at a frequency of 6% (based on

(Kato 1976; Kanizay *et al.* 2013a)), and Ab10 *trkin*(-) starts as a single copy. Ab10 *trkin*(-) introduced as a single copy would often be lost due to drift in a stochastic model (Haldane 1927). Figure 3.10a shows that the more the Ab10 *Trkin*(+) allele reduces drive, the more likely the Ab10 *trkin*(-) is to escape stochastic loss and replace Ab10 *Trkin*(+). However, in actual populations Ab10 *trkin*(-) exists so it must have escaped stochastic loss at some point (Swentowsky *et al.* 2020). Figure 3.10b shows the distribution for time to loss of Ab10 *Trkin*(+), given a rare Ab10 *trkin*(-) allele introduced in an Ab10 *Trkin*(+) population at equilibrium for Ab10 *Trkin*(+), K10L2, and N10 where Ab10 *trkin*(-) escaped stochastic loss. The mean time for loss of Ab10 *Trkin*(+), or the time it takes for Ab10 *trkin*(-) to replace Ab10 *Trkin*(+), is less than 500 generations. This is true if the reduction in drive is more than ~0.01 (our empirical estimates suggest the value is more like 0.1) (Figure 3.9a). Therefore we concluded that Ab10 *trkin*(-) should replace Ab10 *Trkin*(+) in less than 500 generations for most parameter combinations resembling the empirical system.

The results presented above fail to explain the long-term persistence of Ab10 *Trkin*(+). They suggest that either Ab10 *trkin*(-) is very young (less than 500 generations) and is currently replacing Ab10 *Trkin*(+) or that Ab10 *Trkin*(+) confers some fitness advantage that we did not observe.

Discussion

Despite examples of *Trkin* being encoded in all three common Ab10 variants and in K10L2 (Swentowsky *et al.* 2020) and conservation of *TR-I* knobs as far as *Tripsacum*, our data provide no evidence that *Trkin* provides a selective advantage to Ab10. Instead, under the conditions we tested, Ab10 *Trkin* slightly reduces Ab10 drive and acts as an efficient suppressor

of drive in the presence of K10L2. Since we only tested the function of Ab10 *Trkin1*, we cannot rule out the possibility that *Trkin1* has a positive fitness effect only in the presence of functional *Trkin2*. We can, however, confidently conclude that Ab10 *Trkin1* is sufficient to activate TR-1 neocentromeres and allow K10L2 to compete with Ab10 independently of *Trkin2*. Modeling demonstrates that, under our current understanding of the system, Ab10 *Trkin1*(+) *trkin2*(-) would not persist in the population if Ab10 *trkin1*(-) *trkin2*(-) were present. We propose two theories for the existence of *Trkin* on the Ab10 haplotype: an advantage either smaller than could be detected here or only apparent in untested circumstances, or that *Trkin* is in the process of being purged from the Ab10 population.

Our best estimate of *Trkin* prevalence in the Ab10 population places it at around 50% (Swentowsky *et al.* 2020). It is possible that Ab10 *trkin*(-) is a new development. Perhaps in the past, *Trkin* served a function that has been lost in the last ~500 years and is now slowly being purged from the population. It may be that *Trkin* provides benefits to Ab10 in teosinte, but not in maize. However, maize was domesticated from teosinte ~8700 years ago (Piperno *et al.* 2009) which our models suggest would have been ample time for *Trkin* to have been purged from the population (Figure 3.10b). To explain the continued presence of Ab10 *Trkin*(+) in maize it would need to be reintroduced via gene flow from teosinte, which is plausible (Yang *et al.* 2023). It is also possible that gene conversion or illegitimate recombination between Ab10 and K10L2 continuously reintroduces *Trkin* to Ab10.

K10L2 is a relatively common variant of chromosome 10 (Kato 1976; Kanizay *et al.* 2013a) and is known to function as a suppressor of Ab10 drive (Kanizay *et al.* 2013a). Our data demonstrate that the *Trkin* gene is specifically responsible for the ability of K10L2 to suppress Ab10 drive. The evolution of a suppressor on the disadvantaged allele is common in drive

systems (Price *et al.* 2020). However, it is unusual and apparently paradoxical (as far as we know this is the first example) for a driving haplotype to encode its own, albeit context dependent, suppressor. The Ab10 and K10L2 drive systems are clearly complex and have had a major impact on the evolution of maize. Our data suggest that we do not yet understand the full range of contexts where Ab10 either has historically functioned or is currently functioning as a meiotic driver. Further studies of Ab10 and other chromosome 10 variants in teosinte may help provide new leads, and help us better understand the functions of *Trkin* in natural Ab10 populations.

Methods

Assembly of K10L2

CI66 (PI 587148) seed was ordered from the Germplasm Resources Information Network in Ames, Iowa, and grown in the UGA Botany greenhouse in Athens, GA. Leaf tissue was sent to the Arizona Genomics Institute for DNA extraction using a CTAB method (Doyle and Doyle 1987). The sequencing library was constructed using SMRTbell Express Template Prep kit 3.0. The final library was size selected on a Blue Pippin (Sage Science) with 10-25 kb size selection. Sequencing was performed on a PacBio Revio system in CCS mode for 30 hours. We filtered reads to a quality of 0.99 or greater and converted them to fastq format using bamtools v2.5.2 and bedtools 2.30.0 respectively (Quinlan and Hall 2010; Barnett *et al.* 2011). We ran hifiasm v0.19.6 with post joining disabled to assemble the raw reads into contigs (Cheng *et al.* 2021). We identified the K10L2 haplotype by using BLAST v 2.13.0 to identify the contig with homology to the *Trkin* cDNA sequence (Swentowsky *et al.* 2020). Using BLAST v 2.13.0 we determined that the contig bearing *Trkin* also contained two large TR-1 knobs. Using the integrated genome

viewers (IGV) motif finder we determined that the *Trkin* bearing contig ended in 7,674 bp of telomere sequence indicating it was fully assembled (Thorvaldsdóttir *et al.* 2013). The *Trkin* bearing contig had no homology to the *colored1* gene, which marks the beginning of the Ab10 haplotype. To ensure all the chromosome 10 haplotypes were comparable we chose to manually merge the *colored1* gene bearing contig with the contig containing the otherwise complete K10L2 haplotype. Using BLAST v 2.13.0 we identified the contig bearing the *colored1* gene (B73 v5 Zm00001eb429330) and merged it to the *trkin* bearing contig with an interceding 100N gap using RagTag v2.1.0 (Alonge *et al.* 2022). All other contigs were left unaltered.

Assembly of B73-Ab10 v2

We chose to generate a new Ab10 assembly as there had been significant methodological advances since the generation of the first assembly (Liu *et al.* 2020). We used the same high molecular weight genomic DNA that was used in the B73-Ab10 v1 assembly (Liu *et al.* 2020). The sequencing library was constructed using SMRTbell Express Template Prep kit 2.0. The sequencing library was prepared for sequencing with the PacBio Sequel II Sequencing kit 2.0 for HiFi libraries and sequenced in CCS mode at the UGA Georgia Genomics and Bioinformatics Core facility. This data was integrated into the previously published assembly pipeline to produce the v2 assembly (Liu *et al.* 2020).

Comparison of the B73-Ab10 v1 and B73-Ab10 v2 Haplotypes

B73-Ab10 v1 and B73-Ab10 v2 were compared using Mummer v4.0.0 with a minimum length (-m) of 300 and computed all matches not only unique ones (--maxmatch) (Marçais *et al.* 2018; Liu *et al.* 2020). Plots were generated using R v4.3.1.

Annotation of Ab10 and K10L2

The assemblies described above were annotated for repeats and masked using RepeatMasker v4.1.5 in conjunction with the maize repeat library (Smit AFA., Hubley R., Green P. 2015; Ou 2020). All available short read mRNA sequencing data was downloaded for Ab10 (Liu *et al.* 2020) and K10L2 (Swentowsky *et al.* 2020) respectively. Reads were trimmed with Trimmomatic v0.39 (Bolger *et al.* 2014). These reads were then aligned to their respective genomes using HiSat2 v3n-20201216 (Kim *et al.* 2019). The resulting files were converted to a bam format and sorted using samtools v1.17 (Kim *et al.* 2019; Danecek *et al.* 2021). These alignments were used as expression evidence and the Viridiplantae partition of OrthoDB was used as protein evidence in an annotation using BRAKER v3.0.8 (Kuznetsov *et al.* 2023; Gabriel *et al.* 2024). Trinity v2.15.1 and StringTie v2.2.1 were used to assemble a de novo and reference guided transcriptome from the compiled RNAseq data for Ab10 and K10L2 respectively (Haas *et al.* 2013; Perteau *et al.* 2015). These transcriptomes were combined and converted to a comprehensive transcriptome database using PASA v2.5.3 (Haas *et al.* 2003). The resulting comprehensive transcriptome database was used to polish and add UTRs to the BRAKER derived gene annotation file in three rounds of PASA v2.5.3 (Haas *et al.* 2003). We found that the *Trkin* bearing region on Ab10 and K10L2 has an average percent identity of 98.5% for aligned regions (Figure 3.2). However, the annotated genes were quite different. In order to improve the annotations we used Liftoff v1.6.3 to reciprocally update the annotations in the *Trkin* bearing region on both haplotypes (Shumate and Salzberg 2021). We then extracted only genes that were included in the liftoff annotation using bedtools v2.31.0 and incorporated them (Quinlan and Hall 2010). Genes added in this way have names starting with gA in the K10L2 annotation and

gK in the Ab10 annotation. We extracted the CDS and cDNA sequences for both haplotypes using AGAT v1.1.0 (Dainat 2020) Finally, we extracted and functionally annotated the final protein sets using EnTAP v1.0.0 with the nr, Refseq, and Uniprot databases (O’Leary *et al.* 2016; Hart *et al.* 2020; Sayers *et al.* 2022; UniProt Consortium 2023).

Determination of Ab10 knob180 Knob Size

We obtained illumina sequence reads for terminal deletions of Ab10 in the W23 inbred background that either did or did not contain the large knob180 knob on the distal most end (Brady *et al.* 2024). We quantified knob180 repeat abundance in raw illumina short reads as described in (Hufford *et al.* 2021). In brief, we used seqtk v 1.2 to convert the read files to fasta format, used BLAST v2.2.26 to identify reads with homology to knob180, and bedtools merge v2.30.0 to combine overlapping hits (Quinlan and Hall 2010; Camacho *et al.* 2023; “seqtk” 2023). Using a custom R script, we filtered to hits 30 bp or longer, summed the lengths of all hits and divided that value by the average coverage of the library to obtain the Mb value of knob180 in each library. We then subtracted the value of the intact W23-Ab10 from the sample which did not contain the large knob180 knob to obtain the estimated size of the knob180 knob on Ab10. We repeated this process for TR1 and CentC as negative controls.

Comparison of Sequence Homology Between Ab10 and K10L2

All possible pairwise comparisons of chromosome 10 haplotypes were made using Mummer v4.0.0 with a minimum length (-m) of 300 and computed all matches, not only unique ones (--maxmatch). Self by self comparisons were run using the --nosimplify flag (Marçais *et al.* 2018). Plots were generated using R v4.3.1.

To assess the completeness of the nrpd2/e2 gene homologs we extracted all annotated copies coding sequence using AGAT v1.1.0 (Dainat 2020). We then aligned all copies to the nrpd2/e2 coding sequence from the B73v5 assembly using Geneious Prime v 2022.0.2 geneious algorithm (“Geneious 2022.0.2” 2022) (Zm00001eb068960) (Hufford *et al.* 2021). We identified functional domains in the nrpd2/e2 coding sequence using NCBI conserved domain search (Wang *et al.* 2023).

Comparison of Trkin CDS

The newly annotated *Trkin* gene was identified by overlap with the BLAST v 2.13.0 hits for *Trkin* cDNA (Swentowsky *et al.* 2020) against the newly assembled references (Camacho *et al.* 2023). The associated CDS was extracted from the CDS file for the respective genomes produced using AGAT v1.1.0 (Dainat 2020). The CDS sequences were aligned using Geneious Prime v 2022.0.2 geneious algorithm (“Geneious 2022.0.2” 2022). Protein domain locations were determined using NCBI conserved domain search, the cNLS mapper, and the MPI Bioinformatics toolkit (Kosugi *et al.* 2009; Gabler *et al.* 2020; Wang *et al.* 2023).

To better understand the relationship between the *Trkin* alleles we chose to make a phylogenetic tree using the protein motor domain. Unfortunately, TRKIN does not share sufficient homology with similar proteins to use its entire length. (Swentowsky *et al.* 2020). We used NCBI conserved domain search (Wang *et al.* 2023) to identify the motor domain in all the *Trkin* alleles as well as *Drosophila melanogaster* Ncd (Uniprot P20480) and *Zea mays* Dv1 (B73 v5 annotation Zm00001eb069600). We selected *Zea mays* Dv1 as it is the most closely related gene to *Trkin* (Swentowsky *et al.* 2020). We selected *Drosophila melanogaster* Ncd to act as an outgroup. We used geneious prime v2022.0.2 to (“Geneious 2022.0.2” 2022) perform a

MUSCLE alignment of all 4 motor domains and used the geneious tree builder to create a Neighbor-Joining tree using the Jukes-Cantor model. We set *Ncd* as the outgroup and performed 10000 bootstrap replicates. Numbers at nodes indicate the percent of replicate trees supporting that node.

Comparison of Gene Orthologs

Gene orthology between the three variants of the chromosome 10 haplotype was compared as described in (Brady *et al.* 2024). For the purposes of this analysis, the beginning of each haplotype was determined to be the location of the *colored1* gene. Plots were generated using R v4.3.1.

GO term enrichment analysis

We isolated the non-shared region, defined as those areas with no consistent synteny or homology to N10 as determined by the gene ortholog analysis and sequence comparisons, for both Ab10 and K10L2. These genes were tested against the remaining portions of the genome for GO term enrichment using topGO (Adrian Alexa 2024). The Ab10 non-shared region contains several known duplicated genes that heavily influence the results. All known arrayed gene duplicates were collapsed down to a single copy. The two copies of *Trkin* were both included.

Expression of Trkin

We obtained RNA sequencing data for Ab10 and K10L2 from (Swentowsky *et al.* 2020). We trimmed reads using Trimmomatic v0.39 (Bolger *et al.* 2014) and aligned them to the Ab10

v1 reference (Liu *et al.* 2020) using HiSat2 (Kim *et al.* 2019) and processed the output using Samtools v1.9 (Danecek *et al.* 2021). We used the R package featureCounts to determine the expression for each annotated gene (Liao *et al.* 2014). We then calculated the transcripts per million (TPM) for Ab10 *Trkin1* and Ab10 *Trkin2* in all samples requiring a mapping quality of 20. We summed the TPM of Ab10 *trkin1* and *trkin2* for easy comparison between Ab10 and K10L2.

To assess the expression of Ab10 *Trkin1* and *Trkin2* separately we assessed expression at the individual exon level. We obtained RNA sequencing data for 10 tissues of the B73-Ab10 inbred (Liu *et al.* 2020). We aligned them to the Ab10 v2 reference generated here using HiSat2 (Kim *et al.* 2019). We filtered the alignments to a mapping quality of 20 and required no mismatches. We then used the R package featureCounts to determine the expression of each annotated exon (Liao *et al.* 2014). We then calculated the TPM for only the *Trkin* exons containing SNPs (7 and 8) in all samples (Figure 3.3). We used a Welch two sample t-test to determine statistical significance between the two alleles.

Construction and transformation of a plasmid expressing Cas9 and guide RNAs

A CRISPR plasmid expressing Cas9 and three guide RNAs targeting *trkin* was constructed using a pTF101.1 binary plasmid (Paz *et al.* 2004) with similar components as previously used for gene editing in maize (Wang *et al.* 2021). In particular, it utilizes 1991 bp of a maize polyubiquitin promoter and UTR region (GenBank, S94464.1) to drive expression of Cas9 from *Streptococcus pyogenes* flanked by an N-terminal SV40 NLS and a C-terminal VirD2 NLS and followed by a polyadenylation signal provided by a *nopaline synthase* (*NOS*) terminator sequence from *Agrobacterium tumefaciens*. The Cas9 DNA sequence was codon

optimized for maize as described previously except that it did not include the potato ST-LS1 intron (Svitashev *et al.* 2015). The three guide RNAs were transcribed by three individual U6 promoters from maize and rice with two gRNAs targeting *Trkin* exon 3 (GTCTGGAGGCCAATGAGCACG and GAAAGCTTTTGCGGCCTCTGG) and one targeting exon 4 (GCCTACACAAGTAAACAGAT). These target sequences were selected using CHOPCHOP v3 (Labun *et al.* 2019). Gene synthesis and cloning was performed by GenScript (www.genscript.com), and transformation was performed by the Iowa State University Plant Transformation Facility.

Genotyping for trkin mutants

All genotyping DNA extractions were performed using a CTAB protocol (Clarke 2009). Polymerase chain reactions were performed using Promega GoTaq Green Master Mix (M7123). The Ab10 *trkin1* and K10L2 *trkin* edits were identified using the same primers (trkin_EX3 and trkin_EX4), Ab10 *trkin2* was detected using a separate pair of primers (Ptrkin_EX3, Ptrkin_EX4) (Table S3.6). Edits were confirmed by purifying the PCR reaction via Omega Bio-Tek Mag-Bind RxnPure Plus beads (M1386-01) using a 1:1 ratio and Sanger sequencing by Eton Biosciences. The competition assay plants were genotyped using primers specific to an indel in an intron of the *Trkin* gene (K10L2) (Table S3.6). All lines were checked for Cas9 using specific primers (Table S3.6). All reactions were conducted with slightly different temperature profiles and concentrations detailed in Table S3.6.

Immunofluorescence and FISH

Both Immunofluorescence and FISH were performed as described in (Swentowsky *et al.* 2020).

Competition Assay

To assess the effect of *Trkin* on the ability of K10L2 to suppress Ab10 drive we used plants in the same background that had one copy of Ab10 and one copy of K10L2 with varying *trkin* genotypes. In all cases Ab10 was marked by a dominant functional allele of *the colored 1* (*R1*) and K10L2 was marked by a recessive mutant allele (*r1*). We crossed these plants as the female to an *r1/r1* male and scored segregation of the *R1* allele. The background used contained the *C1* allele and was thus appropriate for tracking the *R1* allele. All experiments were conducted in the UGA Botany greenhouse (Athens, GA) across 3 seasons. In the case of K10L2 *trkin*(-) one season of the experiment had Cas9 segregating thus making it impossible to determine what *trkin* mutation was present. These are indicated in (Figure S3.9).

Results were analyzed using an ANOVA. Plots were generated using R v4.3.1.

Assessment of Ab10 Heterozygous Drive and Fitness

To determine the effect of *Trkin* on Ab10 drive we generated plants heterozygous for Ab10 and N10 with various *trkin* genotypes in the same genetic background. Friendly Isles Growing planted all plants in Molokai Hawaii in randomized rows of 15 kernels with every other row being an *r1/r1* male. No border corn was used, but edge effects were included in the final statistical model. All Ab10 bearing plants were detasseled, and allowed to open pollinate with the

r1/r1 males. Upon completion of the growing season Friendly Isles Growing harvested all female plants and sent them to the University of Georgia for processing. All ears were scored for defective kernels, a proxy for aborted kernels, defined as clearly defective kernels surrounded by otherwise healthy kernels with no other explanation. These criteria were selected to exclude insect damage, vivipary, and kernel loss during shipment. We shelled the ears and sorted them by color (dark pigmented *R1* and yellow *r1*). The seeds in each packet were counted using an International Marketing and Design Corp. Programmable Packeting Model 900-2 seed counter with the fast set to 7.2 and the slow set to 0.

The meiotic drive data were found to violate the criteria for an ANOVA, so we square root transformed the data to improve its fit which did not fully satisfy the statistical assumptions for a linear relationship, skew, and kurtosis, but came reasonably close. We chose to proceed with the ANOVA as the residuals appeared normally distributed and alternative statistical methods didn't offer the ability to account for the necessary number of variables. We included the following covariates in the model: field x coordinate, field y coordinate, edge of field, individual who sorted the kernels. The kernel abortion data was very far from a normal distribution so a kruskal-wallis test was used. The total kernel number data were analyzed using an ANOVA and met all assumptions. We included the following covariates in the model: field x coordinate, field y coordinate, edge of field, individual who sorted the kernels. Refer to Figure 3.9 for the full model used for each test.

Assessment of Ab10 homozygous fitness

To assess the effect of *Trkin* on Ab10 fitness we created an F2 mapping population segregating for Ab10 *Trkin1*(+) *trkin2*(-) and Ab10 *trkin1*(-) *trkin2*(-). We grew 39 F2 plants and

scored them for their *trkin1* genotype. We used a chi square test to check for deviation from a Mendelian segregation pattern. Plants were placed in a randomized order and grown to maturity in the UGA Botany greenhouse. They were allowed to open-pollinate amongst themselves. We measured plant height, and average kernel weight as proxies for plant fitness. We also scored total kernel count, but the experiment was underpowered to detect an effect of any magnitude. All data was analyzed using an ANOVA. Plots were generated using R v4.3.1.

Effect of Trkin on male meiotic errors

We scored Ab10 homozygous plants with different *trkin* genotypes for meiotic errors using the slides prepared for FISH as described above. A meiotic error was defined as a micronucleus in a dyad or tetrad, or a microcyte in a dyad or tetrad (Figure S3.11). Counts of meiotic errors were normalized against the total count of same stage cells observed. Results were analyzed using an ANOVA. Plots were generated using R v4.3.1.

Effect of Trkin on unlinked mixed knob

We ordered a line carrying a marker gene expressing GFP from a zein promoter (Li *et al.* 2013) that is closely linked to the knob on chromosome 4L (tdsgR106F01) from the Maize Genetics Cooperation Stock Center, Urbana, Illinois. We generated lines heterozygous for Ab10 or K10L2 with various *trkin* genotypes where the GFP insertion was linked to the knob and the opposite chromosome 4L was from the inbred Ms71 (PI 587137), which lacks a knob on 4L (Albert *et al.* 2010). Cas9 was segregating in the families used for these experiments so it wasn't possible to determine the exact allele used. However, all plants were derived from an individual with a *trkin* null mutation making it extremely likely that all plants, even those carrying Cas9,

carry a *trkin* null mutation as well. We then crossed these lines as the female to Ms71 and scored the resulting kernels for GFP fluorescence under visible blue light using a Dark Reader Hand Lamp and Dark Reader Glasses (Clare Chemical Research #HL34T). All data were analyzed using an ANOVA. Plots were generated using R v4.3.1.

Modeling the effect of trkin on Ab10 population dynamics

We model the system as a single locus where four alleles (Ab10 *Trkin*(+), Ab10 *trkin*(-), K10L2 and N10) are segregating. We initially assumed finite population sizes, discrete non-overlapping generations, diploid organisms, a single panmictic population, and that all individuals have the same number of offspring. We introduced stochasticity later. We assumed the N10/N10 homozygote is the wild-type genotype and has maximal fitness. We assumed that all heterozygotes experience drive during ovule production; pollen production follows Mendelian transmission and Ab10 *Trkin*(+), Ab10 *trkin*(-) and K10L2 alleles bear a fitness cost (3.1, Table 3.2). Ab10 drives against N10 (drive strength: d_1) and K10L2 (drive strength: d_3). K10L2 drives against N10 (drive strength: d_2). The *Trkin*(+) allele suppresses Ab10 drive by an amount of δ_1 ($0 < \delta_1 < d_1$).

Let p_m^+ , p_f^+ , p_m^- , p_f^- , q_m , and q_f denote the frequencies of the Ab10 *Trkin*(+), Ab10 *trkin*(-), and K10L2 alleles in pollen and ovules respectively in one generation. Then, the frequencies of the alleles in the next generation can be given by –

$$\begin{aligned}
p_m^{+'} = \frac{1}{\bar{W}} & ((1-a)p_f^+ p_m^+ + \frac{1}{2}(1-a)(p_f^+ p_m^- + p_f^- p_m^+) + \frac{1}{2}(1-a h_a)(p_m^+(1-p_f^- - p_f^+ \\
& - q_f) + p_f^+(1-p_m^- - p_m^+ - q_m)) + \frac{1}{2}(1-a h_a)(1-k h_k)(p_m^+ q_f \\
& + p_f^+ q_m))
\end{aligned} \tag{1}$$

$$\begin{aligned}
p_f^{+'} = \frac{1}{\bar{W}} & ((1-a)p_f^+ p_m^+ + \frac{1}{2}(1-a)(p_f^+ p_m^- + p_f^- p_m^+) \\
& + \frac{1}{2}(1+d_3)(1-a h_a)(1-k h_k)(p_m^+ q_f + p_f^+ q_m) \\
& + \frac{1}{2}(1-a h_a)(p_m^+(1-p_f^- - p_f^+ - q_f) + p_f^+(1-p_m^- - p_m^+ - q_m))(1 \\
& + d_1 - \delta_1))
\end{aligned} \tag{2}$$

$$\begin{aligned}
p_m^{-'} = \frac{1}{\bar{W}} & ((1-a)p_f^- p_m^- + \frac{1}{2}(1-a)(p_f^+ p_m^- + p_f^- p_m^+) \\
& + \frac{1}{2}(1-a h_a)(p_m^-(1-p_f^- - p_f^+ - q_f) + p_f^-(1-p_m^- - p_m^+ - q_m)) \\
& + \frac{1}{2}(1-a h_a)(1-k h_k)(p_m^- q_f + p_f^- q_m))
\end{aligned} \tag{3}$$

$$\begin{aligned}
p_f^{-'} = \frac{1}{\bar{W}} & ((1-a)p_f^- p_m^- + \frac{1}{2}(1-a)(p_f^+ p_m^- + p_f^- p_m^+) + \frac{1}{2}(1+d_1)(1-a h_a)(p_m^-(1 \\
& - p_f^- - p_f^+ - q_f) + p_f^-(1-p_m^- - p_m^+ - q_m)) + \frac{1}{2}(1+d_3)(1-a h_a)(1 \\
& - k h_k)(p_m^- q_f + p_f^- q_m))
\end{aligned} \tag{4}$$

$$\begin{aligned}
q_m^{'} = \frac{1}{\bar{W}} & ((1-k)q_f q_m + \frac{1}{2}(1-a h_a)(1-k h_k)(p_m^- q_f + p_f^- q_m) + \frac{1}{2}(1-a h_a)(1 \\
& - k h_k)(p_m^+ q_f + p_f^+ q_m) + \frac{1}{2}(1-k h_k)(q_f(1-p_m^- - p_m^+ - q_m) + (1 \\
& - p_f^- - p_m^+ - q_f)q_m))
\end{aligned} \tag{5}$$

$$\begin{aligned}
q_f' = \frac{1}{\bar{W}} & \left((1-k)q_fq_m + \frac{1}{2}(1-d_3)(1-a h_a)(1-k h_k)(p_m^-q_f + p_f^-q_m) + \frac{1}{2}(1 \right. \\
& - d_3)(1-a h_a)(1-k h_k)(p_m^+q_f + p_f^+q_m) + \frac{1}{2}(1+d_2)(1 \\
& \left. - k h_k)(q_f(1-p_m^- - p_m^+ - q_m) + (1-p_f^- - p_f^+ - q_f)q_m) \right)
\end{aligned} \tag{6}$$

Here, the mean fitness $\bar{W}$ can be calculated using –

$$\begin{aligned}
\bar{W} = & (1-a)p_f^-p_m^- + (1-a)p_f^+p_m^+ + (1-a)(p_f^+p_m^- + p_f^-p_m^+) + (1-a h_a)(p_m^-(1 \\
& - p_f^- - p_f^+ - q_f) + p_f^-(1-p_m^- - p_m^+ - q_m)) + (1-a h_a)(p_m^+(1-p_f^- \\
& - p_f^+ - q_f) + p_f^+(1-p_m^- - p_m^+ - q_m)) + (1-p_f^- - p_f^+ - q_f)(1-p_m^- \\
& - p_m^+ - q_m) + (1-k)q_fq_m + (1-a h_a)(1-k h_k)(p_m^-q_f + p_f^-q_m) \\
& + (1-a h_a)(1-k h_k)(p_m^+q_f + p_f^+q_m) + (1-k h_k)(q_f(1-p_m^- - p_m^+ \\
& - q_m) + (1-p_f^- - p_f^+ - q_f)q_m)
\end{aligned} \tag{7}$$

The frequency of N10 allele in pollen and ovules can be calculated using $(1 - p_m^- - p_m^+ - q_m)$ and $(1 - p_f^- - p_f^+ - q_f)$ respectively. We track the frequencies separately in the two sexes such that the frequencies in males and females each add up to 1, and the population always has equal sex-ratios.

We use a subset of parameters for the simulations based on empirical observations from the maize system – $h_a = 0.25$, $h_k = 0.2$, $a = 0.6$, $k = 0.225$, $d_1 = 0.4$ (drive strength of Ab10 against N10 = 70%), $d_2 = 0.1$ (drive strength of K10L2 against N10 = 55%), $d_3 = 0.1$ (drive strength of Ab10 against K10L2 = 55%) (Kanizay *et al.* 2013a; Higgins *et al.* 2018).

At this parameter subset, at $\delta_1=0$, at equilibrium, both Ab10 and K10L2 persist at a frequency of 5% each and the frequencies of Ab10 *Trkin*(+) and Ab10 *trkin*(-) are equal (deterministically).

*Testing the range of d_1 where Ab10 *Trkin*(+) and Ab10 *trkin*(-) can invade a population*

We ran these simulations deterministically for a range of δ_1 ($0 < \delta_1 < 0.4$) using an effective population size, N_e of 10,000 (Tittes *et al.* 2021) for 5000 generations (sufficient to reach equilibrium) with initial frequencies of Ab10 *Trkin*(+) and K10L2 at 5%, and Ab10 *trkin*(-) at $1/N_e$ (equal frequencies in both sexes). At any $\delta_1 > 0$, Ab10 *trkin*(-) always invades the population and replaces Ab10 *Trkin*(+).

We also tested for the invasion of Ab10 *Trkin*(+) similarly by starting the simulations with initial frequencies of Ab10 *trkin*(-) and K10L2 at 5%, and Ab10 *trkin*(+) at $1/N_e$ (equal frequencies in both sexes). For any value δ_1 , Ab10 *Trkin*(+) could never invade the population.

This suggests that the selection against Ab10 *Trkin*(+) is strong to prevent its invasion in a population containing Ab10 *trkin*(-) and Ab10 *trkin*(-) can invade a population containing Ab10 *Trkin*(+) and replace it.

Testing the strength of selection for a range of d_1 and calculating the selection coefficients such that $2N_e s < 1$ (nearly neutral zone)

For the calculation of the relative selective benefit (s) for Ab10 *trkin*(-), we ran the simulations for a range of δ_1 ($0 < \delta_1 < 0.4$) for 5000 generations (sufficient to reach equilibrium) with initial frequencies of Ab10 *Trkin*(+) and K10L2 at $1/N_e$, and Ab10 *trkin*(-) at 0. Then, after 5000 generations, we introduced Ab10 *trkin*(-) at a frequency of $1/N_e$ (only in females) into the

population at equilibrium. Then, we ran the simulation for one more generation and calculated the relative selective benefit of Ab10 *trkin*(-), *s* using allele frequencies after generation 5000 using –

$$s = \left(\frac{p_m^- + p_f^-}{p_m^- + p_f^-} \right) / \left(\frac{p_m^+ + p_f^+}{p_m^+ + p_f^+} \right) - 1 \quad [8]$$

This ‘*s*’ was used to calculate the $2N_e s$ parameter for a range of values of N_e ($10^2 < N_e < 10^4$) and δ_1 ($0 < \delta_1 < 0.4$). We found that $2 N_e s < 1$ only for a very small subset where $d_1 < 0.01$ and $N_e \sim 100$ (The approximate value of δ_1 from empirical observations in the maize system should be ~ 0.1)(Figure 3.9a). This suggests that selection against Ab10 *Trkin*(+) is strong and it could not be maintained in the population by drift (since $2 N_e s \gg 1$). This would imply that Ab10 *Trkin*(+) could not persist in the population in the presence of Ab10 *trkin*(-). Ab10 *Trkin*(+) is probably older than Ab10 *trkin*(-) and could be in the process of being replaced from the populations by invasion from Ab10 *trkin*(-).

Testing how long Ab10 Trkin(+) can persist in a population that is being invaded by Ab10 trkin(-)

We ran these simulations stochastically (modelling drift following a multinomial distribution) at $N_e=10,000$ and for a range of δ_1 ($0 < \delta_1 < 0.4$) (Tittes *et al.* 2021). We started our populations at an initial frequency of 6% for Ab10 *Trkin*(+) and K10L2 and $1/N_e$ for Ab10 *trkin*(-) (equal frequencies in both sexes). For each parameter value, each simulation was run 10,000 times, as Ab10 *trkin*(-) was often lost due to drift.

For the subset of simulations where Ab10 *trkin*(-) could successfully invade and replace Ab10 *Trkin*(+), we looked at the time taken for loss of Ab10 *Trkin*(+) from the population (Figure 3.10 B). For most values of δ_1 , Ab10 *Trkin*(+) was lost within 500 generations. From empirical estimates, $\delta_1 \sim 0.1$, thus, Ab10 *Trkin*(+) would be expected to persist for ~ 200 generations (Figure 3.9a).

We also looked at the proportion of times Ab10 *trkin*(-) (escaping stochastic loss due to drift) could successfully invade the population and outcompete Ab10 *Trkin*(+) (Figure 3.10 A). This proportion was small and for $\delta_1 \sim 0.1$, about 2.5% of the times Ab10 *trkin*(-) could escape stochastic loss and outcompete Ab10 *Trkin*(+).

Author Contributions

MJB and RKD designed the research. JIG assisted in some greenhouse experiments and CRISPR plasmid design. KWS performed CRISPR plasmid design. RLU and AG performed all work related to modeling. MJB performed all other research. MJB and RKD wrote the paper. All authors agreed that this work may be included in this dissertation.

References

- Adrian Alexa J. R., 2024 *topGO: Enrichment Analysis for Gene Ontology. R package*.
- Agren J. A., and A. G. Clark, 2018 Selfish genetic elements. *PLoS Genet.* 14: e1007700.
- Albert P. S., Z. Gao, T. V. Danilova, and J. A. Birchler, 2010 Diversity of chromosomal karyotypes in maize and its relatives. *Cytogenet. Genome Res.* 129: 6–16.
- Alonge M., L. Lebeigle, M. Kirsche, K. Jenike, S. Ou, *et al.*, 2022 Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 23: 258.
- Barnett D. W., E. K. Garrison, A. R. Quinlan, M. P. Strömberg, and G. T. Marth, 2011 BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27: 1691–1692.
- Bolger A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Brady M. J., M. Cheam, J. I. Gent, and R. K. Dawe, 2024 The maize striate leaves2 (sr2) gene encodes a conserved DUF3732 domain and is homologous to the rice yss1 gene. *Plant Direct* 8: e567.
- Bravo Núñez, María Angélica, Lange, Jeffrey J, Zanders, Sarah E, 2018 A suppressor of a wtf poison-antidote meiotic driver acts via mimicry of the driver's antidote. *PLoS Genet.* 14: e1007836.
- Buckler E. S., T. L. Phelps-Durr, C. S. Keith-Buckler, R. K. Dawe, J. F. Doebley, *et al.*, 1999 Meiotic Drive of Chromosomal Knobs Reshaped the Maize Genome. *Genetics* 153: 415–426.
- Burt A., and R. Trivers, 2008 *Genes in Conflict The Biology of Selfish Genetic Elements*. Harvard University Press, Cambridge, MA and London, England.
- Camacho C., G. M. Boratyn, V. Joukov, R. Vera Alvarez, and T. L. Madden, 2023 ElasticBLAST: accelerating sequence search via cloud computing. *BMC Bioinformatics* 24: 117.
- Cheng H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li, 2021 Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18: 170–175.
- Clark F. E., and T. Aker, 2021 Unravelling the mystery of female meiotic drive: where we are. *Open Biol.* 11: 210074.
- Clarke J. D., 2009 Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harb. Protoc.* 2009: db.prot5177.
- Dainat J., 2020 *AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format*.
- Danecek P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, *et al.*, 2021 Twelve years of SAMtools and BCFtools. *Gigascience* 10. <https://doi.org/10.1093/gigascience/giab008>
- Dawe R. K., E. G. Lowry, J. I. Gent, M. C. Stitzer, K. W. Swentowsky, *et al.*, 2018 A Kinesin-14 Motor Activates Neocentromeres to Promote Meiotic Drive in Maize. *Cell* 173: 839–850 e18.
- Dawe R. K., 2022 The maize abnormal chromosome 10 meiotic drive haplotype: a review. *Chromosome Res.* <https://doi.org/10.1007/s10577-022-09693-6>
- Doyle J. J., and J. L. Doyle, 1987 A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem Bull* 19: 11–15.
- Emms D. M., and S. Kelly, 2019 OrthoFinder: phylogenetic orthology inference for comparative

- genomics. *Genome Biol.* 20: 238.
- Fishman L., and J. K. Kelly, 2015 Centromere-associated meiotic drive and female fitness variation in *Mimulus*. *Evolution* 69: 1208–1218.
- Gabler F., S.-Z. Nam, S. Till, M. Mirdita, M. Steinegger, *et al.*, 2020 Protein sequence analysis using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinformatics* 72: e108.
- Gabriel L., T. Brůna, K. J. Hoff, M. Ebel, A. Lomsadze, *et al.*, 2024 BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *bioRxiv.org*. <https://doi.org/10.1101/2023.06.10.544449>
- Geneious 2022.0.2, 2022
- Haas B. J., A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr, *et al.*, 2003 Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31: 5654–5666.
- Haas B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, *et al.*, 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8: 1494–1512.
- Haldane J. B. S., 1927 A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Math. Proc. Camb. Philos. Soc.* 23: 838–844.
- Hall D. W., and R. K. Dawe, 2018 Modeling the evolution of female meiotic drive in maize. *G3 (Bethesda)* 8: 123–130.
- Hart A. J., S. Ginzburg, M. S. Xu, C. R. Fisher, N. Rahmatpour, *et al.*, 2020 EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Mol. Ecol. Resour.* 20: 591–604.
- Hartl D. L., 1970 Analysis of a general population genetic model of meiotic drive. *Evolution* 24: 538.
- Hartl D. L., and A. G. Clark, 2007 *Principles of population genetics*. Oxford University Press, New York, NY.
- Higgins D. M., E. G. Lowry, L. B. Kanizay, P. W. Becraft, D. W. Hall, *et al.*, 2018 Fitness Costs and Variation in Transmission Distortion Associated with the Abnormal Chromosome 10 Meiotic Drive System in Maize. *Genetics* 208: 297–305.
- Hon T., K. Mars, G. Young, Y.-C. Tsai, J. W. Karalius, *et al.*, 2020 Highly accurate long-read HiFi sequencing data for five complex genomes. *Sci Data* 7: 399.
- Hufford M. B., A. S. Seetharam, M. R. Woodhouse, K. M. Chougule, S. Ou, *et al.*, 2021 De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* 373: 655–662.
- Kanizay L. B., P. S. Albert, J. A. Birchler, and R. K. Dawe, 2013a Intragenomic conflict between the two major knob repeats of maize. *Genetics* 194: 81–89.
- Kanizay L. B., T. Pyhajarvi, E. G. Lowry, M. B. Hufford, D. G. Peterson, *et al.*, 2013b Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in *Zea mays*. *Heredity* 110: 570–577.
- Kato Y. T. A., 1976 Cytological studies of maize (*Zea mays* L.) and teosinte (*Zea mexicana* Schrader Kuntze) in relation to their origin and evolution. *Mass. Agric. Exp. Stn. Bull* 635: 1–185.
- Kim D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg, 2019 Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37: 907–915.
- Kosugi S., M. Hasebe, M. Tomita, and H. Yanagawa, 2009 Systematic identification of cell

- cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U. S. A.* 106: 10171–10176.
- Kuznetsov D., F. Tegenfeldt, M. Manni, M. Seppey, M. Berkeley, *et al.*, 2023 OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* 51: D445–D451.
- Labun K., T. G. Montague, M. Krause, Y. N. Torres Cleuren, H. Tjeldnes, *et al.*, 2019 CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 47: W171–W174.
- Lampson M. A., and B. E. Black, 2017 Cellular and Molecular Mechanisms of Centromere Drive. *Cold Spring Harb. Symp. Quant. Biol.* 82: 249–257.
- Li Y., G. Segal, Q. Wang, and H. K. Dooner, 2013 Gene tagging with engineered Ds elements in maize. *Methods Mol. Biol.* 1057: 83–99.
- Liao Y., G. K. Smyth, and W. Shi, 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930.
- Lindholm A. K., K. A. Dyer, R. C. Firman, L. Fishman, W. Forstmeier, *et al.*, 2016 The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends Ecol. Evol.* 31: 315–326.
- Liu J., A. S. Seetharam, K. Chougule, S. Ou, K. W. Swentowsky, *et al.*, 2020 Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol.* 21: 121.
- Marçais G., A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, *et al.*, 2018 MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* 14: e1005944.
- O’Leary N. A., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, *et al.*, 2016 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44: D733–45.
- Ou S., 2020 maizeTE02052020. MTEC.
- Paz M. M., H. Shou, Z. Guo, Z. Zhang, A. K. Banerjee, *et al.*, 2004 Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant. *Euphytica* 136: 167–179.
- Pertea M., G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell, *et al.*, 2015 StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33: 290–295.
- Piperno D. R., A. J. Ranere, I. Holst, J. Iriarte, and R. Dickau, 2009 Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl. Acad. Sci. U. S. A.* 106: 5019–5024.
- Price T. A. R., N. Windbichler, R. L. Unckless, A. Sutter, J. N. Runge, *et al.*, 2020 Resistance to natural and synthetic gene drive systems. *J. Evol. Biol.* 33: 1345–1360.
- Quinlan A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
- Sayers E. W., E. E. Bolton, J. R. Brister, K. Canese, J. Chan, *et al.*, 2022 Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 50: D20–D26.
- seqtk, 2023
- Shumate A., and S. L. Salzberg, 2021 Liftoff: accurate mapping of gene annotations. *Bioinformatics* 37: 1639–1643.
- Smit AFA., Hubley R., Green P., 2015 RepeatMasker Open-4.0
- Svitashev S., J. K. Young, C. Schwartz, H. Gao, S. C. Falco, *et al.*, 2015 Targeted Mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiol.* 169: 931–945.

- Swentowsky K. W., J. I. Gent, E. G. Lowry, V. Schubert, X. Ran, *et al.*, 2020 Distinct kinesin motors drive two types of maize neocentromeres. *Genes Dev.* 34: 1239–1251.
- Thorvaldsdóttir H., J. T. Robinson, and J. P. Mesirov, 2013 Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14: 178–192.
- Tittes S., A. Lorant, S. McGinty, J. B. Holland, J. de J. Sánchez-González, *et al.*, 2021 Not so local: the population genetics of convergent adaptation in maize and teosinte. *bioRxiv*.
- Tørresen O. K., B. Star, P. Mier, M. A. Andrade-Navarro, A. Bateman, *et al.*, 2019 Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res.* 47: 10994–11006.
- UniProt Consortium, 2023 UniProt: The universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51: D523–D531.
- Wang N., J. I. Gent, and R. Kelly Dawe, 2021 Haploid induction by a maize *cenH3* null mutant. *Science Advances* 7: eabe2299.
- Wang J., F. Chitsaz, M. K. Derbyshire, N. R. Gonzales, M. Gwadz, *et al.*, 2023 The conserved domain database in 2023. *Nucleic Acids Res.* 51: D384–D388.
- Yang N., Y. Wang, X. Liu, M. Jin, M. Vallebuena-Estrada, *et al.*, 2023 Two teosintes made modern maize. *Science* 382: eadg8940.

Table 3.1. **Ab10 *trkin* Modeling.** Fitness and proportion of ovules and pollen produced by each genotype.

Genotype	Proportion ovules				Proportion pollen				Fitness
	N10	Ab10 <i>Trkin</i> (+)	Ab10 <i>trkin</i> (-)	K10L2	N10	Ab10 <i>Trkin</i> (+)	Ab10 <i>trkin</i> (-)	K10L2	
N10 N10	1				1				1
N10 Ab10 <i>trkin</i> (-)	$(1-d_1)/2$		$(1+d_1)/2$		1/2		1/2		$1-h_a a$
N10 Ab10 <i>Trkin</i> (+)	$(1-d_1+\delta_1)/2$	$(1+d_1-\delta_1)/2$			1/2	1/2			$1-h_a a$
N10 K10L2	$(1-d_2)/2$			$(1+d_2)/2$	1/2			1/2	$1-h_k k$
Ab10 <i>trkin</i> (-) Ab10 <i>trkin</i> (-)			1				1		$1-a$
Ab10 <i>Trkin</i> (+) Ab10 <i>trkin</i> (-)		1/2	1/2			1/2	1/2		$1-a$
Ab10 <i>Trkin</i> (+) Ab10 <i>Trkin</i> (+)			1				1		$1-a$
K10L2 K10L2				1				1	$1-k$
K10L2 Ab10 <i>Trkin</i> (+)	$(1-d_3)/2$	$(1+d_3)/2$			1/2	$\frac{1}{2}$			$(1-h_a a)^*(1-h_k k)$
K10L2 Ab10 <i>trkin</i> (-)	$(1-d_3)/2$		$(1+d_3)/2$		1/2		$\frac{1}{2}$		$(1-h_a a)^*(1-h_k k)$

Table 3.2: **Ab10 *Trkin* Model Parameters** Parameters used in the model (All parameters range between 0-1 except δ_1 , δ_1 ranges between 0- d_1).

Variable/ Parameter	Description
d_1	Drive strength of Ab10 against N10
δ_1	Amount of Ab10 drive suppressed by <i>trkin</i> (+)
d_2	Drive strength of K10L2 against N10
d_3	Drive strength of Ab10 against K10L2
a	Fitness cost of Ab10 homozygote
h_a	Dominance coefficient for Ab10/N10
k	Fitness cost of K10L2 homozygote
h_k	Dominance coefficient for K10L2/N10

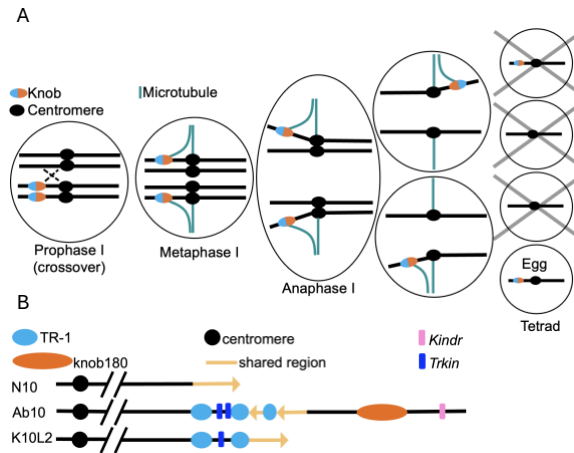


Figure 3.1: **Diagram of Maize Chromosome 10 Haplotypes.** A. Diagram of the structure of three chromosome 10 haplotypes. The orientation of the shared region on K10L2 was unknown prior to this study, the orientation we determined is shown. B. Model of Ab10 meiotic drive. For Ab10 drive to occur during female meiosis, the plant must be heterozygous for Ab10. Then recombination must occur between the centromere and the beginning of the Ab10 haplotype. During metaphase TRKIN associates with TR-1 knobs and KINDR associates with knob180 knobs. Both kinesin-14 proteins then drag the knobs ahead of the centromere during anaphase I and II causing their preferential transmission to the top and bottom cells of the meiotic tetrad. Since only the bottom-most cell becomes the egg cell, Ab10 is overrepresented in progeny (Dawe *et al.* 2018; Swentowsky *et al.* 2020).

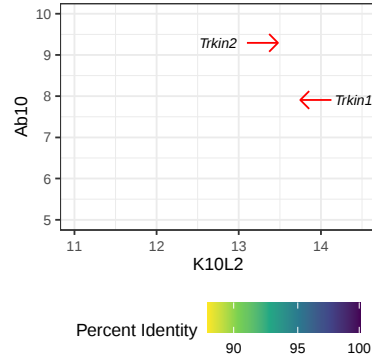


Figure 3.2: **Sequence Comparison of *Trkin* Bearing Region on Ab10 and K10L2.** Each dot marks the start of a maximal unique match (MUM) of at least 300bp long between the Ab10 and K10L2 haplotype, which begin at the *colored1* gene (Marçais *et al.* 2018). Coordinates start at the *colored1* gene. The color of each dot represents the percent identity of that match. All large knob arrays were removed for the sake of clarity. Both Ab10 *Trkin* genes are marked. The K10L2 and Ab10 assemblies refer to the assemblies generated in this work.

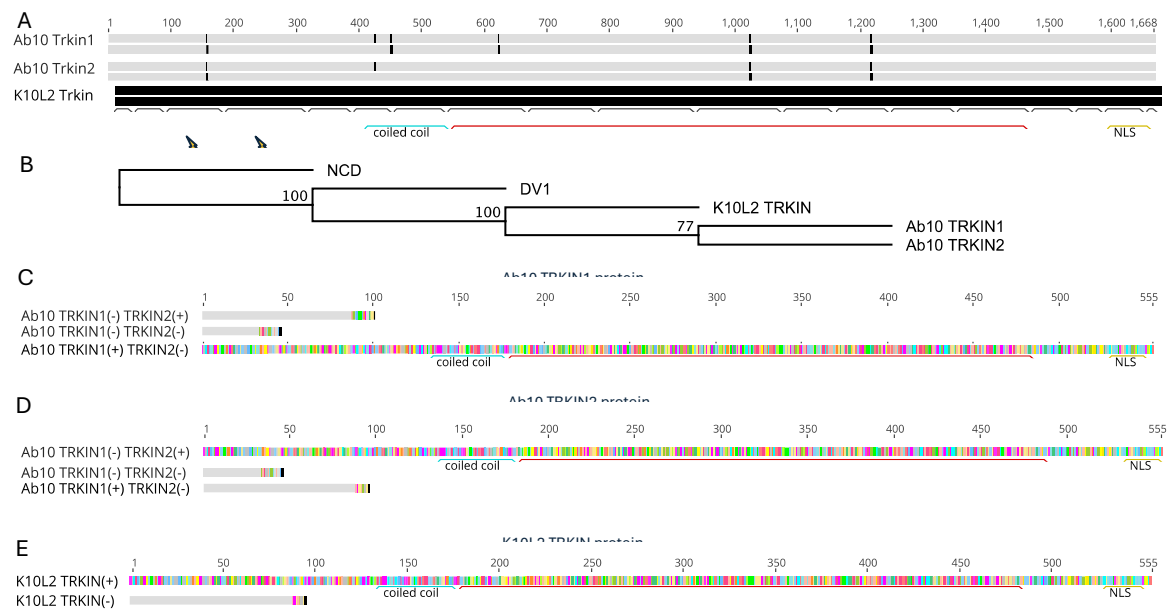


Figure 3.3: **Comparison of *trkin* and Mutants.** A. A coding sequence alignment (top bar) and protein translation (bottom bar) of all three *Trkin* sequences. Grey indicates sequence that is identical to the K10L2 *Trkin*, black indicates sequence that is different from the K10L2 *Trkin*. Exon boundaries are marked by numbered grey boxes. Protein domains are marked by colored boxes and labeled by domain type. NLS = nuclear localization signal (Swentowsky *et al.* 2020). Lightning bolts indicate exons that Cas9 was targeted to. B. Neighbor joining consensus tree using Jukes-Cantor model and 1000 bootstraps of protein motor domain for all TRKIN alleles, the most closely related *Zea mays* gene *Dv1*, and the *Drosophila melanogaster* *Ncd* gene as an outgroup (Swentowsky *et al.* 2020). Number at nodes indicate the number of replicate trees supporting that node. C. Ab10 *Trkin1* protein alignment. Grey indicates sequence identical to the intact (+) Ab10 *Trkin1*. Color indicates sequence that is different from the intact (+) Ab10 *Trkin1*. Ab10 *trkin1*(-) *Trkin2*(+) and Ab10 *trkin1*(-) *trkin2*(-) are truncated as a result of stop codons. D. Ab10 TRKIN2 protein alignment. Grey indicates sequence identical to the intact (+) Ab10 TRKIN2. Color indicates sequence that is different from the intact (+) Ab10 TRKIN2. Ab10 TRKIN1(+) TRKIN2(-) and Ab10 TRKIN1(-) TRKIN2(-) are truncated as a result of the introduction of a stop codon. E. K10L2 TRKIN protein alignment. Grey indicates sequence identical to the intact (+) K10L2 TRKIN. Color indicates sequence that is different from the intact (+) K10L2 TRKIN. K10L2 TRKIN(-) is truncated as a result of the introduction of a stop codon. C, D, E. Protein domains are marked by colored boxes labeled by domain type. NLS = nuclear localization signal (Swentowsky *et al.* 2020).

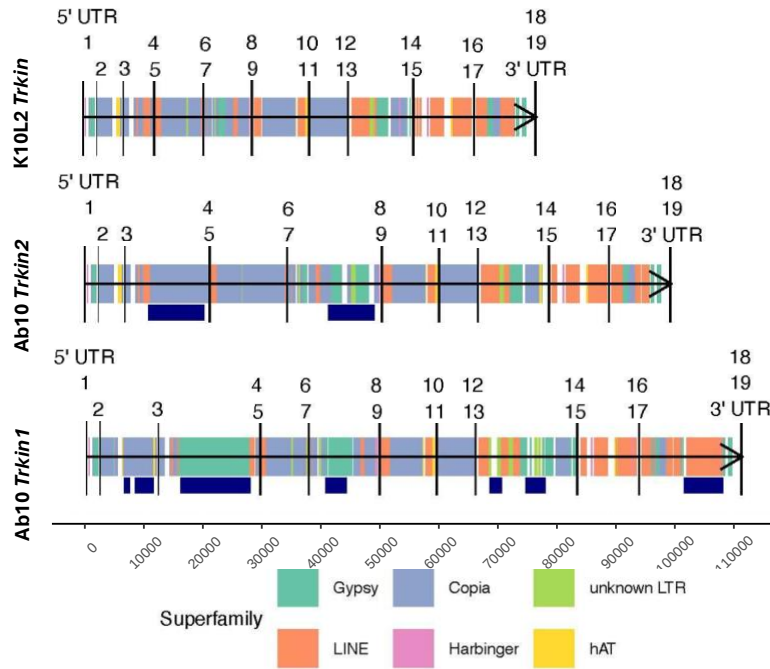


Figure 3.4: **Comparison of Transposable Element (TE) Composition Between All *Trkin* genes.** Genomic sequences for all three *Trkin* alleles, represented by a horizontal black line, are shown from Ab10 and K10L2. Vertical long black lines indicate *Trkin* exons. Short colored boxes centered on the horizontal black line indicate annotated transposable elements colored by their superfamily. Navy bars below the annotated TE blocks indicate insertions unique to that *Trkin* allele.

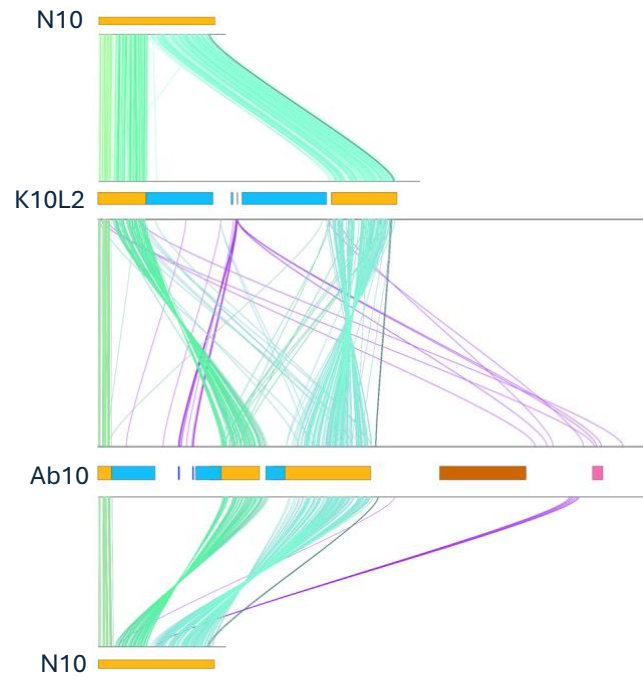


Figure 3.5: **Gene Ortholog Comparisons Among Chromosome 10 Haplotypes.** Each line represents a gene ortholog pair as determined by OrthoFinder (Emms and Kelly 2019). Shades of green represent gene ortholog pairs in the shared region. Purple represents gene ortholog pairs outside of the shared region. Relevant regions of each haplotype are marked by colored bars: gold = shared, light blue = TR1 knob, dark blue = *Trkin*, dark orange = knob180 knob, pink = *Kindr*. K10L2 and Ab10 refer to the assemblies generated in this work. N10 refers to the B73 v5 assembly (Hufford *et al.* 2021).

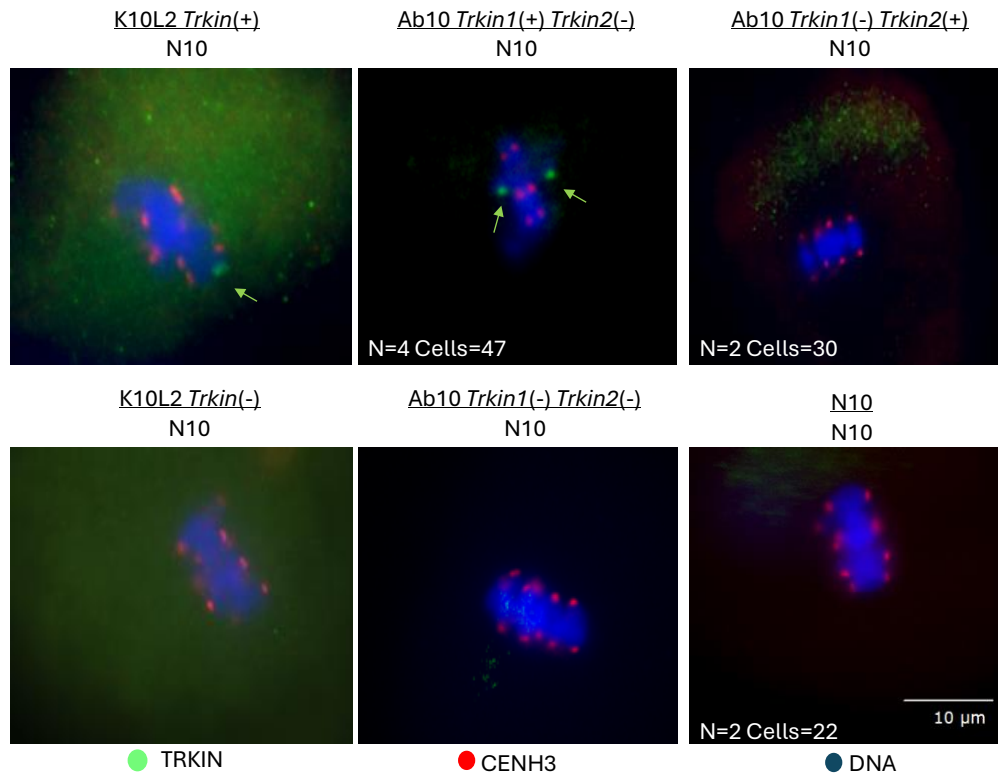


Figure 3.6: **TRKIN Immunofluorescence In Various *trkin* Genotype Male Meiocytes.** All images show metaphase I except for the Ab10 *trkin1*(+) *trkin2*(-) which represents metaphase II. N indicates the number of individual plants observed, cells indicate the number of appropriately staged same phenotype cells observed. CENH3 is in red, TRKIN in green, and DNA in blue. Green arrows show TRKIN staining.

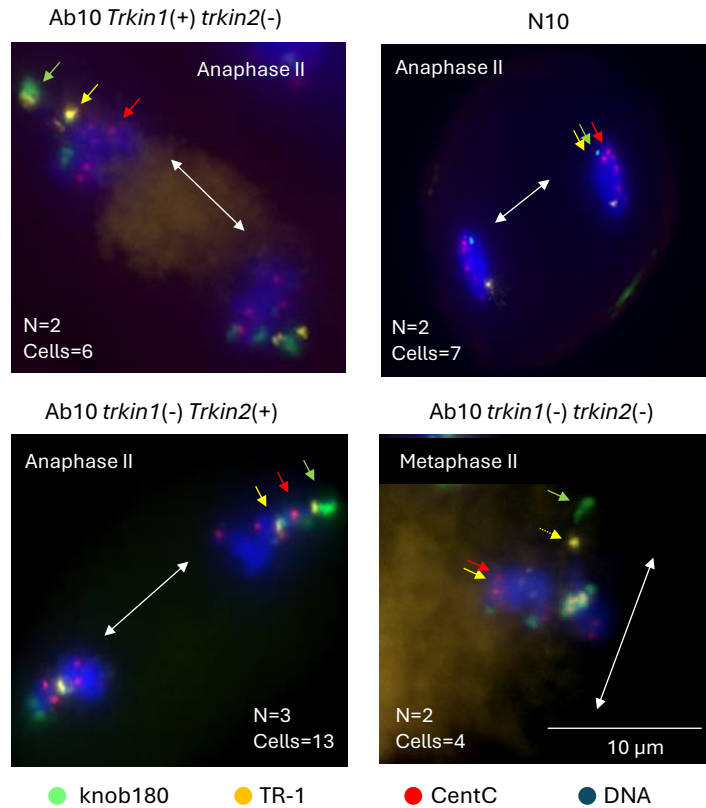


Figure 3.7: **FISH for Neocentromere Activity in Various *trkin* Genotypes Male Meiocytes.**

All plants were homozygous for their respective genotype. All images represent male meiotic anaphase II except the Ab10 *trkin1*(-) *trkin2*(-) which represents male meiotic metaphase II. TR-1 and knob180 neocentromeres are known to appear in these stages (Dawe 2022). Red marks CentC, green marks knob 180, yellow marks TR-1, blue marks DNA. The white double-sided arrows indicate the spindle axis, showing which way the chromosomes were moving at the time of fixation. In the absence of TRKIN activity, TR-1 (small yellow arrows) should be located behind the centromeres (small red arrows). The yellow dot that is off the metaphase plate in the lower right panel (dotted yellow arrow) is being pulled by the large knob180 knob (this is likely Ab10 itself). N indicates the number of individual plants observed, cells indicates the number of appropriately staged same phenotype cells observed.

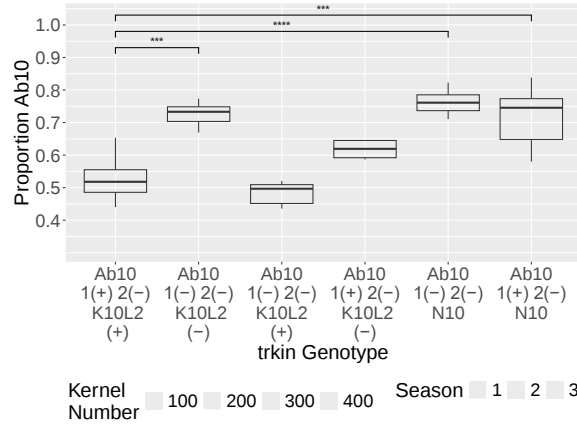


Figure 3.8: Effect of *Trkin* on the meiotic drive of Ab10 when paired with K10L2. The plot shows meiotic drive as measured by the percentage of kernels carrying the *R1* allele linked to Ab10. All plants were grown in the greenhouse in Athens, GA. Each dot represents an individual plant. Season refers to a group of plants grown at the same time. Seasons 1 and 2 were conducted in the same background while Season 3 was conducted in a different background. Season 1 and 2 of the Ab10 *trkin1*(-) *trkin2*(-) and K10L2 *trkin*(-) had *cas9* segregating, refer to Figure S3.9 for details. The multi-way ANOVA model was Proportion Ab10 ~ *Cas9* genotype + season + *trkin* genotype. *Cas9* genotype = F(1,63)=9.656, p=0.00; Season = F(2,63)=0.520 p=0.59726; *trkin* genotype= F(5,63)=19.495, p= 1.11e-11. Tukey's HSD Test for multiple comparisons found that the mean value of Ab10 *Trkin1*(+) *trkin2*(-) / K10L2 *Trkin*(+) was significantly different from Ab10 *trkin1*(-) *trkin2*(-) / K10L2 *trkin*(-) (p=7.364643e-04, 95% C.I.=[3.836241-20.005130), Ab10 *trkin1*(-) *trkin2*(-) / N10 (p=9.996369e-09, 95% C.I.=[13.392222-31.564484) and Ab10 *Trkin1*(+) *trkin2*(-) / N10 (p=1.775886e-04, 95% C.I.=[5.723491-24.404879]). Only significant relationships to Ab10 *Trkin1*(+) *trkin2*(-) / K10L2 *Trkin*(+) are shown, refer to Figure S3.9 for all significant relationships. *=<0.05, **=<0.01, ***, <0.001, ***=0.

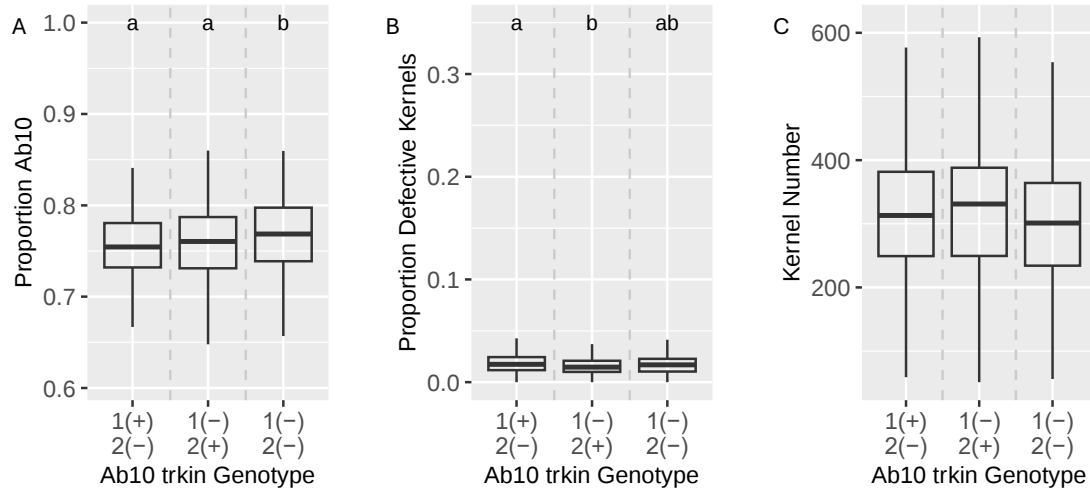


Figure 3.9: Ab10 Drive and Plant Fitness Effects of *Trkin* In Ab10 Heterozygotes. The plot shows meiotic drive as measured by the percentage of kernels carrying the *R1* allele linked to Ab10. Plants were grown in randomized order in a field in Molokai Hawaii. Each dot represents an individual plant. A. Drive: Multi-way ANOVA model was $\text{sqr}(\text{Proportion Ab10-I}) \sim \text{field x} + \text{field y} + \text{field edge} + \text{kernel sorter} + \text{trkin genotype}$. Field x = $F(1,941)=0.331$, $p=0.56$; field y = $F(1,941)=0.135$, $p=0.71$; field edge = $F(1,941)=5.475$, $p=0.02$; kernel sorter = $F(4, 941)=1.392$, $p=0.23$; *trkin* genotype = $F(2,941)=6.986$, $p=0.00$. B. Proportion of Defective kernels as a proxy for kernel abortion. Defective Kernels: Kruskal Wallis test model was: $\text{Proportion Defective Kernels} \sim \text{trkin Genotype}$. $H(2)=10.642$, $p=0.00$. Wilcoxon rank sum test found Ab10 *Trkin1*(+) *trkin2*(-) (mean 0.214) was significantly different from Ab10 *trkin1*(-) *Trkin2*(+) (mean = 0.0173, $p=0.0036$), but was not significantly different from Ab10 *trkin1*(-) *trkin2*(-) (mean = 0.0181, $p=0.1781$). C. Kernel Number: Multi-way ANOVA model was $\text{Kernel Number} \sim \text{field x} + \text{field y} + \text{field edge} + \text{kernel sorter} + \text{trkin genotype}$. Field x = $F(1,941)=1.785$, $p=0.18$; field y = $F(1,941)=3.538$, $p=0.06$; field edge = $F(1,941)=12.734$, $p=0.00$; kernel sorter = $F(4,941)=2.188$, $p=0.07$; *trkin* genotype = $F(2,941)=1.726$, $p=0.18$.

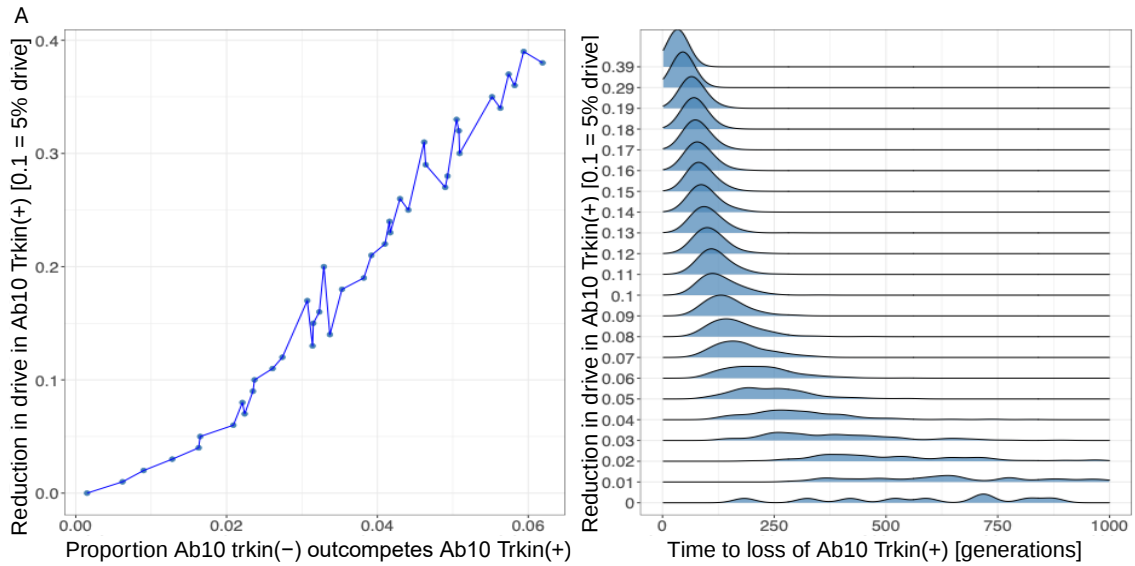


Figure 3.10: **How long can Ab10 *Trkin*(+) persist in a population being invaded by Ab10 *trkin*(-)?** Simulations were run stochastically, modelling drift following a multinomial distribution, at an initial frequency of 6% for Ab10 *Trkin*(+) and K10L2 and $1/N_e$ for Ab10 *trkin*(-) using $N_e=10,000$ and for $0 < \delta_1 < 0.4$. Each simulation was iterated 10,000 times. A. Proportion of realizations Ab10 *trkin*(-) successfully invades into the population and replaces Ab10 *Trkin*(+). The parameter on the y-axis is represented by δ_1 in the model. Note that these proportions are small since Ab10 *trkin*(-) was often lost due to drift. B. Density distribution for the number of generations Ab10 *Trkin*(+) can persist in a population upon invasion by Ab10 *trkin*(-). The parameter on the Y-axis is represented by δ_1 in the model.

CHAPTER 4

GENETIC AND ENVIRONMENTAL INFLUENCES ON THE GEOGRAPHIC DISTRIBUTION OF THREE CHROMOSOMAL DRIVE HAPLOTYPES IN MAIZE

M. J. Brady & R. K. Dawe, Genetic and environmental influences on the geographic distribution of three chromosomal drive haplotypes in maize.

To be submitted to *PNAS*.

Abstract

Meiotic drive elements are features of the genome that are transmitted to progeny at frequencies that exceed Mendelian expectations, often to the detriment of the organism. In maize there are three prevalent chromosomal drive systems known as abnormal chromosome 10 (Ab10), K10L2, and the B chromosome. There has been much speculation about how these drivers might interact with each other and the environment in traditional maize landraces and their teosinte ancestors. Here we used genotype by sequencing (GBS) data to score more than 10,000 maize and teosinte lines for the presence or absence of each driver. Using a GWAS approach we identified unlinked genetic modifiers that both negatively and positively associate with selfish genetic elements. We then assessed the contributions of population structure, genetic modifiers, and the environment on the distribution of each chromosomal driver. We found that the distribution of Ab10 is influenced primarily by genetic modifiers, while the distributions of K10L2 and the B chromosome are affected by all classes of variables. While each meiotic driver has a unique relationship to its genomic and abiotic environment, the drivers are ubiquitous among maize landrace and teosinte populations and likely will persist without active selection against them. Most modern maize inbreds have already been selected to lack these selfish elements.

Introduction

While most genes in most species are transmitted in predictable Mendelian patterns, there are striking exceptions. Genes, gene complexes, sections of chromosomes, and entire chromosomes have evolved mechanisms that ensure they are transmitted at higher frequencies than would be expected based on chance (1–4). These genetic elements generally confer no

selective advantages to the species, and are often deleterious, but are nevertheless maintained in populations based on their selfish properties. They are often described with the catch-all term meiotic drive [\(5\)](#), though only a subset of meiotic drivers affect meiosis. Meiotic drivers that manipulate meiosis often do so in species where female meiosis results in only one functional egg cell. An example is the preferential transmission of larger centromeres towards the egg cell in some mice lines [\(6\)](#). A more common class of meiotic driver interferes with the function of male gametes, often by setting up a dynamic where sperm or pollen are killed by a toxin unless they inherit the antidote present on the driving chromosome [\(7\)](#). The term meiotic drive is also used to describe the maintenance of inert B chromosomes [\(8\)](#), as well as many other varied phenomena, including biased gene conversion processes [\(9\)](#), mobile toxin-antidote systems [\(10\)](#), and engineered gene drive systems based on CRISPR methods [\(11\)](#).

The maize genome contains at least three meiotic drive elements that distort transmission of large chromosomal regions: Abnormal Chromosome 10 (Ab10), K10L2, and the B chromosome. Ab10 is a large variant of normal chromosome 10 (N10) that acts as a female meiotic driver. Approximately 14% of the maize genome is composed of tandem repeats arrays called knobs [\(12\)](#). They come in two classes defined by their repeat element: TR-1 and knob180. The Ab10 haplotype contains knobs of both classes with the knob180 knob being one of the largest knobs in the maize genome [\(13, 14\)](#). The Ab10 haplotype encodes two kinesin proteins, KINDR and TRKIN, which interact with knob180 and TR-1 knobs respectively. The kinesins pull the knobs ahead of the centromeres during meiotic anaphase resulting in their preferential transmission to the egg cell [\(15, 16\)](#). By this mechanism the Ab10 haplotype as well as knobs throughout the genome are preferentially transmitted to ~60-80% of the next generation [\(16\)](#). Ab10 is present at frequencies of 5-13% in teosinte (the ancestor of maize) and maize landraces

(17–19), but is prevented from going to fixation because it impairs fitness when it is homozygous (15). Prior data suggest that overall knob content and Ab10 may be correlated with lower altitudes and latitudes (20, 21). Ab10 is recognized as an important driver of maize genome evolution (20, 22).

The K10L2 variant was originally described as one of the many knobs in the maize genome (18). It was not known to have any special properties until FISH studies revealed that it is composed of two closely spaced, unusually large TR-1-rich knobs. Its location on chromosome 10L in the same general region as the Ab10 haplotype raised suspicion that it might be related to Ab10 in some way. Detailed analyses revealed that K10L2 encodes TRKIN and shows 1-2% meiotic drive when paired with normal chromosome 10 (N10) (23). These and other data suggest that the TRKIN/TR-1 system likely evolved as an independent drive system (16). When Ab10 is paired with K10L2, Ab10 drive is severely suppressed, demonstrating K10L2 not only can drive itself but compete with the stronger Ab10 drive system. More recent data demonstrate that the region between the two TR-1-rich knobs is similar in sequence to a portion of at least one structural variant of Ab10 (Ab10 Type I), suggesting that Ab10 may have subsumed the K10L2 haplotype in recent evolutionary history (13). K10L2 is present at lower frequencies in maize landraces than Ab10 (17), but the fact that at least one traditional inbred line is homozygous for K10L2 suggests that it may not have severe effects on fitness when homozygous (23).

The B chromosome is a ~150 Mb supernumerary chromosome composed primarily of transposable elements (TEs), organellar sequences, and a B-specific repeat element (24). The B chromosome can accumulate to high copy numbers by a mechanism that takes advantage of the fact that there are two sperm in each pollen grain; one fertilizes the egg cell and the other

fertilizes the central cell that gives rise to the starchy endosperm. The B chromosome normally non-disjoins at the second pollen mitosis and the sperm carrying two copies of the B chromosome preferentially fertilizes the egg (25). There is known variation among lines for the efficiency of the second step. Most lines allow the B chromosome to preferentially fertilize the egg, but multiple lines do not (8), or even reverse it, such that the sperm carrying the B chromosome preferentially fertilizes the central cell (26). In natural populations, B chromosomes are found in about 8% of landraces with copy numbers that are usually between 1-3 but may be as high as 14 (17, 27). Under experimental conditions as many as 30 copies of the B chromosome have been observed in a single corn plant, but at higher than 15 copies the plants display reduced seed set and pollen viability (28). In Argentina and Bolivia, B chromosomes were found to be more prevalent at high altitudes, but in Arizona this trend was not observed (27, 29–31).

Little is known about how different meiotic drive systems within the same species might interact with the environment and each other. We know that all three of the chromosomal drivers can substantially increase genome size (20, 22, 23, 28), which may be associated with slower growth (21). Prior data show that maize genome size is negatively correlated with altitude and latitude (21, 32, 33). A general expectation would be that there are fewer drivers, perhaps with non-overlapping distributions, at high altitudes. Supporting this view is the fact that negative correlations have been noted between B copy number and knob content in several studies (31, 34–36). The drivers may also interact with the environment in a more direct manner. For example, the *Drosophila neotestacea* Sex ratio meiotic driver's distribution is closely associated with winter temperature (37). The three maize drivers may also interact with each other in unpredictable ways. While K10L2 acts as its own driver, it also reduces the drive of Ab10 when

paired in opposition (13, 23). In some backgrounds, the B chromosome causes the breakage of chromosomes at knobs, including Ab10 (38, 39). Environmental specialization and negative interactions among drivers would be expected to skew their distributions on a spatial scale.

Genetic variation outside of the drive haplotypes is also expected to alter the frequencies of drive systems in nature. There is extensive genetic variation for traits associated with meiosis and fertilization. Alleles that reduce the efficiency of drive (suppressors) should be selected for when the fitness consequences of the drive system are high (2, 40–42). It is possible, for instance, that the resistance of some lines to the preferential fertilization of B chromosomes reflects ongoing selection against high B chromosome copy numbers (8). We might also expect maize to segregate for suppressors of Ab10 drive, which skews genetic segregation across the genome and reduces plant fitness when homozygous. In past studies, suppressors have been identified by their phenotypes, but in principle they could also be detected using genome-wide association studies (GWAS) as loci that are negatively associated with drive haplotypes.

Here we developed a high throughput scalable method to detect large structural variants, like chromosomal drive haplotypes (CDH), in genotype by sequencing data. Using genotype data from ~10,000 individuals we assessed the frequency of each CDH individually as well as in combination. We then determined how their distribution relates to genetic background, population structure, and the environment. In open pollinated teosintes and maize landraces, the distributions of all three of CDHs are influenced by genetic modifiers, with the environment having a significant effect on K10L2 and the B chromosome but not Ab10. The CDHs are almost completely absent from modern inbred lines, which is a likely result of selection against their negative fitness properties.

Results

Genotype By Sequencing Data Can Reliably Detect Large Structural Variants

Previously the only ways to detect chromosomal drive haplotypes (CDH) were by cytological methods or PCR assays (17–19). A scalable way to identify CDH would be by sequencing. Fortunately, there is a wealth of publicly available genotype-by-sequencing data (GBS) for diverse maize landraces and teosinte (43–45). GBS is a form of reduced representation sequencing where only a relatively small portion of the genome is sequenced at modest coverage (46).

We speculated that the low coverage sequence data (not the SNPs) from prior GBS studies might be useful for identifying large chromosomal drive haplotypes. To test the feasibility of this idea, we first generated GBS data from a collection of control lines carrying different isolates of Ab10, K10L2, and B chromosomes, as well as associated controls. The GBS data was then mapped to reference genomes carrying each drive haplotype: Ab10 (47), K10L2 (48), and the B chromosome (24). We then computed a tag index, which is a function of both the number of tag sites and the read depth over tag sites, in 1 Mb windows (Figure S4.1a). When the tag index is plotted as a heat map, the presence or absence of a CDH becomes visually apparent (Figure 4.1a). We then automated the scoring of CDHs using an iterative k-means clustering approach (Figure 4.1a and Figure 4.1b) and achieved 100% accurate discrimination of the presence and absence of each CDH in our control dataset. To estimate CDH copy number, we normalized CDH tag depth by the average tag depth across all single copy core genes (12) and correctly identified all Ab10 and K10L2 homozygous samples in our control set (Figure 4.2c). We applied this same method to estimate the copy numbers of B chromosomes, though in this

case we did not know the copy number in our control samples (so we refer to our estimates as pseudo-copy number).

To detect CDHs in maize lines of unknown CDH status we used a similar stepwise k-means clustering approach. Our Ab10 detection pipeline cannot detect K10L2, and our K10L2 pipeline cannot distinguish Ab10 from K10L2. Therefore we ran the Ab10 pipeline first, and then ran samples called as negative through the K10L2 detection pipeline. Similarly, to accurately differentiate high copy B chromosome lines from low copy lines, we ran two different models in sequence (Figure 4.1b). To ensure that clustering was driven by our control samples, we used roughly equal numbers (CDH positive and CDH negative) of relevant controls available for each CDH and randomly introduced between 10% and 25% of the experimental samples. Every sample was assayed for each CDH 125x to gain an estimate of call confidence. We obtained confident calls (>95% calls) for more than 99% of samples assayed for each CDH (Figure 4.2b, Figure S4.1b).

When developing our set of Ab10 control samples we included isolates from three major cytological types, known as Ab10-I, Ab10-II, and Ab10-III, that differ in the appearance of the major knobs on the haplotype (19). We developed a random forest model to detect Ab10 type from this control data. When this model was applied to experimental samples it became clear that there is more diversity in natural Ab10 samples than is present in our three control types (Figure 4.1a, Figure 4.2b). Using a confidence threshold that maintained the visual differences between types, only 47 of 352 experimental samples were classable (Figure S4.2a). To explore the variation in Ab10 type further, we performed a PCA on the scaled tag index of diagnostic regions. The results suggest that many of the unclassable Ab10 haplotypes are intermediates between the known types (Figure S4.2b). Ab10-I and Ab10-II have been shown to recombine

with each other in experimental populations (39), and some of the tax index patterns suggest recombinants are also present in natural populations (Figure S4.2b).

Frequencies of Three CDH in Maize

Using our methods, all three CDHs are present at similar frequencies in maize landraces and teosintes as previously reported based on smaller sample sizes (Figure 4.2a). Interestingly, we find that Ab10 and the B chromosome as well as K10L2 and the B chromosome occur together roughly as frequently as expected by chance (Table S4.1). This suggests that neither the chromosome breakage phenotype observed in some B chromosome lines (25) nor the opposing effect of Ab10 and the B chromosome on genome size (20, 27) seem to be driving their joint distribution. Unfortunately, we could not score Ab10 and K10L2 in the same sample, so we are unable to estimate how often they occur together.

We identified all three CDHs in inbred maize lines (45), though at very low frequencies (Table S4.1). K10L2 had previously been detected in an inbred line (23), while Ab10 had not, and was presumed to be too deleterious to be tolerated in an inbred condition (16). B chromosomes are also thought to be absent from inbred lines (28). We obtained seeds from two inbred lines scored as positive for each CDH and discovered that in all cases the lines were segregating for the presence or absence of the CDH (that is, while these lines are homozygous for most of the genome, the CDH chromosome is not) (Table S4.2).

Relationship of Three CDH In Maize to Genetic Variants

We next wanted to understand the relationship between each of our CDHs and genetic variation in the genome. To do this we isolated maize landraces and performed a genome wide

association study (GWAS) on high confidence SNPs that did not overlap any CDH or transposable element. The SNPs were identified from the GBS data by aligning reads to the Mo17 reference using high stringency criteria (mapping quality ≥ 20 , minor allele frequency ≥ 0.05 , and others), resulting in ~50,000 usable SNPs (Figure S4.3). Given our binary traits, relatively large number of individuals, and expectation of few large effect loci we chose to impose a very low significance threshold of 5×10^{-8} .

We found that Ab10 was positively associated with seven SNPs on chromosomes 3, 4, 8, and 9 and negatively associated with four SNPs on chromosome 1, 3, and 10. The positively associated SNP on chromosome 9 is likely an alignment artifact, as it occurs within a sequence that has homology within the Ab10-I v1 reference (47) (see methods; this SNP was not used in subsequent analysis). K10L2 was positively associated with 5 SNPs on chromosomes 1, 4, 5, and 6 and negatively associated with two SNPs on chromosome 8 (Table S4.3). The B chromosome was positively associated with four SNPs on chromosomes 3, 4, and 6, and negatively associated with one SNP on chromosome 3 (Figure 4.3a, Table S4.3). Two positively associated SNPs on chromosomes 3 and 6 are likely artifacts (see methods). B chromosome copy number was 4 SNPs on chromosomes 2, 3, and 6. The associated SNP on chromosome 6 is likely an artifact and was excluded from further analyses (see methods). Of the total 27 associated SNPs, 16 overlap annotated genes (Figure 4.3a, Table S4.3, Figure S4.4). Henceforth we refer to loci associated with each CDH as putative modifiers (suppressors/enhancers) for ease of interpretation.

Many of the associated loci are represented by a single SNP instead of the often observed clusters of SNPs. This is an outcome of the relatively low number of SNPs used in our association study (~50,000), which was a result of the low coverage of GBS data and strict

filtering parameters we used to exclude reads that might be derived from large chromosomal drive haplotypes themselves. Since linkage disequilibrium in maize decays on the order of ~2-3 kb (49), we would only rarely expect to find two or more SNPs in association with a single modifier locus. It is also likely that modifiers were missed in our analysis due to low marker density.

Combined effects of Genetic variants and Environment on the distribution of CDH

To test the impact of location and environment on CDH distribution, we chose ~ 5000 maize landrace and teosinte lines that were confidently identified as CDH positive or negative and had GPS coordinates for their collection location (Figure 4.2 A, Table S4.1). We first assessed the effects of elevation while also accounting for population structure. In contrast to previous studies (19, 27), we found no relationship between Ab10 and B chromosome presence/absence and elevation or B chromosome copy number and elevation (Figure S4.5). However, we did identify a weak but significant positive relationship between K10L2 and elevation.

We went on to develop models to test the effects of specific climatic variables (50) and soil conditions (51). We began each model using elevation, Mean Temperature of Warmest Quarter, Precipitation of Warmest Quarter, Temperature Seasonality, Precipitation Seasonality, average annual solar radiation, average annual wind, average annual water vapor pressure (50), soil nutrient availability, soil rooting conditions, soil oxygen availability to roots, soil excess salts, and soil toxicity (51). We also included B chromosome presence/absence for Ab10 chromosome models to verify that there is no relationship as suggested in Table S4.1. For the B chromosome, we tested the relationship between Ab10 and K10L2 in separate models due to the

intricacies of their detection. For each CDH we generated a simplified model by removing variables that were not significantly associated (p value greater than 0.01) one at a time until all variables were significantly associated (p value of less than 0.01). The model we simplified for the B chromosome did not include Ab10 or K10L2. In all cases we found that the presence of other CDHs did not have a significant effect as suggested in Table S4.1. Additionally, we found that elevation was not significantly associated with any CDH when considering other environmental variables (Figure S4.6).

We repeated the above process including genetic modifiers in the initial models as well as performed logistic regressions to identify associations between each CDH and the three classes of variables — population structure, genetic modifiers, and environment. Population structure was included to illustrate the impact of genetic relatedness on the distributions of CDHs. We tested all 10 principal components of population structure that were included in the GWAS analysis (Figure 4.3a). We calculated the amount of deviance explained by the full model (Figure 4.3b) and each of the variables separately (Figure 4.3c). Due to interactions among the variables (which we did not pursue here), the deviation explained by each class of variables individually does not sum to the deviation explained by the full model.

Ab10

Ab10 significantly associates with 6 genetic modifiers and four principal components of population structure accounting for 18.4% of the deviance (Figure 4.3b,c). The environment seems to have little if any effect on the distribution of Ab10.

K10L2

K10L2 significantly associates with population structure, both genetic modifiers, and environmental factors accounting for 18.5% of the deviance (Figure 4.3c). It is associated with lower water vapor pressure. Further we found that K10L2 was overrepresented in poor quality soil specifically with respect to excess soil salts, soil rooting conditions, and soil oxygen (Figure 4.3b).

B Chromosome

The B chromosome significantly associates with population structure, genetic modifiers and environmental factors accounting for 21.9% of the deviance. The B chromosome is more likely to occur in regions with higher temperature seasonality and solar radiation (Figure 4.3b,c). We also found the B chromosome was overrepresented in soil with severely growth limiting excess soil salt.

The B chromosome was exceptionally strongly associated with Chr3 SNP 2 (5366738) suggesting it is an artifact derived from the B chromosome itself. However, it does not share homology with the assembled B chromosome (24). We did not include it in our main analyses due to the very high likelihood that it is an artifact (Figure 4.3). Given the very small chance that this SNP is not an artifact and is instead a crucial part of the B chromosomes accumulation mechanism we chose to run separate analyses including it (Figure S4.7). In these analyses we found similar results with the B chromosome significantly associating with population structure, genetic modifiers and environmental factors however they accounted for 56.7% of the deviance. Again, we believe the results presented in Figure 4.3 are exceptionally more likely, though those presented in Figure S4.7 remain a possibility. In these analyses we found the B

chromosome was associated with higher temperature and precipitation seasonality and was overrepresented in soil with severely growth limiting oxygen levels. We present these results for completeness but believe that those shown in Figure 4.3 are more accurate.

B chromosome copy number is associated with population structure, two genetic modifiers, and higher temperature seasonality. These factors account for 17.4% of the deviance.

Discussion

In this work we used GBS data to identify chromosomal drive haplotypes in over 10,000 maize and teosinte accessions with the aim of better understanding how drive haplotypes interact with each other and the environment. While GBS was developed as a method to score SNPs, our approach using k-means clustering illustrates that the low coverage sequence data can also be used to identify large haplotypes that may not be present in the primary genome references.

There have been few empirical studies of the interactions between selfish genetic elements in populations, primarily related to mobile elements, indicating both mutually beneficial and antagonistic dynamics ([52–54](#)). The effects of both B chromosomes and Ab10 on increasing genome size ([20, 25](#)) and oft-observed inverse correlation between B chromosomes and knobs ([31, 34–36](#)) suggested that Ab10 and B chromosomes might be antagonistic, and rarely occur together. However, we found that they appear together almost exactly as frequently as expected by chance. We also anticipated that both Ab10 and B chromosomes would occur on elevational clines however, we observed no correlation with elevation for either driver. The prior literature has generally been interpreted in the context of an assumption that genome size imparts a selective burden on maize plants ([21](#)). While this is true, the level of selection may be weaker than is commonly assumed. Recent results suggest that in the large-genome maize plant, a gain

of 14 Mb results in a 0.1% reduction yield (55). By this reasoning, a single ~30 Mb knob (14), the ~85 Mb Ab10 haplotype (13), or the ~150 Mb B chromosome would be expected to result ~<1% drop in yield, which may not be sufficient to counteract the selfish properties of these powerful drivers.

The maize chromosomal drivers have effects beyond simply increasing genome size. Ab10 is likely to have fitness costs associated with skewing segregation across the genome, as well as its severe effects when homozygous (15, 22). If left unchecked, the B chromosome can accumulate to 15-30 copies and visibly impair plant performance (28). When fitness is impaired by a selfish element, suppressors are expected to evolve (42). Here we identified twenty four likely modifiers of drive using a GWAS approach. Interestingly, they both positively and negatively associate with CDHs which we refer to as enhancers and suppressors respectively. The preferential fertilization mechanism of the B chromosomes is known to be affected by alleles elsewhere in the genome (25), and we identified multiple loci that both positively and negatively associate with B chromosomes. For K10L2, which has weak drive and likely minor fitness effects when homozygous (23), we identified four enhancers. For Ab10, a strong driver with sweeping impacts on the maize genome, we found four suppressors of drive as well as six enhancers. Four of these enhancers overlap the same gene, a C3H-transcription factor 350, and presumably functionally behave as one enhancer for a total of 3. The enhancers may be ameliorating the negative fitness effects of Ab10 (for instance when it is homozygous), allowing it to be better tolerated.

Selfish genetic element distribution has generally been thought to be limited by genetic factors with occasional relationships to the environment (37, 42, 56). Here we assessed the importance of population structure, genetic modifiers, and the environment to the distribution of

CDHs in an unusually large sample. Ab10 distribution appears to be influenced only by genetic modifiers and population structure, reflecting its antagonistic relationship to the rest of the genome. In contrast, the distributions of K10L2 and the B chromosome are explained by population structure, genetic modifiers, and the environment in roughly equal proportions. If anything, population structure is the largest contributor. While each element has a unique relationship to genetic and environmental variables, the effects are relatively small, suggesting that these ancient and thoroughly distributed drivers may have largely evaded these constraints. The major limit to spread has been the shift from propagating maize as open pollinated landraces to inbred lines, where the fitness burdens associated with CDH are not tolerated. Outside the single inbred line where K10L2 was discovered (CI66), we did not find any other inbred fixed for a chromosomal drive haplotype amongst the ~4500 inbreds assayed.

Methods

GBS Sequencing Controls

Our control GBS data were obtained from two different sources, Cornell and CD Genomics. For the Cornell dataset, plants known to be heterozygous for Ab10-I-MMR or Ab10-II-MMR were self crossed to create populations segregating for either Ab10 structural variant. Ab10 was marked by an allele of the *colored1* gene (*R1*) which makes the kernels purple (Table S4.4). There is an approximately 2% chance of recombination between Ab10 and *R1* (57). We extracted genomic DNA from Ab10 positive and N10 kernels using a CTAB extraction (58). Using these DNA samples, GBS libraries were prepared and sequenced on an Illumina HiSeq 2000 in accordance with (46) by the Genomic Diversity Facility, Cornell University (this facility is no longer in operation).

For the CD genomics dataset, we grew plants from 49 Ab10 controls from 11 genetic backgrounds, 13 K10L2 controls from 2 genetic backgrounds, 18 B chromosome controls from 5 genetic backgrounds, and 18 no CDH controls from 7 genetic backgrounds (Table S4.4). We first verified that the controls were CDH positive or negative by extracting DNA using a CTAB extraction (58) and performing PCR for *Kindr*, *Trkin*, or the B repeat (Table S4.5). We then sent leaf tissue to CD Genomics (Shirley, NY) who extracted DNA using QIAgen DNeasy Plant Kits. They prepared GBS libraries as in Elshire et al (2011) with minimal modification. Basically this involved digesting DNA with ApeKI (New England Biolabs, Ipswich, MA), adding barcoded adapters, and sequencing the libraries on an Illumina NovaSeq6000 using a 150×2 paired-end sequencing protocol.

After receiving the data we identified several lines that were misclassified (W23_AB10-I.11.DC1, W23_AB10-I.13.DC1, W23_AB10-II.36.DC1, W23_N10.14.DC1 NSL-2833_B-Chrom.2.DC2, B542C_L289_B-Chrom.1.DC2); these were either reclassified or excluded from further analysis.

Obtaining GBS Data

We obtained existing GBS sequence reads from the authors of three prior publications (43–45). These data were generated following the protocol of (46). The data from (44) were in the format of demultiplexed qualified reads; we converted them to a format usable for TASSEL using custom R v4.3.1 code and barcode faker (59). The data from each plant described in (44) was split into approximately four libraries as technical replicates, and these were summed during analysis (see below).

K-means Clustering of Controls

We first needed to establish that it is possible to differentiate Ab10, K10L2, and N10 from each other as well as B chromosome presence/absence from genotype by sequencing data alone. We began by mapping the full set of GBS data described above to the B73-Ab10 v2 (13) genome with the B chromosome appended (24) and the CI66-K10L2 genome (48) using TASSEL v5.2.44 (59) and BWA v 0.7.17 (60). Using TASSEL v5.2.44 (59), we obtained the presence of each tag and the number of associated reads in each sample for both the B73-Ab10/BChromosome assembly and the CI66-K10L2 assembly (TagByTaxa Table). We converted the alignments to a bed file using samtools v0.1.20 (61), and bedtools v2.29.2 (62) to locate the position of each tag in the TagByTaxa table. For each assembly, we summed the tag counts for all technical replicates per biological individual for (44). Unless otherwise noted all further steps were carried out using custom R v4.3.1 code. In order to normalize across libraries of varying size, we calculated reads per million for each tag in each individual sample in both assemblies separately. We calculated the minimum proportion of missing data for blank samples (where no genomic DNA was added; this represents sequencing background), and subtracted 0.001. We then removed any sample with more missing data than this cut off, as well as any tag with a BWA mapping quality of less than 20. We verified that all datasets were affected similarly by these filters and extracted all tags on each CDH. We then calculated the tag index in non overlapping 1 Mb bins across all CDHs ($\sqrt{c} + d$), where c is the count of tags mapped to that bin and d is the sum of the read depth of all tags in that bin (Figure S4.1a). Then we isolated our control data with known CDH status and visualized the tag index in a min/max scaled heat map. We did not observe any visual distinction between our two sets of control data (the Cornell and CD genomics datasets), indicating that this method is robust to differences in sequencing.

This is important as the experimental data set is pooled from multiple data sources. We found that the CDH positive and negative lines were visually very distinct (Figure 4.1a).

We needed to verify that we could correctly and automatically detect CDH presence or absence in our control data set. We chose to use an iterative k-means clustering method on the scaled tag index in order to do this. The entire pipeline outlined below was performed using custom R v4.3.1 code unless otherwise stated. We selected only the CDH specific portions, or those regions that showed a stark difference between CDH positive and negative controls (Figure 4.1a). For each CDH we had high and low copy number controls. For Ab10 and K10L2 high copy number controls were homozygous plants with two copies of the CDH and low copy number controls were heterozygous plants with one copy of the CDH. For the B chromosome, the exact copy number was unknown and they were divided into high and low copy number controls by visual comparison of the min/max scaled tag index heat maps (Figure 4.1a). We performed SGE detection with the high and low copy number controls separately in order to ensure that clustering was based on the distinction between presence and absence rather than copy number. First, we split our group of control samples into three randomly selected groups. On each subsample we performed k-means clustering ($k=2$). If a cluster was made of at least 80% CDH (Ab10, K10L2, B chromosome) or non CDH (N10, no B chromosome) samples it was assigned as such (this is the naming step in Figure 4.1b). The k-means cluster assignment was then compared to the true CDH status of that sample in order to determine if the k-means clustering assigned the sample correctly. This was repeated 100 times for each sample, where each iteration involved a different, randomly selected set of control individuals (Figure 4.1b). For the B chromosome low copy number model we didn't have adequate samples to break them into three subsamples so they were clustered as a single sample. Using this method we were able to

correctly identify the CDH status of all of the experimental samples 100% of the time, regardless of where the GBS data were acquired (either from Cornell or CD genomics).

Use of K-means Clustering on Experimental Samples

Having established that the method correctly identifies each CDH in control data, we then extended it to our experimental samples ([43–45](#)). We identified the chromosome 10 CDHs and the B chromosomes in two separate workflows before finally estimating their copy number (Figure 4.1b). The entire pipeline outlined below was performed using custom R v4.3.1 code unless otherwise stated.

The general approach was to select roughly equal numbers of the appropriate controls (positive and negative) for each CDH and then randomly add a small number of experimental samples. For Ab10 and K10L2, the number of experimental samples added was 25% the number of controls, for the B chromosome, the number of experimental samples added was 10% the number of controls. Then we performed k-means clustering (Figure S4.1b). If a cluster was made of at least 80% CDH (Ab10, K10L2, B chromosome) or non CDH (N10, no B chromosome) samples it was assigned as such. We verified that all control samples were correctly identified. If they were not, we repeated the k-means clustering until all controls were correctly identified. We then assigned all experimental samples the class of their k-means cluster. We repeated this workflow until all experimental samples had undergone one round of k-means clustering. Then we repeated the entire process 125 times to obtain 125 independent calls for the CDH class per experimental sample. To make the final CDH class calls, we required that the experimental sample be called the same class 95% of the time. All other samples were labeled ambiguous (Figure S4.1b).

Our Ab10 model is unable to distinguish K10L2 from N10, while our K10L2 model is unable to distinguish Ab10 from K10L2. Therefore we employed them one after the other. We ran the Ab10 model first and identified 394 Ab10 positive samples (Figure 4.2a, Table S4.1). We then isolated the samples called as N10 and ran the K10L2 model, identifying 310 K10L2 positive samples (Figure 4.2a, Table S4.1). We plotted all the Ab10 and K10L2 positive samples in single heat maps with ward.D clustering (Figure 4.2b).

The variability in B chromosome copy number in experimental samples sometimes caused our k-means clustering pipeline to fail (lines with many copies of the B chromosome sometimes formed their own cluster). Therefore we used a two-step process. First we extracted all high copy number experimental samples using high copy number controls. Then we took all samples not identified as B chromosome positive in the high copy number iteration and ran them through the same model using the low copy number B chromosome controls. In this way we were able to extract all B chromosome positive samples without introducing unnecessary variation in the k-means clustering (Figure 4.1b). We identified 773 B chromosome positive samples (Figure 4.2a,b, Table S4.1). We plotted all B chromosome positive samples in a single heat map with ward.D clustering (Figure 4.2b).

Random Forest Modeling

We then attempted to differentiate the Ab10 types within our experimental classes. We first trained a random forest model on 70% of the Ab10 control data with known types (15). We checked the random forest models performance using the remaining 30% of the Ab10 control data. It correctly predicted type 100% of the time. We then applied that random forest model to all of our experimental samples. We required that 65% of decision trees call the same Ab10 type,

and all other samples were classed as ambiguous (Figure S4.2a). We selected this confidence threshold as it preserved the visually apparent difference between types when plotted as a heat map (Figure 4.1a). However, only 11.9% of Ab10 samples were classable in this manner. To better explore the Ab10 type we extracted the bins with the highest mean decreasing Gini in the random forest model, meaning the model suffered the most when these variables were excluded, and performed a principal coordinate analysis (Figure S4.2b).

Estimating the Copy Number of CDHs

In order to estimate the copy number for each CDH we needed an estimate of what the tag index of a single copy gene was. In order to do this we lifted over annotations from the B73 v5 reference (12) genome onto both the Ab10 and K10L2 assemblies (13) using liftoff v1.6.3 (63). We then extracted all single copy core genes, and calculated their tag index in 1 Mb bins (Figure S4.1a). Note that the 1 Mb bins refer to 1 Mb of single copy core gene sequence and not true genomic coordinates. Then we calculated the average tag index across all single copy core gene bins for each sample. We divided the average CDH specific tag index value by that sample's average single copy core gene tag index value. We refer to the ratio of CDH/single copy core gene tag index as pseudo copy number (Figure 4.2c).

PCR verification of CDHs in inbred lines

The frequency of CDHs varies in open pollinated populations like landraces and teosinte and are rarely if ever fixed (19, 23, 25). Thus while we may have scored one plant from a landrace as positive for a CDH, it is unlikely that the next plant we scored from the same population would have the CDH. However, we identified several maize inbred lines containing Ab10, K10L2, and B chromosomes (Table S4.1). We ordered two inbred lines called as positive

for each CDH from the Germplasm Resources Information Network (GRIN). We extracted DNA using a CTAB extraction (58) and performed PCR for *Kindr*, *Trkin*, or the B repeat to verify the CDHs presence (Table S4.2, Table S4.5).

GWAS

We generated artificial reference genomes with Mo17 (14) chromosome 1-10 and Ab10 (13), K10L2 (13), or the B chromosome (24). For the Mo17 Ab10 and K10L2 reference genomes we used samtools v1.18 (61) to truncate Mo17 chr10 at the beginning of the *colored1* gene, which traditionally defines the beginning of the Ab10 and K10L2 haplotypes (16). We then isolated the CDHs beginning at the *colored1* gene using samtools v1.18 (61) and SeqKit v0.16.1 (64). For the B chromosome we left all Mo17 chromosomes intact and appended the B chromosome (24). We modified the key for all samples such that all technical replicates from (44) were read into a single biological sample. We then used TASSEL v5.2.44 (59) and BWA v0.7.17 (60) to align reads from all samples to the Mo17 + CDH references. We filtered mapped tags to a mapping quality of 20 using samtools v1.18 (61) and called SNPs using TASSEL v5.2.44 (59). We then extracted SNPs on chromosomes 1-10 and the relevant CDH using bcftools v1.15.1 (61). We isolated only maize landraces using bcftools v1.15.1(61) because we believed genetic modifiers might be less frequent in inbred lines due to their very low frequency of CDHs across many generations (Table S4.1). Then we applied the following filters: Read depth >3 and <20, minor allele frequency >= 0.05, genotype quality >60, and a per sample missingness of 75% or less. We used BEAGLE v5.4 to impute missing data based on haplotypes found in our data (65). We did not use a reference panel due to concerns about the maintenance of genetic modifiers for CDHs in inbred lines. Using PLINK v1.9 (66) we removed plants with

more than 10% missing data. We were left with ~50,000 SNPs. We carried out a principal component analysis on the whole genome non-CDH SNPs to identify population structure in the data, and included the top 10 principal components in an association test as covariates using PLINK v1.9 (66). For Ab10, K10L2, and B chromosome presence/absence we used a logistic regression for B chromosome pseudo copy number we used a linear regression. We plotted the output using custom R v4.3.1 code.

GBS tags are just 64 bp (59). We know that some regions of each CDH are homologous to chromosomes 1-10 (13, 16, 23). It seemed possible that a GBS tag originating from a CDH could erroneously map to chromosomes 1-10 and create an erroneous association. We identified genes orthologous between the CDH and the Mo17 genome using OrthoFinder v2.5.5 (67) and identified any associated loci overlapping them using bedtools v2.31.0 (62). We removed any associated loci overlapping a CDH homolog. Additionally we removed any SNP overlapping an annotated transposable element (14) using bedtools v2.31.0 (62). Finally, we extracted 64 bp upstream and downstream of the SNP using Samtools v1.18 (61) and used BLAST v2.13.0 to compare the region to B73-Ab10 v2 reference (13). We removed any SNP that had a hit to the CDH of at least 62bp with a percent identity of 80% or greater. However, three loci had suspiciously high associations with Ab10 and the B chr. For Ab10, we used maizeGDB BLAST (68) to compare the SNP locus for Chr9 SNP 2 (29026173) to B73-Ab10 v1(47) and found it had 83% percent identity to Ab10. We suspect this associated SNP is an artifact and excluded it from further analysis. We repeated the same process for the two SNPs suspiciously highly associated with the B chromosome [Chr3 SNP 2 (5366738) and Chr6 SNP 1(124809142)]. While the locus surrounding Chr6 SNP 1 did not have homology to the B chromosome we found the gene it falls within (Zm00014ba298980) has a homolog on the B chromosome. We believe this SNP is an

artifact and excluded it from further analyses for the B chromosome and B chromosome copy number, which it was also positively associated with (Table S4.3). Chr3 SNP 2 (5366738) did not overlap with any annotated genes or the B chromosome, however the very high $-\log(p)$ value strongly suggests it is an artifact. We chose to exclude this SNP from the main analyses presented in Figure 4.3. However, we repeated the analyses including it in Figure S4.7 out of an abundance of caution.

We repeated the above procedure using all the same criteria on SNPs on the CDH as a control for loci linked to the CDH (Figure S4.8). We removed SNPs that did not occur in at least 75% of the samples, so any locus specific to the CDH would have been removed. Loci present in the inverted shared region of Ab10 could exist in individuals carrying Ab10 or N10 allowing them to pass the missing data filters. While K10L2 also has a shared region it is not inverted ([13](#)) and is known to recombine with N10 ([23](#)) so while many SNPs passed the filtering we would expect fewer of them to strongly associate with K10L2 presence. The B chromosome does not have a shared region thus most SNPs on it occur in less than 75% of samples and were removed by the missing data filter (Figure S4.8, Table S4.1).

GLM models on All CDHs

We obtained climatic data from WorldClim2 ([50](#)) and soil quality data from the FOA Harmonized world soil database ([51](#)). We chose to begin each model with environmental variables known to be associated with either maize or the CDHs: Elevation ([19, 27](#)), Mean Temperature of Warmest Quarter, Precipitation of Warmest Quarter, Temperature Seasonality, Precipitation Seasonality, average annual solar radiation, average annual wind, average annual water vapor pressure ([50](#)), soil nutrient availability, soil rooting conditions, soil oxygen

availability to roots, soil excess salts, and soil toxicity ([51, 69](#)). For Ab10 we included the B chromosome result as well. We could not include K10L2 because we cannot detect K10L2 in lines where Ab10 is present. For the B chromosome we first ran the full model including either Ab10 or K10L2 presence/absence to test for an association. We did not include Ab10 or K10L2 presence/absence in the model that was simplified due to the fact that we cannot detect K10L2 in lines carrying Ab10. We extracted environmental data for each collected sample using the raster package in R v4.3.1 ([70](#)). For solar radiation, average annual wind, and average annual water pressure we used custom R v4.3.1 code to generate the average value for each location from monthly data ([50](#)). We tested for collinearity between the environmental variables using custom R v4.3.1 code and found that elevation and soil nutrient availability had a greater than 70% correlation to other variables (Figure S4.9).. We excluded soil nutrient availability from all starting models due to its collinearity. We chose to include elevation in the starting models because it has been investigated previously ([19, 71](#)). We ensured all models were robust to variable order. We also included the first 10 genome wide SNP principal components generated as part of the GWAS. For the CDH presence absence we used binomial family models. For B chromosome pseudo copy number we log transformed the variable to make it normally distributed and used a gaussian family model. From the largest model we performed stepwise model simplification using custom R v4.3.1 code. We tested model fit using the DHARMa R package before and after simplification ([72](#)). We determined model fit was acceptable in all cases. Our models represented a relatively low proportion of the total deviance, thus we chose to apply an alpha value of 0.01.

To assess the relative contribution of population structure, genetic loci, and the environment we selected all significant loci from the GWAS using custom R v4.3.1 code.

Because some loci had more than one alternate allele, we began by coding these as factors to check if any of the second minor alleles had a significant relationship to any CDH. We found that they did not and removed them, coding the remaining minor allele additively. We then performed stepwise model simplification. We plotted the results using custom R v4.3.1 code. To partition the variance from the entire model into population structure, genetic loci, and environmental variables we first used custom R v4.3.1 code to remove all missing data to ensure the null model was identical between all runs. Then we ran individual models with only the remaining population structure, genetic loci, or environmental variables. We calculated the amount of deviance explained by each of these models and plotted them using custom R v4.3.1 code.

Author Contributions

MJB and RKD designed the research. MJB performed all research. MJB and RKD wrote the paper. All authors agreed that this work may be included in this dissertation.

References

1. J. H. Werren, Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl 2**, 10863–10870 (2011).
2. A. K. Lindholm, *et al.*, The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends Ecol. Evol.* **31**, 315–326 (2016).
3. J. A. Ågren, A. G. Clark, Selfish genetic elements. *PLoS Genet.* **14**, e1007700 (2018).
4. F. E. Clark, T. Akera, Unravelling the mystery of female meiotic drive: where we are. *Open Biol.* **11**, 210074 (2021).
5. L. Sandler, E. Novitski, Meiotic Drive as an Evolutionary Force. *Am. Nat.* **91**, 105–110 (1957).
6. T. Akera, E. Trimm, M. A. Lampson, Molecular Strategies of Meiotic Cheating by Selfish Centromeres. *Cell* **178**, 1132–1144.e10 (2019).
7. C. Courret, X. Wei, A. M. Larracuente, New perspectives on the causes and consequences of male meiotic drive. *Curr. Opin. Genet. Dev.* **83**, 102111 (2023).
8. R. N. Douglas, J. A. Birchler, “B Chromosomes” in *Chromosome Structure and Aberrations*, (Springer India, 2017), pp. 13–39.
9. B. L. Stoddard, Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* **19**, 7–15 (2011).
10. S. J. Saupe, H. Johannesson, On the mechanistic basis of killer meiotic drive in fungi. *Annu. Rev. Microbiol.* **76**, 305–323 (2022).
11. E. Bier, Gene drives gaining speed. *Nat. Rev. Genet.* **23**, 5–22 (2022).
12. M. B. Hufford, *et al.*, De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).
13. M. J. Brady, *et al.*, Antagonistic kinesin-14s within a single chromosomal drive haplotype. *bioRxiv* 2025.02.05.636711 (2025).
14. J. Chen, *et al.*, A complete telomere-to-telomere assembly of the maize genome. *Nat. Genet.* **55**, 1221–1231 (2023).
15. D. M. Higgins, *et al.*, Fitness Costs and Variation in Transmission Distortion Associated with the Abnormal Chromosome 10 Meiotic Drive System in Maize. *Genetics* **208**, 297–305 (2018).
16. R. K. Dawe, The maize abnormal chromosome 10 meiotic drive haplotype: a review. *Chromosome Res.* (2022). <https://doi.org/10.1007/s10577-022-09693-6>.

17. Á. T. Yamakake Kato, “Cytological studies of maize (*Zea mays* L.) and teosinte (*Zea mexicana* (Schraeder) Kuntze) in relation to their origin and evolution,” University of Massachusetts, Amherst, MA, USA. (1975).
18. B. McClintock, T. Yamakake, A. Blumenschein, *Chromosome Constitution of Races of Maize: its significance in the interpretation of relationships between races and varieties in the Americas* (Colegio de Pos graduados, Escuela Nacional de Agricultura, 1981).
19. L. B. Kanizay, *et al.*, Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in *Zea mays*. *Heredity* **110**, 570–577 (2013).
20. E. S. Buckler, *et al.*, Meiotic Drive of Chromosomal Knobs Reshaped the Maize Genome. *Genetics* **153**, 415–426 (1999).
21. P. Bilinski, *et al.*, Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLoS Genet.* **14**, e1007162 (2018).
22. R. K. Dawe, *et al.*, A Kinesin-14 Motor Activates Neocentromeres to Promote Meiotic Drive in Maize. *Cell* **173**, 839–850.e18 (2018).
23. L. B. Kanizay, P. S. Albert, J. A. Birchler, R. K. Dawe, Intragenomic conflict between the two major knob repeats of maize. *Genetics* **194**, 81–89 (2013).
24. N. Blavet, *et al.*, Sequence of the supernumerary B chromosome of maize provides insight into its drive mechanism and evolution. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).
25. J. A. Birchler, H. Yang, The supernumerary B chromosome of maize: drive and genomic conflict. *Open Biol.* **11**, 210197 (2021).
26. W. R. Carlson, Reversal of preferential fertilization. *Maize Genet. Coop. News Lett.* **73**, 39 (1999).
27. M. Rosato, A. M. Chiavarino, C. A. Naranjo, J. C. Hernandez, L. Poggio, Genome Size and Numerical Polymorphism for the B chromosome in Races of Maize (*Zea Mays* ssp. *Mays*, Poaceae). *American Journal of Botany* **85**, 168–174 (1998).
28. L. F. Randolph, Genetic characteristics of the B chromosomes in maize. *Genetics* **26**, 608–631 (1941).
29. H. L. Porter, A. L. Rayburn, B-chromosome and C-band heterochromatin variation in Arizona maize populations adapted to different altitudes. *Genome* **33**, 659–662 (1990).
30. L. Poggio, M. Rosato, A. M. Chiavarino, C. A. Naranjo, Genome Size and Environmental Correlations in Maize (*Zea mays* ssp. *mays*, Poaceae). *Ann. Bot.* **82**, 107–115 (1998).
31. G. E. González, L. Poggio, Intragenomic conflict between knob heterochromatin and B chromosomes is the key to understand genome size variation along altitudinal clines in maize. *Plants* **10**, 1859 (2021).

32. A. L. Rayburn, J. A. Auger, Genome size variation in *Zea mays* ssp. *mays* adapted to different altitudes. *Züchter Genet. Breed. Res.* **79**, 470–474 (1990).
33. C. M. Díez, *et al.*, Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytol.* **199**, 264–276 (2013).
34. A. E. Longley, Chromosomes of maize from North America. *J. Agric. Res.* **56**, 177–195 (1938).
35. A. Bianchi, M. V. Ghatnekar, A. Ghidoni, Knobs in Italian maize. *Chromosoma* **14**, 601–617 (1963).
36. P. K. Bretting, M. M. Goodman, Karyotypic variation in Mesoamerican races of maize and its systematic significance. *Econ. Bot.* **43**, 107–124 (1989).
37. K. A. Dyer, Local selection underlies the geographic distribution of sex-ratio drive in *Drosophila neotestacea*: Sex-ratio drive in *D. neotestacea*. *Evolution* **66**, 973–984 (2012).
38. M. M. Rhoades, E. Dempsey, On The Mechanism of Chromatin Loss Induced By The B Chromosome of Maize. *Genetics* **71**, 73–96 (1972).
39. M. M. Rhoades, E. Dempsey, “Structural heterogeneity of chromosome 10 in races of maize and teosinte” in *Plant Genetics*, M. Freeling, Ed. (Alan R. Liss, New York, 1985), pp. 1–18.
40. T. W. Lyttle, Experimental population genetics of meiotic drive systems II. Accumulation of genetic modifiers of Segregation Distorter (SD) in laboratory populations. *Genetics* **91**, 339–357 (1979).
41. M. Cazemajor, C. Landré, C. Montchamp-Moreau, The sex-ratio trait in *Drosophila simulans*: genetic analysis of distortion and suppression. *Genetics* **147**, 635–642 (1997).
42. T. A. R. Price, *et al.*, Resistance to natural and synthetic gene drive systems. *J. Evol. Biol.* **33**, 1345–1360 (2020).
43. K. Swarts, *et al.*, Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* **357**, 512–515 (2017).
44. J. A. Romero Navarro, *et al.*, A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* **49**, 476–480 (2017).
45. M. C. Roday, *et al.*, Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **14**, 1–18 (2013).
46. R. J. Elshire, *et al.*, A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6**, e19379 (2011).
47. J. Liu, *et al.*, Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol.* **21**, 121 (2020).

48. M. J. Brady, et al., Antagonistic kinesin-14s within a single chromosomal drive haplotype. *bioRxiv* 2025.02.05.636711 (2025).
49. D. L. Remington, et al., Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11479–11484 (2001).
50. S. E. Fick, R. J. Hijmans, WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
51. FAO, IIASA, *Harmonized World Soil Database Version 2.0* (FAO ; International Institute for Applied Systems Analysis (IIASA), 2023).
52. E. Gluck-Thaler, et al., Giant Starship elements mobilize accessory genes in fungal genomes. *Mol. Biol. Evol.* **39** (2022).
53. M. C. Stitzer, S. N. Anderson, N. M. Springer, J. Ross-Ibarra, The genomic ecosystem of transposable elements in maize. *PLoS Genet.* **17**, e1009768 (2021).
54. J. Engelstädter, H. Montenegro, G. D. D. Hurst, To what extent do different types of sex ratio distorters interfere? *Evolution* **58**, 2382–2386 (2004).
55. M. C. Stitzer, et al., Transposable element abundance subtly contributes to lower fitness in maize. *Evolutionary Biology* (2023).
56. A. Manser, A. K. Lindholm, B. König, H. C. Bagheri, Polyandry and the decrease of a selfish genetic element in a wild house mouse population. *Evolution* **65**, 2435–2447 (2011).
57. M. M. Rhoades, Preferential Segregation in Maize. *Genetics* **27**, 395–407 (1942).
58. J. J. Doyle, J. I. Doyle, A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem Bull.* **19**, 5 (1987).
59. J. C. Glaubitz, et al., TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* **9**, e90346 (2014).
60. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
61. P. Danecek, et al., Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).
62. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
63. A. Shumate, S. L. Salzberg, Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).
64. W. Shen, B. Sipos, L. Zhao, SeqKit2: A Swiss army knife for sequence and alignment processing. *Imeta* **3**, e191 (2024).

65. B. L. Browning, X. Tian, Y. Zhou, S. R. Browning, Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).
66. S. Purcell, *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
67. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
68. J. L. Portwood 2nd, *et al.*, MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Res.* **47**, D1146–D1154 (2019).
69. S. Ali, *et al.*, Modeling current and future potential land distribution dynamics of wheat, rice, and maize under climate change scenarios using MaxEnt. *Land (Basel)* **13**, 1156 (2024).
70. R. Hijmans, *raster: Geographic Data Analysis and Modeling* (2025).
71. V. V. Lia, V. A. Confalonieri, L. B. Poggio, B Chromosome Polymorphism Maize Landraces: Adaptive vs. Demographic Hypothesis Clinal Variation. *Genetics* **177**, 895–904 (2007).
72. F. Hartig, *DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models* (2024).

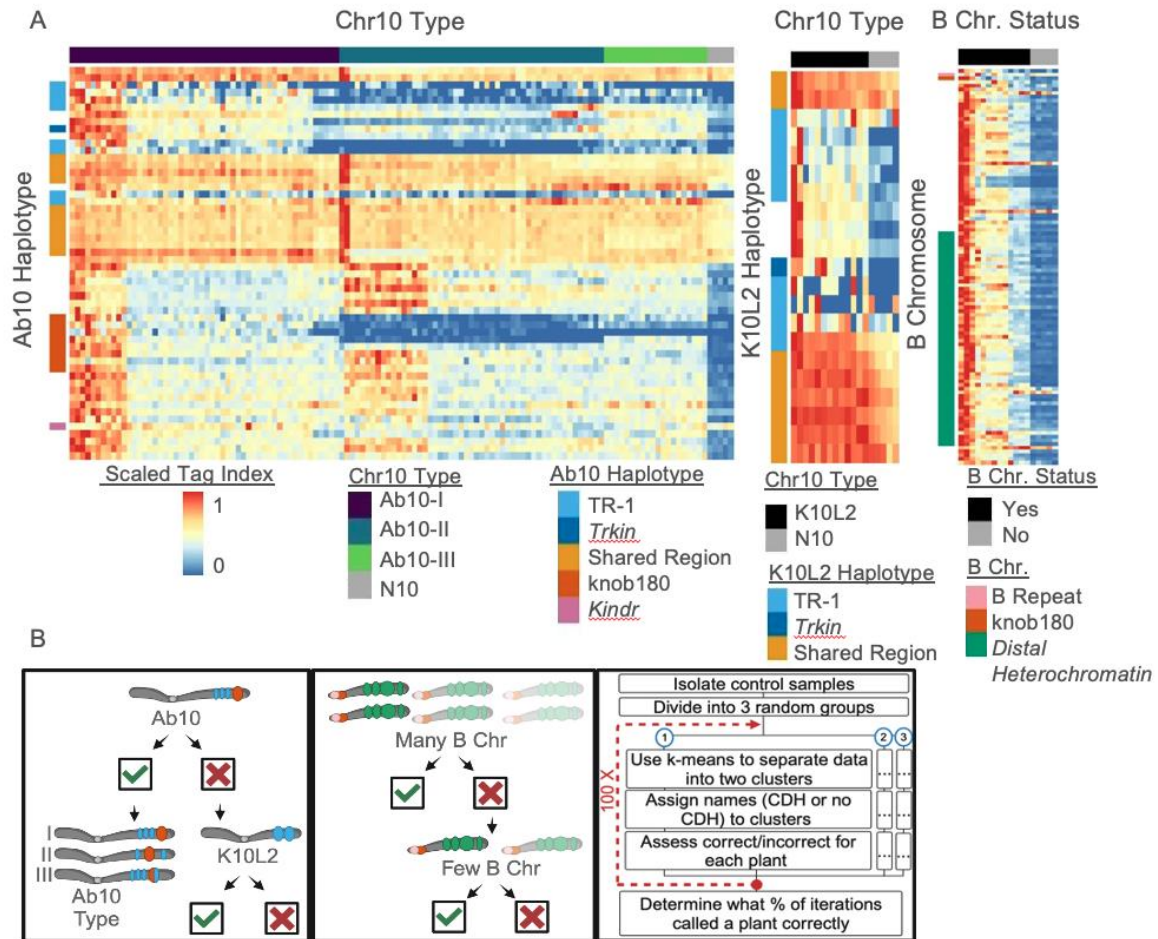
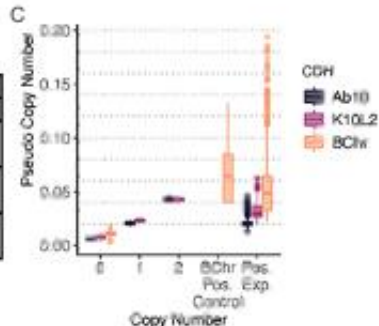


Figure 4.1 **Detection of CDHs in genotype by sequencing data.** A. Min/Max scaled tag index for abnormal chromosome 10 (Ab10) (13), K10L2 (13), and the B chromosome (24). Relevant features of each CDH are highlighted on the y axis. The CDH status of each control sample is highlighted on the x axis. B. Workflow for the automatic detection of CDHs via CDH specific scaled tag index bins. Check indicates present, x indicates absence.

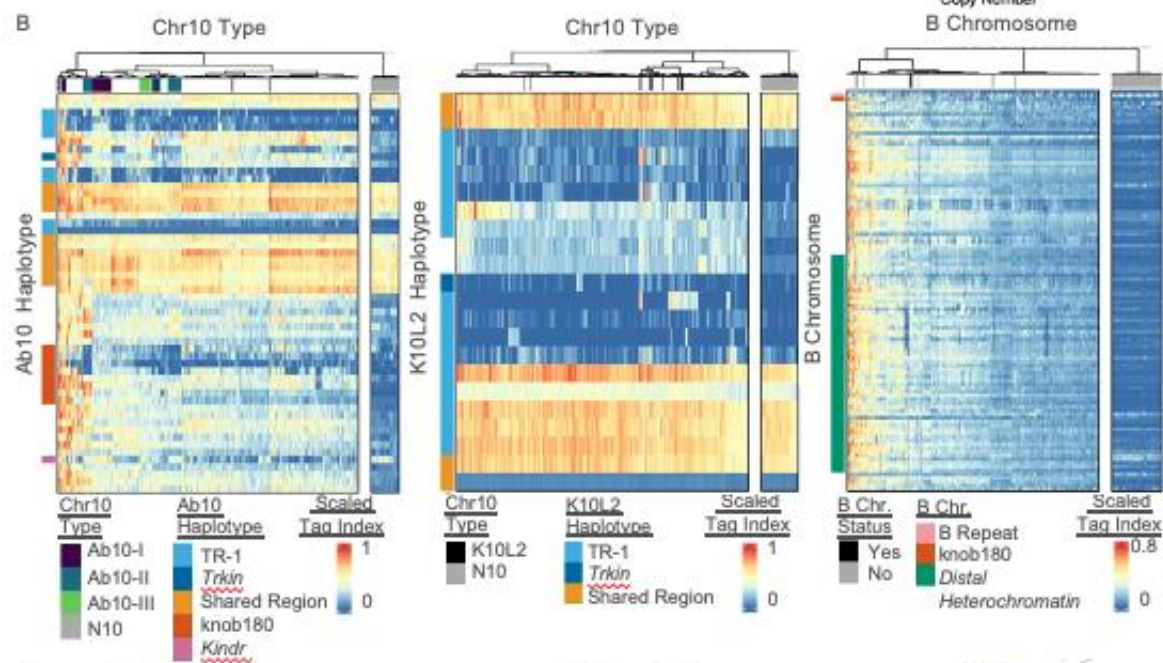
A

Paper	Geographic Range	Maize			Teosinte		
		Ab10	K10L2	B Chr	Ab10	K10L2	B Chr
Yamakake 1975 & McClintock et al. 1981	Americas	5.34% (6324)	4.86% (6251)	9.80% (6324)	9.35% (310)	15.48% (310)	6.77% (310)
Kanizay et al. 2013	Central and South America	13% (519)	NA	NA	15% (240)	NA	NA
Present	Americas	6.32% (5334)	5.16% (5334)	12.28% (5334)	4.14% (314)	3.82% (314)	1.59% (314)

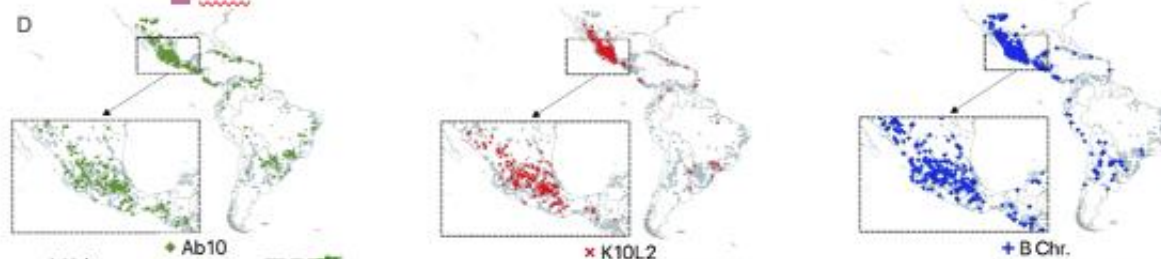
C



B



D



E

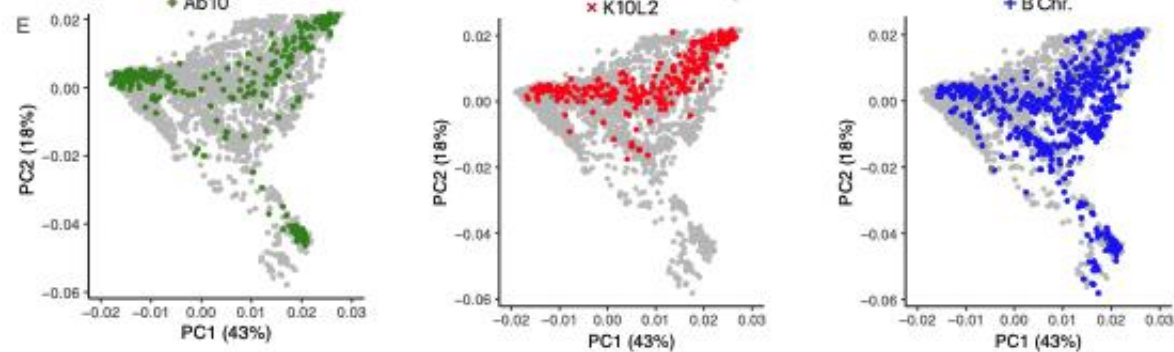


Figure 4.2. Detection of CDHs in Experimental Samples. A. Summary table for previous studies of CDH distribution as well as the results from the work presented here for comparison. B. Scaled tag index for all CDH positive controls and experimental samples for abnormal chromosome 10 (Ab10), K10L2, and the B chromosome. CDH positive controls and experimental samples and N10 control samples and plotted separately, each group is independently ward.D clustered. Ab10 and K10L2 tag indexes are min/max scaled. The B chromosome tag indexes are log scaled. Relevant features of each CDH are highlighted on the y axis. The CDH status of each control sample is highlighted on the x axis. C. Pseudo copy number of all three CDHs in control and experimental samples. B Chr Pos. Control refers to samples that are B chromosome positive via PCR but the copy number is unknown. Pos. Exp. refers to all CDH positive experimental samples. Dotted lines indicate approximate values for one copy as determined by the relationship between one and two copies of control samples. These are shown as a guideline only and are not meant to imply true copy number. D. Maps showing the location of all maize landraces assayed and those determined to be CDH positive. E. First two principal components from a whole genome SNP principle coordinate analysis with CDH positive individuals marked. Numbers in parentheses indicate the percent of variation that principal coordinate accounts for.

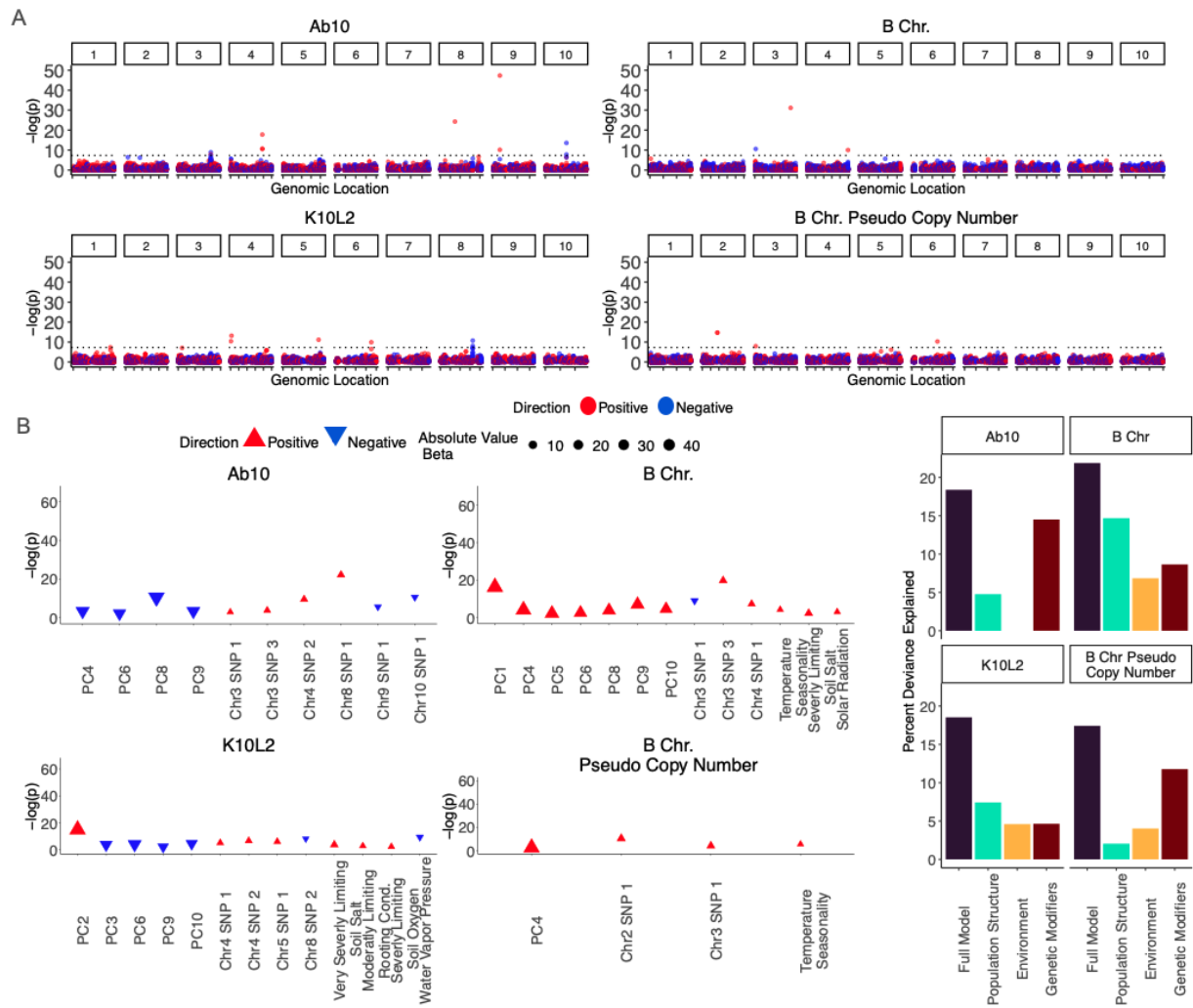


Figure 4.3 Relationship of CDH to the Environment and Genetic Loci A. Manhattan plots for genome wide association study results for each CDH and B chromosome copy number. Dotted black line indicates a p value of 5×10^{-8} . The Ab10 associated SNP on chromosome 9 is believed to be an artifact. The B chromosome plot excludes two highly associated SNPs that are likely artifacts, see Figure S4.4 for full plot. B. Plots of fully simplified generalized linear models for each CDH including population structure, genetic loci, and environmental variables. Shape color and orientation indicate the direction of the relationship to the CDH. Shape size represents the effect size, units are not always comparable between variables. Very severely and severely limiting soil salt refers to growth limiting excess soil salts. Moderately Limiting rooting cond refers to moderately growth limiting rooting conditions. Severely limiting soil oxygen refers to severely growth limiting soil oxygen. All soil variables are from (51) C. Partitioned deviance of each model shown in Figure 4.3b. The partitions do not sum to the full model due to shared variation between the partitions but indicate a relative relationship.

CONCLUSIONS

Selfish genetic elements (SGEs), specifically chromosomal drive haplotypes (CDHs), have played a significant role in the evolution of maize as well as many other species (Herrmann and Bauer 2012; Courret et al. 2019; Dawe 2022; Birchler and Yang 2021). They bent the guard rails provided by Mendel's first law adding nuance to the field of genetics. We, and others, have used ecology as a useful framework for understanding the way that SGEs relate to their genomic and abiotic environment (Agren and Clark 2018; Venner, Feschotte, and Biémont 2009). On occasion, some scientists push back against the application of ecology to non-independent entities like SGEs. Afterall, ecology is defined as the study of the relationships between living organisms and their physical environment ("What Is Ecology?" 2025), and an SGE is simply not a living organism. Its fate is intertwined with the rest of the genome in a way that is analogous to an ecosystem but still wholly unique. And yet, the framework provided by ecology offers structure to the rapidly evolving field of selfish genetic element ecology much in the way that Mendell's first law did for genetics. We hope that the field of ecology won't mind our borrowing them while we build our own.

Here we have demonstrated that the structure of Abnormal chromosome 10 (Ab10) is more complicated than previously understood and contains a small internal duplication. It is possible that there are additional duplications on Ab10 that could possibly help explain its large size (Brady et al. 2024). This duplication allowed us to identify a classical maize marker gene as a regulator of chloroplast transcription. Further we searched for an explanation for the existence

of the *Trkin* gene on the Ab10 haplotype, finally concluding it likely does not have one and appears to be deleterious (Brady et al. 2025). This illustrates the lack of optimization in evolution specifically of selfish genetic elements: it only needs to work, it does not have to be pretty. Finally, we thoroughly explored the ecology of three CDHs (Ab10, K10L2, and the B chromosome) in maize finding that each CDH has a unique relationship to its biotic and abiotic environment.

Future Directions

The mechanisms of Ab10, K10L2, and the B chromosome have been particularly well studied (Dawe 2022; Birchler and Yang 2021). However, there are still some remaining open questions. Ab10 is known to increase recombination as a fundamental part of its drive mechanism (Dawe 2022). Anecdotally, the effect is as large as 50% in the peri-centromere. Low recombination areas are particularly troublesome in crop breeding, making the Ab10 recombination effect very intriguing. A significant obstacle to the identification of the recombination effect is the lack of mutants. It is possible that the Ab10 isolates discovered in Chapter 4 could be useful in either population genetic or experimental approaches to identifying the Ab10 recombination effect. While no modifiers of Ab10 drive have been identified before, we have previously observed that certain inbreds are more tolerant of Ab10 than others. We would be interested to learn if the enhancers found in Chapter 4, or other unknown enhancers, might be contributing to this observation. Further, the B chromosome preferential fertilization mechanism has yet to be identified (Birchler and Yang 2021). Experimental studies of the enhancers identified in Chapter 4 may be useful.

All of the CDHs we discuss here occur at relatively low, but highly variable, frequencies in populations (Yamakake Kato 1975; McClintock, Yamakake, and Blumenschein 1981). The

data we had access to generally only included one individual per population. While a random sample like this was adequate for the questions that we posed, population level data would have been helpful. For example, there were certainly populations we identified as CDH negative that were segregating for the CDH. Additionally, population level data would have allowed us to explore the antagonistic dynamics, detailed in Chapter 3, between Ab10 and K10L2. Our data simply did not allow us to scratch the surface on this question. The decreasing cost of deep short read whole genome sequencing and possibly even assembly will greatly enhance future work on the distribution, evolution, and ecology of CDHs.

In the studies presented here we did not explore the evolution of the CDH haplotypes directly. Unfortunately, some questions must always be left for those who come after us. However, it is entirely possible to explore the divergence time and phylogenies of Ab10 and K10L2 using the current assemblies and genotyping data presented in Chapter 3 and 4. Though, assembly of multiple CDH haplotypes would add considerably to an analysis like this. We are particularly interested in learning the elusive origin of Ab10. My personal prediction, supported by no evidence, is that the Ab10 haplotype began as an inversion and slowly accumulated small duplications from other regions of the genome over millennia. The timing of Ab10s origin and how many times it occurred is also of great interest. The prevalence of knob sequences of both types across the grass family suggest it may be incredibly ancient.

In the vein of Ab10 evolution, we have long wondered where the fitness effects associated with the Ab10 haplotype come from. Theory predicts they should come from the shared region as a result of an accumulation of deleterious alleles due to the lack of recombination, but we have no evidence for this (Higgins et al. 2018). Anecdotally, we have observed that Ab10 homozygous plants become healthier when the distal tip is absent. We

believe that a substantial portion of Ab10 fitness effects may come from the KINDR/knob180 drive system encoded on the distal tip of Ab10. *Kindr* is known to be expressed in non-meiotic tissues such as leaf (Dawe et al. 2018). However, its effect on mitosis has never been investigated. This would be an interesting future area of research and would contribute to our understanding of both maize biology and selfish genetic element biology.

References

- Agren, J. A., and A. G. Clark. 2018. “Selfish Genetic Elements.” *PLoS Genetics* 14 (11): e1007700.
- Birchler, J. A., and H. Yang. 2021. “The Supernumerary B Chromosome of Maize: Drive and Genomic Conflict.” *Open Biology* 11 (11): 210197.
- Brady, Meghan J., Maya Cheam, Jonathan I. Gent, and R. Kelly Dawe. 2024. “The Maize Striate Leaves2 (Sr2) Gene Encodes a Conserved DUF3732 Domain and Is Homologous to the Rice Yss1 Gene.” *Plant Direct* 8 (2): e567.
- Brady, Meghan J., Anjali Gupta, Jonathan I. Gent, Kyle W. Swentowsky, Robert L. Unckless, and R. Kelly Dawe. 2025. “Antagonistic Kinesin-14s within a Single Chromosomal Drive Haplotype.” *BioRxiv*. <https://doi.org/10.1101/2025.02.05.636711>.
- Courret, C., C. H. Chang, K. H. Wei, C. Montchamp-Moreau, and A. M. Larracuente. 2019. “Meiotic Drive Mechanisms: Lessons from *Drosophila*.” *Proceedings. Biological Sciences / The Royal Society* 286 (1913): 20191430.
- Dawe, R. K. 2022. “The Maize Abnormal Chromosome 10 Meiotic Drive Haplotype: A Review.” *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*. <https://doi.org/10.1007/s10577-022-09693-6>.
- Dawe, R. K., E. G. Lowry, J. I. Gent, M. C. Stitzer, K. W. Swentowsky, D. M. Higgins, J. Ross-Ibarra, et al. 2018. “A Kinesin-14 Motor Activates Neocentromeres to Promote Meiotic Drive in Maize.” *Cell* 173 (4): 839-850 e18.
- Herrmann, B. G., and H. Bauer. 2012. “The Mouse *t*-Haplotype: A Selfish Chromosome – Genetics, Molecular Mechanism, and Evolution.” In *Evolution of the House Mouse*, edited by Miloš Macholán, Stuart J. E. Baird, Pavel Munclinger, and Jaroslav Piálek, 297–314. Cambridge University Press.
- Higgins, D. M., E. G. Lowry, L. B. Kanizay, P. W. Bercraft, D. W. Hall, and R. K. Dawe. 2018. “Fitness Costs and Variation in Transmission Distortion Associated with the Abnormal Chromosome 10 Meiotic Drive System in Maize.” *Genetics* 208 (1): 297–305.
- McClintock, B., T. Yamakake, and A. Blumenschein. 1981. *Chromosome Constitution of Races of Maize: Its Significance in the Interpretation of Relationships between Races and Varieties in the Americas*. Chapingo, Edo. Mexico: Colegio de Pos graduados, Escuela Nacional de Agricultura.
- Venner, Samuel, Cédric Feschotte, and Christian Biémont. 2009. “Dynamics of Transposable Elements: Towards a Community Ecology of the Genome.” *Trends in Genetics: TIG* 25 (7): 317–23.
- “What Is Ecology?” 2025. <https://esa.org/about/what-does-ecology-have-to-do-with-me/>.
- Yamakake Kato, Ángel Takeo. 1975. “Cytological Studies of Maize (*Zea Mays* L.) and Teosinte (*Zea Mexicana* (Schraeder) Kuntze) in Relation to Their Origin and Evolution.” Doctor of Philosophy, Amherst, MA, USA: Univeristy of Massachusetts.

APPENDIX A

SUPPLEMENTARY TABLE AND FIGURES – CHAPTER 2

Table S2.1. **Orthologs outside of regions of known Homology between Ab10 and N10.**

B73 Gene Name	Ab10 Gene Name	Gene Number
Zm00001eb431630	Zm00043a049151	NA ¹
Zm00001eb431660	Zm00043a049258	NA
Zm00001eb433140	Zm00043a051035	NA
Zm00001eb434390	Zm00043a050572	NA
Zm00001eb434490	Zm00043a050219	Gene 1
Zm00001eb434500	Zm00043a050220	Gene 2
Zm00001eb434520	Zm00043a050224	Gene 4
Zm00001eb434530	Zm00043a050225	Gene 5
Zm00001eb434540	Zm00043a050226	Gene 6

¹ Genes labeled NA were not considered *sr2* candidates because they are located outside of the mapped *sr2* region (the purple lines in Figure 1).

Table S2.2. Gene specific primers¹.

Name	F sequence	Target	Name	R sequence	Target
Sr_490_F	GTCGAGATAGAAGGGGCGAG	Zm00001eb434490	Sr_490_R	GCGAGCTCCGGAATACTCTC	Zm00001eb434490
Sr_500_F	CAACATGTGGAGCGACCAAC	Zm00001eb434500	Sr_500_R	GGAACCCGATCTAAGAGGCG	Zm00001eb434500
Sr_520_F	CAGGAAGAGCACACGGAAGA	Zm00001eb434520	Sr_520_R	CAGACAAGCGCGATTGACCT	Zm00001eb434520
Sr_530_F	GTAGTCTGGAGGGCACGTTG	Zm00001eb434530	Sr_530_R	GAAAGCGTAAAGACGTGGCG	Zm00001eb434530
JIG-419	TAGGCTTCCTCCTGCTCTC	Zm00001eb231550 4th exon	JIG-420	TGCTTGAGTTGGTCGTCATC	Zm00001eb434490 4th exon
JIG-421	CTAACGGCTGCTGGTCTG	unique to sr2-ref transcript	JIG-422	GCTACCCAGCCACTGAAAT	Zm00001eb434490 3rd exon
JIG-423	TGCTGGTCTGCGAGGAT	unique to sr2-ref transcript	JIG-424	GCCACTGAAATTTGTGCTCTTG	Zm00001eb434490 3rd exon

¹ All primers are shown in the 5' to 3' orientation.

Table S2.3. **Mu alleles used in this study.**

Gene Number	B73 Gene Name	Uniform Mu Stock	Mu of interest	Location
1	Zm00001eb434490	UFMu-08182	mu1058934	1st exon
2	Zm00001eb434500	UFMu-12931	mu1089572	1st exon
4	Zm00001eb434520	UFMu-03495	mu1037653	1st exon
5	Zm00001eb434530	UFMu-03745	mu1037651	1st exon

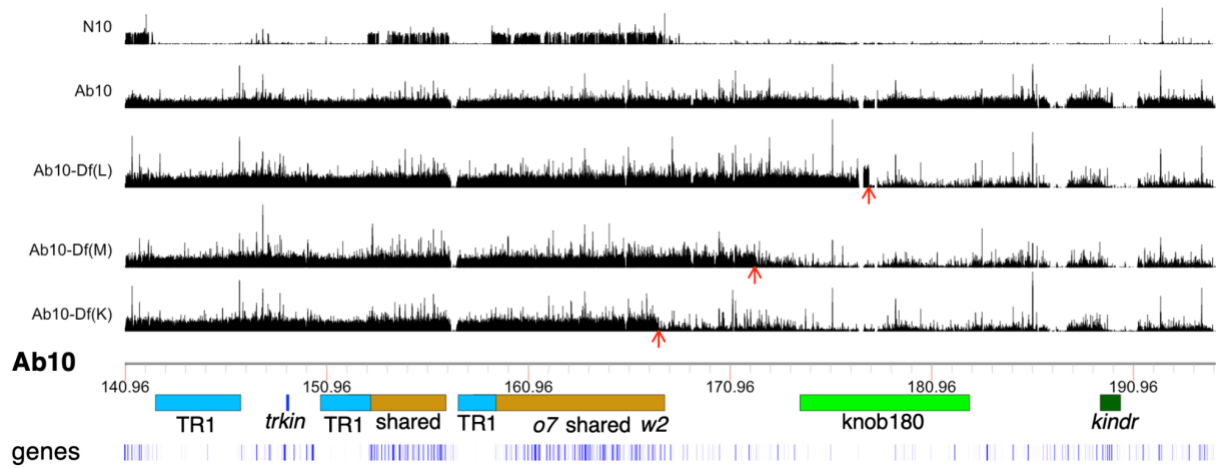


Figure S2.1. **Ab10 deletion breakpoints.** Black bar plots show average read count calculated in 1 kb windows for Illumina paired end reads mapped to the Ab10 reference using BWA-mem filtered to a MAPQ of 20 and only primary alignments. Red arrows indicate the presumed breakpoint. Location of genes with known physical position, *o7* and *w2*, are shown (Wang et al., 2011; Udy et al., 2012). TR1 and knob180 are maize knob types. *trkin* and *kindr* are kinesin proteins responsible for the preferential transmission of Ab10 (Dawe et al., 2018; Swentowsky et al., 2020). Shared indicates regions of known homology between Ab10 and N10. Blue bars indicate annotated genes (Liu et al., 2020).

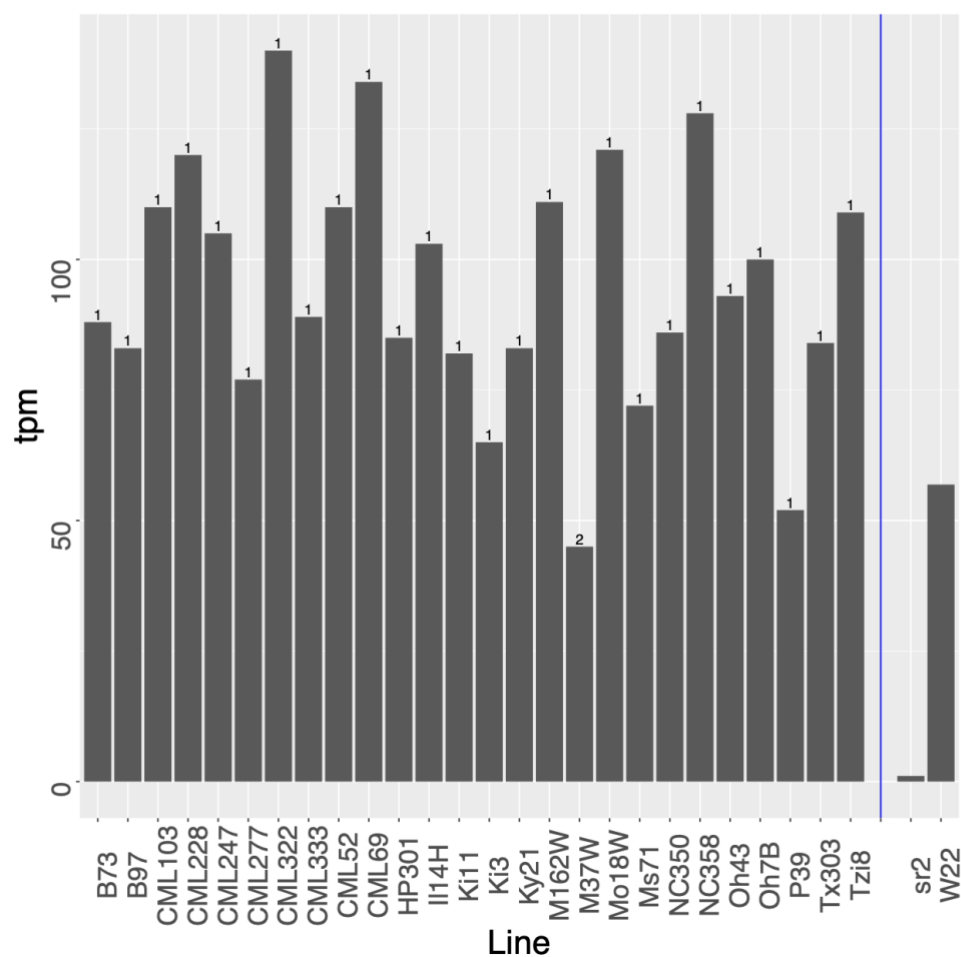


Figure S2.2. **Expression of *sr2*.** Leaf tip expression values (bar height) and copy number (number above bar) for *sr2* in all NAM founders (Hufford et al., 2021). Data to the right of the blue line are leaf tip expression data from this project, copy number is unknown for these lines.

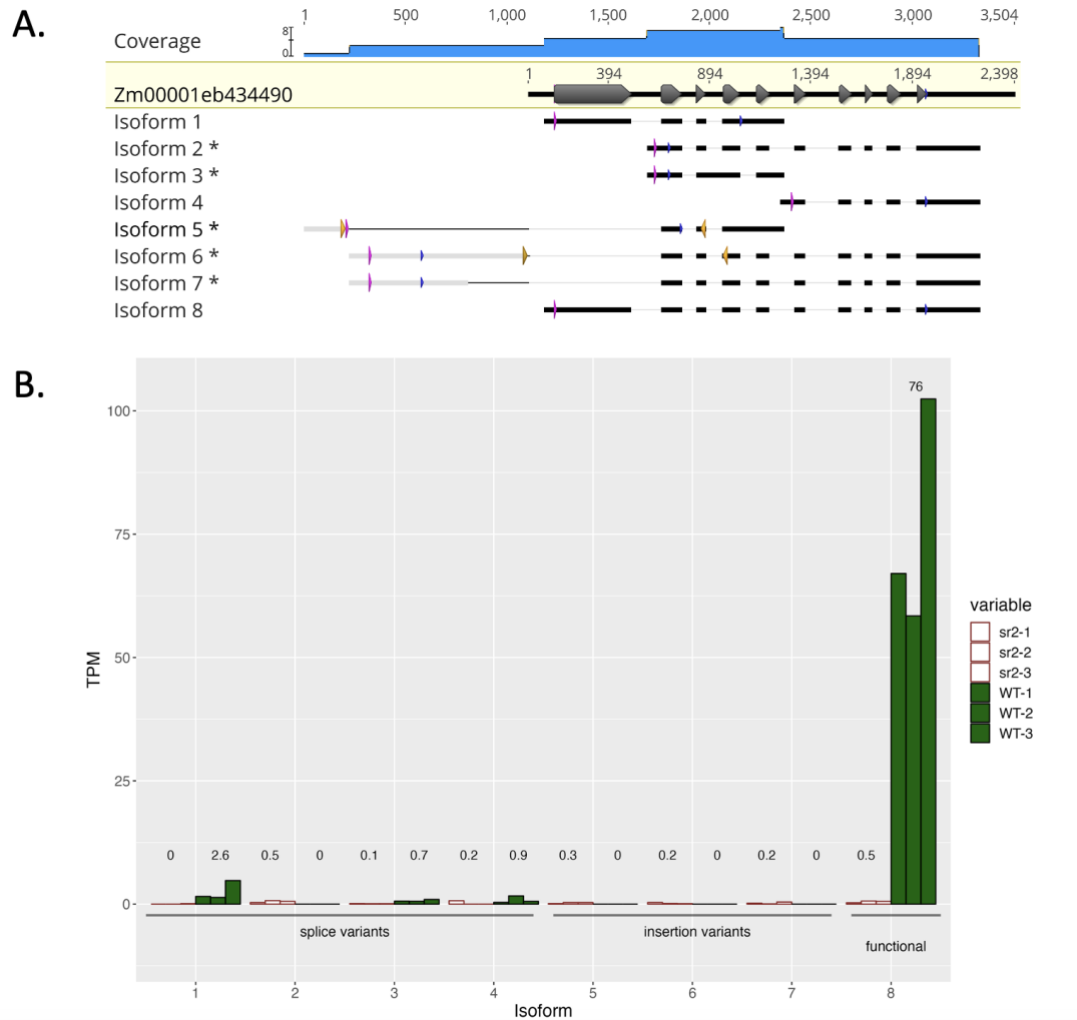


Figure S2.3. **The *sr2-ref* allele contains an insertion.** A. The first line shows the gene model from B73, where grey arrows indicate exons. Isoforms 1 through 8 are derived from a Trinity de novo transcriptome assembly of *sr2-ref* (Haas et al., 2013). * indicates there is a frame shift relative to the wild type in that isoform. Pink segments indicate the first ATG, blue segments indicate the first in-frame stop codon, gold segments indicate primers. Light grey boxes indicate sequence that is different from the *sr2* B73 reference, black indicates strong homology to *sr2*. Isoform 5 includes sequence with homology to the DNA transposon Zm10271_AC186904_1. Isoforms 6 and 7 include sequence with homology to CASP-like protein 4A2 (Zm00001eb231550). None of the isoforms include the 3' TIR of Zm10271_AC186904_1, indicating it is reproducibly spliced out of the transcript (Figure S2.4). B. Transcripts per million (TPM) for each isoform in each biological replicate of *sr2-ref* and wild type leaf tissue. The number above each set of three bars represents the average.

JIG-419

TAGGCTTCCCTCTGCTCTCCTCCAGGTACGGGAAGTCCTCGGTGAAGTGCAGGCCGCGGCT
 CTCGCGCTTGGCCAGCGCGCTCCTGACGACGTGCAGACGAGCGCCGCGCTCCTGCTCGGGC
 GCCGGCTCGGCACCTGCTCGGGCCGCGGCTCGCTCGTCAGGGAGAGAGCGCTGGGGGAGAG
 GAGCGCCCGTGGATCCGTCGTAAGAAACGGACTGCCGTAACACCTAGGTTAAATCGAGCTAA
 ATCTGACTCAGATACTACAATTCCAGAGTTCAACGGATTATAATACTAAAATTCCAAATGTG
 ATGTTGGGAATACCTATTTTCAGAAGCCCAAGTCTGACGATATCACTGTTCCAAAAAGCTCC
 ATCCCGAGCTTTTTGGAACCGTGATATCCCAGCTTAATATTTTCTTTCTTTCTGTCTCAC
 AAATGAAATCAAGCCAAACCAACATTGTGATATGCAGATCGACGACACCAACAACCGCCCCCT
 TGGAGTGCATAATCAGGAGAGTATTCCGGAGCTCGCAGGACCACGACTGCATGCTTCTCTGC
 CCAGTTGACATGTAAGTGACACTCTCTCTTTTTTGTACTCATACTCATCGCAGACACCCAAG
 TCACACTCTCTCTCTCTTACAGGCCTGTGTCAGGTCTCTCAAGAGCACAAATTTAGTGGCTGG
 GTAGCTGTATGACTTGGCCCTCATATACATTCTTCAACTGATTATTTCAATTTATTTCCCAT
 GGTCTAATACTGTTTATTTATTTAGGTCGATGACGACCAACTCAAGCA

JIG-420

Figure S2.4. The *sr2-ref* insertion is in the first exon. Sanger sequenced amplicon of JIG-419 and JIG-420 (Table S2.4) from *sr2-ref* DNA. Blue solid underlined text indicates homology to CASP-like protein 4A2 (Zm00001eb231550). Green dotted underlined text indicates homology to the DNA transposon Zm10271_AC186904_1. Orange not underlined text indicates homology to *sr2* (Zm00001eb434490). Bold text indicates sequence in *sr2-ref* isoform 6 (Figure S2.3). Italicized text indicates sequence that was expected in the amplicon, but did not appear in the Sanger sequence. Grey highlighted text indicates primers.

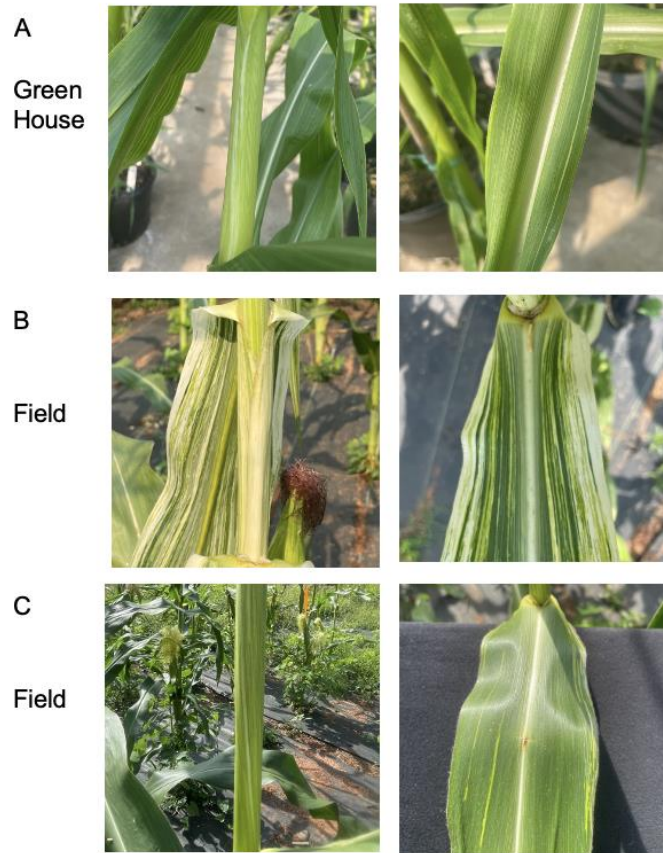


Figure S2.5. **The *sr2-ref* phenotype varies with environment and between individuals.** All individuals shown are *sr2-ref* homozygous full siblings grown at the same time in Athens GA, USA. A. Plant grown in the greenhouse with weak leaf striping. This phenotype is representative of three individuals. B. Plant grown in the field with strong leaf striping. C. Plant grown in the field with moderate leaf striping. B and C, two additional individuals grown in the field that had an intermediate leaf striping phenotype.

APPENDIX B

SUPPLEMENTARY TABLES AND FIGURES – CHAPTER 3

Table S3.1: **Ab10 non-shared genes functional annotation.** Only genes with an EggNOG description are shown.

Gene	EggNOG Description	Gene	EggNOG Description
g5481.t1	Reverse transcriptase (RNA-dependent DNA polymerase)	g5853.t1	DNA-directed RNA Polymerase
g5482.t1	DUF789	g5854.t1	DNA-directed RNA Polymerase
g5483.t1	ribosomal protein	g5856.t1	DNA-directed RNA Polymerase
g5484.t1	Pfam:hATC	g5858.t1	DNA-directed RNA Polymerase
g5486.t1	K10405 kinesin family member C1	g5860.t1	DNA-directed RNA Polymerase
g5487.t1	DUF716	g5862.t1	DNA-directed RNA Polymerase
g5488.t1	Oxidoreductase, 2OG-Fe(II) oxygenase family	g5863.t1	DNA-directed RNA Polymerase
g5489.t1	cell division	g5867.t1	DnaJ (Hsp40) homolog
g5491.t1	K10405 kinesin family member C1	g5870.t1	UPF0161 protein
g5809.t1	LIM domain kinase	g5871.t1	isoamylase
g5810.t1	Coiled-coil domain	g5872.t1	Reverse transcriptase (RNA-dependent DNA polymerase)
g5811.t1	Cold shock domain protein 2	g5873.t1	K10405 kinesin family member C1
g5812.t1	DUF3727	g5874.t1	K10405 kinesin family member C1
g5813.t1	zinc finger	g5875.t1	FHA domain
g5814.t1	Signal peptidase, S26	g5881.t1	Z1C alpha zein protein
g5815.t1	thioesterase	g5882.t2	Z1C alpha zein protein
g5827.t1	serine threonine-protein kinase	g5883.t1	vac14 homolog (S. cerevisiae)
g5828.t1	ribosomal protein	g5884.t1	vac14 homolog (S. cerevisiae)
g5830.t1	atp sulfurylase	g5885.t1	Z1C alpha zein protein
g5833.t1	nuclear pore protein 84 107	g5886.t1	Z1C alpha zein protein
g5838.t1	Sodium hydrogen exchanger	g5887.t1	DUF716
g5839.t1	Zinc knuckle	g5888.t1	histone h2a
g5840.t1	phosphatase	g5889.t1	histone h2a
g5842.t1	40s ribosomal protein	g5890.t1	CCAAT enhancer binding protein (C EBP), zeta
g5844.t1	nuclear pore protein 84 107	g5891.t1	WD repeat domain 1
g5845.t1	WD repeat	g5892.t1	DUF724
g5846.t1	Myb-like DNA-binding domain	g5894.t1	DNA polymerase
g5847.t1	kinesin motor domain	g5895.t1	SAM domain (Sterile alpha motif)
g5848.t1	1-aminocyclopropane-1-carboxylate oxidase homolog	g5896.t1	NEDD4 binding protein
g5849.t1	DNA-directed RNA Polymerase	g5897.t1	Kinesin motor domain
g5851.t1	DNA-directed RNA Polymerase	g5898.t1	starch synthesis
		g5899.t1	Kinesin motor domain

Table S3.2: **K10L2 non-shared genes functional annotation.** Only genes with an EggNOG description are shown.

Gene	EggNOG Description
g246.t1	Multicopper oxidase
g247.t1	complex ard1 subunit
g250.t1	Reverse transcriptase (RNA-dependent DNA polymerase)
g254.t1	Nicotiana lesion-inducing like
g255.t2	FES
g256.t2	Oxidoreductase, 2OG-Fe(II) oxygenase family
g257.t1	cell division
g258.t1	K10405 kinesin family member C1
g262.t1	Pentatricopeptide repeat-containing protein
g263.t1	complex ard1 subunit
g264.t1	Tyrosine 3-monooxygenase tryptophan 5-monooxygenase activation protein
g266.t1	Coiled-coil domain-containing protein
g267.t1	Protein of unknown function (DUF674)
g268.t1	HLH
g269.t2	Cold shock domain protein 2-like protein
g270.t1	organic cation
g272.t1	NmrA-like family

Table S3.3: Gene ortholog pairs on the non-shared regions of both Ab10 and K10L2.

Ab10 ID	K10L2 ID	EggNOG Description
g5475	gA5475	NA
g5479	g251	NA
g5486	g258	K10405 kinesin family member C1
g5487	gA5487	Family of unknown function (DUF716)
g5488	g256	Oxidoreductase, 2OG-Fe(II) oxygenase family
g5489	g257	cell division
g5491	g258	K10405 kinesin family member C1
g5847	g258	kinesin motor domain containing protein
g5867	g269	DnaJ (Hsp40) homolog, subfamily B, member 11
g5873	g258	K10405 kinesin family member C1
g5874	g258	K10405 kinesin family member C1
gK255	g255	FES

Table S3.4: Gene ortholog pairs on the shared region of K10L2 and the non-shared region of Ab10.

Ab10 ID	K10L2 ID	EggNOG Description
g5845	g106	WD repeat-containing protein
g5848	g300	1-aminocyclopropane-1-carboxylate oxidase homolog
g5872	g75	Reverse transcriptase (RNA-dependent DNA polymerase)
g5891	g208	WD repeat domain 1

Table S3.5: Gene ortholog pairs on N10 and the non-shared region of Ab10.

Ab10 ID	B73 ID	EggNOG Description
g5820	Zm00001eb430320	NA
g5849	Zm00001eb431870	DNA-directed RNA Polymerase
g5851	Zm00001eb431870	DNA-directed RNA Polymerase
g5853	Zm00001eb431870	DNA-directed RNA Polymerase
g5854	Zm00001eb431870	DNA-directed RNA Polymerase
g5856	Zm00001eb431870	DNA-directed RNA Polymerase
g5858	Zm00001eb431870	DNA-directed RNA Polymerase
g5860	Zm00001eb431870	DNA-directed RNA Polymerase
g5862	Zm00001eb431870	DNA-directed RNA Polymerase
g5863	Zm00001eb431870	DNA-directed RNA Polymerase

Table S3.6: Primers and Reaction Parameters Used For Genotyping.

Primer Name	Primer Sequence	Primer Name	Primer Sequence
Ptrkin1_EX3_F1	GGG GAG GTC TGG CTG CTA G	Ptrkin1_EX3_R1	AGG AGA GAG GCC TGC GA
Ptrkin1_EX4_F1	CCT TTC AAC GGC CAA TCC G	Ptrkin1_EX4_R1	AAC AGT GGC CCC TAG CTG CC
K10L2_F3	CAG CGT TAC CCC TTG CGA TT	K10L2_R3	GGA TTG GGG CGG TGA ACA TA
trkin_EX3_F2	GCT GGA AGA AGT AGC TCG CCG	trkin_EX3_R1	GCA TGC GAC TAG GGA CTG GG
trkin_EX4_F1	CTA CAT GAC GGC CAA TCC G	trkin_EX4_R1	AAC AGT GGC CCC TAG TTG CC
CAS9_F1	ACG AGA AGT ACC CGA CAA TCT ACC	CAS9_R1	TGA TTT GAA GTT CGG CGT CAG G
Primer Pair	Temperature Profile	Reaction Concentrations	
Ptrkin1_EX3	hold 95°C, 2 min 95°C, 33(30 s 95°C, 30 s 61°C, 55 s 72°C), 5 min 72°C	manufacturers recommendation	
Ptrkin1_EX4		manufacturers recommendation	
K10L2	hold 95°C, 2 min 95°C, 30(30 s 95°C, 30 s 58°C, 32 s 72°C), 5 min 72°C	manufacturers recommendation	
trkin_EX3	Ab10: hold 95°C, 2 min 95°C, 33(30 s 95°C, 30 s 64°C, 45 s 72°C), 5 min 72°C	manufacturers recommendation	
	K10L2: hold 95°C, 2 min 95°C, 33(30 s 95°C, 30 s 62°C, 1m 72°C), 5 min 72°C	3.3X Primer and DNA concentration from manufacturers recommendation	
trkin_EX4	Ab10: hold 95°C, 2 min 95°C, 33(30 s 95°C, 30 s 62°C, 45 s 72°C), 5 min 72°C	manufacturers recommended	
	K10L2: hold 95°C, 2 min 95°C, 33(30 s 95°C, 30 s 64°C, 1 m 72°C), 5 min 72°C	3.3X Primer and DNA concentration from manufacturers recommendation	
CAS9	hold 95°C, 2 min 95°C, 30(30 s 95°C, 30 s 58°C, 30 s 72°C), 5 min 72°C	manufacturers recommendation	

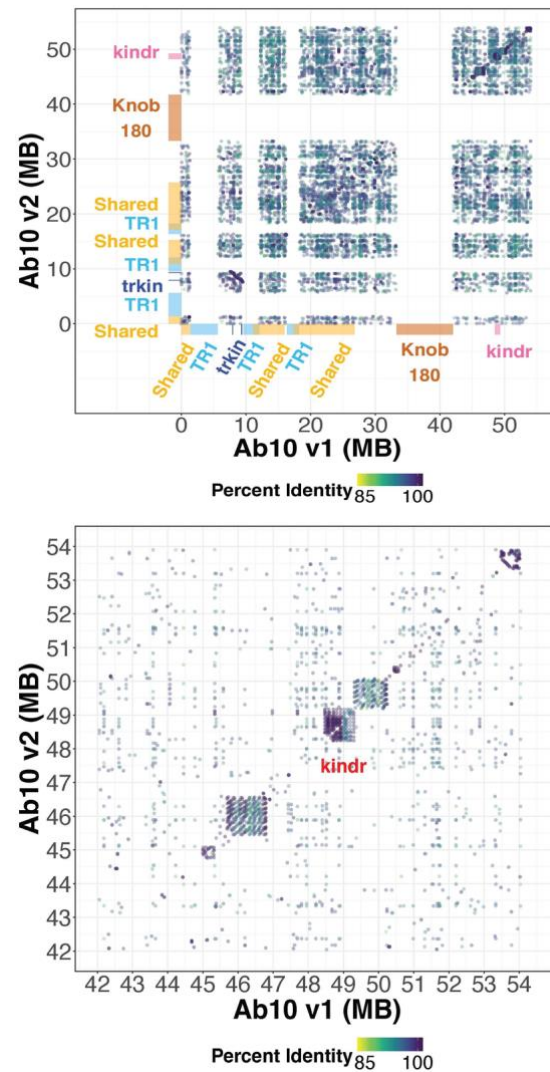


Figure S3.1: **Sequence Comparison of Chromosome 10 Haplotypes.** Each dot marks the start of a maximal unique match (MUM) of at least 300bp long between the B73-Ab10 v1 and B73-Ab10 v2 genomes Ab10 haplotype, which begin at the *colored1* gene (Marçais *et al.* 2018). B73-Ab10 v1 refers to the first Ab10 assembly (Liu *et al.* 2020), B73-Ab10 v2 refers to the assembly generated here. The color of each dot represents the percent identity of that match. All large knob arrays were removed for the sake of clarity. Relevant regions of each genome are marked.

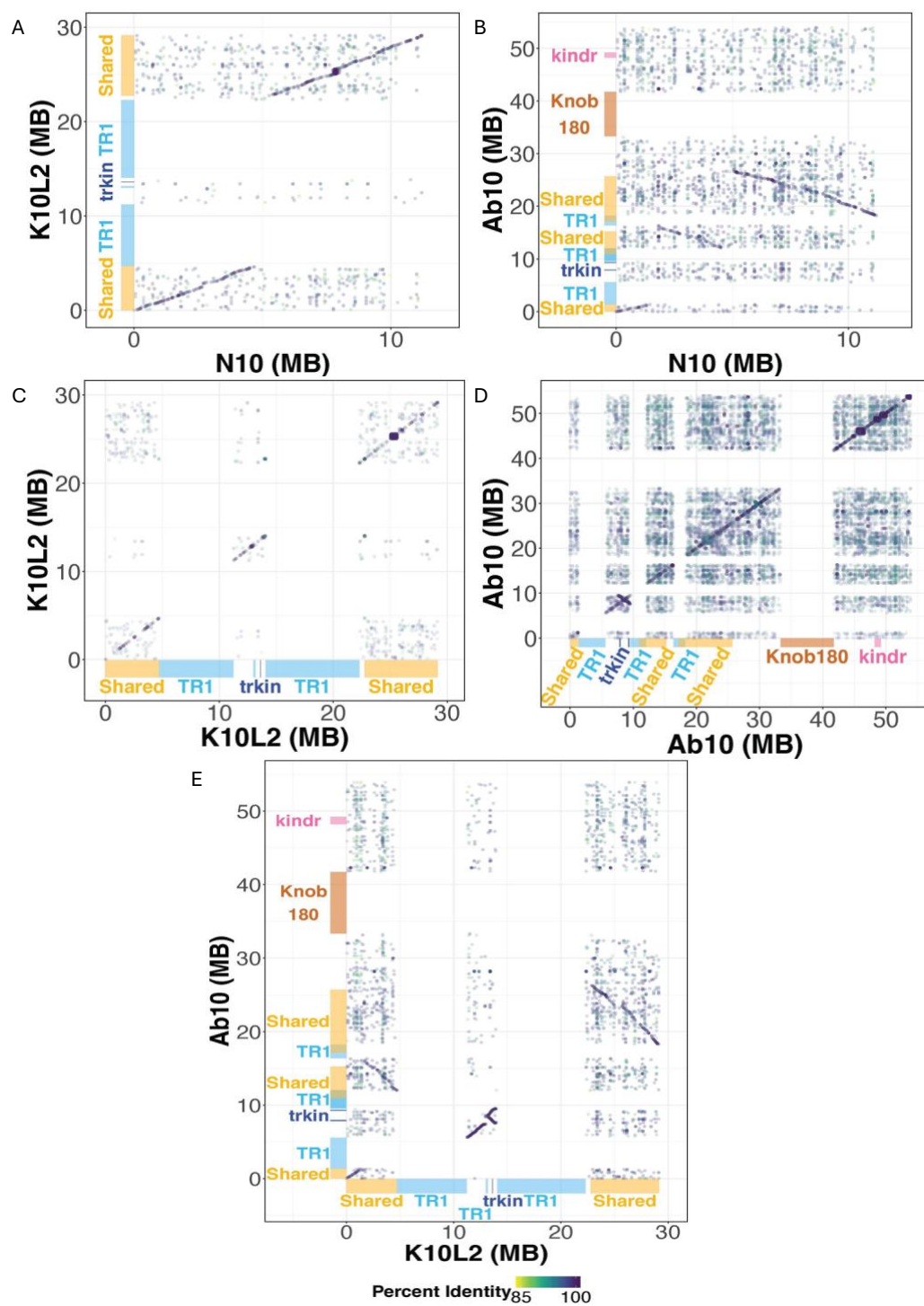


Figure S3.2: **Sequence Comparison of Chromosome 10 Haplotypes.** Each dot marks the start of a maximal unique match (MUM) of at least 300 bp long between various chromosome 10 haplotypes all of which begin at the *colored1* gene (Marçais *et al.* 2018). N10 refers to the B73 v5 assembly (Hufford *et al.* 2021), Ab10 and K10L2 refer to the assemblies generated in this work. The color of each dot represents the percent identity of that match. All large knob arrays were removed for the sake of clarity. Relevant regions of each haplotype are marked.

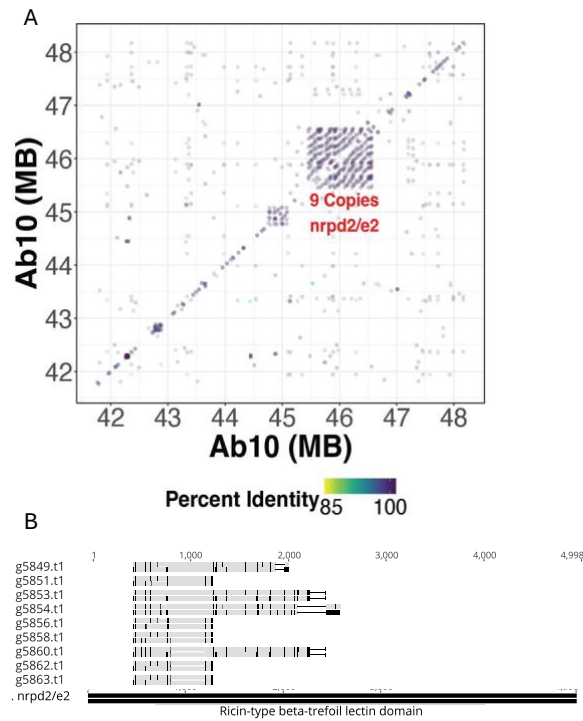


Figure S3.3: **Subset of Sequence Comparison of Ab10 to Ab10.** A. Each dot marks the start of a maximal unique match (MUM) of at least 300 bp long between various chromosome 10 haplotypes all of which begin at the *colored1* gene (Marçais *et al.* 2018). The color of each dot represents the percent identity of that match. All large knob arrays were removed for the sake of clarity. Ab10 refers to the assembly generated in this work. Array with 9 copies of the nrpd2/e2 homolog is marked. B. Alignment of the coding sequence of the 9 copies of the nrpd2/e2 homolog to the coding sequence of their closest homolog with the only protein functional domain marked. For all genes the top bar indicates the DNA coding sequence and the bottom line represents the protein translation. Grey indicates sequence that is identical to the nrpd2/e2 reference and black indicates sequence that is different from nrpd2/e2 the reference.

Abnormal Chromosome 10

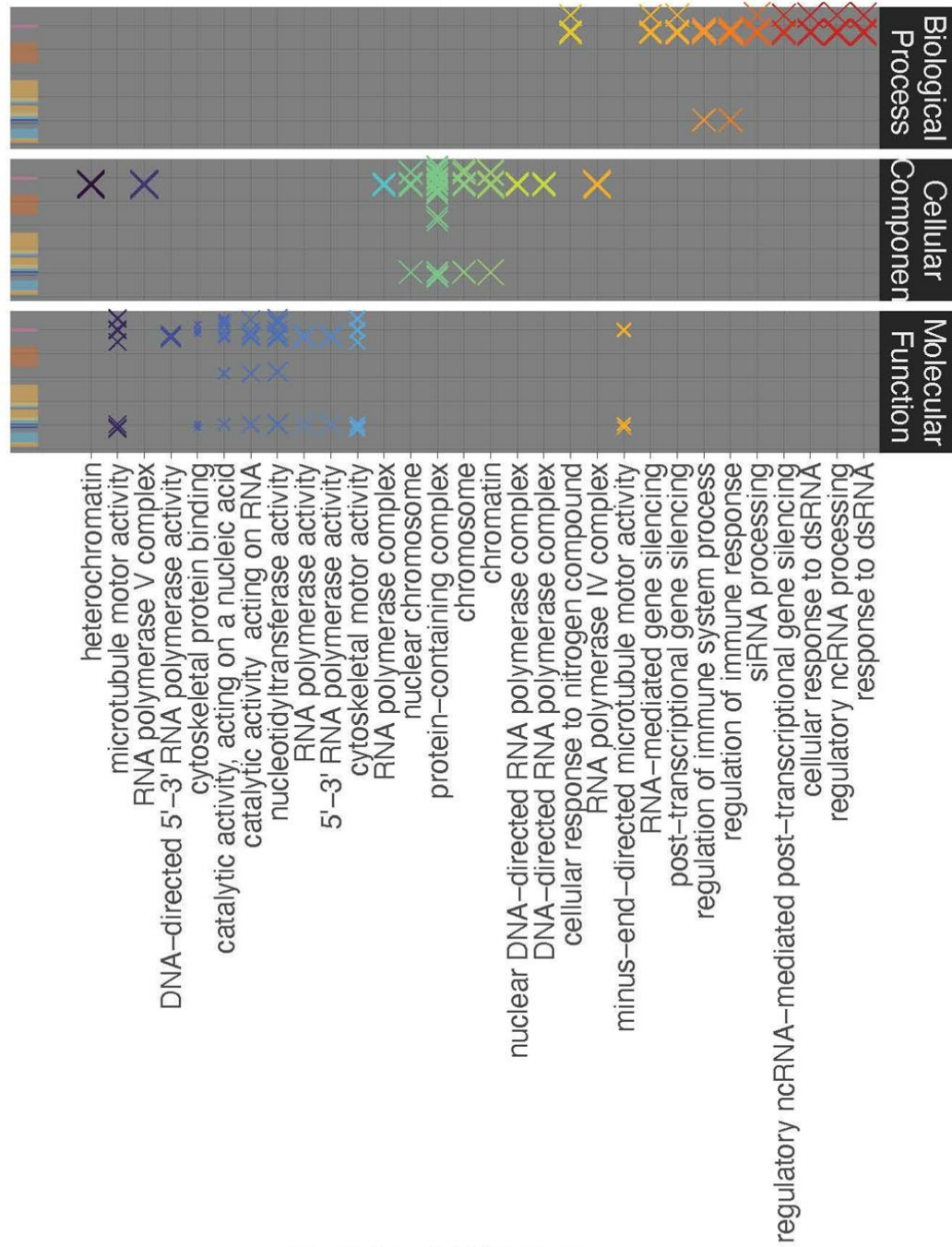


Figure S3.4: **GO term enrichment on Ab10 non-shared Regions.** Non-shared region refers to the regions with no consistent homology to normal chromosome 10. The y axis represents significantly enriched GO terms, the x axis indicates where genes associated with that GO term are located on the Ab10 haplotype. Color of each X represents fold enrichment, size represents statistical significance of enrichment. Relevant regions are marked by colored boxes: gold = shared, light blue = TR1 knob, dark blue = *trkin*, dark orange = knob180 knob, pink = *kindr*.

Abnormal Chromosome 10

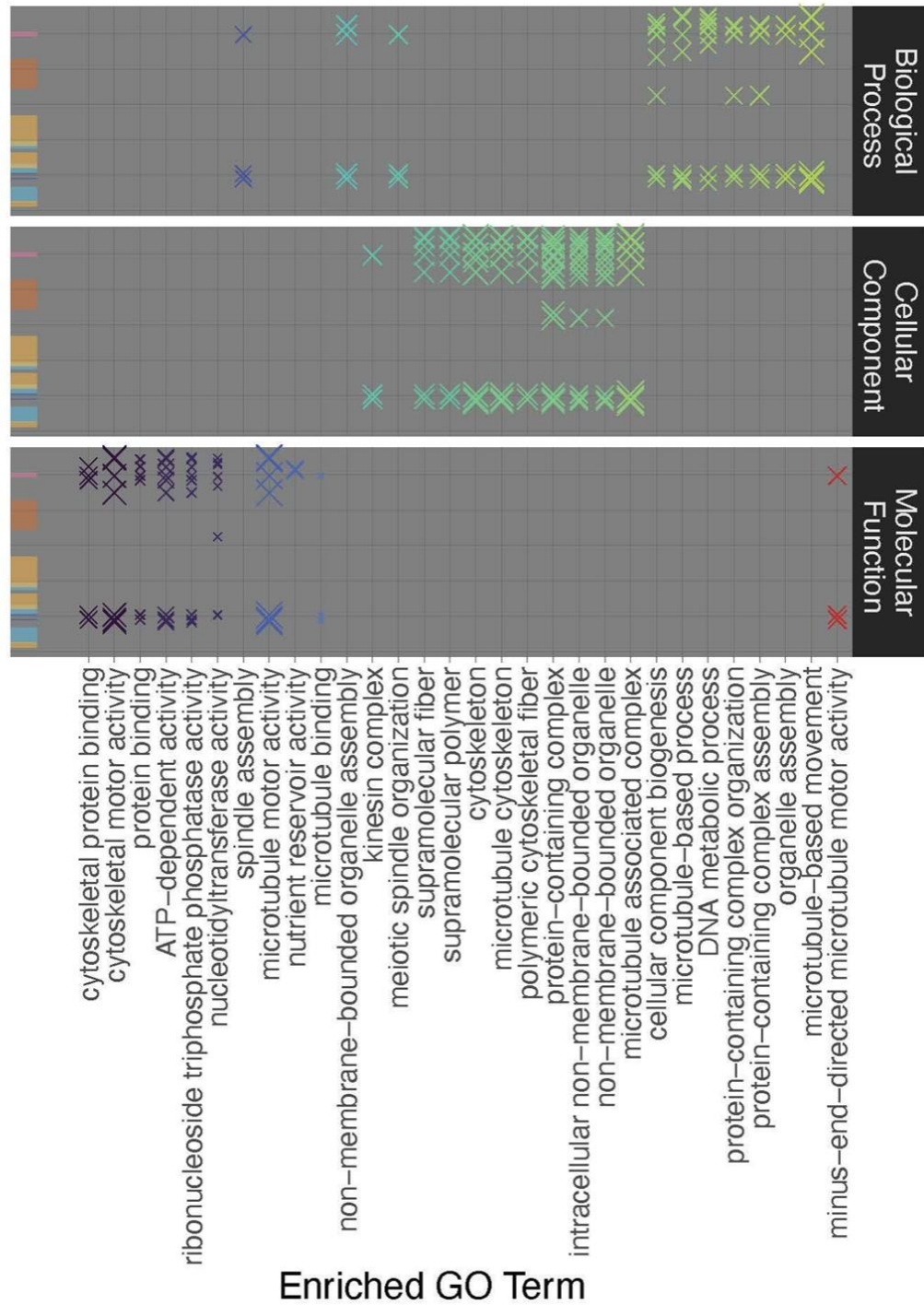


Figure S3.5: GO term enrichment on Ab10 non-shared Regions Without Duplicate Genes.

Non-shared region refers to the regions with no consistent homology to normal chromosome 10.

The y axis represents significantly enriched GO terms, the x axis indicates where genes associated with that GO term are located on the Ab10 haplotype. Color of each X represents fold enrichment, size represents statistical significance of enrichment. Relevant regions are marked by colored boxes: gold = shared, light blue = TR1 knob, dark blue = *trkin*, dark orange = knob180 knob, pink = *kindr*.

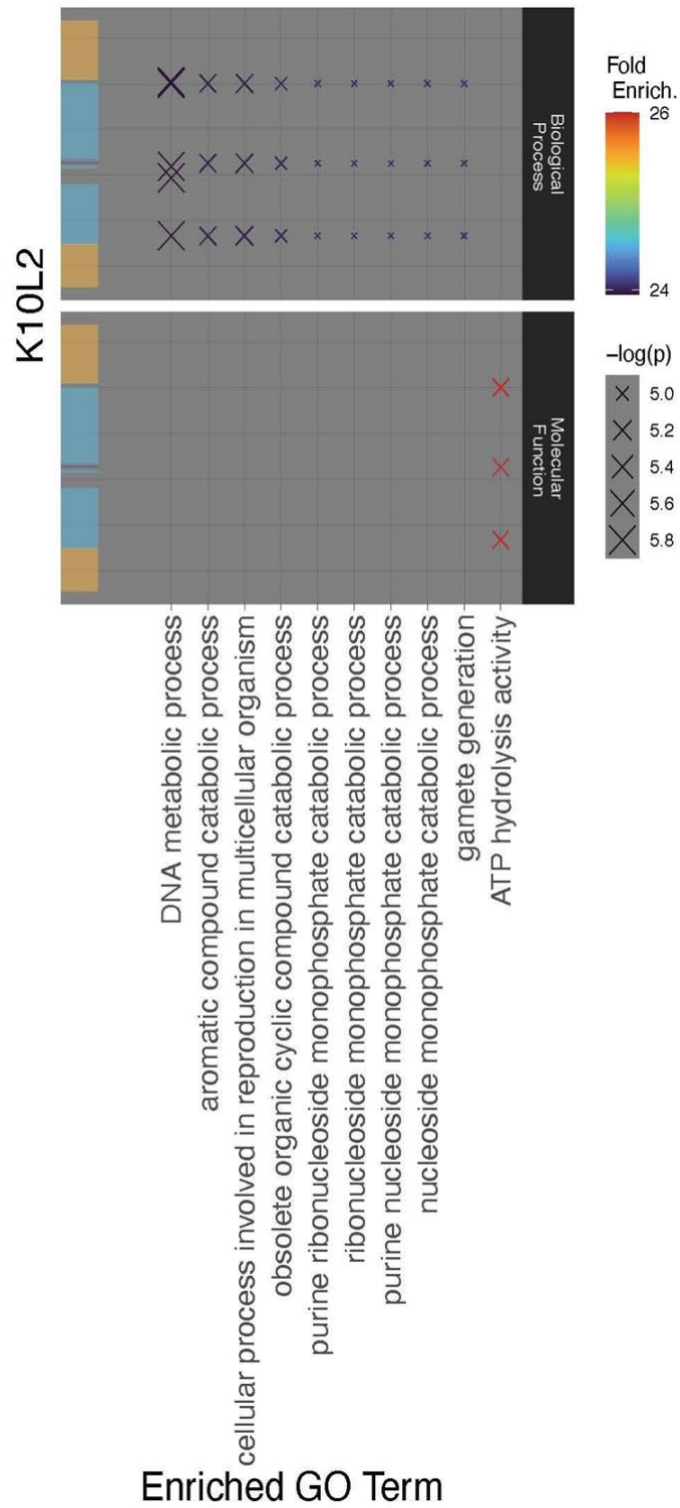


Figure S3.6: **GO term enrichment on K10L2 non-shared Regions.** Non-shared region refers to the regions with no consistent homology to normal chromosome 10. The y axis represents significantly enriched GO terms, the x axis indicates where genes associated with that GO term are located on the K10L2 haplotype. Color of each X represents fold enrichment, size represents statistical significance of enrichment. Relevant regions are marked by colored boxes: gold = shared, light blue = TR1 knob, dark blue = *trkin*.

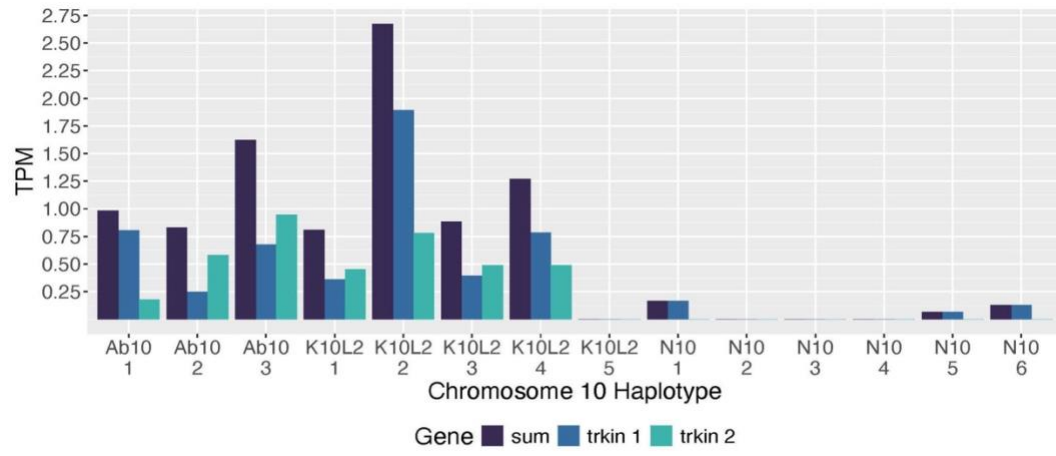


Figure S3.7: **Expression of *Trkin* in Ab10, K10L2, and N10.** Transcripts per million (TPM) for Ab10 *Trkin1* and *Trkin2* as well as their sum from mRNA sequencing data (Swentowsky *et al.* 2020).

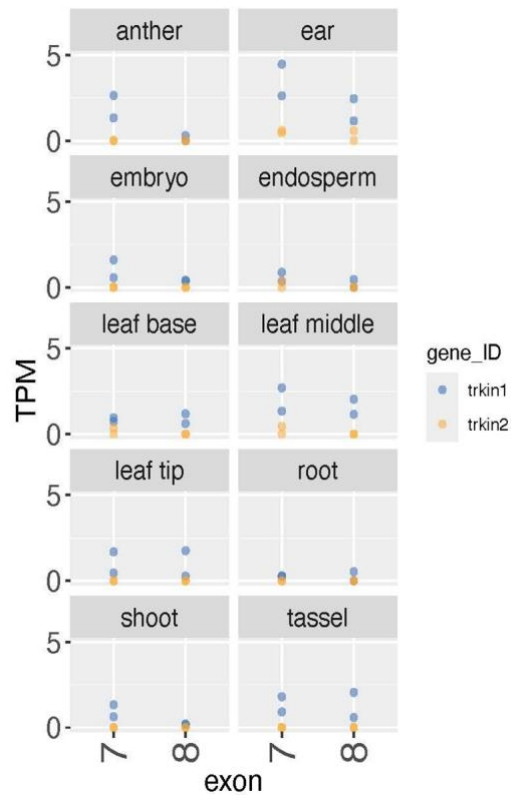
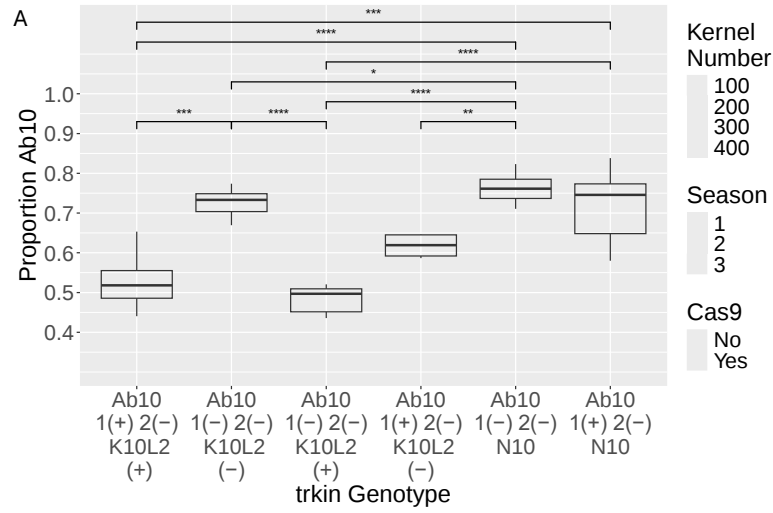


Figure S3.8. **Expression of Ab10 *Trkin1* and *Trkin2*.** TPM indicates transcripts per million.

Each tissue was sequenced in two replicates indicated by two points per gene (color). Only exons 7 and 8 are differentiable between Ab10 *Trkin1* and *Trkin2*, so only those were compared. Ab10 *Trkin1* mean expression was 1.078 TPM and Ab10 *trkin2* mean expression was 0.069 TPM. Welch t sample t test revealed the expression of exons 7 and 8 were different between Ab10 *Trkin1* and Ab10 *Trkin2* were significantly different ($t = 6.5734$, $df = 41.476$, $p\text{-value} = 6.286 \times 10^{-8}$).



B

Comparison	diff	lwr	upr	p adj	sig
Ab10 trkin(-) K10L2 trkin(-) / Ab10 trkin(+) K10L2 trkin(+)	11.92	3.84	20.01	0.00	***
Ab10 trkin(-) K10L2 trkin(+) / Ab10 trkin(+) K10L2 trkin(+)	-8.22	-19.15	2.70	0.25	
Ab10 trkin(+) K10L2 trkin(-) / Ab10 trkin(+) K10L2 trkin(+)	4.99	-8.64	18.62	0.89	
Ab10 trkin(-) N10 / Ab10 trkin(+) K10L2 trkin(+)	22.48	13.39	31.56	0.00	****
Ab10 trkin(+) N10 / Ab10 trkin(+) K10L2 trkin(+)	15.06	5.72	24.40	0.00	***
Ab10 trkin(-) K10L2 trkin(-) / Ab10 trkin(-) K10L2 trkin(-)	-20.14	-31.23	-9.06	0.00	****
Ab10 trkin(+) K10L2 trkin(-) / Ab10 trkin(-) K10L2 trkin(-)	-6.93	-20.69	6.82	0.68	
Ab10 trkin(-) N10 / Ab10 trkin(-) K10L2 trkin(-)	10.56	1.28	19.83	0.02	*
Ab10 trkin(+) N10 / Ab10 trkin(-) K10L2 trkin(-)	3.14	-6.38	12.67	0.93	
Ab10 trkin(+) K10L2 trkin(-) / Ab10 trkin(-) K10L2 trkin(+)	13.21	-2.39	28.81	0.14	
Ab10 trkin(-) N10 / Ab10 trkin(-) K10L2 trkin(+)	30.70	18.87	42.54	0.00	****
Ab10 trkin(+) N10 / Ab10 trkin(-) K10L2 trkin(+)	23.29	11.26	35.32	0.00	****
Ab10 trkin(-) N10 / Ab10 trkin(+) K10L2 trkin(-)	17.49	3.13	31.86	0.01	**
Ab10 trkin(+) N10 / Ab10 trkin(+) K10L2 trkin(-)	10.08	-4.45	24.61	0.33	
Ab10 trkin(+) N10 / Ab10 trkin(-) N10	-7.41	-17.80	2.97	0.30	

Figure S3.9: Effect of *Trkin* on K10L2 Ab10 Competition showing all comparisons and significance values. All plants were grown in the greenhouse in Athens, GA. Each dot represents an individual plant. Season refers to a group of plants grown at the same time. Seasons 1 and 2 were conducted in the same background while Season 3 was conducted in a different background. Season 1 and 2 of the Ab10 *trkin1*(-) *trkin2*(-) and *K10L2 trkin*(-) had *Cas9* segregating. The multi-way ANOVA model was Proportion Ab10 ~ *Cas9* genotype + Round + *trkin* genotype. *Cas9* genotype = F(1,63)=9.656, p=0.00; Round = F(2,63)=0.520 p=0.59726; *trkin* genotype= F(5,63)=19.495, p= 1.11e-11. B. Results for Tukey's HSD Test for multiple comparisons between all genotypes. diff=estimate of effect size, lwr = lower bound of 95% confidence interval, upr= upper bound of 95% confidence interval, p adj = p value adjusted for multiple comparisons, sig = symbol used. *=<0.05, **=<0.01, ***, <0.001, ***=0 *=<0.05, **=<0.01, ***, <0.001, ***=0. A. All significant relationships shown. B. Statistical output for all comparisons made.

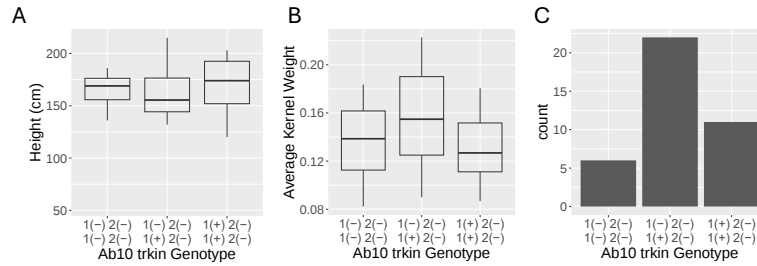


Figure S3.10: Plant fitness Effects of *trkin* in Ab10 Homozygotes. Plants were grown in the green house in Athens GA in a fully randomized order. A. Height: Fitted linear regression model was: Height ~ Pot + Position in Greenhouse + *trkin* Genotype $R^2=0.1202$, $F(17,21)=1.305$, $p=0.2782$. B. Average Kernel Weight: Fitted linear regression model was: Average Kernel Weight ~ Pot + Position in Greenhouse + Silking time + Anthesis Time + *trkin* Genotype $R^2=0.3251$ $F(20,10)=1.723$, $p=0.1891$. D,E. C. Transmission: Chi squared test was used to determine if the observed segregation of the *trkin* genotypes fit with Mendelian segregation $\chi^2(2)=1.2671$, $p=0.5307$.

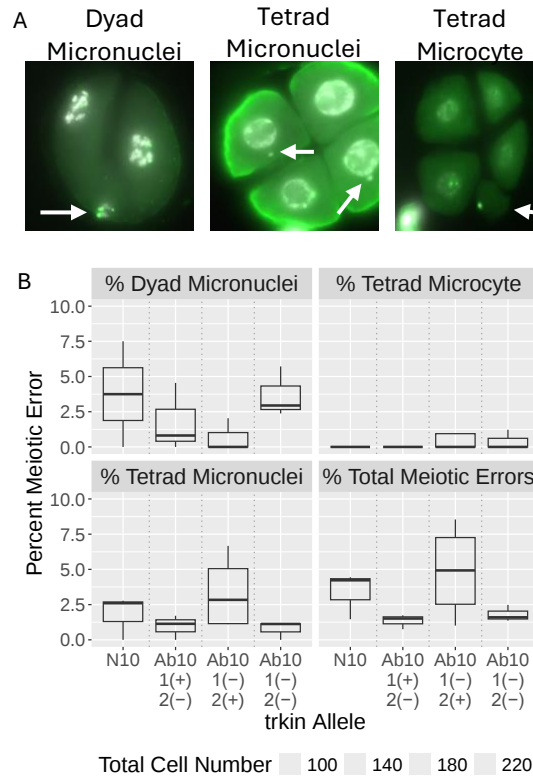


Figure S3.11: Meiotic Errors in Male Meiocytes of Various *Trkin* Genotypes. Meiotic errors were scored on Ab10 or N10 homozygous plants stained for FISH (Figure 7) with the indicated *trkin* genotypes. Dyad micronuclei refers to a lost chromosome at the conclusion of meiosis I. Tetrad micronuclei refers to a lost chromosome at the conclusion of meiosis II. Tetrad microcyte refers to an additional small cell containing DNA likely representing a lost chromosome at the conclusion of meiosis II. A. Contains examples of all scored meiotic errors. Cells with meiotic errors were normalized against the total number of same stage cells observed. Each dot represents an individual plant. One Way ANOVA determined no statistical difference in any class of meiotic error between *trkin* genotypes: % Dyad Micronuclei ($F(3,9)=0.413$, $p=0.748$, % Tetrad Micronuclei ($F(3,9)=1.552$, $p=0.268$, % Tetrad Microcyte ($F(3,9)=0.549$, $p=0.661$, % Total Meiotic Errors ($F(3,9)=1.89$, $p=0.202$).

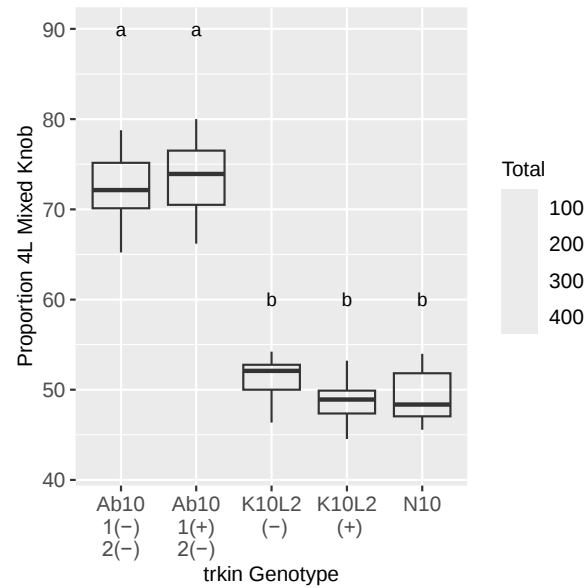


Figure S3.12: **Effect of *Trkin* on Segregation of a Mixed Knob Not on Chromosome 10.** All plants were grown in the greenhouse in Athens, GA. Each dot represents an individual plant. One way ANOVA model was Proportion 4L Mixed Knob $\sim$ *trkin* genotype $F(4,41)= 101, p=<2e-16$. Tukey's HSD Test for multiple comparisons found that the mean value of all Ab10 bearing lines were significantly different from all K10L2 and N10 bearing lines (all p values = 0.00). Ab10 lines were not significantly different from each other. K10L2 and N10 lines were not significantly different from each other.

APPENDIX C

SUPPLEMENTARY TABLES AND FIGURES – CHAPTER 4

Table S4.1. **Summary table of all CDHs identified.**

Variety	Total	Ab10 Positive	K10L2 Positive	B Chr Positive	Ab10 B Chr Positive (expectation =0.81%)	K10L2 B Chr Positive (expectation = 0.66%)
Inbred	4525	2 (0.04%)	23 (0.51%)	16 (0.35%)	0 (0%)	0 (0%)
Landrace	5334	337 (6.32%)	275 (5.16%)	655 (12.28%)	42 (0.79%)	55 (1.03%)
Teosinte - Mexicana	146	7 (4.79%)	7 (4.79 %)	4 (2.74%)	0 (0%)	0 (0%)
Teosinte - Parviglumis	168	6 (3.57%)	5 (2.98%)	1 (0.60%)	0 (0%)	0 (0%)

Table S4.2. Inbred lines Positive for CDHs and those used to verify CDH detection method.

Accession	CDH	Total	Positive
PI-639505	Ab10	11	6
NSL-68244	Ab10	5	2
PI-702819 and NSL437882.1	K10L2	3	1
PI-601300	K10L2	9	8
Ames22440.3	K10L2		
Ames22440.5	K10L2		
Ames26795.1	K10L2		
Ames28954.1	K10L2		
CI1872.1	K10L2		
CI1872.2	K10L2		
CI66.1	K10L2		
CI66.2	K10L2		
CI90C.1	K10L2		
CI90C.2	K10L2		
CIze44.2	K10L2		
CIze44.3	K10L2		
CIze90.1	K10L2		
CIze91.1	K10L2		
CIze91.2	K10L2		
CML320.1	K10L2		
CML416.1	K10L2		
CML416(Krakowsky).1	K10L2		
CML454.1	K10L2		
PI587148.1	K10L2		
PI641061.1	K10L2		
PI-198903	B Chr	9	7
PI-186184	B Chr	6	1
Ames21512	B Chr		
Ames23506	B Chr		
Ames26125	B Chr		
NSL22631	B Chr		
PI186229	B Chr		
PI198889	B Chr		
PI200196	B Chr		
PI340812	B Chr		
PI641259	B Chr		
TIL06	B Chr		
TIL06	B Chr		
TIL06	B Chr		
TIL06	B Chr		
TIL06	B Chr		

Table S4.3. All CDH associated SNPs and genes they overlap.

SNP (Mo17 coordinate)	CDH	Direction	Mo17 Gene ID	B73 Gene ID	Gene Description
Chr3 SNP 1 (192484757)	Ab10	Negative	Zm00014ba176840	Zm00001eb149620	ppr189 - pentatricopeptide repeat protein189
Chr3 SNP 2 (192486778)	Ab10	Positive	Zm00014ba176850	Zm00001eb149630	Probable zinc-ribbon domain
Chr3 SNP 3 (192562582)	Ab10	Negative	NA		
Chr4 SNP 1 (193239638)	Ab10	Positive	Zm00014ba223110	Zm00001eb195700	C3H-transcription factor 350
Chr4 SNP 2 (193239646)	Ab10	Positive	Zm00014ba223110	Zm00001eb195700	C3H-transcription factor 350
Chr4 SNP 3 (193239658)	Ab10	Positive	Zm00014ba223110	Zm00001eb195700	C3H-transcription factor 350
Chr8 SNP 1 (73817047)	Ab10	Positive	Zm00014ba366470	Zm00001eb343470	
Chr9 SNP 1 (29026157)	Ab10	Positive	NA		
*Chr9 SNP 2 (29026173)	Ab10	Positive	NA		
Chr10 SNP 1 (71197305)	Ab10	Negative	NA		
Chr10 SNP 2 (71198303)	Ab10	Negative	NA		
Chr1 SNP 1 (277994336)	K10L2	Positive	Zm00014ba054920	Zm00001eb056120	FHA-transcription factor 4
Chr4 SNP 1 (2010582)	K10L2	Positive	Zm00014ba192070	Zm00001eb165250	Pentatricopeptide repeat-containing prot
Chr4 SNP 2 (5674254)	K10L2	Positive	Zm00014ba193770	Zm00001eb167120	cpps4 - copalyl diphosphate synthase4
Chr5 SNP 1 (190520703)	K10L2	Positive	Zm00014ba275220	Zm00001eb248620	Shortage in chiasmata 1
Chr6 SNP 1 (175749381)	K10L2	Positive	Zm00014ba311330	Zm00001eb287290	Methyltransferase type 11 domain-containing protein
Chr8 SNP 1 (163979957)	K10L2	Negative	NA		
Chr8 SNP 2 (164038154)	K10L2	Negative	Zm00014ba384340	Zm00001eb361900	Uncharacterized protein
Chr3 SNP 1 (5366715)	B Chr.	Negative	NA		
**Chr3 SNP 2 (5366738)	B Chr.	Positive	NA		
Chr3 SNP 3 (211286197)	B Chr.	Positive	Zm00014ba181860	Zm00001eb154570	flz14 - FCS-like zinc finger14
Chr4 SNP 1 (248470517)	B Chr.	Positive	Zm00014ba236070	Zm00001eb400700	denosylmethionine decarboxylase
*Chr6 SNP 1 (124809142)	B Chr.	Positive	Zm00014ba298980	Zm00001eb274250	Protein JASON
Chr2 SNP 1 (90555553)	B Copy	Positive	NA		
Chr2 SNP 2 (90555560)	B Copy	Positive	NA		
Chr3 SNP 1 (5366738)	B Copy	Positive	NA		
*Chr6 SNP 1 (124809142)	B Copy	Positive	Zm00014ba298980	Zm00001eb274250	Protein JASON

* Very likely artifact excluded from all analyses

** Suspected artifact included in only Supplementary Figure 7

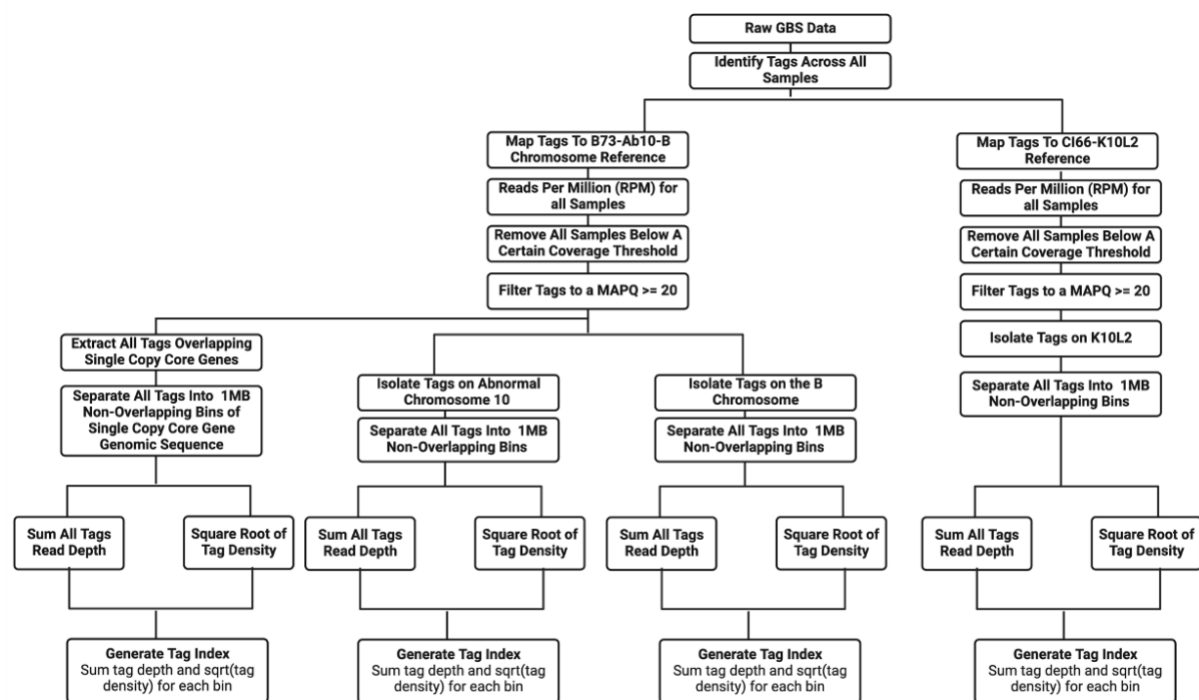
Table S4.4. Genotype By Sequencing (GBS) control lines.

CDH	Type	Copy Number	ID	Source	Number Sequenced
Ab10	I	1	PI 628445	GRIN	5
Ab10	I	1	Marcus Rhoades Collection	GRIN	5
Ab10	I	unknown	Marcus Rhoades Collection	GRIN	41
Ab10	II	1	Ames 21826	GRIN	5
Ab10	II	1	Marcus Rhoades Collection	GRIN	5
Ab10	II	unknown	Marcus Rhoades Collection	GRIN	37
Ab10	II	1	Ames 21826	GRIN	5
Ab10	III	1	PI 444296	GRIN	5
Ab10	III	1	PI 444834	GRIN	5
Ab10	III	1	PI490825	GRIN	5
Ab10	III	1	Ames 19980	GRIN	5
Ab10	I	2	NA - B73 Ab10 bc 6x into TX40JB	Maize GDB	3
Ab10	unknown	1	PI-483314	GRIN	1
K10L2	NA	2	NA- CI66 backcrossed 6x into B73	GRIN	2
K10L2	NA	1	NA- CI66 backcrossed 6x into B73	GRIN	9
K10L2	NA	2	PI-483314	GRIN	1
K10L2	NA	1	PI-483314	GRIN	1
B Chr	NA	unknown	B542C L289 B chromosomes	Maize GDB	4
B Chr	NA	unknown	NA - B-peru marked B chromosome	Birchler Lab	4
B Chr	NA	unknown	NSL- 2833	GRIN	4
B Chr	NA	unknown	PI-490921	GRIN	3
None	NA	0	W23	Dawe lab stocks	19
None	NA	0	B73	Dawe lab stocks	3
None	NA	0	B542C L289 B chromosomes	Maize GDB	3
None	NA	0	TX40JB	Maize GDB	3
None	NA	0	PI-217409	GRIN	3
None	NA	0	PI-444954	GRIN	3

Table S4.5. **Primers used for genotyping.**

Primer Name	Sequence
Kindr F1	TACTGGTGGGCTCTTTCGAC
Kindr R1	GTTCTTAGGAGCGGGGGTT
Trkin F1	TTCCACGCGTGTACTGG
Trkin R2	ATTAGATGAATGTGAATTGCGG
B Chrom Repeat F1	CCAGTGTCCGGTGGTCTTAG
B Chrom Repeat R2	AATAAGCCAGCGGCGATTCT
K10L2 F1 (followed by bccI digestion)	CACCGTCGATCGTCAGAA
K10L2 R1 (followed by bccI digestion)	CGCATATCCTCCATCTCACTA

A



B

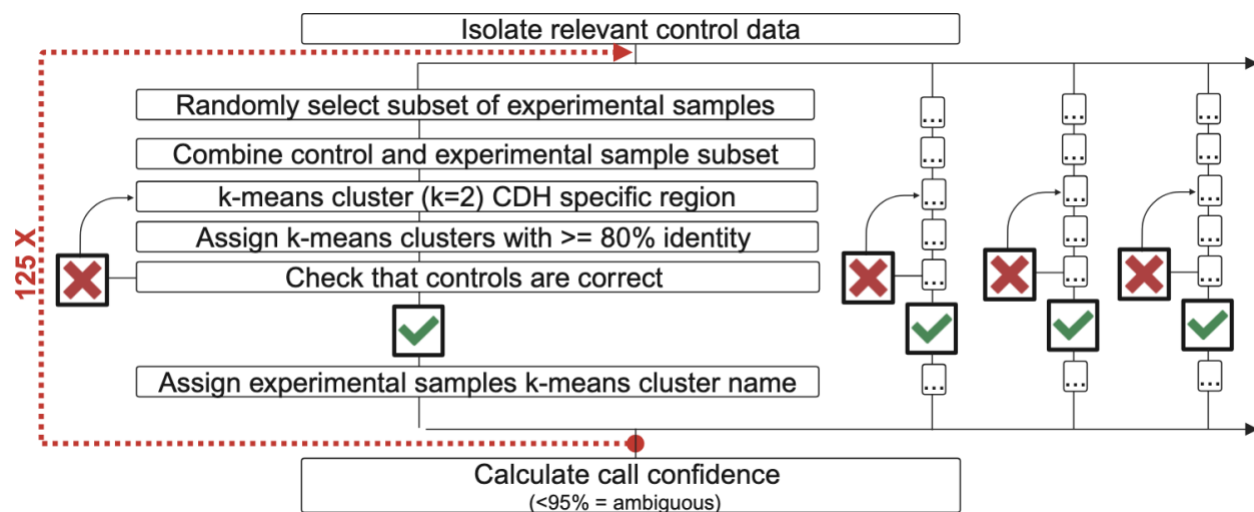
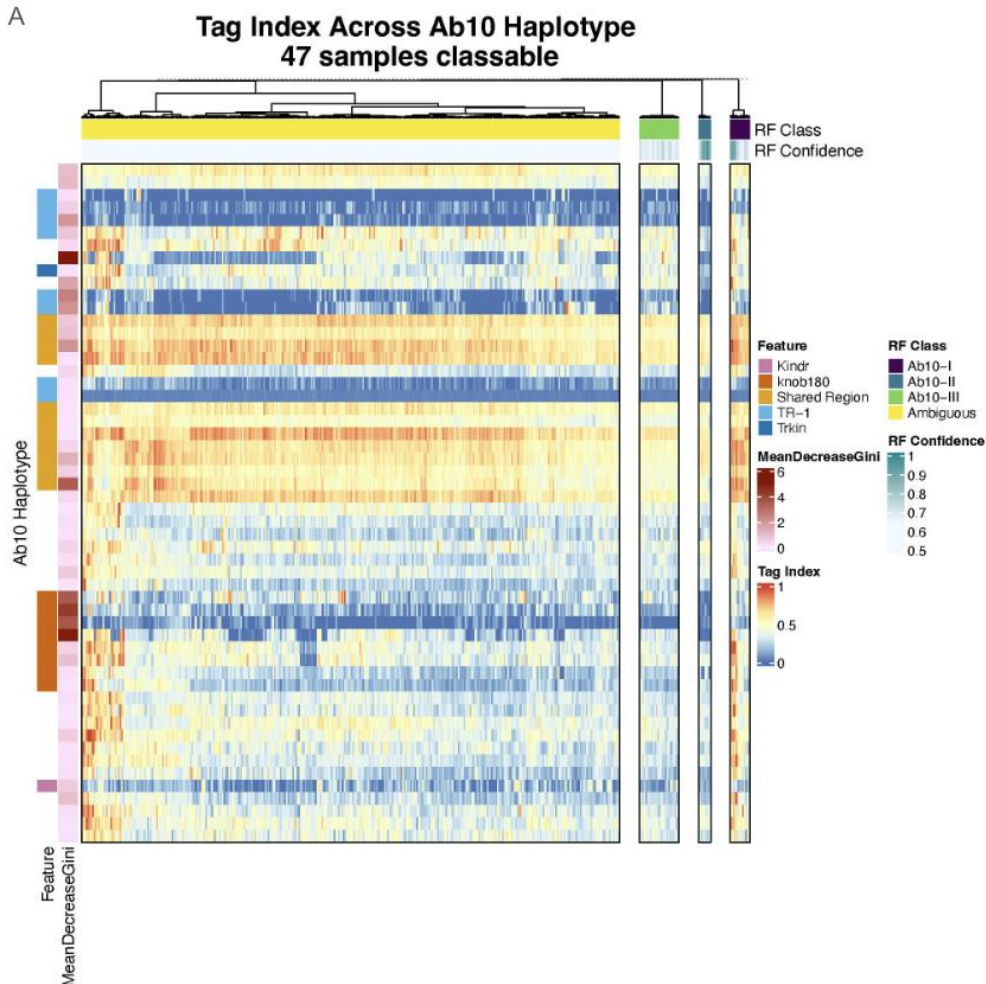


Figure S4.1. **Workflow Diagrams.** **Workflow Diagrams.** A. Workflow diagram for the generation of the tag index in all CDHs as well as the single copy core genes (12). B. Diagram of the general workflow for detecting CDHs in experimental samples. Check indicates passing, x indicates failing.

A



B

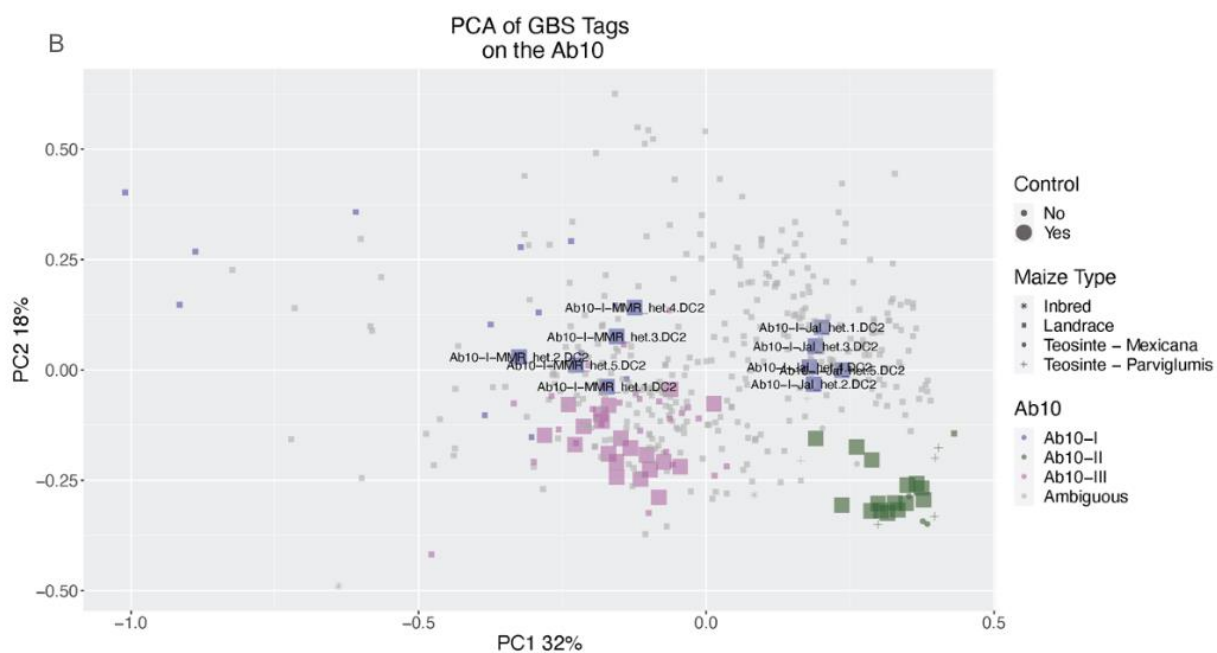


Figure S4.2. **Identification of Ab10 Type.** Min/Max scaled tag index for all Ab10 positive control and experimental samples. Random forest models determined Ab10 types are plotted separately. Each group is ward.D clustered. The x axis represents individual samples, the y axis represents the Ab10 haplotype. The data source of origin, the known Ab10 status of control lines, and the class that the random forest model identified as well as how confident that call is are on the x axis. The RF confidence value indicates the proportion of decision trees that are called the predominant Ab10 type. On the y axis relevant features of the Ab10 haplotype are indicated and their importance to determining Ab10 type in the random forest model (mean decreasing Gini).

B. A PCA of all the Ab10 positive samples scaled tag index with controls and their type indicated.

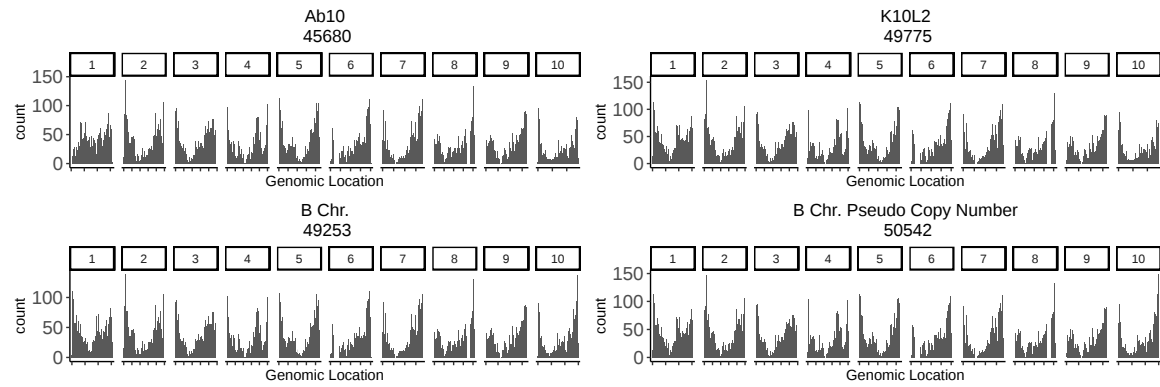


Figure S4.3. **Location of SNPs Used for GWAS.** Numbers below the CDH name indicate the total number of SNPs.

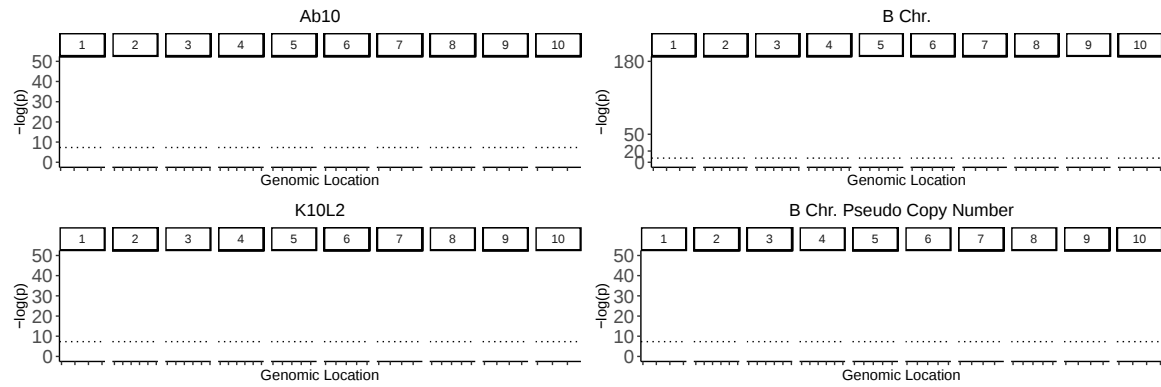


Figure S4.4. **Relationship of CDH to Genetic Loci** Manhattan plots for genome wide association study results for each CDH and B chromosome copy number. Dotted black line indicates a p value of 5×10^{-8} . The Ab10 associated SNP on chromosome 9 is believed to be an artifact. The B chr. associated SNPs on chromosomes 3 and 6 are believed to be artifacts.

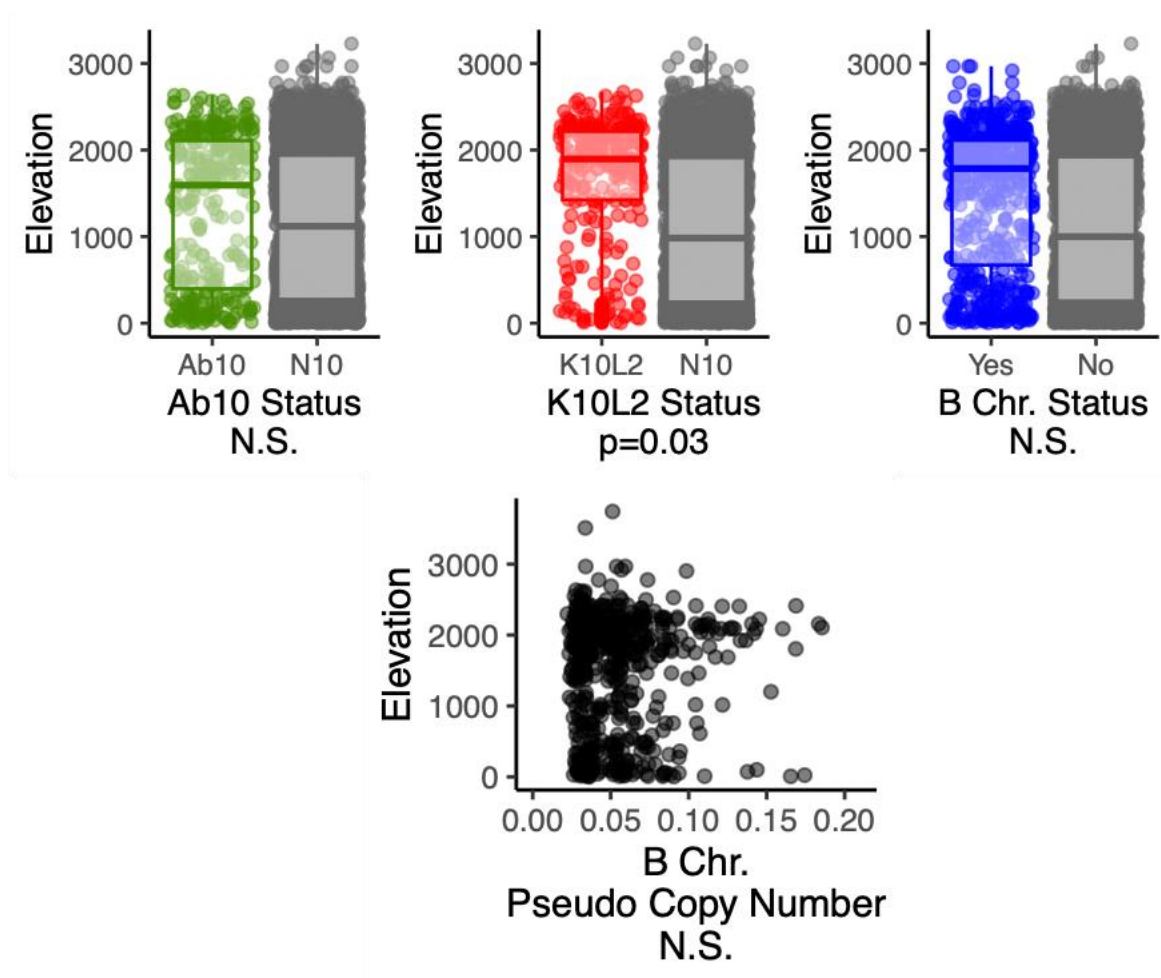


Figure S4.5. **Relationship of CDHs to Elevation.**

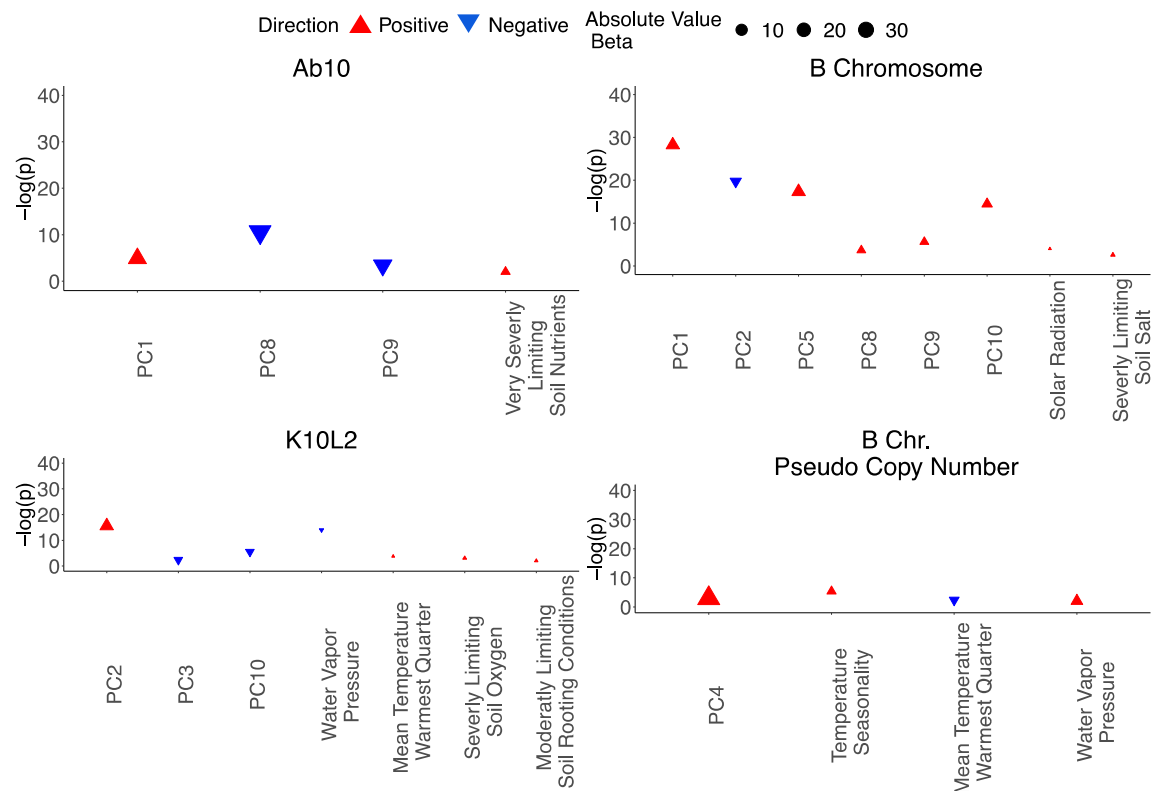


Figure S4.6. **Relationship Between CDHs and the Environment.** Plots of fully simplified generalized linear models for each CDH including population structure and environmental variables. Shape color and orientation indicate the direction of the relationship to the CDH. Share size represents the effect size, units are not always comparable between variables.

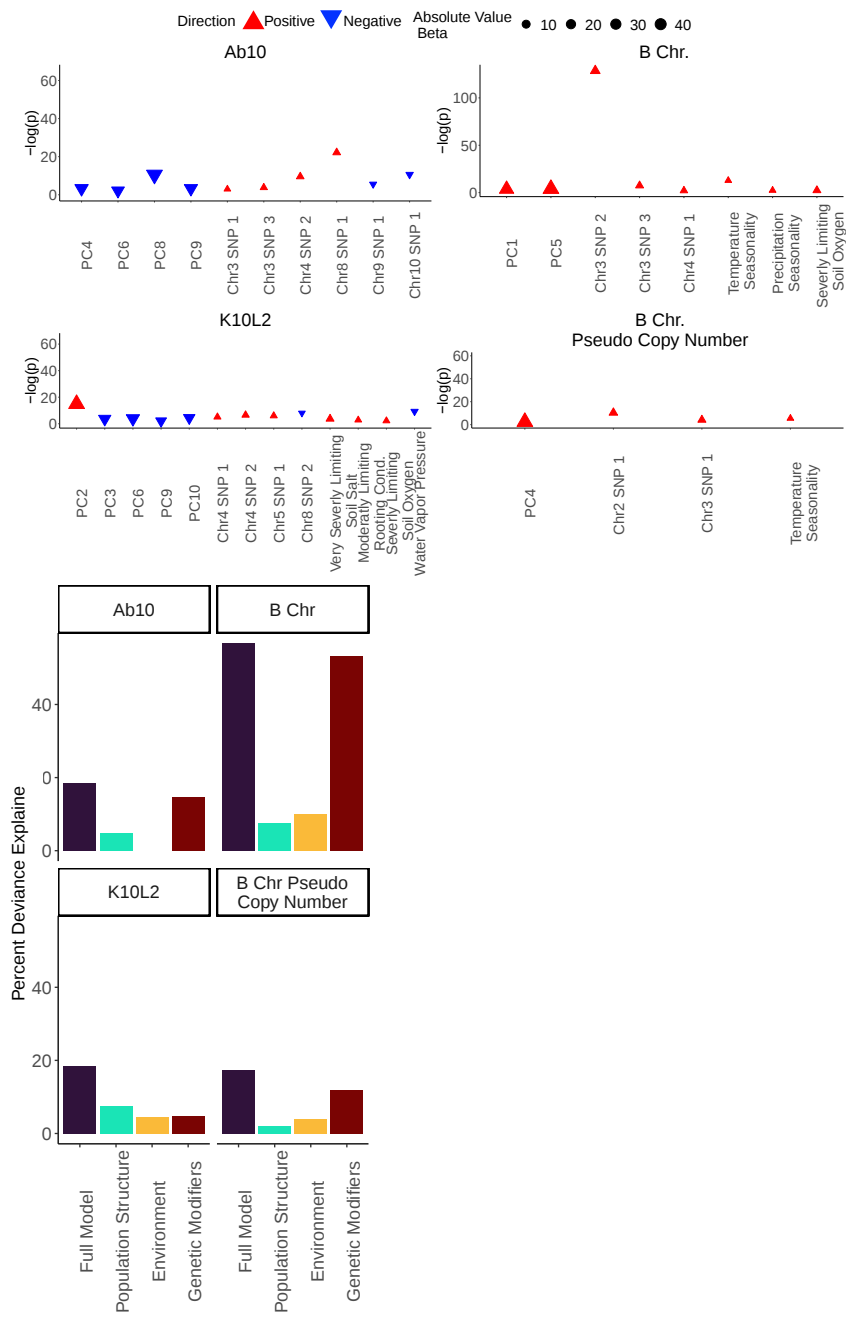


Figure S4.7. Relationship of CDH to the Environment and Genetic Loci Including Suspected Artifact. A. Plots of fully simplified generalized linear models for each CDH including population structure, genetic loci, and environmental variables. Shape color and orientation indicate the direction of the relationship to the CDH. Shape size represents the effect size, units are not always comparable between variables. All soil variables are from [\(51\)](#) C. Partitioned deviance of each model shown. The partitions do not sum to the full model due to shared variation between the partitions but indicate a relative relationship.

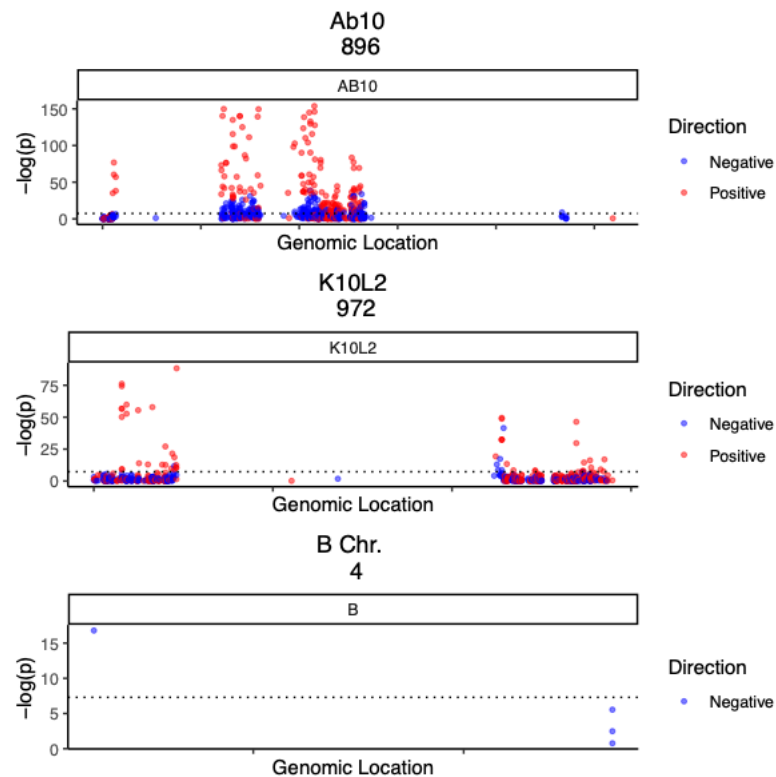


Figure S4.8. **CDH Control Manhattan Plots.** Manhattan plots for genome wide association study results for each CDH and B chromosome copy number for SNPs mapped to the CDH. Pos = positive association, neg = negative association. Top dotted grey line indicates a p value of 5×10^{-9} and bottom dotted grey line indicates a p value of 5×10^{-8} .

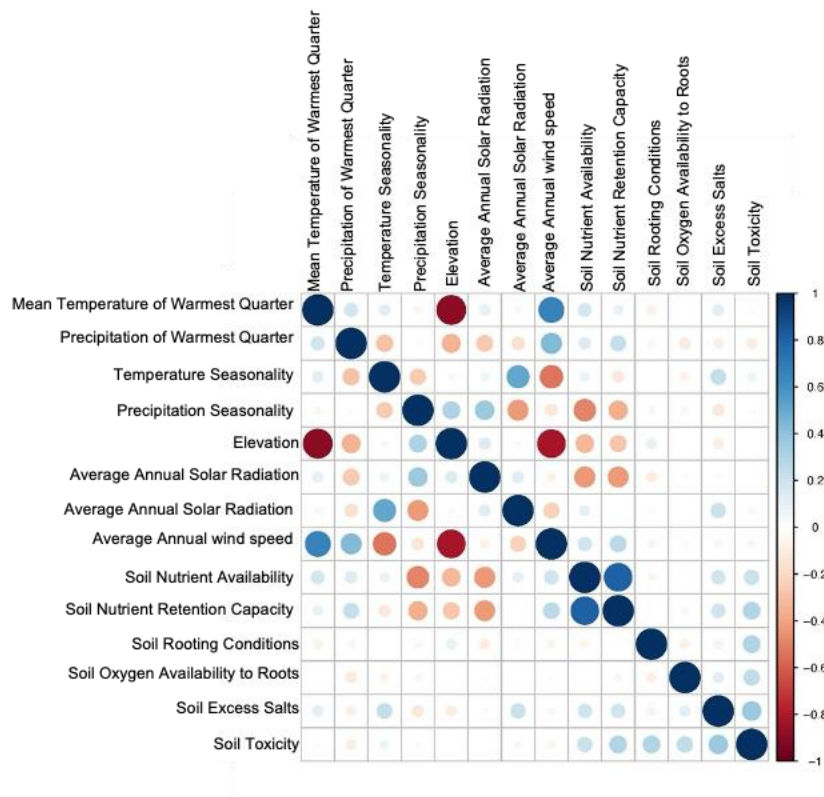


Figure S4.9. **Correlation Matrix for Environmental Variables.** Correlation matrix for selected environmental variables from (50, 51).