COUNTING ON CROSSOVERS: INSIGHTS INTO GENE MAPPING AND CONTROLLED RECOMBINATION FOR ALLOPOLYPLOID PLANT BREEDING

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Ella Taagen
May 2022

COUNTING ON CROSSOVERS: INSIGHTS INTO GENE MAPPING AND CONTROLLED RECOMBINATION FOR ALLOPOLYPLOID PLANT BREEDING

Ella Taagen, Ph. D.
Cornell University 2022

Plant breeders rely on heritable genetic variation for trait improvement, and the two primary novel sources of this variation are recombination and mutation of genetic material during meiosis. These independent processes can have overlapping significance for breeders, as the genetic resolution generated by recombination influences our ability to locate mutations. In addition, population improvement often results in inbreeding, which reduces the effective recombination and increases the likelihood of deleterious mutation hitchhiking via repulsion linkages. This is especially true in regions of chromosomes with low recombination rates, like the pericentromere. Polypoid crops pose an additional challenge to pinpointing the genetic control of desirable traits, as the phenotypic consequence of a single-variable locus (e.g., genomic structural variation) can be masked by the redundant copies of other homoeologous genomes. Here I address how geneticists can improve the accuracy of identifying targets for positional cloning, and whether breeders should seek to manipulate recombination to further harness the power of selection, using hexaploid wheat (*Triticum aestivum* L.) as a model.

First, I present a study with a traditional approach to positionally clone a quantitative trait locus for yield components on wheat chromosome arm 5A. Leveraging a fine-mapping population, genomic data, phenotypic associations, early grain development transcriptome profiles, and predicted gene function, it was

determined the quantitative trait locus was a result of strong linkage disequilibrium with a large deletion on wheat chromosome arm 5AS. This study highlights the phenotypic resiliency of polyploids harboring structural variants and actionable recommendations to increase the likelihood of identifying causal variants in wheat.

Second, I used simulation models with empirical data to assess the potential of controlled recombination for genomic selection breeding programs. While controlled recombination remains under research and development, initial reports have prompted interest in evaluating increased recombination in the pericentromere to disrupt repulsion linkages. Comparing high and low values for a range of simulation parameters identified that few combinations under increased recombination retained greater genetic variation and fewer still achieved higher genetic gain. More recombination was associated with loss of genomic prediction accuracy, which often outweighed the benefits of disrupting repulsion linkages. Irrespective of recombination frequency and distribution and QTL location, enhanced response to selection under increased recombination largely depended on quantitative trait architecture, high heritability, more repulsion than coupling linkages, and greater than six cycles of genomic selection. Altogether, the results discourage a controlled recombination approach to genomic selection in wheat as a more efficient path to retaining genetic variation and increasing genetic gains compared to existing breeding methods.

BIOGRAPHICAL SKETCH

Ella Taagen is a quantitative geneticist and data scientist, motivated by solving research bottlenecks to genetic gain, food security, and sustainable food production. She grew up in Seattle, Washington and was first introduced to the world of scientific research during a high school internship with the Center for Global Infectious Disease Research. She went on to attend the University of Washington and in 2016 completed a Bachelor of Science in Molecular, Cellular, and Developmental Biology, with a minor in Nutritional Sciences. During her undergraduate education she worked in multiple academic research labs and discovered her love for molecular mechanisms and studying plants in Dr. Takato Imaizumi's lab, assisting then post-doc Dr. Akane Kubota. In 2016 she interned under then PhD candidate Dr. Bethany Econopouly in Dr. Stephen Jones' lab at Washington State University and was encouraged to pursue graduate school herself. Ella came to Ithaca, New York in 2017 to start a PhD in the Small Grains Breeding and Genetics program at Cornell University under the advisement of Dr. Mark Sorrells. During her PhD, Ella explored research interests spanning applied breeding, molecular biology, gene editing, meiotic recombination, and theoretical quantitative genetics. A plant breeder's reliance on natural recombination rose to the forefront of her research interests after experiencing her own struggles to achieve genetic resolution while fine-mapping a quantitative trait locus in wheat. This inspired her doctorate's focus on the range and limit of recombination, which she has demonstrated may more efficiently harness the power of selection for plant breeding and influence the development of novel gene editing approaches. In 2021 Ella participated in a six-month internship with Bayer Crop Science and will resume working with the company as a Data Scientist at the intersection of variant discovery and gene editing strategy development in July 2022.

ACKNOWLEDGMENTS

I would like to wholeheartedly thank my advisor, Dr. Mark Sorrells, for his continuous support, trust, and guidance. During my doctorate he challenged me to be a better scientist every day and imparted foundational mentorship skills. Through our shared values of independence and responsibility, he granted me the time and freedom necessary to identify the motivations driving my career centered on crop genetic improvement. I am also incredibly grateful for the advice, enthusiasm, and honesty of my committee members and co-authors. Dr. Adam Bogdanove has been instrumental in growing my creative and critical thinking skills and merging the disciplines of applied plant breeding and DNA targeting. Dr. Margaret Smith has provided excellent technical advice, career guidance, and helped me find clarity at numerous points of personal uncertainty, especially as a woman in science. Dr. Jean-Luc Jannink's mentorship has refined my quantitative genetics and data science skillset, and imparted the confidence to question results and test assumptions.

My research would not have been possible without the support of the Small Grains program technicians David Benscher, Amy Fox, and James Tanaka. I would like to express my gratitude for Synapsis members and lab-mates past and present who have shaped the scientist I am today and will be lifelong colleagues and friends, especially Dr. Shantel Martinez, Dr. Daniel Sweeney, Karl Kunze, Will Stafstrom, and Taylor Ferebee. Lastly, I would like to thank my wonderful family for encouraging my education, and always being one phone call away.

Funding for this research, tuition, and a stipend was provided by the United States Department of Agriculture National Institute of Food and Agriculture grant 2017-67007-25939 and 2022-68013-36439 (Wheat CAP) and Hatch Project 149-945. Funding for spring 2021 was provided by the Biology teaching assistantship for BIOMG 2801, supervised by Dr. Kristina Blake-Hodek.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANOVA, analysis of variance

ARF, auxin response factor

AsA, ascorbate

BLUP, best linear unbiased prediction

Chr, Chromosome (map type)

CO, crossover

CV, causal variant (relationship matrix)

dCas9, nuclease-deactivated Cas9

DEG, differentially expressed gene

DH, doubled haploid

dHJ, double Holliday junction

DMC1, disrupted meiotic cDNA1

DSB, double-stranded breaks

DV, deleterious variant (QTL)

GBS, genotyping by sequencing

GC, gene conversion

GFD, grain fill duration

GL, grain length

GO, gene ontology

GW, grain width

GW, genomewide (relationship matrix)

HC, high confidence (gene)

HD, heading date

HIF, heterogenous inbred family

HR, homologous recombination

HT, height

HWE, Hardy-Weinberg equilibrium

H3K9me2, histone 3 lysine 9 dimethylation

IWGSC, International Wheat Genome Sequencing Consortium

KASP, kompetitive allele specific PCR

LC, low confidence (gene)

LD, linkage disequilibrium

LOD, logarithm of odds

NCO, non-crossover

NHEJ, non-homologous end joining

NIL, near isogenic line

Peri, Pericentromere (map type)
PPKL, protein phosphatases with Kelch-like domains
QTL, quantitative trait loci
QTgw.cnl-5A+, Opata allele
QTgw.cnl-5A-, W7984 allele
R, random (QTL)
RIL, recombinant inbred line
RNAi, RNA interference
SDSA, synthesis dependent strand annealing
SPO11, SPORULATION-DEFICIENT11
SPS, spikelets per spike
SSR, simple sequence repeats
SynOpDH, W7984 x Opata M85 doubled-haploid reference population
TALEN, transcription activator-like effector nuclease
TGW, thousand grain weight
TP, training population
WT, wild type
ZFN, zinc-finger nuclease
5AS+, chromosome 5A short arm presence
5AS-, chromosome 5A short arm absence

CHAPTER 1

INTRODUCTION

Rationale and significance

Mutations and recombination between homologous chromosomes that occur during meiotic segregation are essential for generating heritable diversity and evolutionary change among sexually reproducing eukaryotes. One crossover (CO) for each homologous chromosome pair is required for proper segregation, and recombination rates above this minimum can vary among and within species (Henderson & Bomblies, 2021). However, the distribution of COs in many species, including plants, is strongly biased toward subtelomeric regions and away from the pericentromere. In addition to the rate of recombination, this distribution bias limits the genetic variation accessible to plant breeders and impacts the efficiency of response to artificial selection.

There are evolutionary advantages as well as costs associated with variation in recombination and mutation. The benefits of recombination are described by two models in population and evolutionary genetics: the Hill-Robertson effect and Muller's Ratchet (Hill & Robertson, 1966; Muller, 1964). In a finite population, recombination can break down the linkage between favorable and deleterious loci, aiding in the efficiency of selection (i.e., The Hill-Robertson effect). Without recombination the deleterious load of a population will steadily increase and can never be less than the lowest load in the original population (i.e., Muller's Ratchet). For example, in regions of low recombination like the pericentromere, deleterious loci are more frequent and are likely to become linked in repulsion with positive loci (Rodgers-Melnick et al., 2015; Jordan et al., 2018). Significantly increased recombination can come at a cost though, leading to decreased fitness by breaking up beneficial linkages and reduced fertility (Charlesworth & Barton, 1996; Mieulet et al., 2018).

Mutations provide another source of genetic diversity and can have a range of fitness effects from beneficial to lethal. The rate of new mutations in eukaryotes is estimated to be at least 1 x 10^{-8} / base pair / meiosis (i.e., progeny will have a handful of mutations not present in the parents), and it is predicted that new mutations in coding regions will be deleterious in some of the environments the species inhabits (Ohta, 1972, 1992; Baer et al., 2007). Artificial selection and population improvement (i.e., increased fitness) often result in inbreeding, which reduces the effective recombination rate and under the cost of domestication hypothesis increases the likelihood of deleterious variants hitchhiking via linkage disequilibrium (Moyers et al., 2018). Natural recombination frequency and distribution, as well as mutation rates, have traditionally required plant breeders to work with large populations over many generations to capture desirable haplotypes and select new cultivars, which may take 10 to 15 generations for annual row crops.

To sustainably support the world’s growing population and farm profitability under a changing climate, a primary focus of breeding in the 21st century has been improving the speed and precision of cultivar development. The affordability of high-throughput genotyping and the adoption of genomic prediction are expected to lead to greater genetic gains per year in plant breeding programs (Meuwissen et al., 2001;

Heffner et al., 2009; Sweeney et al., 2020). The increased availability of annotated crop reference genomes and genome editing such as the CRISPR/Cas system have introduced precision-based approaches to breeding, and renewed interest in causal variant identification (Atkins & Voytas, 2020; Khan et al., 2020). While these technological advances generally represent two schools of thought, respectively prediction and causal variant-based breeding approaches, their utility to breeders both depend on genetic resolution which is inextricably linked to the frequency and distribution of recombination.

The complementation of genomic to phenotypic data that allows breeders to make predictive selections based on genetic markers associated with a phenotype of interest has been referred to as Breeding 3.0 (Wallace et al., 2018). Markers in regions of chromosomes where historical recombination has created sufficient genetic resolution to identify quantitative trait loci (QTL) are more likely to be subject to selection. However, the selected markers are not expected to be causal and rather are in linkage with the causative variant. Breeders can make successful selections based on linkage, but in broad genomic regions (e.g., spanning the pericentromere) these selections can result in false positives, are insensitive to repulsion linkages with deleterious loci, and fail to detect small or large mutations (e.g., structural variation deletions, duplications, insertions, inversions, and translocations). We are now entering Breeding 4.0, where the complementation of genome resequencing and genome editing can unveil a more precise resolution of genomic variation, and facilitate identifying, controlling, and ultimately repairing deleterious loci (Wallace et al., 2018).

This dissertation explores challenges and opportunities for the future of breeding imposed by structural variation and recombination, using common wheat (*Triticum aestivum* L.) as a model. Wheat is an annual allopolyploid (AABBDD, 2n = 42) grass that was domesticated during the origins of agriculture approximately 10,000 years ago in the Fertile Crescent of western Asia (Preece et al., 2017). Today wheat remains a staple crop grown globally, delivering 20% of daily calories and protein to the human population (FAOSTAT, 2021). Increasing wheat productivity under fewer production hectares is a primary objective for breeders, however the large, redundant polyploid genome, the polygenic nature of grain yield, and natural recombination and mutation rates in wheat collectively pose unique challenges for trait improvement.

To actualize our ability to manipulate individual base pairs and transition wheat genetic improvement schemes from Breeding 3.0 to Breeding 4.0 will require strategies that are sensitive to novel genetic variation that has a high probability of favorably affecting a phenotype. Evaluations of the recombination landscape in diverse wheat populations reveal that over 75% of the recombination events fall within 10% of the distal ends of chromosomes, and that the mutation rate ranges from 0 to 4.97×10^{-3} / base pair / meiosis (Raquin et al., 2008; Jordan et al., 2018). In addition, a higher density of putative deleterious variants was detected in the pericentromere versus distal regions of chromosomes. Given the tight relationship between the frequency and distribution of recombination, mutations, and genetic resolution, and the relatively recent publication of an annotated wheat reference genome (2017), there are many questions that remain to be answered about the genetic diversity that wheat harbors. The ensuing chapters seek to address two primary topics, respectively related to causal variant and prediction-based approaches to breeding – 1) how geneticists can improve the accuracy of identifying targets for positional cloning, and 2) whether breeders should seek to manipulate recombination to further harness the power of selection.

Objectives

This work encompasses in-vivo and in-silico approaches to identifying, controlling, and purging deleterious loci from wheat, with an emphasis on improving productivity, as well as overall recommendations for the near and distant future of polyploid breeding programs. My objectives were to:

1. Characterize a large deletion on wheat chromosome arm 5AS that was previously misidentified as a QTL on chromosome arm 5AL harboring a single gene for increased grain weight.

2. Identify methods for more efficient detection of structural variation and robust positional cloning experimental design in polyploids.

3. Review the potential for genome editing reagents to "control recombination" in plant breeding programs.

4. Simulate a genomic selection breeding program and identify the parameter space in which increased recombination maintains genetic diversity and increases genetic gain.

REFERENCES

Atkins, P.A., & Voytas, D.F. (2020). Overcoming bottlenecks in plant gene editing. *Current Opinion in Plant Biology*, *54*, 79–84. https://doi.org/10.1016/J.PBI.2020.01.002

Baer, C.F., Miyamoto, M.M., & Denver, D.R. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics 2007 8:8*, *8*, 619–631. https://doi.org/10.1038/NRG2158

Charlesworth, B., & Barton, N.H. (1996). Recombination load associated with selection for increased recombination. *Genetical research*, *67*, 27–41. https://doi.org/10.1017/S0016672300033450

FAOSTAT. (2021). *FAOSTAT*. http://www.fao.org/faostat/en/#data (accessed January 17, 2020).

Felsenstein, J. (1974). The evolutionary advantage of recombination'. *Genetics* (pp. 737–756).

Heffner, E.L., Sorrells, M.E., & Jannink, J.-L.L. (2009). Genomic Selection for Crop Improvement. *Crop Science*, . https://doi.org/10.2135/cropsci2008.08.0512

Henderson, I.R., & Bomblies, K. (2021). Evolution and Plasticity of Genome-Wide Meiotic Recombination Rates. https://doi.org/10.1146/annurev-genet-021721

Hill, W.G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research*, *8*, 269–294. https://doi.org/10.1017/S0016672300010156

Jordan, K.W., Wang, S., He, F., Chao, S., Lun, Y., Paux, E., Sourdille, P., Sherman, J., Akhunova, A., Blake, N.K., Pumphrey, M.O., Glover, K., Dubcovsky, J., Talbert, L., & Akhunov, E.D. (2018). The genetic architecture of genome-wide recombination rate variation in allopolyploid wheat revealed by nested association mapping. *The Plant Journal*, *95*, 1039–1054. https://doi.org/10.1111/tpj.14009

Khan, A.W., Garg, V., Roorkiwal, M., Golicz, A.A., Edwards, D., & Varshney, R.K. (2020). Super-Pangenome by Integrating the Wild Side of a Species for Accelerated Crop Improvement. *Trends in Plant Science*, *25*, 148–158. https://doi.org/10.1016/J.TPLANTS.2019.10.012

Meuwissen, T.H.E., Hayes, B.J., & Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, . https://doi.org/11290733

Mieulet, D., Aubert, G., Bres, C., Klein, A., Droc, G., Vieille, E., Rond-Coissieux, C., Sanchez, M., Dalmais, M., Mauxion, J.-P., Rothan, C., Guiderdoni, E., & Mercier, R. (2018). Unleashing meiotic crossovers in crops. *Nature Plants*, *4*, 1010–1016. https://doi.org/10.1038/s41477-018-0311-x

Moyers, B.T., Morrell, P.L., & McKay, J.K. (2018). Genetic costs of domestication and improvement. *Journal of Heredity*, *109*, 103–116. https://doi.org/10.1093/jhered/esx069

Muller, H.J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *1*, 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Ohta, T. (1972). Population size and rate of evolution. *Journal of Molecular Evolution 1972 1:4*, *1*, 305–314. https://doi.org/10.1007/BF01653959

Ohta, T. (1992). The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics*, *23*, 263–286. https://doi.org/10.1146/ANNUREV.ES.23.110192.001403

Preece, C., Livarda, A., Christin, P.A., Wallace, M., Martin, G., Charles, M., Jones, G., Rees, M., & Osborne, C.P. (2017). How did the domestication of Fertile Crescent grain crops increase their yields?. *Functional Ecology*, *31*, 387. https://doi.org/10.1111/1365-2435.12760

Raquin, A.L., Depaulis, F., Lambert, A., Galic, N., Brabant, P., & Goldringer, I. (2008). Experimental Estimation of Mutation Rates in a Wheat Population With a Gene Genealogy Approach. *Genetics*, *179*, 2195–2211. https://doi.org/10.1534/GENETICS.107.071332

Rodgers-Melnick, E., Elshire, R.J., Li, Y., Bradbury, P.J., Mitchell, S.E., Li, C., Glaubitz, J.C., Buckler, E.S., & Acharya, C.B. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences*, *112*, 3823–3828. https://doi.org/10.1073/pnas.1413864112

Sweeney, D.W., Rutkoski, J., Bergstrom, G.C., & Sorrells, M.E. (2020). A connected half-sib family training population for genomic prediction in barley. *Crop Science*, *60*, 262–281. https://doi.org/10.1002/CSC2.20104

Wallace, J.G., Rodgers-Melnick, E., & Buckler, E.S. (2018). On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. *https://doi-org.proxy.library.cornell.edu/10.1146/annurev-genet-120116-024846*, *52*, 421–444. https://doi.org/10.1146/ANNUREV-GENET-120116-024846

CHAPTER 2

POSITIONAL-BASED CLONING ‘FAIL-SAFE’ APPROACH IS OVERPOWERED BY WHEAT CHROMOSOME STRUCTURAL VARIATION [1]

Abstract

Positional based cloning is a foundational method for understanding the genes and gene networks that control valuable agronomic traits, such as grain yield components. In this study, we sought to positionally clone the causal genetic variant of a thousand grain weight (TGW) quantitative trait loci (QTL) on wheat chromosome arm 5AL. We developed heterogenous inbred families (HIFs) (> 5,000 plants) for enhanced genotypic resolution and fine-mapped the QTL to a ten Mbp region. The transcriptome of developing grains from positive and negative control HIF haplotypes revealed presence/absence chromosome arm 5AS structural variation, and unexpectedly no differential expression of genes within the chromosome arm 5AL candidate region. Evaluation of genomic, transcriptomic, and phenotypic data, and predicted function of genes, identified that the 5AL QTL was in fact the result of strong linkage disequilibrium with chromosome arm 5AS presence/absence (HIF $r^2 = 0.91$). Structural variation is common in wheat, and our results highlight that the redundant polyploid genome’s masking of such variation is a significant barrier to positional cloning. We propose recommendations for more efficient and robust detection of structural variation, including transitioning from a SNP to a haplotype-based approach to identify positional cloning targets. We also present nine candidate genes for grain yield components based on chromosome arm 5AS presence/absence, which may unveil hidden variation of homoeolog dosage dependent genes across the group five chromosome short arms. Taken together, our discovery demonstrates the phenotypic resiliency of polyploid genomic structural variation and highlights a considerable challenge to routine positional cloning in wheat.

[1] Originally published as Taagen, E., Tanaka, J., Gul, A. & Sorrells, M. E. (2021). Positional-based cloning 'fail-safe' approach is overpowered by wheat chromosome structural variation. *Plant Genome.* 2021;14:e20106. https://doi.org/10.1002/tpg2.20106

Ella Taagen contributions: Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing-original draft; Writing-review & editing.

Introduction

Common wheat (*Triticum aestivum* L.) and durum wheat [*Triticum turgidum* L. subsp. *Durum* (Desf.) van Slageren] deliver more than 20% of the daily calories and protein consumed by the human population (FAOSTAT, 2021). To sustainably support the world's growing population and farm profitability, wheat productivity must increase under fewer production hectares. Despite wheat's crucial bearing on global food security, the genes and gene networks controlling wheat grain yield remain poorly understood. Grain yield is a highly polygenic trait that is influenced by genetic and environmental factors at every stage of plant growth (Slafer, 2003). An additional challenge to pinpointing the genetic control of grain yield is posed by the large polypoid wheat genome, where the phenotypic consequence of a single variable locus can be masked by the redundant copies of other homoeologous genomes (Borrill et al., 2018). Given the complexity of identifying genetic controls of grain yield, a reductionist approach that considers highly heritable yield components is a valuable strategy to improve our understanding of underlying genes and gene networks (Brinton and Uauy, 2018; Zhang et al., 2018a).

Total grain yield is a balancing act between yield components such as spikes per unit of area, grain number per spike, and grain weight. QTL for yield components have been identified on every wheat chromosome, but many of these QTL span broad genomic regions and offer limited impact for breeding. Some of the latest advances in wheat genomics now allow us to go beyond QTL mapping and invest in positional cloning, such as the advent of an annotated reference genome and gene editing techniques (Borrill et al., 2018). Positional based cloning identifies a gene through fine-mapping, sequencing, and functional validation. Fine-mapping resolution is limited by the low frequency and uneven distribution of crossovers between linked genetic markers. A “fail-safe” approach is to develop heterogenous inbred families (HIFs) or near isogenic lines (NILs), which share a highly inbred and homogenous genome but are segregating for the genomic region of interest, such as a QTL (Tuinstra et al., 1997; Brinton et al., 2017; Kuzay et al., 2019). Continuous inbreeding of several thousand HIF progeny will produce distinct crossovers that delimit the QTL to a gene variant that can be confidently associated with the plant’s phenotype, for example using TILLING populations or gene editing.

We set out to positionally clone the causal variant underlying a thousand grain weight (TGW) and grain morphology QTL on chromosome arm 5AL, *QTgw.cnl-5A*, as previously identified in the Synthetic W7984 x Opata M85 spring wheat doubled-haploid reference population (herein abbreviated ‘W7984’, ‘Opata’, and ‘SynOpDH’) (Breseghello and Sorrells, 2006; Sorrells et al., 2011; Williams et al., 2014). Strong associations between grain weight or morphology and markers in this region have also been reported in (Kato et al., 2000; Brinton et al., 2017; Sukumaran et al., 2018). In this study we leveraged the SynOp recombinant inbred line (‘SynOpRIL’) population for HIF development by identifying two founder F_6 lines segregating for the *QTgw.cnl-5A* flanking markers (*5A_283300187* and *5A_482369161*). We screened more than 5,000 progenies over five generations (2017 – 2019 field and greenhouses) and narrowed *QTgw.cnl-5A* to a ten Mbp region flanking 37 high confidence (HC) genes. This fine-mapping advancement coincided with the publication from

(Gutierrez-Gonzalez et al., 2019) that identified the short arm of chromosome 5A was missing from the W7984 parent and prompted us to evaluate the implications for our positional cloning.

The W7984 chromosome arm deletion may not have been detected by the larger wheat community, or earlier in our research, in part due to no aberrant morphological variation or infertility, which is characteristic of whole chromosome arm deletions (Zhang et al., 2019, 2020). Here, we used a subset of both SynOpDH and SynOpRIL populations to (i) significantly and consistently map *QTgw.cnl-5A* across four environments; (ii) develop HIFs for fine-mapping *QTgw.cnl-5A* and disrupt linkage disequilibrium (LD) with chromosome arm 5AS presence (+) and absence (-); (iii) conduct grain growth rate analysis for greater phenotypic resolution; (iv) measure the transcriptome of chromosome arm 5AS+ / *QTgw.cnl-5A*+ and chromosome arm 5AS- / *QTgw.cnl-5A*- HIFs; and (v) conduct gene ontology (GO) term enrichment analysis. Based on genetic data, phenotypic associations, early grain development expression profiles, and predicted function of genes, we suggest that *QTgw.cnl-5A* is the result of strong LD with chromosome arm 5AS presence and absence (SynOpDH $r^2 = 0.95$, HIF $r^2 = 0.91$). The resources invested for positional cloning *QTgw.cnl-5A* were overpowered by chromosomal structural variation, and we discuss the challenges that persist for identifying gene function in wheat.

We also present nine candidate genes on chromosome arm 5AS that may impact yield components including TGW, grain length (GL), grain width (GW), and spikelets per spike (SPS). These results lay the foundation for identifying hidden variation of homoeolog dosage dependent and functionally redundant genes on the group five chromosome short arms. Altogether, our findings highlight the phenotypic resiliency of polyploid genomic structural variation and present recommendations for future approaches to positional cloning.

Materials and Methods

QTL validation plant materials

A synthetic hexaploid wheat, generated by crossing the durum wheat 'Altar 84' (AABB) with an *Ae. tauschii* (DD) accession, crossed with the spring wheat cultivar 'Opata 85', were used to generate 215 DHs (SynOpDH) via chromosome doubling, and 2,039 RILs (SynOpRIL) (Sorrells et al., 2011). Both populations segregate for the absence of W7984 chromosome arm 5AS, or presence of Opata, but the structural variation remains to be characterized in all 2,039 SynOpRILs (Supplementary file_S1.5.csv). A subset of each mapping population was used for this study. 149 entries from SynOpDH, along with parental checks and two commercial checks Glenn and Tom, were grown in two replicated and randomized one-meter, single-row plots in Ithaca, NY during the field seasons (April – August) of 2016 (Caldwell field), 2017 (Caldwell field), and 2018 (Caldwell and Helfer field). An unbalanced set of thirteen additional entry observations were included during BLUP phenotype calculations, for 162 total entries. All field trials were non-irrigated. All of the spikes in each one-meter row were hand harvested and threshed with a belt thresher (Almaco, Nevada, IA). Heading date (HD), TGW, GL, and GW were measured and used to validate the QTL across years and environments, as well as identify the flanking marker positions of the QTL based on Chinese Spring genome assembly released by the International Wheat Genome Sequencing Consortium (IWGSC), henceforth RefSeq v1.0 (International Wheat Genome Sequencing Consortium (IWGSC), 2018). Gene annotations presented in this study are based on RefSeq v1.1 annotation. While RefSeq v2.0 assembly was available at the time of analysis, the annotation was still under development.

Development of HIF-derived fine-mapping population

The fine-mapping population was constructed using HIFs derived from two F_6 SynOpRIL founder entries (7-956 and 7-1201) heterozygous for markers flanking the QTL on chromosome arm 5AL (*5A_283300187* and *5A_482369161; marker origin is RefSeq v1.0 and the syntax is chromosome_RefSeq v1.0 position*). In order to increase the genetic resolution of the QTL and maintain an isogenic background genome, individual progeny from these two entries were inbred X generations ($F_{6:X}$) and genotyped to screen for recombinants between the QTL flanking markers. Heterozygous entry advancement, recombinant evaluation, and progeny testing took place 2016-2019 until the $F_{6:4}$ or $F_{6:5}$ generation. Inbreeding cycled between Snyder or Caldwell fields in Ithaca, NY and Cornell University Guterman greenhouse. Under field evaluation, 100 progenies of each heterozygous entry were advanced, and 20 progenies from a recombinant entry were advanced for validation. Greenhouse evaluation space was limited, and this environment was only used for recombinant validation testing between field cycles. The greenhouse environment was supplemented with artificial lighting to obtain a sixteen-hour day / eight-hour night photoperiod, with 21-23 °C day and 15-17 °C night temperatures. Individual plant identity was tracked throughout HIF development and all of the spikes of each plant were hand harvested into a coin envelope and belt threshed. For example, HIF entry 7-956-2-89-2-17-01 started with F_6 founder entry 7-956 and out of the 100 progenies planted, the 2nd plant was heterozygous for the QTL flanking markers and was selected from $F_{6:1}$ planting, the 89th plant was selected from $F_{6:2}$ planting, the 2nd plant was selected from $F_{6:3}$ planting, at $F_{6:4}$ the 17th plant recombined between the QTL flanking markers and was validated at the $F_{6:5}$ generation. Homozygous recombinant and sister entries, Opata (+) and W7984 (-) allele controls without a crossover between QTL flanking markers, were selected and evaluated for pre- and post-harvest phenotypes in field experiments.

In total more than 5,000 progenies were screened for recombinants between the QTL flanking markers and 109 ($F_{6:4}$ or $F_{6:5}$) recombinant haplotypes were selected for the fine-mapping population. In addition, nine sister entries with Opata alleles and eleven with W7984 alleles spanning *QTgw.cnl-5A* (+ / - controls) were selected. The 129 fine-mapping population entries were genotyped with 31 kompetitive allele specific PCR (KASP) markers spanning *QTgw.cnl-5A*. 129 entries, along with parental checks and two commercial checks Glenn and Tom, were grown in two replicated and randomized one-meter, single-row plots in Ithaca, NY during the field seasons of 2019 (Snyder field) and 2020 (Caldwell and Helfer field). All of the spikes in each one-meter row were hand harvested and threshed with a belt thresher. HD, HT, grain fill duration (GFD), SPS, TGW, GL, and GW were measured and used for t-test comparisons between recombinant haplotypes and control haplotypes to narrow the QTL flanking markers (Supplementary file_S2.7.csv).

Phenotypic data and statistical analysis

The pre-harvest phenotypes measured in this study include HD, HT, DPA grain morphology, and GFD; post-harvest include SPS, TGW, GL, and GW. The HD was recorded as Julian date of 50% spike emergence. Plant HT was measured as the average height (cM) from the ground to spike tip in a one-meter, single-row plot. DPA was recorded by tagging individual spikes at 0-DPA, when anthers at the center spikelet are light green and pollination can be confirmed within 24-hours of tagging. GFD was recorded as the time between 0-DPA and physiological maturity (when the peduncle turned yellow) for a one-meter, single-row plot.

The spikes from every plot in this study were hand harvested with sickles or scissors. SPS was recorded as the average spikelet number from ten random spikes in a plot. GL and GW were measured using > 150 grains per sample on a WinSEEDLE STD4800 system flatbed scanner. The number of grains was recorded, and then weighed as a proxy for TGW. This process was replicated without resampling for a given plot, and the average GL, GW, and TGW was recorded. Phenotypes for the SynOpDH across four field-year environments and the fine-mapping population across three field-year environments can be found in Supplementary file_S1.1.csv and file_S2.1.csv, respectively.

In order to evaluate *QTgw.cnl-5A+/-* haplotypes and phenotypes across field and year combinations, univariate linear models with random genotype and environment effects, and correlated information (HD, chromosome arm 5AS+/-, or HIF RIL founder) fixed effects were fitted with the *R/lme4* package (Supplementary file_S1.8.xlsx / script_S1.md and file_S2.9.csv / script_S2.md) (Bates et al., 2015; R

Core Team, 2020). Models were evaluated based on their broad-sense heritability and/or Akaike information criterion. Fixed effect significance was assessed with *R/car* package Wald chi-square test (Fox and Sanford, 2019). Broad-sense heritability estimates (H^2) for HD, TGW, GL and GW in the SynOpDH population were calculated by

$$H^2 = \sigma_G^2 \Big/ \sigma_G^2 + \frac{\sigma_{GxE}^2}{l} + \frac{\sigma_E^2}{rl}$$

where σ_G^2 is the genetic variance, σ_{GxE}^2 is the GxE variance, σ_E^2 is the residual variance, *r* is the number of replications and *l* is the number of environments. Best linear unbiased predictions (BLUPs) were obtained from the univariate linear models for SynOpDH and HIF phenotypes. The SynOpDH BLUP phenotypes were used for QTL mapping with the *R/qtl* package (Broman 2003). The HIF BLUP phenotypes were used for Welch two-sample t-tests with control and recombinant haplotypes, and chromosome arm 5AS+/- haplotypes (Supplementary file_S2.7.csv, script_S2.md) (R Core Team, 2020).

For the grain growth rate analysis, grain length, width, fresh weight and dry weight were measured at 0, 4, 10, 16, and 22-DPA in a replicated subset of ten fine-mapping haplotypes (five *QTgw.cnl-5A-* and five *QTgw.cnl-5A+*, Supplementary file_S3.1.csv) during the 2019 field season. The experiment was replicated in the Cornell University Guterman greenhouse during spring 2020 for four HIF haplotypes (same entries as RNA-seq experiment), and also included measurements at 28-DPA and senescence (Supplementary file_S3.2.csv). For every haplotype and replicate, ten primary spikes were used per timepoint. From each of the ten spikes, ten primary developing grains in florets 1 and 2 were taken only from the central spikelets and the average length, width, fresh weight, and dry weight (5-days in a 30 °C dryer) were measured. Single trait mixed models with a fixed interaction between haplotype and DPA, and a random effect of plot (field) or tube (greenhouse) nested within haplotype were fitted with the *R/lme4* package (Bates et al., 2015). Post hoc comparison of least-squares means for haplotype and DPA was performed within each model using the *R/emmeans* package, in addition to multiple-test correction *P*-values (Supplementary script_S3.md) (Lenth et al., 2019).

Genetic Map Construction and QTL Mapping

A genetic map for the 162 entry SynOpDH subset was constructed with 1,551 polymorphic Genotyping by Sequencing (GBS) and simple sequence repeats (SSR) markers that were previously published, and using the R package 'qtl' (Broman et al., 2003; Sorrells et al., 2011; Poland et al., 2012). The function 'estmap' was used to estimate the genetic distances using the 'kosambi' mapping function, followed by maximum likelihood analysis of marker order on each chromosome, using the function 'ripple'. Any missing genotypes were imputed with the function 'fill.geno' and 'imp' method. QTL were identified by a single QTL model genome scan using the function 'scanone' with the Haley-Knott regression method. A 0.05 significance LOD threshold for each phenotype was determined with the function 'scanone' and 'n.perm = 1000'. Next, the percent variance explained by each significant QTL was calculated. The RefSeq v1.0 physical position of flanking markers was determined for each significant QTL as well. Later a consensus marker for chromosome arm 5AS presence and absence was added for QTL mapping. The code for the genetic map and QTL mapping can be found in Supplementary script_S1.md.

High Resolution QTgw.cnl-5A Genetic Mapping

Seedling leaf tissue samples for DNA extraction were collected in the field or greenhouse from a single plant in each plot or pot, with two replicates of the parents per 96-well plate. The DNA was extracted from lyophilized leaf tissue using a modified cetyl trimethylammonium bromide extraction (Doyle and Doyle, 1990). KASP assays were developed to screen recombinant entries and their progeny, as well as identify sister entries. The KASP markers were generated from polymorphisms identified using the wheat exome-capture and regulatory-capture sequence of parental entries of the Wheat-CAP project (https://www.triticeaecap.org/wheatcap-germplasm-list/) (Gardiner et al., 2019; He et al., 2019a). We also used the 10 + Genome Project data repository to BLAST our KASP marker sequences and confirmed genome and chromosome specificity and the marker order across Chinese Spring and eleven diverse wheat cultivars; ‘ArinaLrFor’, ‘Jagger’, ‘Julius’, ‘Lancer’, ‘Landmark’, ‘Mace’, ‘Norin61’, ‘Stanley’, ‘SY-Mattis’, ‘Zavitan’ and ‘Spelt’ (Supplementary file_S2.6.csv, script_S2.md) (Walkowiak et al., 2020). The KASP assay procedure followed the methods outlined in (Makhoul et al., 2020), including PACE-IR Genotyping Master Mix with a low ROX level and thermo-cycling conditions according to 3CR bioscience protocols. On each SNP reaction plate the parent lines, check lines, and at least one water sample were included as controls. All experiments were repeated at least twice. If clear genotyping clusters were not obtained, the KASP marker was abandoned. The clustering patterns of the three KASP markers that span *QTgw.cnl-5A* (ten Mbp region) are available as Supplementary file_2.13.pdf. *QTgw.cnl-5A* flanking KASP markers (*5A_283300187* and *5A_482369161*) were used to screen the HIF progeny, and all 31 KASP markers were used to select the fine-mapping population (129 entries). Later, the fine-mapping population was genotyped with three SSR markers for chromosome arm 5AS presence and absence (Supplementary file_S2.5.csv). The homogeneous genetic background of the HIFs and the high heritability of the traits allowed us to differentiate their association with *QTgw.cnl-5A* from the causal effects of chromosome arm 5AS structural variation.

Gene Expression

We used RNA-seq to compare the levels of expression of genes within the candidate *QTgw.cnl-5A* region, across chromosome arm 5AS, and to evaluate the HIF isogenic background genome. The plant tissue was sampled from developing grains at 4 and 8-DPA from four HIF haplotypes: 7-956-2-19-1-31-03 (Opata control), 7-956-2-19-1-44 (W7984 control), 7-956-2-19-1-31-05 (recombinant I), 7-956-2-12-1-69-07 (recombinant II). The four haplotypes were grown in a completely randomized design in the greenhouse during spring 2020. We sampled 500 mg of whole grain tissue per biological replicate. At the 4-DPA timepoint, 40 primary spikes were collected for each haplotype and ten developing primary grains from the central spikelets of ten randomly sampled spikes (four biological replicates, 100 hundred grains / biological replicate) were immediately frozen in liquid nitrogen and stored at -80 °C. The process was repeated at the 8-DPA timepoint, but only two randomly sampled spikes per biological replicate were necessary due to the rapidly growing grains (four biological replicates, 20 grains / biological replicate). The tissue was ground with liquid nitrogen and mortar and pestle and total RNA was extracted using a modified hot borate protocol (Wan and Wilkins, 1994). Quantity and quality of the isolated total RNA was determined using a Biotek Epoch 2 Spectrophotometer with Nanodrop functionality and gel electrophoresis. Three of the four biological replicates were sent to Novogene for further quality screening and non-directional 150-bp paired-end reads mRNA sequencing (four haplotypes, three biological replicates at two timepoints: 24 samples). The raw sequence reads were submitted to the SRA of NCBI under BioProject ID: PRJNA693003.

The bioinformatics pipeline and scripts for gene expression analysis can be found in Supplementary script_S4.md. Paired-end reads from raw sequence data were imported to the Cornell University BioHPC server and all computational analysis was performed within the command line or R. Imported reads quality was checked with 'FastQC' (Andrews, 2010). The IWGSC RefSeq v1.0 assembly and RefSeq v1.1 high and low confidence gene annotations were downloaded from https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v1.0/ and https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.1/. The genome was indexed, and the sequence reads were aligned with STAR for both low confidence (LC) and HC RefSeq v1.1 gene annotation (Dobin et al., 2013). Batch effects were accounted for using the 'ComBat_seq' function in the *R/sva* package, and differential expression between haplotypes was analyzed with the *R/DESeq2* package (Love et al., 2014; Leek et al., 2020). A false discovery rate cut-off value of 0.01 and log2 fold change threshold of two was used to select the list of differentially expressed genes at 4 and 8-DPA between the haplotype contrasts. Variance stabilizing transformation of the DESeq2 adjusted counts was used for principal component analysis and heatmaps of the count matrix. SNP variants in the *QTgw.cnl-5A* region were identified with the SAMtools BCFtools command ‘mpileup’ (Li et al., 2009).

Gene ontology (GO) term enrichment

At this point we excluded low confidence gene models from further analysis. The list of 556 HC genes differentially expressed between all three positive haplotypes (Opata control, recombinant I and recombinant II) and the negative haplotype (W7984 control) at 4 and 8-DPA can be found in Supplementary file_S4.20.csv. We obtained RefSeq v1.1 HC gene annotation GO terms from the Wheat@URGI JBrowse portal (Alaux et al., 2018). In total we extracted GO terms for 93,141 genes, which included all 556 differentially expressed genes (DEGs). The 93,141 genes served as the universe of genes for a GO term enrichment study with the R package ‘GOseq’ to test over-representation of GO terms among the DEGs (Young et al., 2010). We considered over-represented GO terms with Benjamini–Hochberg false discovery rate adjusted *P*-value < 0.05 to be significant.

We also identified DEGs with GO terms known to be associated with spike architecture and early grain development, including cell proliferation / division / growth / differentiation, meristem initiation / maintenance / transition, flowering time, floral organ development, mitosis, regulation of gene expression, DNA methylation and gene silencing, auxin, cytokinin, response to stress, phosphorylation, photosynthesis, nucleosome assembly, starch, nucleic acid metabolic process, protein metabolic process, glucose / sucrose, ubiquitin pathway, brassinosteroid, microtubule, and endoreduplication (Wang et al., 2017; Brinton and Uauy, 2018; Li et al., 2018). Arabidopsis and rice (*Oryza* sativa L.) orthologs were identified using *Ensembl*Plants (https://plants.ensembl.org/index.html), arabidopsis.org TAIR Gene Search, and funricegenes (Yao et al., 2018).

Results

A QTL on chromosome 5A is associated with increased grain weight

A genetic map was developed for the SynOpDH population comprising of 1,551 polymorphic molecular markers (Supplementary file_S1.9.csv, script_S1.md). Using the SynOpDH phenotype BLUP values, calculated from four field-year observations, significant QTL for TGW were identified on chromosomes 2D (*QTgw.cnl-2D*), 5A (*QTgw.cnl-5A*), and 6A (*QTgw.cnl-6A*). *QTgw.cnl-2D* and *QTgw.cnl-6A* colocalized with QTL for GL, and *QTgw.cnl-5A* colocalized with a QTL for GW. A QTL for HD on the long arm of chromosome 5A was also detected, spanning the *VRN-1* gene (Yan et al., 2003). The flanking and peak markers, logarithm of odds (LOD) score, positions in the genetic map and wheat reference genome, and grain weight percent variation explained by each QTL are described in Supplementary script_S1.md. A univariate linear model for TGW with random entry and environment effects, and fixed interaction effects between peak markers of the three TGW QTL showed highly significant effects for all three QTL, but no significant interactions (Supplementary script_S1.md). These results indicated that *QTgw.cnl-5A*, the focus of this study, could be mapped regardless of combinations with *QTgw.cnl-2D* and *QTgw.cnl-6A.*

The *QTgw.cnl-5A* flanking markers were *wmc705* (0.1 cM, LOD 3.86) and *synopGBS284* (4.49 cM, LOD 3.49), and accounted for 10.39% of the grain weight and 37.9% of the grain width variation. Upon inspection of the flanking marker RefSeq v1.0 physical positions (*wmc705,* 290 Mbp and *synopGBS284,* 487 Mbp), it became clear that *QTgw.cnl-5A* mapped to chromosome arm 5AL and the genetic map lacked markers on chromosome arm 5AS. Our QTL mapping occurred before (Gutierrez-Gonzalez et al., 2019) published dense GBS linkage maps that indicated that the SynOpDH population segregated for the presence or absence of the short arm on chromosome 5A. Later, we adjusted the marker order to reflect physical positions and incorporated a single marker for QTL mapping that represented the chromosome arm 5AS structural variation (Figure 2.1A).

Phenotypic distributions and ANOVA tests of the SynOpDH population revealed Opata provides the increased TGW and GW allele (Figure 2.1B, Table 2.1). We hypothesized that the increased grain weight is mediated by differences in grain width rather than length, which are under independent genetic control (Gegas et al., 2010). The broad sense heritability for TGW, GL, GW, and HD were 0.68, 0.75, 0.81, and 0.78 respectively. SynOpDH entries with the *QTgw.cnl-5A* Opata allele (+) were, on average, 6.4% heavier and 4.2% wider than the W7984 allele (-), and significantly different across all environments (Table 2.1).

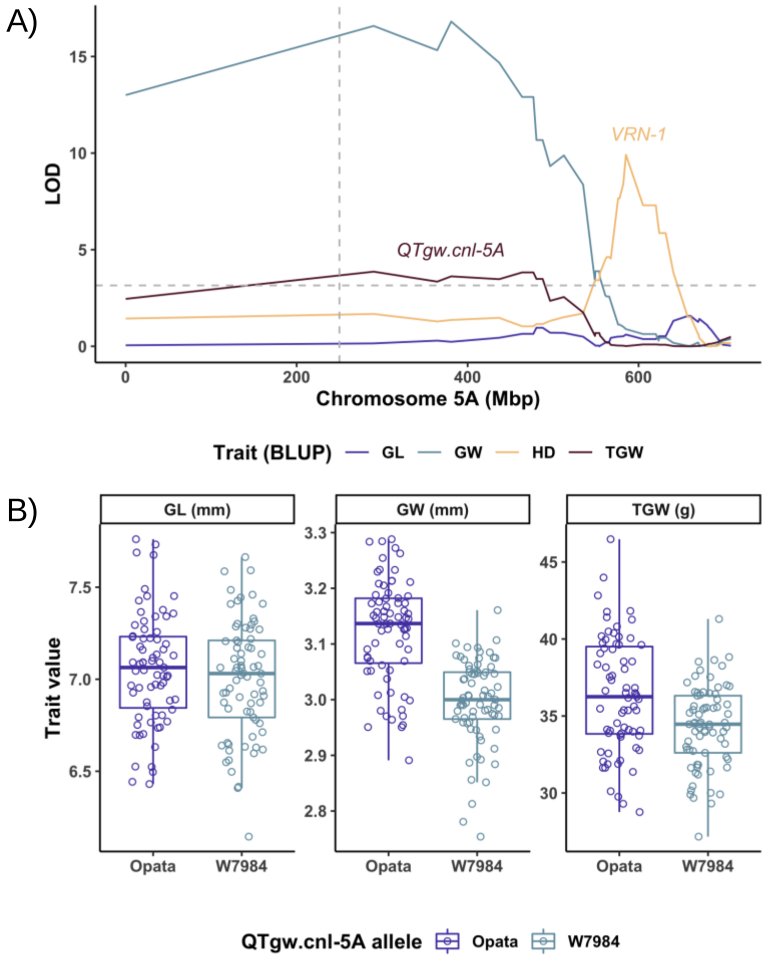

Figure 2.1: (A) QTL detected for TGW, GW, and HD in the SynOpDH population on chromosome 5A. A single marker was used to represent the chromosome arm 5AS structural variation. Values on the X-axis represent RefSeq v1.0 physical position, and the vertical dashed line represents the centromere. The dashed line on the Y-axis represents the LOD significance threshold. (B) BLUP phenotype distribution for SynOpDH entries subset by *QTgw.cnl-5A* allele, based on flanking marker genotypes.

Year	Location	Genotype	TGW (g)	GL (mm)	GW (mm)	HD (julian)
2016	Caldwell	*QTgw.cnl-5A+*	39.19	7.16	3.26	
		QTgw.cnl-5A-	36.49	7.09	3.08	
			7.4% ***	0.98%	5.7% ***	
2017	Caldwell	*QTgw.cnl-5A+*		6.58	2.84	
		QTgw.cnl-5A-		6.53	2.68	
				0.81%	6.1% ***	
2018	Caldwell (A)	*QTgw.cnl-5A+*	40.36	7.25	3.26	184.71
		QTgw.cnl-5A-	38.37	7.28	3.14	182.64
			5.2% **	-0.39%	3.7% ***	1.1% *
2018	Caldwell (B)	*QTgw.cnl-5A+*	39.97	7.24	3.26	185.2
		QTgw.cnl-5A-	37.8	7.25	3.13	182.89
			5.7% **	-0.21%	4.04% ***	1.3% **
2018	Helfer (A)	*QTgw.cnl-5A+*	31.04	7.08	3.08	187.05
		QTgw.cnl-5A-	28.55	6.97	2.92	184.8
			8.7% ***	1.70%	5.4% ***	1.2% *
2018	Helfer (B)	*QTgw.cnl-5A+*	31.61	7.03	3.09	186.91
		QTgw.cnl-5A-	29.77	6.97	2.96	184.79
			6.2% **	0.81%	4.5% ***	1.1% *
BLUP		*QTgw.cnl-5A+*	36.46	7.07	3.12	186.02
		QTgw.cnl-5A-	34.26	7.004	2.996	183.77
			6.4% ***	0.92%	4.2% ***	1.2% **
H^2			0.68	0.75	0.81	0.78

Table 2.1: Mean thousand grain weight (TGW), grain length (GL), grain width (GW), and heading date (HD) of SynOpDH entries. Percentages (%) refer to the trait value gained in SynOpDH *QTgw.cnl5A+* (Opata allele) entries as compared to *QTgw.cnl5A-* (W7984 allele) entries. The *QTgw.cnl5A+* and *QTgw.cnl5A-* allele categories were determined by genotype at the peak QTL marker, *5A_341510829*. Broad sense heritability (H^2) considered trait observations across all locations. Replicates grown in locations with spatial variation are reported independently and denoted with a letter in parentheses. Asterisks indicate significance determined by ANOVA for each location, or best linear unbiased prediction (BLUP). Key: no symbol, nonsignificant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

HIFs differing for QTgw.cnl-5A show a 21.3% difference in TGW

To further investigate the effect of *QTgw.cnl-5A* on TGW, HIF populations were generated from two F_6 SynOpRIL entries (7-956 and 7-1201) heterozygous for flanking KASP markers *5A_283300187* and *5A_482369161*. The fine-mapping population selection was based on recombination between flanking markers and sister line genotypes, as individual plant phenotypes for TGW were too variable. Our fine-mapping population development occurred before (Gutierrez-Gonzalez et al., 2019) was published.

In 2016 we grew a seed increase of the SynOpRIL population ($F_{6:1}$) and identified HIF founder lines 7-956 and 7-1201, heterozygous for *QTgw.cnl-5A* flanking markers. We screened 600 $F_{6:1}$ progeny ($F_{6:2}$) during the 2017 field season (Snyder field) with *QTgw.cnl-5A* flanking KASP markers and identified no crossovers within the maker interval and 45 plants that remained heterozygous. 900 progenies from the $F_{6:2}$ heterozygous plants ($F_{6:3}$) were evaluated in a greenhouse during the winter of 2017, and 44 plants with crossovers within the flanking KASP interval, seven W7984 sister lines (recombination outside of flanking KASP interval), and six Opata sister lines were identified. Five additional KASP markers were developed across the target region based on SNPs identified in parental exome-capture data (He et al., 2019a). The $F_{6:4}$ generation, validated during the 2018 field season (Caldwell field), consisted of inbreeding 1,620 entries with a single heterozygous flanking marker and progeny tests of the 44 recombinants. We identified 65 new homozygous recombinant plants within the flanking KASP interval, four W7984 sister lines, and three Opata sister lines. An additional 24 KASP markers were developed between the flanking markers based on SNPs identified in the exome-capture and regulatory-capture data (Gardiner et al., 2019). Progeny tests of the 65 recombinant plants ($F_{6:5}$) as well as replicates of the 44 $F_{6:4}$ validated recombinants and 20 sister lines were planted in a greenhouse during winter 2018 and characterized with the full set of 31 KASP markers. We screened an additional 1,514 segregating $F_{6:6}$ plants during the 2019 field season (Snyder) but did not find a crossover event that reduced the target region. The 109 recombinants and 20 sister lines were selected for the fine-mapping population and field based phenotypic evaluation.

The fine-mapping population was grown in 2019 (Snyder), and 2020 (Caldwell and Helfer) and evaluated for HD, HT, GFD, SPS, TGW, GL, and GW (Table 2.2). Significant differences in TGW and GW associated with the peak marker *5A_341510829* narrowed the *QTgw.cnl-5A* candidate region to a ten Mbp interval containing 37 HC and 84 LC genes, flanked by markers *5A_339757917* and *5A_349628635* (Figure 2.2, Supplementary script_S2.md). The Opata allele frequency at SNP *5A_341510829* was 0.58, and W7984 allele frequency was 0.42. The lower variability of the HIFs and homogenous background genome increased the resolution of additional yield component quantitative traits, including significant associations for GL, SPS, and HT with *QTgw.cnl-5A* (Table 2.2). There was no significant difference in HD or GFD, suggesting the difference in grain weight may be due to the grain filling rate rather than duration. Given the additional phenotypes associated with the candidate gene region, we decided to explore the transcriptome of *QTgw.cnl-5A*+ and *QTgw.cnl-5A*- haplotypes.

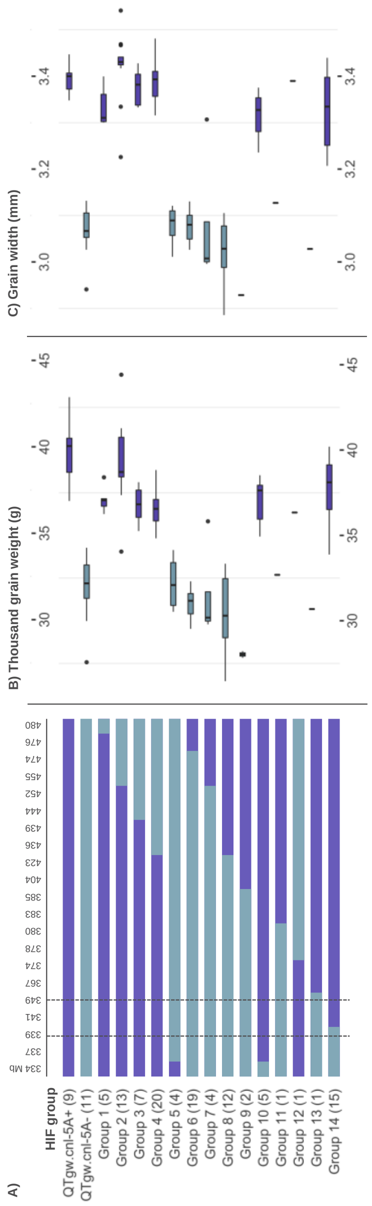

Figure 2.2: Fine-mapping of HIF recombinant entries and trait variation pinpoints *QTgw.cnl-5A* to a ten Mbp interval on wheat chromosome arm 5AL. (A) Graphical genotypes of 129 HIFs grouped by shared recombination intervals (KASP marker positions not to scale, purple Opata SNPs, blue W7984 SNPs). Number of HIFs in each group noted within parentheses. (B/C) BLUP phenotype boxplot distribution of each HIF group. Boxes are colored based on Opata- or W7984- like phenotype, determined by t-test (Supplementary script_S2.md).

Year	Location	Genotype	HD	GFD	HT	SPS	TGW	GL	GW
			(julian)	(days)	(cM)		(g)	(mm)	(mm)
2019	Snyder (A)	QTgw.cnl-5A+	172.19	30.9	77.55	15.03	37.49	6.54	3.24
		QTgw.cnl-5A-	172.57	30.98	74.96	14.76	31.22	6.4	2.93
			-0.22%	-0.26%	0.35%	1.80%	20.08% ***	2.2% ***	10.4% ***
2019	Snyder (B)	QTgw.cnl-5A+	171.59	30.4	81.99	15.11	38.69	6.58	3.28
		QTgw.cnl-5A-	172	30.82	78.67	14.57	31.87	6.41	2.95
			-0.24%	-1.40%	4.20%	3.7% **	21.4% ***	2.7% ***	11.1% ***
2019	Snyder (C)	QTgw.cnl-5A+	173.56		83.81	15.59	38.33	6.45	3.24
		QTgw.cnl-5A-	171.67		79.83	14.9	32.58	6.53	2.95
			1.10%		4.90%	4.60%	17.6% ***	-1.20%	9.9% ***
2020	Caldwell	QTgw.cnl-5A+	174.21	29.36	66.24	12.71	38.21	6.92	3.53
		QTgw.cnl-5A-	173.81	29.46	62.94	12.43	31.41	6.81	3.21
			0.23%	-0.35%	5.2% *	2.20%	21.7% ***	1.6% **	10.2% ***
2020	Helfer	QTgw.cnl-5A+	172.13	29.23	58.73	11.71	36.58	6.8	3.45
		QTgw.cnl-5A-	171.93	28.91	55.59	11.43	29	6.67	3.11
			0.12%	1.10%	5.7% **	2.40%	26.1% ***	1.9% ***	11.02% ***
BLUP		QTgw.cnl-5A+	172.57	29.32	71.2	13.62	37.72	6.71	3.37
		QTgw.cnl-5A-	172.63	29.39	68.45	13.43	31.09	6.58	3.06
			-0.04%	-0.26%	4.02% **	1.4% ***	21.3% ***	1.99% ***	10.3% ***

Table 2.2: Mean heading date (HD), grain fill duration (GFD), plant height (HT), spikelets per spike (SPS), thousand grain weight (TGW), grain length (GL), and grain width (GW) of SynOp HIF entries. Percentages (%) refer to the trait value gained in SynOp HIF *QTgw.cnl5A+* (Opata allele) entries as compared to *QTgw.cnl5A-* (W7984 allele) entries. The *QTgw.cnl5A+* and *QTgw.cnl5A-* allele categories were determined by genotype at the peak QTL marker, *5A_341510829*. Replicates grown in locations with spatial variation are reported independently denoted with a letter in parentheses. Asterisks indicate significance determined by ANOVA for each observation or best linear unbiased prediction (BLUP). Key: no symbol, nonsignificant; *, P < 0.05; **, P < 0.01; ***, P < 0.001.

HIF variation in grain weight and morphology was significantly associated with early grain development

In order to better understand the mechanism driving the grain weight and morphology phenotype, and to identify the optimum timepoints for our transcriptome study, we conducted a DPA analysis of the developing grains in five *QTgw.cnl-5A*+ and five *QTgw.cnl-5A*- HIF haplotypes. Grains of ten replicated haplotypes were sampled from the 2019 field season at 0, 4, 10, 16, and 22-DPA. The first significant difference in grain length, grain width and fresh weight was measured at 10-DPA, with *QTgw.cnl-5A*+ grains 4.6% longer ($P < 0.001$), 5.7% wider ($P < 0.001$), and 18.9% ($P < 0.001$) heavier than *QTgw.cnl-5A*- grains (Figure 2.3). The first significant difference in dry weight was measured at 16-DPA, with *QTgw.cnl-5A*+ grains 25.2% heavier ($P < 0.001$) than *QTgw.cnl-5A*- grains (Figure 2.3). These effects increased and were maintained at 22-DPA, and senescence (data not shown).

The experimental design was repeated in a greenhouse environment and included additional measurements at 28-DPA and senescence with four HIF entries that would be selected for RNA-seq (Opata control, W7984 control, recombinant I, recombinant II). The trend remained that the first significant difference in grain width and fresh weight was measured at 10-DPA and dry weight at 16-DPA, for Opata control versus W7984 control grains. However, for comparisons between recombinant I and W7984 control, and recombinant II and W7984 control, the difference for any phenotype was first measurable at 16-DPA (Supplementary script_S3.md). Differences in grain development between the field and greenhouse could be due to the growing environment, time of year, HIF haplotype, or too few experimental entry comparisons.

Grain development begins after pollination, starting with rapid proliferation of cells that form the outer layer of the grain, followed by the endosperm's cell division and expansion at approximately 6 DPA (Li and Li, 2016; Brinton and Uauy, 2018). Our DPA study had consistent association with phenotypic variation at 10 to 16-DPA, and in conjunction with the 2019 and 2020 field data that found no significant difference in grain fill duration among *QTgw.cnl-5A+* and *QTgw.cnl-5A-* HIFs, suggests the difference in grain width and weight is driven by the rapid endosperm cell division and expansion during early grain development.

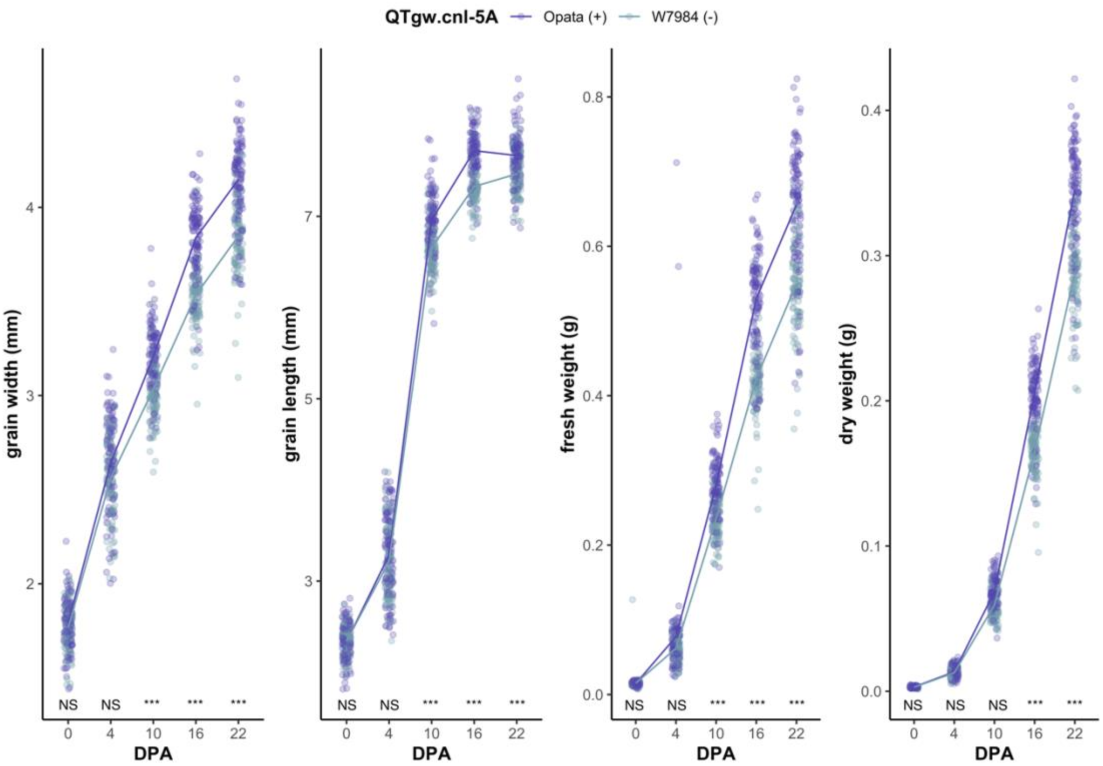

Figure 2.3: Developing grain phenotypes measured across 22 days post anthesis (DPA) for five *QTgw.cnl-5A+* (purple) and five *QTgw.cnl-5A-* (blue) HIFs. Samples taken at 0, 4, 10, 16, and 22-DPA during 2019 field season. Key: NS, nonsignificant; ***, P < 0.001.

QTgw.cnl-5A in linkage disequilibrium with chromosome arm 5AS structural variation

Upon learning from (Gutierrez-Gonzalez et al., 2019) that the SynOpDH population was segregating for the presence or absence of the short arm on chromosome 5A, we genotyped our fine-mapping population with three SSR markers for chromosome arm 5AS presence and absence. The frequency of chromosome arm 5AS presence was 0.57, and absence was 0.43. Of the 129 fine-mapping entries, 126 had either haplotype chromosome short arm 5A+ / *QTgw.cnl-5A+* or chromosome short arm 5A- / *QTgw.cnl-5A-*, with only one entry chromosome short arm 5A+ / *QTgw.cnl-5A-* and two entries chromosome short arm 5A- / *QTgw.cnl-5A+*. The skewed haplotype distributions revealed chromosome arm 5AS structural variation and *QTgw.cnl-5A* alleles are in very strong LD, with a correlation coefficient of 0.91. We tested for an interaction between the two loci using a univariate linear model for TGW with entry as a random effect and a fixed interaction between chromosome arm 5AS and peak marker *5A_341510829*. A Wald chi-square test showed highly significant effects for chromosome arm 5AS and *5A_341510829*, but no significant interaction. Phenotypic distributions of the fine-mapping population indicated that trait values of the three recombinant HIFs that broke the LD (chromosome short arm 5A+ / *QTgw.cnl-5A-*, n = 1, and chromosome short arm 5A- / *QTgw.cnl-5A+*, n = 2) are associated with the chromosome arm 5AS structural variant, rather than *QTgw.cnl-5A* allele (Figure 2.4A). In the context of Figure 2.2, the entry with chromosome short arm 5A+ / *QTgw.cnl-5A-* recombination belongs to group 7 and the two entries with chromosome short arm 5A- / *QTgw.cnl-5A+* recombination belong to group 14 (Supplementary file_S2.12.xlsx). Finally, we included the chromosome arm 5AS structural variant as a fixed effect in the univariate models for calculating HIF phenotype BLUPs. The new BLUP phenotypic distributions for HIF entries showed no significant difference for presence and absence of chromosome arm 5AS or the *5A_341510829* allele, suggesting the phenotypic variation was largely explained by the chromosome arm 5AS structural variation, and *QTgw.cnl-5A* is a result of linkage (Figure 2.4B & C, Supplementary script_S2.md).

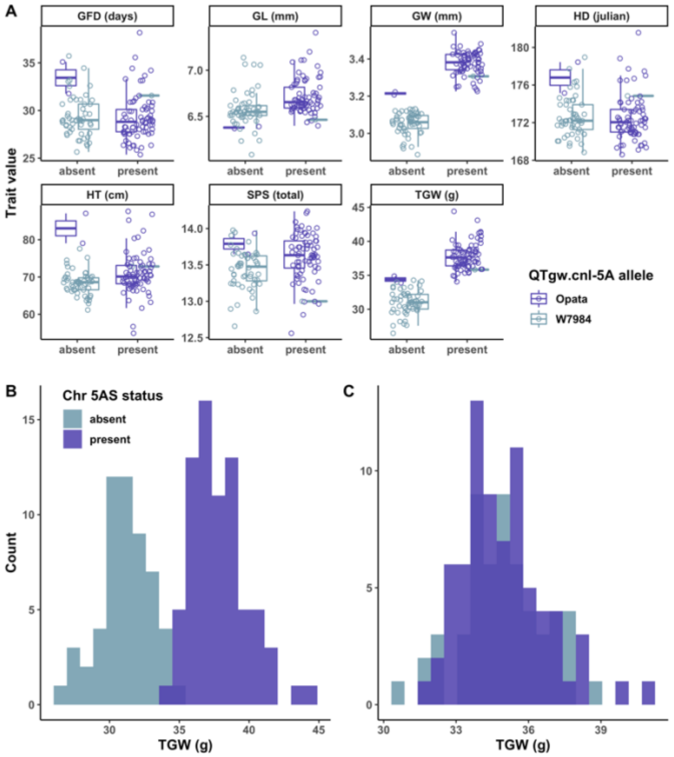

Figure 2.4: (A) BLUP phenotype boxplot distribution for HIF entries subset by chromosome arm 5AS presence or absence, and colored by *QTgw.cnl-5A* allele (KASP *5A_341510829*, Opata: purple, W7984: blue). Of the 129 fine-mapping entries, 73 had haplotype chromosome short arm 5A+ / *QTgw.cnl-5A*+, 53 had haplotype chromosome short arm 5A- / *QTgw.cnl-5A*-, one had haplotype chromosome short arm 5A+ / *QTgw.cnl-5A*-, and two had haplotype chromosome short arm 5A- / *QTgw.cnl-5A*+. (B) Original BLUP TGW histogram, subset by chromosome arm 5AS presence (purple) or absence (blue). (C) Chromosome arm 5AS structural variant fixed effect BLUP TGW histogram, subset by presence (purple) or absence (blue).

RNA-sequencing of QTgw.cnl-5A HIFs confirmed the significance of chromosome arm 5AS structural variation

We used RNA-seq to investigate three questions with four HIF haplotypes (Table 2.3), i) are genes in the *QTgw.cnl-5A* candidate region differentially expressed, ii) are there genes on chromosome arm 5AS that are differentially expressed, and iii) do the HIFs share an isogenic background genome. We hypothesized that gene expression changes influencing phenotypic differences occurred before the first significant difference in grain morphology was detectable at 10-DPA and sampled whole grains at 4 and 8-DPA. We measured over 621 M reads across the 24 samples, with individual sample reads ranging from 21.8 M to 37 M and an average of 25.8 M reads (standard error 0.7 M) per sample (Supplementary file_S4.02.csv, file_S4.03.csv). We aligned the reads to RefSeq v1.1 HC and LC annotations independently, and on average across samples 89.4 $^{+}_{-}$ 0.6% of reads aligned to the HC gene annotation, and 89 $^{+}_{-}$ 0.62% aligned to the LC gene annotation (for read counts see Supplementary file_S4.md).

HIF [1]	Name	chr 5AS	*QTgw.cnl-5a*	TGW (g)	GW (mm)
7-956-2-19-1-31-03	Opata control	present	+	37.895	3.408
7-956-2-19-1-44	W7984 control	absent	-	29.983	3.091
7-956-2-19-1-31-05	recombinant I	present	+	36.341	3.39
7-956-2-12-1-69-07	recombinant II	present	+	40.199	3.44

[1] HIF entries were chosen to prioritize isogenic lineage and QTL resolution, over chr 5AS and *QTgw.cnl-5A* recombinants.

Table 2.3: Heterogenous inbred family (HIF) haplotypes selected for RNA-sequencing. Chromosome arm 5AS structural variation (chr 5AS), *QTgw.cnl-5A* genotype (marker 5A_341510829) where Opata is (+) and W7984 is (-), thousand grain weight (TGW), grain width (GW).

We performed differential expression comparisons between the HIFs at both timepoints. Independent comparison between HIFs for Opata control, recombinant I, and recombinant II versus W7984 control did not identify any DEGs at 4 and 8-DPA in the *QTgw.cnl-5A* candidate region (chromosome 5A *339757917 – 349628635 bp*) for the HC or LC gene annotation. However, there were 532 significant DEGs in common among these HIF comparisons at 4-DPA, 469 at 8-DPA, and 556 in total at either timepoint for the HC gene annotation. Of the 556 DEGs, fifteen were on chromosome 1A, one on 1B, one on 3A, 532 on 5A short arm, three on 5A long arm, one on 5D, two on 7A, and one on 7B. The 535 genes on chromosome 5A were all differentially expressed due to no expression in the W7984 control HIF. None of the DEGs homoeologous copies were differentially expressed. Given only 21 of the 556 DEGs were on chromosomes other than chromosome 5A, the differential expression comparisons between HIFs also validated the fine-mapping population's isogenic background genome (Figure 2.5 and Supplementary script_S4.md).

These results prompted us to evaluate the considerable resources we had invested in an otherwise reliable positional cloning approach for *QTgw.cnl-5A*. It was possible we sampled tissue at the wrong timepoints, or *QTgw.cnl-5A* association with grain weight and morphology was driven by a post-transcriptional or post-translational modification. We investigated the ten Mbp candidate region of the 24 samples' raw reads for SNP variants from the reference genome and identified only one variant. Among the three W7984 control biological replicates a 3' untranslated region mutation (T → C) was present in *TraesCS5A02G160900* (Supplementary file_S4.18.xlsx). This gene codes for the third largest subunit of RNA polymerase II and the SNP, termed BA00617686, had previously been identified by the CerealsDB Axiom 820K and 35K SNP Array (Winfield et al., 2016). The SNP has not been associated with any phenotype, and there is no literature on the Arabidopsis and rice orthologs that implicate the gene in grain development.

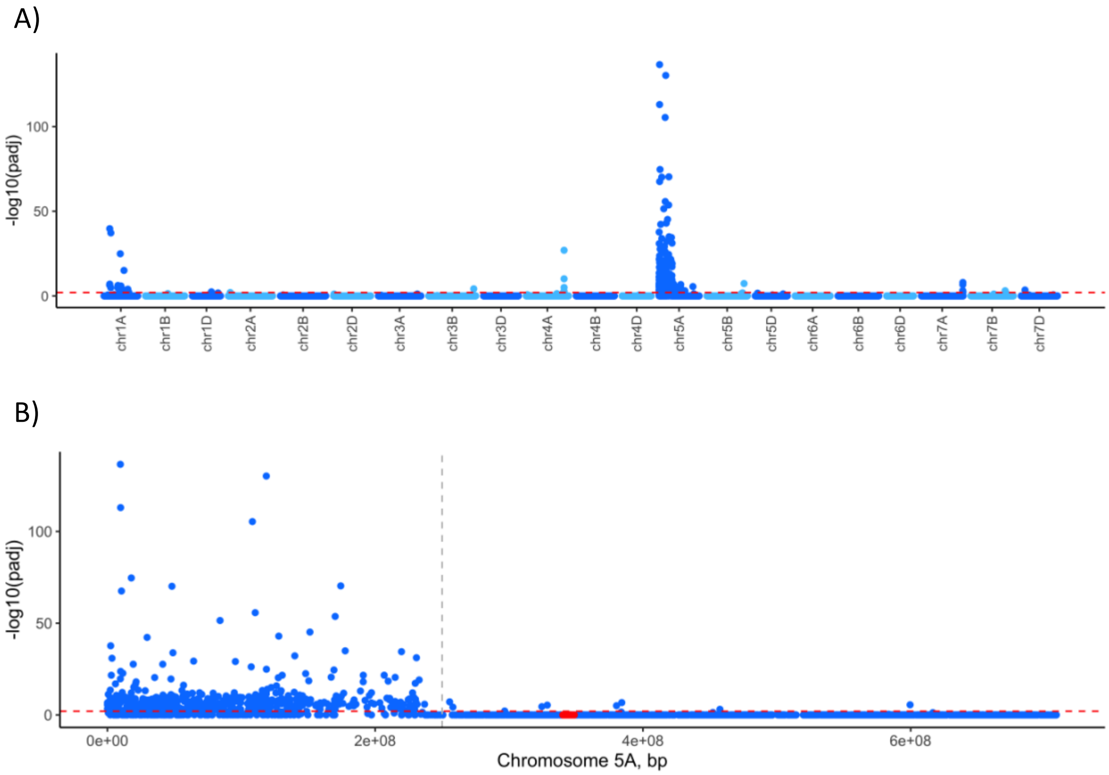

Figure 2.5: Differentially expressed genes (A) genome wide, and (B) on chromosome 5A, for Opata versus W7984 control haplotypes at 8 DPA. Each point represents a gene, ordered by RefSeq v1.0 physical position along the X-axis. The Y-axis represents the differential expression significance, -log base 10 adjusted P-value. The red dashed line marks P-value = 0.01. In (B) the grey dashed line represents the centromere, and the red points highlight genes between *QTgw.cnl-5A* flanking KASP markers *5A_339757917* and *5A_349628635*.

Centromeres are vital for proper chromosomal segregation during mitosis and meiosis. Consequentially, experimentally derived chromosome arm deletion lines in wheat often lack a clean break at the centromere (Gill, 1996). The centromere of chromosome 5A is near 250 Mbp in Chinese Spring RefSeq v1.0. Although (Gutierrez-Gonzalez et al., 2019) reported that the entire short arm was missing for W7984 chromosome 5A, our transcriptome study revealed fifteen genes expressed in the W7984 control haplotype 237 - 250 Mbp. None of the W7984 haplotype expressed genes were differentially expressed with the positive haplotypes, which made them unlikely candidates for the observed phenotypic variation. The centromeric position can change among cultivars, but this has not previously been associated with a structural variation event (Walkowiak et al., 2020). We propose that W7984 short arm of chromosome 5A broke approximately thirteen Mbp from the centromere, as compared to Chinese Spring. This remains to be validated with a transcriptome study across additional SynOp entries and parents.

Across the W7984 haplotype samples there were 23 genes that were unexpectedly expressed on chromosome arm 5AS (0-237 Mbp) (Supplementary file_S4.14.xlsx). We investigated the read counts of these genes and identified eleven of the genes were differentially expressed, but still from less expression by W7984 samples. We explored the 23 gene sequences using *EnsemblPlants* and IWGSC BLAST, and on average there was 94.6 $^{+}_{-}$ 1.7% sequence alignment with at least one homoeolog or paralog. On average, reads aligned to multiple loci represent 5.16 $^{+}_{-}$ 0.22% of the total alignment (for read counts see Supplementary file_S4.md). We suggest these 23 genes have multiple sequence alignments and that their expression in the W7984 haplotype samples was incorrect.

Collectively, these findings indicate our approach to positional cloning was overpowered by chromosome arm 5AS structural variation and that *QTgw.cnl-5A* was a result of strong LD. The presence and absence of chromosome arm 5AS was significantly associated with the yield component variation we identified in the fine-mapping population. An in-depth comparison of the 556 DEGs and association with yield components follows in the next section.

Chromosome arm 5AS differential expression provided insight into candidate genes

The phenotypes measured in the fine-mapping population were chosen based on the quantitative phenotypic variation associated with *QTgw.cnl-5A* in the SynOpDH background. In an isogenic fine-mapping background the phenotypic resolution was enhanced, and included significant associations between HT, SPS, TGW, GL, and GW and the presence or absence of chromosome arm 5AS. The percent difference between the *QTgw.cnl-5A* +/- HIFs for these traits is reflective of additive, rather than dominance variation (Table 2.2). There were no other variable phenological or morphological traits among SynOpDH and HIF populations that were obvious during field trials. It is well known that wheat harbors hidden variation due to polyploidy and there are likely non-additive and functionally redundant genes among the DEGs on chromosome 5A.

To better understand the DEGs, we performed GO enrichment analysis with GOseq (Young et al., 2010). The DEGs were associated with 1,919 out of 9,709 unique GO terms from the RefSeq v1.1. After statistical over-representation tests and correction for multiple testing the only significant term was “killing of cells of other organism” (GO:0031640) (Supplementary script_S4.md). There were nine DEGs with GO:0031640, including the physical cluster of genes *TraesCS5A02G018000, TraesCS5A02G018800*, *TraesCS5A02G019000* and *TraesCS5A02G019100* which are orthologs of Arabidopsis bacterium and fungus response gene Osmotin-like protein, OSM34 (Capelli et al., 1997). *TraesCS5A02G077600* is within the interval of a fusarium head blight resistance QTL termed *Qfhs.ifa-5AS*, and is orthologous to the rice Osmotin-like protein Os12g0569300 (NCBI; Steiner et al., 2019). *TraesCS5A02G059000* is orthologous to the rice defense response gene OsPR1b, which is constitutively expressed at background levels (Luan and Zhou, 2015). Although the physical gene cluster of *TraesCS5A02G046300*, *TraesCS5A02G046400*, and *TraesCS5A02G046500* has no known orthologs in Arabidopsis or rice, they share the additional GO terms defense response to fungus (GO:0050832), binding (GO:0005488), peptidase activity (GO:0008233), extracellular region (GO:0005576), cell wall (GO:0005618), integral component of membrane (GO:0016021), cellular component organization (GO:0016043), and cellular metabolic process (GO:0044237).

The relative lack of GO term enrichment among chromosome arm 5AS+ and 5AS- HIF DEGs coupled with phenological variation characteristic of a single gene rather than entire chromosome arm, highlights the challenge of identifying hidden variation in a redundant polyploid genome. Despite the limited GO term enrichment, the positive association of SPS, TGW, GL and GW with chromosome arm 5AS presence prompted us to investigate genes with GO terms related to spike architecture and early grain development. It was unclear if the difference in height was attributable to any DEGs measured at 4 or 8-DPA. There were 292 DEGs that had at least one GO term related to biological, cellular, and developmental processes during spike and early grain growth (file_S4.16.csv). Of these genes, 52 were only differentially expressed at 4-DPA, and ten genes were only differentially expressed at 8-DPA. The majority of the 230 genes that were differentially expressed at both timepoints decreased in expression across time. Only seven of the genes were on chromosomes other than 5A, and are poor candidates based on GO terms and Arabidopsis and rice orthologs. We explored the Arabidopsis and rice orthologs for all 292 genes and present nine genes as candidates for wheat yield components in Table 2.4.

While the genes in Table 2.4 have not previously been associated with wheat yield components, their Arabidopsis and rice orthologs have been shown to regulate inflorescence development or seed / grain size. Based on RefSeq v1.0 alignment and v1.1 annotation, there are eight candidate genes on chromosome arm 5AS and one candidate gene on chromosome arm 1AS, *TraesCS1A02G103900*. *TraesCS1A02G103900* had GO terms related to cell size, cell growth, and epidermal cell differentiation, and was orthologous to HAIKU2 (IKU2), a regulator of endosperm proliferation and cellularization (Luo et al., 2005; Li and Li, 2016). IKU2 loss-of-function is associated with reduced seed size in Arabidopsis (Luo et al., 2005). Notably, the homoeologous copies of *TraesCS1A02G103900*, *TraesCS5B02G012000* and *TraesCS5D02G019400*, map to the group 5 chromosomes, not group 1. The homoeologs, as reported on *EnsemblPlants,* are orthologues to IKU2 as well. Our transcriptome analysis identified three other DEGs on chromosome 1AS clustered near *TraesCS1A02G103900* (9.76-9.98 Mbp), which also have homoeologs on the group 5 chromosome (Supplementary file_S4.19.xlsx). All four of these genes were differentially expressed due to zero read counts from the W7984 haplotype, while the other eleven DEGs identified on chromosome 1A have a mix of expression profiles among the four haplotypes. We submitted the *TraesCS1A02G103900* coding sequence to IWGSC BLAST RefSeq v1.0 and RefSeq v2.0, which identified 100% alignment with RefSeq v1.0 chromosome 1A and 71% with 5A, and 100% alignment with RefSeq v2.0 chromosome 5A and 75% with 1A. We also identified 99% alignment with chromosome 5A of the Durum wheat (cv. Svevo) genome assembly v1. Given the haplotype expression profile, homoeologous copies on group 5 chromosomes, and RefSeq v2.0 and Durum alignment, we believe *TraesCS1A02G103900* was misannotated by RefSeq v1.1 and belongs on chromosome arm 5AS. The eight additional candidate genes on chromosome arm 5AS and their ortholog functions are discussed in the proceeding section *Candidate genes for yield components*.

T. aestivum	RefSeq v1.0	Refseq v2.0	4-DPA *P*-value[1]	8-DPA *P*-value[1]	Selected GO terms[2]	*A. thaliana*	*O. sativa*
TraesCS1A02G103900	1AS	5AS	3.99 x 10^{-8}	1.07 x 10^{-7}	regulation of cell size (GO:0008361) plant-type cell wall organization (GO:0009664) multidimensional cell growth (GO:0009825) epidermal cell differentiation (GO:0009913)	*HAIKU2 (IKU2)*, regulator of seed size	*HAIKU2*
TraesCS5A02G025900	5AS	5AS	5.59 x 10^{-30}	1.34 x 10^{-17}	post-embryonic development (GO:0009791) cell differentiation (GO:0030154) reproductive structure development (GO:0048608)	*YAB2*, abaxial cell fate	*OsYAB6*
TraesCS5A02G030300	5AS	5AS	2.25 x 10^{-9}	1.7 x 10^{-7}	brassinosteroid mediated signaling pathway (GO:0009742)	*BSL2* and *BSL3*, brassinosteroid-insensitive suppressor family	*OsPPKL3*, negative regulator of grain length
TraesCS5A02G037100	5AS	5AS	4.99 x 10^{-4}		stamen development (GO:0048443) regulation of cell size (GO:0008361) cell projection (GO:0042995) auxin-activated signaling pathway (GO:0009734)	*PID*, establishment of bilateral symmetry	*BIF2*, organogenesis during inflorescence
TraesCS5A02G038300	5AS	5AS	1.43 x 10^{-16}	5.75 x 10^{-10}	auxin-activated signaling pathway (GO:0009734) flower development (GO:0009908)	*ARF6*, regulator of flower development / maturation	*OsARF25*, regulator of grain size
TraesCS5A02G103800	5AS	5AS	6.54 x 10^{-11}	7.61 x 10^{-7}	auxin metabolic process (GO:0009850) cytokinin metabolic process (GO:0009690)	*RGLG1* and *RGLG2*, regulator of apical dominance	
TraesCS5A02G106400	5AS	5AS	4.69 x 10^{-12}	4.93 x 10^{-7}	diacylglycerol kinase activity (GO:0004143) intracellular anatomical structure (GO:0005622)	*DGK2*, pollen and seed development	
TraesCS5A02G107800	5AS	5AS	1.37 x 10^{-14}	9.33 x 10^{-14}	response to ethylene (GO:0009723) response to auxin (GO:0009733)	*VTC2* and *VTC5*, ascorbate biosynthesis	*OsGGP*, biomass production
TraesCS5A02G110600	5AS	5AS	1.07 x 10^{-13}	4.18 x 10^{-9}	phragmoplast (GO:0009524) spindle assembly (GO:0051225)	*AUG6*, mitotic and meiotic cell division	*Os02g0329300*, AUGMIN subunit 6

Table 2.4: Candidate *T. aestivum* RefSeq v1.1 genes and *A. thaliana* and *O. sativa* orthologs. RefSeq v1.0 and v2.0 assembly chromosome. [1]GOSeq2 adjusted P-value from Opata vs W7984 HIF haplotype comparisons, 4 and 8-days post anthesis (DPA). All haplotype comparisons can be found in Supplementary script_S4.md. [2]All GO terms are available in Supplementary file file_S4.17.xlsx.

Discussion

In this study we discovered that a stable and robust QLT was confounded by linkage with chromosome structural variation. We confirmed that the SynOpRIL population, in addition to the SynOpDH population described by (Gutierrez-Gonzalez et al., 2019) is segregating for presence of the Opata parent chromosome arm 5AS and absence of the majority of the W7984 parent chromosome arm 5AS. Furthermore, we associated the chromosome arm 5AS structural variation in an isogenic background with yield component phenotypes, characterized the early grain development transcriptome, and propose nine candidate genes for agronomically valuable traits on chromosome arm 5AS. Genomic structural variations are common across polyploids, and this study contributes to our understanding of the complexities associated with fine-mapping in a polyploid species, and discusses a more robust approach to positional cloning (Song et al., 1995; Saxena et al., 2014).

Detecting the chromosome structural variation

SynOpDH and SynOpRIL were two of the most widely referenced mapping populations leading up to the advent of the annotated wheat reference genome. Aside from the phenotypic variation subtleties, why was the W7984 chromosome arm deletion not detected sooner by the wider wheat community? The original W7984 X Opata crosses were developed in the early 1990s and were later reconstructed and expanded in 2011 (Sorrells et al., 2011). The re-released population was genotyped with SSR and DArT markers, which are scored for the presence or absence of genomic fragments and would not have alerted researchers to the segregation of chromosome arm 5AS. When the SynOpDH high-density GBS marker genetic map was developed in 2012 there was no reference genome available to flag that none of the markers mapped to chromosome arm 5AS (Poland et al., 2012). Our research group first noticed an anomaly on chromosome arm 5AS when we were unable to identify genome and chromosome arm specific polymorphic sites among the parent line exome-capture and regulatory-capture sequence for KASP marker development. In 2019, the dense GBS linkage maps created by Gutierrez-Gonzalez et al. confirmed the significant structural variation of chromosome arm 5AS. It is also worth noting that many SynOp studies relied on a subset of either mapping population and may not have had a high enough frequency of the missing arm to detect the abnormality or were not focused on chromosome 5A. We urge previous studies involving SynOp populations and group 5 chromosomes to consider implications of chromosome arm 5AS structural variation on their results and recommend a list of SSR markers for any future studies that need to characterize SynOp entries not included in our current study (Supplementary file_S2.11.csv).

Why did QTgw.cnl-5A map to chromosome arm 5AL?

Fundamentally, our positional cloning difficulties began with a QTL that mapped to the wrong chromosome arm. Our SynOpDH QTL map significantly associated markers on chromosome arm 5AL with grain weight and morphology across four environments, which is consistent with previous publications but inconsistent with our fine-mapping results (Breseghello and Sorrells, 2006; Williams et al., 2014). Even when we were prompted by the findings of (Gutierrez-Gonzalez et al., 2019) to add a chromosome arm 5AS structural variation marker to the genetic map, *QTgw.cnl-5A* still presented on the long arm. Potentially, R/qtl is insensitive to chromosome arm 5AS presence or absence because the *QTgw.cnl-5A* causal variant is very near the centromere or part of the W7984 chromosome arm 5AS thirteen Mbp segment, but none of the genes in this region were differentially expressed. The lack of phenotypic resolution for quantitative traits like TGW and GW in the SynOpDH background may have also reduced our power to detect the association with chromosome arm 5AS, as seen among the minimal DH chromosome arm 5AS and *QTgw.cnl-5A* haplotype recombinants (Supplementary script_S1.md). Alternatively, the genetic map of GBS and SSR markers on chromosome arm 5AL may not be genome specific and false recombination events placed *QTgw.cnl-5A* on chromosome arm 5AL. Finally, as we observed across *QTgw.cnl-2D*, *QTgw.cnl-5A*, and *QTgw.cnl-6A*, GL and GW can independently influence grain weight. Although the GW 5A QTL spanned the centromere, because of its overlap with *QTgw.cnl-5A* we considered the difference in GW and TGW to be a pleiotropic effect. Collectively, these results encouraged us to develop a fine-mapping population targeting *QTgw.cnl-5A* on chromosome arm 5AL.

Candidate genes for yield components

Grain yield is a key economic driver behind the success of wheat production and the value chain. However, grain yield is determined at the end of plant growth and is directly or indirectly impacted by all genes. Balancing the mechanisms driving the source tissue growth, the number of grains produced by a plant, and the weight of those grains are a key challenge for breeders as the constituent yield components compete for resources. Notably, in this study we have identified a positive combination among five traits (HT, SPS, TGW, GL, and GW) under independent genetic control on Opata chromosome arm 5AS. Phenotypic variation associated with chromosome arm 5AS structural variation, significant differential expression during early grain development, GO terms, and orthologs were leveraged to identify nine candidate genes that may impact spike or grain development (Table 2.4). Ultimately, the genomic resolution of this study was limited by presence/absence structural variation and it is premature to conclude that any of the nine candidate genes contribute to the observed spikelet and grain morphology variation. Independent knockout studies of each candidate gene in HIF entries with Opata chromosome arm 5AS are necessary for functional validation. A discussion of each candidate gene follows.

Among the DEGs we identified *TraesCS5A02G025900,* an ortholog to Arabidopsis gene *YAB2* and rice gene *YAB6*, which regulate abaxial cell fate in leaf, sepal, petal, stamen and carpel primordia (Siegfried, 1999). In Arabidopsis, *YAB2* acts redundantly with the larger YABBY gene family, and single loss-of-function plants exhibited no measurable organ polarity defects (Stahle et al., 2009). There are 21 YABBY genes subdivided into six families in wheat but functional validation studies and phenotypic associations remain unexplored (Buttar et al., 2020). Given the functional redundancy identified in Arabidopsis, and large gene family in wheat, the loss of *TraesCS5A02G025900* (TaYABBY5-5A) in our study was likely masked by gene family copies.

Another candidate gene that may impact wheat spike inflorescence development is *TraesCS5A02G037100*. *TraesCS5A02G037100* is orthologous to Arabidopsis gene *PID*, and rice and maize gene *BIF2*, which are involved in inflorescence organogenesis (Benjamins et al., 2001; McSteen et al., 2007; He et al., 2019b). While loss-of-function plants in maize produced fewer spikelets and florets, rice did not suffer from flower initiation defects and indicated *BIF2* function is likely veiled by redundant partners (He et al., 2019b). *TraesCS5A02G037100* has not been associated with SPS in wheat outside of our chromosome arm 5AS structural variation study.

TraesCS5A02G103800 is orthologous to Arabidopsis RING domain Ligase genes *RGLG1* and *RGLG2*, which modulate the directional flow of auxin (Yin et al., 2007). While single mutants in either gene have no apparent phenotypic effect, double mutant plants exhibit loss of apical dominance and altered phyllotaxy. *TraesCS5A02G103800* and homoeologous gene copies may harbor hidden variation regulating spike development in wheat.

TraesCS1A02G103900 is orthologous to *IKU2*, a gene involved in endosperm growth and seed development signaling pathways (Luo et al., 2005). The RefSeq v1.1 annotation maps *TraesCS1A02G103900* to chromosome 1A, however our results indicate this gene was misannotated and maps to chromosome 5A. *SHB1, IKU1, IKU2* and *MINI3* function in the same signaling pathway to control seed size in Arabidopsis, and loss of any gene can reduce seed size (Li and Li, 2016). Loss of the *IKU2* wheat ortholog impact on grain size remains to be functionally validated outside of our chromosome arm 5AS structural variation study.

TraesCS5A02G030300 is orthologous to protein phosphatases with Kelch-like domains (PPKL) genes *BSL2* & *BSL3* in Arabidopsis, and *OsPPKL3* in rice (Kim et al., 2018). PPKL are known positive effectors of brassinosteroid signaling in plants, and brassinosteroid has been shown to regulate seed size and shape in Arabidopsis (Jiang et al., 2013). *OsPPKL3* and its homolog *OsPPKL1* are negative regulators of grain length, but rare allelic variation is associated with extra-large grains and increased yield (Zhang et al., 2012). The wheat ortholog to *OsPPKL1*, *TaGL3-5A*, is associated with longer grain length and has been functionally validated, suggesting *TraesCS5A02G030300* may be a strong candidate for exploring allelic diversity and association mapping (Yang et al., 2019).

Another potential grain size regulator is *TraesCS5A02G038300*, or *TaARF14*, which is orthologous to auxin response factor (ARF) *ARF6* in Arabidopsis, and *OsARF25* in rice (Nagpal et al., 2005; Zhang et al., 2018b; Xu et al., 2020). Dosage effects among ARF knockouts in Arabidopsis found that *ARF6* impacts the timing of flower maturation. *OsARF25* has been functionally validated in an auxin signaling pathway where it binds to the promoter of *OsERF142*, a positive regulator of cell expansion and ultimately grain length (Zhang et al., 2018b). A recent comprehensive atlas of wheat ARF gene expression suggests *TaARF14* is required to promote stamen development, but functional validation outside of chromosome arm 5AS structural variation remains to be determined.

TraesCS5A02G106400, *TaDGK5A*, is the wheat ortholog of the Arabidopsis gene *DGK2*, a diacylglycerol kinase essential for reproductive organ development (Angkawijaya et al., 2020). There are seven DGK genes in Arabidopsis which cumulatively contribute to phospholipid signaling and three are known to impact gametogenesis. A recent pre-print of wheat DGK2 genomic and expression profiles identified 20 genes and their upregulated expression in root tissues under salt and drought stress (Jia et al., 2020). The study did not sample grain tissue for transcriptome analysis. The larger TaDGK family likely masked a phenotypic response to loss of chromosome arm 5AS *TaDGK5A* in our HIF population and functional validation of TaDGK genes in wheat is an outstanding area of research.

Meiotic and mitotic cell divisions are elevated during flowering plant reproductive growth and organogenesis. Among the DEGs we identified *TraesCS5A02G110600*, an ortholog to Arabidopsis and rice subunit 6 of the augmin complex, which is responsible for microtubule nucleation during cell divisions (Hotta et al., 2012; Oh et al., 2016). Knockdown lines of *AUG6* in Arabidopsis disturbed mitotic and meiotic cell divisions due to malformed microtubule arrays, and effected both male and female gamete development (Oh et al., 2016). Functional validation of genes that affect gamete development is challenging due to homozygous lethality. An alternative approach to studying the effect of *TraesCS5A02G110600* on HIF entries could be to measure cell size and number of developing grains, which has previously been associated with grain morphology variation in wheat (Brinton et al., 2017).

Given the dramatic genomic structural variation and relatively few variable visual phenotypes in the HIF population, there is likely more phenotypic variation than what meets the eye. For example, *TraesCS5A02G107800* orthologs *VTC2* and *VTC5* in Arabidopsis, and *OsGGP* in rice catalyze the first step in the ascorbate (AsA) biosynthetic pathway (Gao et al., 2011; Höller et al., 2015; Lim et al., 2016). AsA is an essential signal-transducing molecule for regular plant development, and loss-of-function *OsGGP* plants have significantly reduced biomass (80% loss), panicle number, and panicle weight. Knockout *VTC2* and *VTC5* plants and AsA deficiency have been associated with reduced growth by some studies, or more recently with an independent cryptic mutation (Gao et al., 2011; Lim et al., 2016). While we did not detect biomass variation on the order of magnitude measured in rice, hidden variation or homoeologous masking is likely at play. A future study of signal-transducing molecule concentrations or hormone panels of the HIF population would likely unveil additional candidate genes.

Wheat positional cloning recommendations

Positional cloning in wheat has lagged behind the progress made in other staple crops in large part due to the redundant polyploid genome's masking of quantitative variation and the delayed availability of an annotated genome sequence. Even in an isogenic background the missing chromosome 5A short arm was associated with only subtle phenotypic variation. Wheat deletion lines are classically associated with aberrant morphological variation or infertility, but our results demonstrate that large structural variation can go undetected for years, across environments, research projects, and lab groups. Recent studies of fifteen bread wheat cultivar genome assemblies showed that only 59% of each cultivar's genome was identical-by-state with other sequenced cultivars and detected extensive genomic rearrangements, underscoring the structural diversity of wheat (Brinton et al., 2020; Walkowiak et al., 2020). Our study illustrates that approaching positional cloning based on stable and robust biparental QTL mapping may overlook some hurdles unique to wheat's resilient polyploid genome. If one is going to invest the considerable resources necessary for successful positional cloning in wheat, or other polyploid crops, we propose the following recommendations:

- **Move from a SNP to haplotype-based approach to identify genetic diversity**. Genetic markers are often associated with a trait of interest but are not causal. In a fine-mapping context this can overlook linkage, especially with broad genomic regions spanning the centromere such as the GW 5A QTL, and result in selection of false-positives (Platten et al., 2019). Haplotype blocks can identify genetic structural variation more comprehensively and precisely, as demonstrated by the recent release of fifteen cultivar genome assemblies and accompanying visualization platform (Brinton et al., 2020; Walkowiak et al., 2020). Exome sequence capture data are available for far more wheat cultivars than genome assemblies, but safeguards need to be in place to detect and flag structural variation, for example in W7984 exome and regulatory-capture (Gardiner et al., 2019; He et al., 2019a).
- **Invest in sequencing to detect structural variants.** While the cost of whole genome sequencing in wheat is not yet feasible for individual breeding programs, long-read sequencing and greater fold coverage (e.g., 10) has become increasingly affordable. Longer sequencing reads can detect even small- (30 – 10,000 bp) to mid-scale (10,000 – 30,000 bp) structural variants, which impact trait diversity and are shown to be widespread in polyploid species (Gabur et al., 2019; Mahmoud et al., 2019; Chawla et al., 2021). For

example, a recent study of *Brassica napus* L. found that up to 10% of all genes were affected by small- to mid-scale structural variants, including flowering-time pathway genes which can influence agronomic traits (Chawla et al., 2021). Obtaining long-read sequencing (small- to mid-scale structural variants) or 10-fold coverage (large-scale structural variants) of parental lines for a mapping population could become a pre-requisite first-step in positional cloning population development.

- **Use the transcriptome to identify candidate genes.** RNA-seq of isogenic material can identify differential expression and coding region allelic variation. This method alerted us to variants of interest arguably faster than pursuing additional HIF recombinants in the *QTgw.cnl-5A* ten Mbp region. Measurements across multiple timepoints (ie. > five) can open the door to developing gene networks for candidate gene discovery (Borrill et al., 2020). It is also notable that wheat transcriptome studies have reported large-scale structural variants, including inter-homoeolog exchanges (He et al., 2017).
- **Traits with broad overlapping QTL may not be pleiotropic.** The traditional yield components (spikes/ m2, grain number/spike, and grain weight) are polygenic traits themselves. For example, in our study we showed that GL and GW can independently contribute to TGW. However, our hypothesis that a gene contributing to GW drove the TGW variation on chromosome arm 5AL overlooked the association with GW and chromosome arm 5AS. Eventually, in an isogenic background we identified additive and hidden phenotypic variation associated with structural variation, rather than a single gene. Another recent

example that has disrupted pleiotropic assumptions based on single-marker trait associations was identified in a highly conserved region of chromosome 6A with a haplotype-led approach (Brinton et al., 2020).

With an annotated reference genome and advancing gene editing techniques, positional cloning in wheat may become routine but pinpointing quantitative phenotypic variation and causal genomic loci requires a more tactical approach in polyploids. Incorporating strategies that are sensitive to structural variation with classic positional cloning population development approaches will reduce the likelihood of mapping in the wrong direction. Fewer roadblocks to identifying a candidate gene will also lead to more efficient selection from mutant/TILLING populations or transgenic approaches during sequencing and functional validation (Krasileva et al., 2017). A comprehensive review of the latest advances of genomics and phenomics for trait discovery in polyploid wheat and gene functional characterization is given in (Borrill et al., 2018) and (Adamski et al., 2020)

Conclusion

The outcomes of this research challenge whether a causal gene variant approach to characterizing wheat grain yield components offers an efficient and sustainable path to genetic gains and food security. We set out to identify the causal variant underlying a previously characterized grain weight and morphology QTL, *QTgw.cnl-5A*, using a well vetted positional based cloning approach. We leveraged a HIF population, SNP genomic data, phenotypic associations, early grain development expression profiles, and predicted gene function to determine that *QTgw.cnl-5A* was a result of strong LD with chromosome arm 5AS presence and absence (SynOpDH $r^2 = 0.95$, HIF $r^2 = 0.91$). Our results highlight that chromosome structural variation linkages can overpower the considerable resources required for positional cloning, and that wheat harbors hidden phenotypic variation. Chromosome structural variation is common in among polyploids, and the results and recommendations presented will be immensely useful for aiding future causal variant discovery. We also identified nine candidate genes on chromosome arm 5AS that may impact yield components, however their practical application to breeding remains to be functionally validated. Given the resources required for individual gene validation, uncertain impact on final grain yield, and unknown response in a different genetic background, we argue that causal variant discovery for a complex quantitative trait like wheat yield requires an update to traditional positional based cloning approaches. Altogether, our findings demonstrate the phenotypic resiliency associated with polyploid genomic structural variation and provide recommendations for variant discovery strategies.

Acknowledgements

We thank David Benscher and Amy Fox for planting and harvesting assistance. We are grateful for the contributions of undergraduate research assistants Elizabeth De Meyer, Rui Fengui Zhan, and J. Edward Keating for phenotyping and greenhouse support. We also thank the Eastern Regional Small Grains Genotyping Laboratory for assistance with KASP marker genotyping of HIFs. Funding for this project was provided by USDA National Institute of Food and Agriculture grant 2017-67007-25939 (Wheat-CAP) and Hatch Project 149-945.

Supplemental material

All of the data referenced in this manuscript, supplementary files, and scripts for reproducing results are publicly available at: https://github.com/etaagen/Taagen_2021_TPG.git . Refer to .md files to view analysis output, and .Rmd files to view full script and reproduce results. Additional packages required to reproduce results or figures include *R/tidyverse, R/kintr, R/kableExtra, R/gt, R/rstatix, R/rlang, R/ggpubr, R/vsn, R/PCAtools, R/pheatmap, R/reshape, R/ggbio, R/biovizBase, R/gghighlight, R/inauguration* (Huber et al., 2002; Yin et al., 2012, 2020; Wickham, 2018; Kolde, 2019; Wickham et al., 2019; Blighe and Lun, 2020; Xie, 2020; Yutani, 2020; Zhu, 2020; Henry et al., 2020; Iannone et al., 2020; Kassambara, 2020a; b; Bedford-Petersen, 2021).

REFERENCES

Adamski, N.M., P. Borrill, J. Brinton, S.A. Harrington, C. Marchal, A.R. Bentley, W.D. Bovill, L. Cattivelli, J. Cockram, B. Contreras-Moreira, B. Ford, S. Ghosh, W. Harwood, K. Hassani-Pak, S. Hayta, L.T. Hickey, K. Kanyuka, J. King, M. Maccaferrri, G. Naamati, C.J. Pozniak, R.H. Ramirez-Gonzalez, C. Sansaloni, B. Trevaskis, L.U. Wingen, B.B. Wulff, and C. Uauy. 2020. A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat. Elife 9:1–30. doi:10.7554/eLife.55646

Alaux, M., J. Rogers, T. Letellier, R. Flores, F. Alfama, C. Pommier, N. Mohellibi, S. Durand, E. Kimmel, C. Michotey, C. Guerche, M. Loaec, M. Lainé, D. Steinbach, F. Choulet, H. Rimbert, P. Leroy, N. Guilhot, J. Salse, C. Feuillet, E. Paux, K. Eversole, A.F. Adam-Blondon, and H. Quesneville. 2018. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. Genome Biol. 19. doi:10.1186/s13059-018-1491-4

Andrews, S. 2010. FastQC A Quality Control Tool for High Throughput Sequence Data [Online]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed November 26, 2020).

Angkawijaya, A.E., V.C. Nguyen, F. Gunawan, and Y. Nakamura. 2020. A pair of arabidopsis diacylglycerol kinases essential for gametogenesis and endoplasmic reticulum phospholipid metabolism in leaves and flowers. Plant Cell. doi:10.1105/tpc.20.00251

Bates, D., M. Mächler, B.M. Bolker, and S.C. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw.. doi:10.18637/jss.v067.i01

Bedford-Petersen, C. 2021. Inauguration, R color palette

Benjamins, R., A. Quint, D. Weijers, P. Hooykaas, and R. Offringa. 2001. The PINOID protein kinase regulates organ development in Arabidopsis by enhancing polar auxin transport. Development

Blighe, K., and A. Lun. 2020. PCAtools: PCAtools: Everything Principal Components

Analysis

Borrill, P., S.A. Harrington, J. Simmonds, and C. Uauy. 2020. Identification of Transcription Factors Regulating Senescence in Wheat through Gene Regulatory Network Modelling 1[OPEN]. doi:10.1104/pp.19.00380

Borrill, P., S.A. Harrington, and C. Uauy. 2018. Applying the latest advances in genomics and phenomics for trait discovery in polyploid wheat. Plant J. 97:tpj.14150. doi:10.1111/tpj.14150

Breseghello, F., and M.E. Sorrells. 2006. Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics. doi:10.1534/genetics.105.044586

Brinton, J., R.H. Ramirez-Gonzalez, J. Simmonds, L. Wingen, S. Orford, S. Griffiths, W. Genome Project, G. Haberer, M. Spannagl, S. Walkowiak, C. Pozniak, and C. Uauy. 2020. A haplotype-led approach to increase the precision of wheat breeding. Commun. Biol. in press. doi:10.1038/s42003-020-01413-2

Brinton, J., J. Simmonds, F. Minter, M. Leverington-Waite, J. Snape, and C. Uauy. 2017. Increased pericarp cell length underlies a major quantitative trait locus for grain weight in hexaploid wheat. New Phytol. 215:1026–1038. doi:10.1111/nph.14624

Brinton, J., and C. Uauy. 2018. A reductionist approach to dissecting grain weight and yield in wheat. J. Integr. Plant Biol.. doi:10.1111/jipb.12741

Broman, K.W., H. Wu, ´ Saunak Sen, and G.A. Churchill. 2003. R/qtl: QTL mapping in experimental crosses. Bioinforma. Appl. NOTE 19:889–890. doi:10.1093/bioinformatics/btg112

Buttar, Z.A., Y. Yang, R. Sharif, S. Nan Wu, Y. Xie, and C. Wang. 2020. Genome Wide Identification, Characterization, and Expression Analysis of YABBY-Gene Family in WHEAT (Triticum aestivum L.). Agronomy 10:1189. doi:10.3390/agronomy10081189

Capelli, N., T. Diogon, H. Greppin, and P. Simon. 1997. Isolation and characterization of a cDNA clone encoding an osmotin-like protein from Arabidopsis thaliana. Gene 191:51–56. doi:10.1016/S0378-1119(97)00029-2

Chawla, H.S., H. Lee, I. Gabur, P. Vollrath, S. Tamilselvan-Nattar-Amutha, C. Obermeier, S. V. Schiessl, J. Song, K. Liu, L. Guo, I.A.P. Parkin, and R.J. Snowdon. 2021. Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. Plant Biotechnol. J. 19:240–250. doi:10.1111/pbi.13456

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. doi:10.1093/bioinformatics/bts635

Doyle, J., and J. Doyle. 1990. Isolation of plant DNA from fresh tissue.. Focus (Madison). 12:13–15

FAOSTAT. (2021). *FAOSTAT*. http://www.fao.org/faostat/en/#data (accessed January 17, 2020).

Fox, J., and W. Sanford. 2019. An R Companion to Applied Regression. Third. Sage, Thousand Oaks, CA.

Gabur, I., H.S. Chawla, R.J. Snowdon, and I.A.P. Parkin. 2019. Connecting genome structural variation with complex traits in crop plants. Theor. Appl. Genet. 132:733–750. doi:10.1007/s00122-018-3233-0

Gao, Y., A.A. Badejo, H. Shibata, Y. Sawa, T. Maruta, S. Shigeoka, M. Page, N. Smirnoff, and T. Ishikawa. 2011. Expression analysis of the VTC2 and VTC5 genes encoding GDP-L-galactose phosphorylase, an enzyme involved in ascorbate biosynthesis, in Arabidopsis thaliana. Biosci. Biotechnol. Biochem.. doi:10.1271/bbb.110320

Gardiner, L.J., T. Brabbs, A. Akhunov, K. Jordan, H. Budak, T. Richmond, S. Singh, L. Catchpole, E. Akhunov, and A. Hall. 2019. Integrating genomic resources to present full gene and putative promoter capture probe sets for bread wheat. Gigascience. doi:10.1093/gigascience/giz018

Gegas, V.C., A. Nazari, S. Griffiths, J. Simmonds, L. Fish, S. Orford, L. Sayers, J.H. Doonan, and J.W. Snape. 2010. A Genetic Framework for Grain Size and Shape Variation in Wheat. Plant Cell 22:1046–1056. doi:10.1105/tpc.110.074153

Gill, T.R.E.& B.S. 1996. The Deletion Stocks of Common Wheat. J. Hered. 295–307. doi:10.1093/oxfordjournals.jhered.a023003

Gutierrez-Gonzalez, J.J., M. Mascher, J. Poland, and G.J. Muehlbauer. 2019. Dense genotyping-by-sequencing linkage maps of two Synthetic W7984×Opata reference populations provide insights into wheat structural diversity. Sci. Rep. 9:1793. doi:10.1038/s41598-018-38111-3

He, F., R. Pasam, F. Shi, S. Kant, G. Keeble-Gagnere, P. Kay, K. Forrest, A. Fritz, P. Hucl, K. Wiebe, R. Knox, R. Cuthbert, C. Pozniak, A. Akhunova, P.L. Morrell, J.P. Davies, S.R. Webb, G. Spangenberg, B. Hayes, H. Daetwyler, J. Tibbits, M. Hayden, and E. Akhunov. 2019a. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nat. Genet. 51:896–904. doi:10.1038/s41588-019-0382-2

He, Y., L. Yan, C. Ge, X.-F. Yao, X. Han, R. Wang, L. Xiong, L. Jiang, C.-M. Liu, and Y. Zhao. 2019b. PINOID Is Required for Formation of the Stigma and Style in Rice. Plant Physiol.. doi:10.1104/pp.18.01389

He, Z., L. Wang, A.L. Harper, L. Havlickova, A.K. Pradhan, I.A.P. Parkin, and I. Bancroft. 2017. Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotechnol. J. 15:594–604. doi:10.1111/pbi.12657

Henry, L., H. Wickham, and Y. Collet. 2020. rlang: Functions for Base Types and Core R and “Tidyverse” Features

Höller, S., Y. Ueda, L. Wu, Y. Wang, M.R. Hajirezaei, M.R. Ghaffari, N. von Wirén, and M. Frei. 2015. Ascorbate biosynthesis and its involvement in stress tolerance and plant development in rice (Oryza sativa L.). Plant Mol. Biol.. doi:10.1007/s11103-015-0341-y

Hotta, T., Z. Kong, C.M.K. Ho, C.J.T. Zeng, T. Horio, S. Fong, T. Vuong, Y.R.J. Lee, and B. Liu. 2012. Characterization of the Arabidopsis augmin complex uncovers its critical function in the assembly of the acentrosomal spindle and phragmoplast microtubule arrays.. Plant Cell. doi:10.1105/tpc.112.096610

Huber, W., A. von Heydebreck, H. Sueltmann, A. Poustka, and M. Vingron. 2002. Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression. Bioinformatics S96–S104

Iannone, R., J. Cheng, and B. Schloerke. 2020. gt: Easily Create Presentation-Ready Display Tables

International Wheat Genome Sequencing Consortium (IWGSC). 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome.. Science 361:eaar7191. doi:10.1126/science.aar7191

Jia, X., X. Si, Y. Jia, H. Zhang, S. Tian, W. Li, K. Zhang, and Y. Pan. 2020. Genomic proling and expression analysis of the diacylglycerol kinase gene family in heterologous hexaploid wheat. doi:10.21203/rs.3.rs-49275/v1

Jiang, W.-B., H.-Y. Huang, Y.-W. Hu, S.-W. Zhu, Z.-Y. Wang, and W.-H. Lin. 2013. Brassinosteroid Regulates Seed Size and Shape in Arabidopsis 1[W][OPEN]. Plant Physiol. Ò 162:1965–1977. doi:10.1104/pp.113.217703

Kassambara, A. 2020a. rstatix: Pipe-Friendly Framework for Basic Statistical Tests

Kassambara, A. 2020b. ggpubr: “ggplot2” Based Publication Ready Plots

Kato, K., H. Miura, and S. Sawada. 2000. Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat. Theor. Appl. Genet. 101:1114–1121. doi:10.1007/s001220051587

Kim, E.J., S.H. Lee, C.H. Park, and T.W. Kim. 2018. Functional Role of BSL1 Subcellular Localization in Brassinosteroid Signaling. J. Plant Biol. 61:40–49. doi:10.1007/s12374-017-0363-x

Kolde, R. 2019. pheatmap: Pretty Heatmaps

Krasileva, K. V, H.A. Vasquez-Gross, T. Howell, P. Bailey, F. Paraiso, L. Clissold, J. Simmonds, R.H. Ramirez-Gonzalez, X. Wang, P. Borrill, C. Fosker, S. Ayling, A.L. Phillips, C. Uauy, and J. Dubcovsky. 2017. Uncovering hidden variation in polyploid wheat. Proc. Natl. Acad. Sci.. doi:10.1073/pnas.1619268114

Kuzay, S., Y. Xu, J. Zhang, A. Katz, S. Pearce, Z. Su, M. Fraser, J.A. Anderson, G. Brown-Guedira, N. DeWitt, A. Peters Haugrud, J.D. Faris, E. Akhunov, G. Bai, and J. Dubcovsky. 2019. Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. Theor. Appl. Genet. 1–17. doi:10.1007/s00122-019-03382-5

Leek, J., W. Johnson, H. Parker, E. Fertig, A. Jaffe, Y. Zhang, J. Storey, and L. Torres. 2020. sva: Surrogate Variable Analysis

Lenth, R., H. Singmann, J. Love, P. Buerkner, and M. Herve. 2019. Package “emmeans.” doi:10.1080/00031305.1980.10483031

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352

Li, N., and Y. Li. 2016. Signaling pathways of seed size control in plants This review comes from a themed issue on Cell signalling and gene regulation. Curr. Opin. Plant Biol. 33:23–32. doi:10.1016/j.pbi.2016.05.008

Li, Y., X. Fu, M. Zhao, W. Zhang, B. Li, D. An, J. Li, A. Zhang, R. Liu, and X. Liu. 2018. A Genome-wide View of Transcriptome Dynamics During Early Spike Development in Bread Wheat. Sci. Rep.. doi:10.1038/s41598-018-33718-y

Lim, B., N. Smirnoff, C.S. Cobbett, and J.F. Golz. 2016. Ascorbate-deficient VTC2 mutants in arabidopsis do not exhibit decreased growth. Front. Plant Sci.. doi:10.3389/fpls.2016.01025

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

Luan, Z., and D. Zhou. 2015. Screening of rice (Oryza sativa L.) OsPR1b-interacting factors and their roles in resisting bacterial blight. Genet. Mol. Res. 14:1868–1874. doi:10.4238/2015.March.13.15

Luo, M., E.S. Dennis, F. Berger, W.J. Peacock, and A. Chaudhury. 2005. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis

Mahmoud, M., N. Gobet, D.I. Cruz-Dávalos, N. Mounier, C. Dessimoz, and F.J. Sedlazeck. 2019. Structural variant calling: The long and the short of it. Genome Biol. 20:246. doi:10.1186/s13059-019-1828-7

Makhoul, M., C. Rambla, K.P. Voss-Fels, L.T. Hickey, R.J. Snowdon, and C. Obermeier. 2020. Overcoming polyploidy pitfalls: a user guide for effective SNP conversion into KASP markers in wheat. Theor. Appl. Genet. 1–18. doi:10.1007/s00122-020-03608-x

McSteen, P., S. Malcomber, A. Skirpan, C. Lunde, X. Wu, E. Kellogg, and S. Hake. 2007. barren inflorescence2 encodes a co-ortholog of the Pinoid serine/threonine kinase and is required for organogenesis during inflorescence and vegetative development in maize. Plant Physiol. 144:1000–1011. doi:10.1104/pp.107.098558

Nagpal, P., C.M. Ellis, H. Weber, S.E. Ploense, L.S. Barkawi, T.J. Guilfoyle, G. Hagen, J.M. Alonso, J.D. Cohen, E.E. Farmer, J.R. Ecker, and J.W. Reed. 2005. Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. Development. doi:10.1242/dev.01955

NCBI. LOC4352571 Osmotin-like Protein [Oryza Sativa Japonica Group (Japanese Rice)] - Gene. https://www.ncbi.nlm.nih.gov/gene/4352571 (accessed November 26, 2020).

Oh, S.A., J. Jeon, H.J. Park, P.E. Grini, D. Twell, and S.K. Park. 2016. Analysis of gemini pollen 3 mutant suggests a broad function of AUGMIN in microtubule organization during sexual reproduction in Arabidopsis. Plant J.. doi:10.1111/tpj.13192

Platten, J.D., J.N. Cobb, and R.E. Zantua. 2019. Criteria for evaluating molecular markers: Comprehensive quality metrics to improve marker-assisted selection. PLoS One 14:e0210529. doi:10.1371/journal.pone.0210529

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PLoS One 7:e32253. doi:10.1371/journal.pone.0032253

R Core Team. 2020. R: A Language and Environment for Statistical Computing

Saxena, R.K., D. Edwards, and R.K. Varshney. 2014. Structural variations in plant genomes. Briefings Funct. Genomics Proteomics 13:296. doi:10.1093/bfgp/elu016

Siegfried, K.R. 1999. Members of the YABBY genes specify abaxial cell fate in Arabidopsis

Slafer, G.A. 2003. Genetic basis of yield as viewed from a crop physiologist's perspective. Ann. Appl. Biol. 142:117–128. doi:10.1111/j.1744-7348.2003.tb00237.x

Song, K., P. Lu, K. Tang, and T.C. Osborn. 1995. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc. Natl. Acad. Sci. U. S. A. 92:7719–7723. doi:10.1073/pnas.92.17.7719

Sorrells, M.E., J.P. Gustafson, D. Somers, S. Chao, D. Benscher, G. Guedira-Brown, E. Huttner, A. Kilian, P.E. McGuire, K. Ross, J. Tanaka, P. Wenzl, K. Williams, C.O. Qualset, and A. Van Deynze. 2011. Reconstruction of the Synthetic W7984 × Opata M85 wheat reference population. Genome 54:875–882. doi:10.1139/g11-054

Stahle, M.I., J. Kuehlich, L. Staron, A.G. Von Arnim, and J.F. Golz. 2009. YABBYs and the transcriptional corepressors LEUNIG and LEUNIG-HOMOLOG maintain leaf polarity and meristem activity in Arabidopsis. Plant Cell. doi:10.1105/tpc.109.070458

Steiner, B., M. Buerstmayr, C. Wagner, A. Danler, B. Eshonkulov, M. Ehn, and H. Buerstmayr. 2019. Fine-mapping of the Fusarium head blight resistance QTL Qfhs.ifa-5A identifies two resistance QTL associated with anther extrusion. Theor. Appl. Genet. 132:2039–2053. doi:10.1007/s00122-019-03336-x

Sukumaran, S., M. Lopes, S. Dreisigacker, and M. Reynolds. 2018. Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for grain weight and grain number. Theor. Appl. Genet. 131:985–998. doi:10.1007/s00122-017-3037-7

Tuinstra, M.R., G. Ejeta, and P.B. Goldsbrough. 1997. Heterogeneous inbred family (HIF) analysis: A method for developing near-isogenic lines that differ at quantitative trait loci. Theor. Appl. Genet. 95:1005–1011. doi:10.1007/s001220050654

Walkowiak, S., L. Gao, C. Monat, G. Haberer, M.T. Kassa, J. Brinton, R.H. Ramirez-Gonzalez, M.C. Kolodziej, E. Delorean, D. Thambugala, V. Klymiuk, B. Byrns, H. Gundlach, V. Bandi, J.N. Siri, K. Nilsen, C. Aquino, A. Himmelbach, D. Copetti, T. Ban, L. Venturini, M. Bevan, B. Clavijo, D.-H. Koo, J. Ens, K. Wiebe, A. N'Diaye, A.K. Fritz, C. Gutwin, A. Fiebig, C. Fosker, B.X. Fu, G.G. Accinelli, K.A. Gardner, N. Fradgley, J. Gutierrez-Gonzalez, G. Halstead-Nussloch, M. Hatakeyama, C.S. Koh, J. Deek, A.C. Costamagna, P. Fobert, D. Heavens, H. Kanamori, K. Kawaura, F. Kobayashi, K. Krasileva, T. Kuo, N. McKenzie, K. Murata, Y. Nabeka, T. Paape, S. Padmarasu, L. Percival-Alwyn, S. Kagale, U. Scholz, J. Sese, P. Juliana, R. Singh, R. Shimizu-Inatsugi, D. Swarbreck, J. Cockram, H. Budak, T. Tameshige, T. Tanaka, H. Tsuji, J. Wright, J. Wu, B. Steuernagel, I. Small, S. Cloutier, G. Keeble-Gagnère, G. Muehlbauer, J. Tibbets, S. Nasuda, J. Melonek, P.J. Hucl, A.G. Sharpe, M. Clark, E. Legg, A. Bharti, P. Langridge, A. Hall, C. Uauy, M. Mascher, S.G. Krattinger, H. Handa, K.K. Shimizu, A. Distelfeld, K. Chalmers, B. Keller, K.F.X. Mayer, J. Poland, N. Stein, C.A. McCartney, M. Spannagl, T. Wicker, and C.J. Pozniak. 2020. Multiple wheat genomes reveal global variation in modern breeding. Nat. 2020 1–7. doi:10.1038/s41586-020-2961-x

Wan, C.Y., and T.A. Wilkins. 1994. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (Gossypium hirsutum L.). Anal. Biochem. 223:7–12. doi:10.1006/abio.1994.1538

Wang, Y., H. Yu, C. Tian, M. Sajjad, C. Gao, Y. Tong, X. Wang, and Y. Jiao. 2017. Transcriptome association identifies regulators of wheat spike architecture. Plant Physiol. 175:746–757. doi:10.1104/pp.17.00694

Wickham, H. 2018. reshape: Flexibly Reshape Data

Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. 2019. Welcome to the Tidyverse. J. Open Source Softw. 4:1686. doi:10.21105/joss.01686

Williams, K., M.E. Sorrells, and C. Univ. 2014. Three-Dimensional Seed Size and Shape QTL in Hexaploid Wheat (Triticum aestivum L.) Populations. www.crops.org Crop Sci. 54:98–110. doi:10.2135/cropsci2012.10.0609

Winfield, M.O., A.M. Allen, A.J. Burridge, G.L.A. Barker, H.R. Benbow, P.A. Wilkinson, J. Coghill, C. Waterfall, A. Davassi, G. Scopes, A. Pirani, T. Webster, F. Brew, C. Bloor, J. King, C. West, S. Griffiths, I. King, A.R. Bentley, and K.J. Edwards. 2016. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnol. J. 14:1195–1206. doi:10.1111/pbi.12485

Xie, Y. 2020. knitr: A General-Purpose Package for Dynamic Report Generation in R

Xu, L., D. Wang, S. Liu, Z. Fang, S. Su, C. Guo, C. Zhao, and Y. Tang. 2020. Comprehensive Atlas of Wheat (Triticum aestivum L.) AUXIN RESPONSE FACTOR Expression During Male Reproductive Development and Abiotic Stress. Front. Plant Sci.. doi:10.3389/fpls.2020.586144

Yan, L., A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima, J. Dubcovsky, and S.D. Tanksley. 2003. Positional cloning of the wheat vernalization gene VRN1

Yang, J., Y. Zhou, Q. Wu, Y. Chen, P. Zhang, · Yu'e Zhang, W. Hu, · Xicheng Wang, H. Zhao, L. Dong, J. Han, Z. Liu, and · Tingjie Cao. 2019. Molecular characterization of a novel TaGL3-5A allele and its association with grain length in wheat (Triticum aestivum L.) 132:1799–1814. doi:10.1007/s00122-019-03316-1

Yao, W., G. Li, Y. Yu, and Y. Ouyang. 2018. funRiceGenes dataset for comprehensive understanding and application of rice functional genes. Gigascience 7:1–9. doi:10.1093/gigascience/gix119

Yin, T., D. Cook, and M. Lawrence. 2012. ggbio: an R package for extending the grammar of graphics for genomic data. Genome Biol.. doi:10.1186/gb-2012-13-8-r77

Yin, T., M. Lawrence, and D. Cook. 2020. biovizBase: Basic graphic utilities for visualization of genomic data

Yin, X.J., S. Volk, K. Ljung, N. Mehlmer, K. Dolezal, F. Ditengou, S. Hanano, S.J. Davis, E. Schmelzer, G. Sandberg, M. Teige, K. Palme, C. Pickart, and A. Bachmair. 2007. Ubiquitin lysine 63 chain-forming ligases regulate apical dominance in Arabidopsis. Plant Cell. doi:10.1105/tpc.107.052035

Young, M.D., M.J. Wakefield, G.K. Smyth, and A. Oshlack. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 11:R14. doi:10.1186/gb-2010-11-2-r14

Yutani, H. 2020. gghighlight: Highlight Lines and Points in "ggplot2"

Zhang, A., N. Li, L. Gong, X. Gou, B. Wang, X. Deng, C. Li, Q. Dong, H. Zhang, and B. Liu. 2020. Global Analysis of Gene Expression in Response to Whole-Chromosome Aneuploidy in Hexaploid Wheat 1[OPEN]. doi:10.1104/pp.17.00819

Zhang, J., · Shiferaw, A. Gizaw, · Eligio Bossolini, J. Hegarty, · Tyson Howell, A.H. Carter, · Eduard Akhunov, and J. Dubcovsky. 2018a. Identification and validation of QTL for grain yield and plant water status under contrasting water treatments in fall-sown spring wheats 131:1741–1759. doi:10.1007/s00122-018-3111-9

Zhang, R., S. Geng, Z. Qin, Z. Tang, C. Liu, D. Liu, G. Song, Y. Li, S. Zhang, W. Li, J. Gao, X. Han, and G. Li. 2019. The genome-wide transcriptional consequences of the nullisomic-tetrasomic stocks for homoeologous group 7 in bread wheat. BMC Genomics 20:1–17. doi:10.1186/s12864-018-5421-3

Zhang, X., J. Wang, J. Huang, H. Lan, C. Wang, C. Yin, Y. Wu, H. Tang, Q. Qian, J. Li, and H. Zhang. 2012. Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. Proc. Natl. Acad. Sci. U. S. A. 109:21534–21539. doi:10.1073/pnas.1219776110

Zhang, Z., J. Li, Z. Tang, X. Sun, H. Zhang, J. Yu, G. Yao, G. Li, H. Guo, J. Li, H. Wu, H. Huang, Y. Xu, Z. Yin, Y. Qi, R. Huang, W. Yang, and Z. Li. 2018b. Gnp4/LAX2, a RAWUL protein, interferes with the OsIAA3–OsARF25 interaction to regulate grain length via the auxin signaling pathway in rice. J. Exp. Bot.. doi:10.1093/jxb/ery256

Zhu, H. 2020. kableExtra: Construct Complex Table with “kable” and Pipe Syntax

CHAPTER 3

COUNTING ON CROSSOVERS: CONTROLLED RECOMBINATION FOR PLANT BREEDING [2]

Abstract

Crossovers, the genetic exchange between homologous chromosomes, are strongly biased toward sub-telomeric regions in plant species. Manipulating the rate and positions of crossovers to increase the genetic variation accessible to breeders is a longstanding goal. Use of genome editing reagents that induce double-stranded breaks or modify the epigenome at desired sites of recombination, and manipulation of crossover factors, are increasingly applicable approaches for achieving this goal. These strategies for ‘controlled recombination’ have the potential to reduce the time and expense associated with traditional breeding, reveal currently inaccessible genetic diversity and increase control over inheritance of preferred haplotypes. Considerable challenges to address include translating knowledge from models to crop species and determining best stages of the breeding cycle to control recombination.

[2] Originally published as Taagen, E., Bogdanove, A. J. & Sorrells, M. E. (2020). Counting on Crossovers: Controlled Recombination for Plant Breeding. *Trends in Science*. https://doi.org/10.1016/j.tplants.2019.12.017

Ella Taagen contributions: Conceptualization; Investigation; Writing-original draft; Writing-review & editing.

Counting on Crossovers

Plant breeders rely on natural recombination of genetic information during crossovers (COs) to generate novel and favorable haplotypes. A minimum of one CO for each chromosome pair, termed the obligate CO, is required for proper segregation, and rarely three COs are exceeded per meiosis (Chelysheva et al., 2010; Mercier et al., 2015). In addition to their frequency being tightly regulated, distribution of COs is uneven along chromosomes, regardless of the genome size (Higgins et al., 2014; Darrier et al., 2017; Wang & Copenhaver, 2018). The frequency and position of COs govern a breeder's ability to disrupt linkage drag, that is, to incorporate novel genetic variants without simultaneously introgressing large segments from a donor chromosome. The low-frequency and patterning of COs has traditionally required breeders to work with large populations over many generations to capture desirable haplotypes. Manipulation of pro and anti-CO factors, and use of site-directed nucleases and epigenetic modifiers are a novel and increasingly applicable approaches for manipulating recombination or CO frequency and distribution (Hayut et al., 2017; Corem et al., 2018; Fernandes et al., 2018b; Mieulet et al., 2018; Serra et al., 2018; Underwood et al., 2018). We refer to the use of these approaches collectively as 'controlled recombination'. Controlled recombination has the potential to reduce cost and time for high resolution mapping to identify genes of interest. Likewise, it may facilitate reintroduction of genetic variance at sites of selective sweeps and introgression of diverse alleles from wild crop relatives for varietal development (Tam et al., 2011; Presting, 2018; Rey et al., 2018). In this opinion piece, we review the mechanisms underlying recombination and CO patterns in plants in the context of how they can or might be manipulated, and discuss how controlled recombination could be directly applied to crop breeding programs. Important questions and technical hurdles that need to be addressed before the full potential of controlled recombination of plant breeding can be realized are also presented.

DNA double-stranded breaks and regulation of COs in plants

DNA double-stranded breaks (DSBs) are the precursor to a reciprocal genetic exchange between homologous chromosomes. During prophase I of meiosis, hundreds of DSBs occur along the chromosomes, catalyzed by the evolutionarily conserved SPORULATION-DEFICIENT11 (SPO11) protein and several other associated proteins (Keeney et al., 1997; Wang & Copenhaver, 2018). Homologous recombination (HR) mediated repair of a DSB can result in COs or non-crossovers (NCOs) (Wang & Copenhaver, 2018). The two recombinases RAD51 and Disrupted Meiotic cDNA1 (DMC1) assist the 3' single-strand DNA ends of DSBs to invade either the intact sister chromatid or homologous chromosome for a repair template (Lao et al., 2013; Brown & Bishop, 2014; Singh et al., 2017; Wang & Copenhaver, 2018). This invasion forms a D-loop intermediate and is primed for DNA synthesis using the complementary strand of the invaded chromatid as a template. At this stage the predominant repair mechanism in plants is for the extended invading strand to disassociate and re-anneal to the other end of the original DSB, called synthesis dependent strand annealing (SDSA), which results in NCOs (Wang & Copenhaver, 2018). Alternatively, DNA synthesis proceeds, then second-end capture and double Holliday junction (dHJ) intermediates facilitate reciprocal exchange of DNA between homologous chromosomes, forming a CO (Figure 3.1) (Wang & Copenhaver, 2018). There are two independent pathways that promote CO formation in plants: the major Class I pathway, which is sensitive to interference, and the minor Class II pathway, which is insensitive (Copenhaver et al., 2002; De Los Santos et al., 2003; Wang & Copenhaver, 2018). In addition, there are at least three independent pathways that suppress COs (Table 3.1). Together, pro and anti-CO pathways in plants usually result in one to three crossovers per chromosome per meiosis (Mercier et al., 2015). For a comprehensive review of meiotic recombination in plants see (Wang & Copenhaver, 2018).

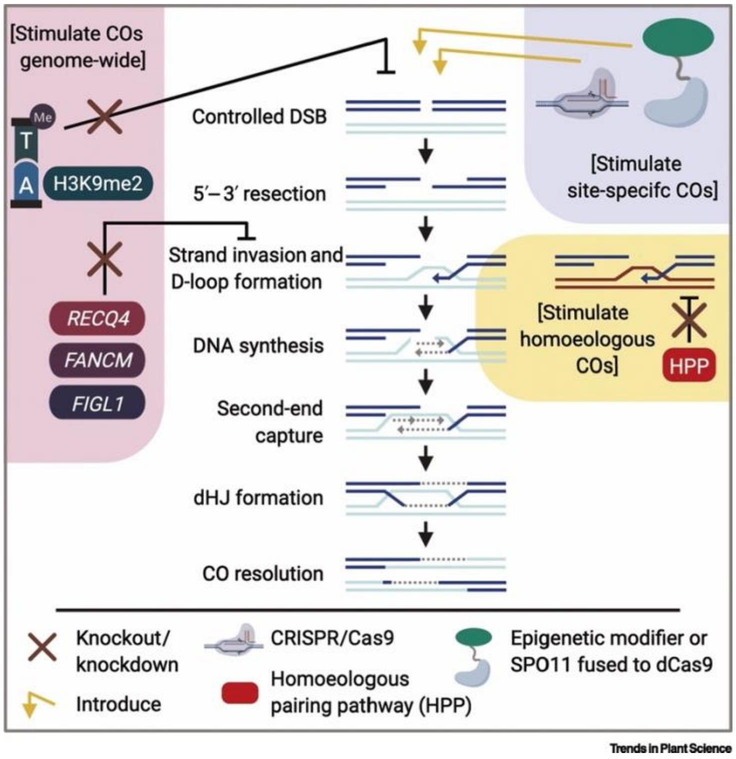
[Stimulate COs genome-wide]
Me
T
A
H3K9me2
RECQ4
FANCM
FIGL1
Controlled DSB
5′–3′ resection
Strand invasion and D-loop formation
DNA synthesis
Second-end capture
dHJ formation
CO resolution
[Stimulate site-specifc COs]
[Stimulate homoeologous COs]
HPP
Knockout/ knockdown
Introduce
CRISPR/Cas9
Homoeologous pairing pathway (HPP)
Epigenetic modifier or SPO11 fused to dCas9
Trends in Plant Science

Figure 3.1: Given a region of heterozygous homologous chromosomes there are various tools and approaches that can be applied to stimulate COs during meiotic recombination. *Arabidopsis* plants with loss of non-CG methylation and H3K9me2 are known to have increased DSBs and COs genome-wide, and notably within pericentromeric regions (Underwood et al., 2018). Disruption of any of the three

characterized anti-CO factors, *RECQ4, FANCM,* and *FIGL1,* through targeted knockouts or knockdowns offers ways to increase COs 2-3 fold in plants (Mieulet et al., 2018). Using a site-directed nuclease such as Cas9, or fusions of epigenetic modifiers or proteins like SPO11 to site-specific DNA binding reagents (e.g. dCas9; see text), allows for precisely targeted stimulation of recombination and increased rate of COs (Sarno et al., 2017). The CRISPR-Cas9 system itself is also an attractive method for programmed positions of DSBs via the Cas9 nuclease (Sadhu et al., 2016; Hayut et al., 2017). Mutating genes involved in homoeologous pairing pathways has been proposed as a method for encouraging COs between more distant crop relatives (Tam et al., 2011; Rey et al., 2018). For a comprehensive review of plant meiotic recombination see (Wang & Copenhaver, 2018). This figure was created using BioRender (https://biorender.com/).

Factor	Interactions	Anti-CO mechanism	References
RECQ4	Forms the BTR complex with TOP3α & RMI1	Unwinds D-loops to suppress Class I COs	(Lande & Thompson, 1990; Hartung & Puchta, 2006; Hartung et al., 2007; Fasching et al., 2015; Séguéla-Arnaud et al., 2015; Fernandes et al., 2018b; Mieulet et al., 2018; De Maagd et al., 2019)
FANCM	Direct DNA-binding cofactor with MHF1 & MHF2	May unwind D-loops, limits Class II COs & promotes SDSA	(Crismani et al., 2012; Girard et al., 2014, 2015; Fernandes et al., 2018b)
FIDGETIN-LIKE-1 (FIGL1)	Forms complex with FIDGETIN-LIKE-1 INTERACTING PROTEIN (FLIP), suppress RAD51 & DMC1	Constrains strand invasion	(Girard et al., 2015; Fernandes et al., 2018a; b; Mieulet et al., 2018)

Table 3.1: Conserved anti-CO factors, studied in *Arabidopsis,* rice, pea & tomato

Genome editing tools for controlled recombination

Given that COs determine the amount of genetic diversity accessible to breeders, manipulation of both meiotic and somatic recombination frequency and patterning is a longstanding research focus. The focus of many plant and animal breeders is quickly turning to precision genome editing tools such as zinc-finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN) and the

CRISPR-Cas system (Klug, 2010; Bogdanove & Voytas, 2011; Jiang & Marraffini, 2015), or variants of these fused to base editors, epigenetic modifiers or other proteins, to manipulate recombination frequency and patterns through DSB mutations (Sadhu et al., 2016; Hayut et al., 2017; Sarno et al., 2017; Mieulet et al., 2018). In addition to their high specificity, these reagents can be designed (e.g., TALENs) or multiplexed (e.g., CRISPER-Cas) to simultaneously target multiple copies of a gene, which is particularly useful for breeding polyploid crops with homoeologous gene copies. The growing collection of tools can be used to target DSBs or epigenetic modifications to sites of desired recombination or to perturb pro or ant-CO pathway genes (Figure 3.1). Simulations of controlled COs facilitated by such tools for inbred and hybrid crops as well as livestock predict doubling of genetic gain in a single breeding cycle (Bernardo, 2017; Gonen et al., 2017; Brandariz & Bernardo, 2019; Ru & Bernardo, 2019). A seminal study of controlled recombination using CRISPR-Cas based editing has been conducted in yeast *(Saccharomyces cerevisiae* Hansen*),* and included the development of high resolution genetic mapping populations from a single mitosis (Sadhu et al., 2016). Many DNA repair proteins are conserved across eukaryotes, yet pathway preference and efficiency differ between yeast and plants. To date, successful plant breeding applications of controlled recombination using genome editing tools have been directed at tomato fruit color quality traits via germinally transmitted targeted somatic HR (Hayut et al., 2017). Thus, while the yeast and initial plant studies suggest exciting potential for controlled recombination in plant breeding using genome editing tools, the extent to which this potential will be fully realized is yet to be determined, and more research is warranted.

Manipulation of CO frequency

A particularly attractive method of increasing the frequency of COs is to knock out CO suppression genes, several of which are listed in Table 3.1. In rice (*Oryza sativa* L.), pea (*Pisum sativum* L.) and tomato (*Solanum lycopersicum* L.), knockouts of orthologs of three anti-CO pathway associated genes, originally characterized in *Arabidopsis*, generated by CRISPR-Cas9 or by TILLING dramatically increased the recombination and CO rate (Mieulet et al., 2018). For example, mutation of *recq4* orthologs increased COs nearly three-fold compared to WT among these species. Notably however, homozygous knockout of *FIGL1* in tomato and pea caused sterility, highlighting the sometimes challenging nature of translating findings from model to crop species (Mieulet et al., 2018). Some anti-CO genes may have pleiotropic effects. For example, *MEICA1*, a recently identified anti-CO factor in rice, was found to be essential for normal meiotic recombination (Hu et al., 2017). Though *meica1* rice plants showed increased frequency of COs, they also exhibited nonhomologous chromosome associations, chromosome fragmentations, and high rates of sterility, which limit practical breeding applications (Hu et al., 2017).

RECQ4, FANCM and FIGL1 regulate HR by different mechanisms and their anti-CO effects have been observed to be synergistic in plants (Table 3.1, Figure 3.1). A comparison study disrupting one, two or all three pathways in pure lines and hybrid *Arabidopsis* plants reported the greatest increased CO effect with combined *figl1* and *req4* mutations (Fernandes et al., 2018b). *Arabidopsis* plants with *figl1* and *flip* mutations exhibited greater meiotic recombination and a substantially greater number of RAD51and DMC1 foci on chromosomes compared to WT (Fernandes et al., 2018a). A double mutant of FIGL1 and FANCM showed a synergistic increase in CO formation compared to the single mutants and WT, suggesting that FIGL1 acts in concert or sequentially with FANCM to limit COs (Girard et al., 2015). Further supporting independent mechanisms of CO regulation by FIGL1 and FANCM, *figl1* increased CO frequency in inbred lines and F1 hybrids, while *fancm* led to significant CO increase only in inbred lines, though the synergistic effect was observed in both inbreds and F1 hybrids (Girard et al., 2015). Additionally, Class I CO frequency can be manipulated by over expressing pro-crossover E3 ligase gene *HEI10* (Ziolkowski et al., 2017). In *Arabidopsis*, increasing copies of *HEI10* and knocking out *RECQ4A* and *RECQ4B* additively leads to a 5-fold and 1.5-fold increase in meiotic recombination in the chromosome arms and pericentromeric heterochromatin, respectively (Serra et al., 2018).

In a breeding program context, manipulating expression of pro and anti-CO genes may be an efficient method for increasing CO frequency during pre-breeding (Figure 3.1 & 3.2A). Beyond the pre-breeding stage though an increased CO frequency could be problematic for elite line development. Suppression of the elevated CO may be necessary, for example via doubled-haploids or as demonstrated in reverse breeding (Figure 3.2B) (Wijnker et al., 2012).

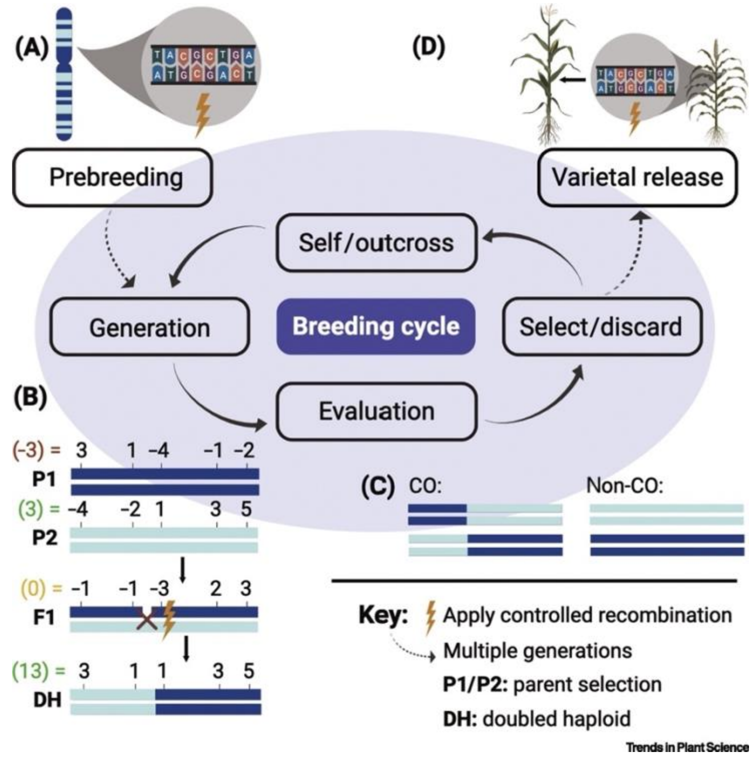

Figure 3.2: Plant breeding involves recurrent cycles of plant generation, evaluation, and selection for genetic improvement resulting in an eventual varietal release. Throughout the breeding cycle, applying controlled recombination tools such as CRISPR-Cas9 and SPO11-dCas9 for induced DSB or knockouts of H3K9me2 and non-CG DNA methylation, homoeologous pairing pathways, or anti-CO genes, has the potential to improve genetic gains and reduce the time and cost associated with a traditional breeding program (Sadhu et al., 2016; Hayut et al., 2017; Sarno et al., 2017; Mieulet et al., 2018; Rey et al., 2018; Underwood et al., 2018; Xue et al., 2018). Controlled recombination may accelerate the identification of genes and beneficial haplotypes and can be used to improve existing elite cultivars**.** **(A)** Manipulating recombination may improve pre-breeding by reducing the number of individuals and generations necessary for high resolution mapping and identification of genes of interest, especially in wild relatives of a crop. **(B)** Parent selection is a critical decision for breeders, and controlled recombination may facilitate optimal recombination of linkage groups into a single haplotype. For inbred and hybrid crops, it may be advantageous to suppress recombination after desired haplotypes are achieved through doubled-haploids or reverse breeding (Wijnker et al., 2012). **(C)** Some controlled recombination methods increase the frequency of DSBs whereas others silence CO inhibitors, and the CO efficiency will vary. The breeder can select true COs with the optimal combination of marker effects or null segregants. **(D)**. Beneficial causal variants identified in controlled recombination experiments might be rapidly and precisely introduced with genome editing reagents, such as CRISPR-Cas9, to improve

existing elite cultivars. This figure was created using BioRender (https://biorender.com/).

Controlled somatic recombination

Repair of spontaneous DSBs continue throughout the plant life cycle in somatic cells. Although non-homologous end joining (NHEJ) is the predominant repair pathway in somatic cells, controlled DSBs can induce HR between homologous chromosomes (Hayut et al., 2017) or between repeats (Puchta, 1999). Induced DSBs and recombination in somatic cells should be explored for direct application by breeders (Sadhu et al., 2016; Hayut et al., 2017).

Fine-mapping of causal genetic variants is limited by COs that disrupt linked markers, which has traditionally required thousands of gametes and multiple generations to achieve (Zhu et al., 2008). In the yeast study cited earlier, mapping panels were built by stimulating HR during mitosis using CRISPR-Cas9 induced DSBs (Sadhu et al., 2016). Three gRNAs were designed to target heterozygous sites across a 2.9 kb manganese-sensitivity QTL on chromosome 7 in biparental diploid cells. During a single mitosis, CRISPR-Cas9 induced DSBs resulted in sufficient COs and loss-of-heterozygosity to identify the causal variant for manganese-sensitivity among 358 lines. The authors note that producing an equivalent number of COs across this locus by random meiotic segregation would have required more than 7,500 lines (Sadhu et al., 2016). In the tomato fruit color study, a fruit color assay was used to measure NHEJ versus HR events from CRISPR-Cas9 targeted somatic cell DSBs in the *PHYTOENE SYNTHASE* (*PSY1*) gene, which is implicated in carotenoid accumulation (Hayut et al., 2017). This method achieved 14% greater HR at DSBs compared to WT (Hayut et al., 2017). Together, the two studies underscore the fact that allele-specific DSBs and HR can occur outside of meiotic COs and that recombination in somatic cells could be explored in a breeding context. More work must be done to determine how broadly applicable this approach could be to different crop species.

Fine tuning introgressions of wild germplasm

Introgression of genetic material from wild crop relatives has been a valuable approach to capture allelic diversity for breeding programs, especially for disease resistance (Badaeva et al., 2007; Faris et al., 2008; Ren et al., 2017; Mammadov et al., 2018). Yet, because COs are suppressed between divergent sequences, a high degree of linkage-drag associated with such introgression has limited its practical application (Mammadov et al., 2018). However, manipulation of DNA mismatch repair and homologous chromosome pairing mechanisms can increase COs during meiosis in hybrids of elite and wild germplasm (Tam et al., 2011; Rey et al., 2018). A nearly 18% increase in CO frequency for an introgression from wild tomato (*Solanum lycopersicoides* L) into a cultivated tomato (*Solanum lycopersicum* L) was facilitated by silencing the DNA mismatch repair system that suppresses homoeologous recombination (Tam et al., 2011). Silencing in this study was achieved by RNA interference (RNAi), but could just as well be achieved today using fusions of nuclease-deactivated Cas9 (dCas9) to transcriptional repressors, also known as CRISPR interference, or CRISPRi (Tam et al., 2011; Lowder et al., 2017). In a recent study, active CRISPR-Cas was successfully used to promote homoeologous COs in an interspecific hybrid between wild and cultivated tomato by knockout of the anti-CO gene *RECQ4* (De Maagd et al., 2019). A related approach was effective in allopolyploid wheat *(Triticum aestivum* L.*).* Recombination between homoeologous chromosomes in wheat is suppressed by the *Pairing homoeologous 1 (Ph1)* locus, resulting in diploid-like behavior during meiosis (Rey et al., 2018). It was recently determined that the major CO gene *ZIP4* is located at the *Ph1* locus, and its disruption in hexaploid wheat with CRISPR-Cas led to significantly increased homeologous CO frequency in hybridizations with rye (Rey et al., 2018). In addition to these experimental studies, simulations of controlled recombination in soybean for introgressing exotic germplasm into elite cultivars have also been promising [S. Ru and R. Bernardo, bioRxiv 701987 unpublished]. Simulated crosses of seven exotic soybeans (*Glycine max* L) to the elite line IA3032 predicted that introgressions of specific chromosome segments facilitated by controlled recombination could boost yields 8-25% over the IA3032 yield, while the best predicted recombinant inbred lines produced without controlled recombination from each biparental cross had negative or negligible yield gains over IA3032 [S. Ru and R. Bernardo, bioRxiv 701987 unpublished]. Though work needs to be done to better understand regulatory mechanisms for chromosome pairing across species, the further development of strategies to increase CO frequency between elite and wild relatives in different crops has exciting potential for increasing genetic gain and accelerating finer introgressions of diverse alleles for breeding (Figure 3.2).

Enhancing recombination in heterochromatic regions

COs occur at low and variable rates along chromosomes, and this limits the genetic variation captured during a breeding cycle. In most organisms DSBs can occur across the entire chromosome, but 80% of COs are concentrated in approximately 25% of the genome (Blary & Jenczewski, 2019; Fernandes et al., 2019). Recombination "hotspots" occur most often in the gene promoters and terminators in regions of euchromatin (Choi et al., 2018). How specific sites become recombination hotspots is an active area of research. Hotspots can be suppressed by DNA methylation (Shilo et al., 2015; Yelina et al., 2015; He et al., 2017; Corem et al., 2018). Genome-wide DSB maps in yeast*, Arabidopsis* and maize (*Zea mays L.*), indicate that chromatin accessibility contributes substantially to DSB formation (He et al., 2017; Tock & Henderson, 2018; Wolde et al., 2019).

Most plant species have strong CO frequency biases away from the heterochromatic centromere and telomeres (Higgins et al., 2014; Fernandes et al., 2019). Suppression of COs in proximity to the centromere is in part due to the centromere's conserved kinetochore assembly function, as centromeric COs have been associated with increased rates of mis-segregation and aneuploidy in multiple species (Fernandes et al., 2019). Recombination "cold regions" are typically associated with inaccessible chromatin and epigenetically silenced pericentromeric regions, although this gradient varies across crop species. For example, 82% of COs are concentrated on the distal ends of wheat chromosome 3B, 19% of the chromosome total length (Darrier et al., 2017). In maize, genes near the centromere are prone to selective sweeps due to extremely low rates of recombination and strong selection for specific alleles (Presting, 2018). As much as 20% of the maize and barley (*Hordeum vulgare L.)* total gene content is located in pericentromeric regions, and this limits the pool of allelic diversity available to breeders (Consortium, 2012; Bauer et al., 2013). Cold regions pose additional challenges for breeders of outcrossing species as deleterious mutations are known to accumulate across low recombination regions (Rodgers-Melnick et al., 2015). The practicality of breeding is directly linked to recombination patterns, and controlled recombination to expand the distribution of COs to cold regions would be a desirable approach to harness the power of selection.

A potential strategy to explore recombination in heterochromatic regions is manipulation of the positions of SPO11-induced DSBs. Plants have three SPO11 paralogs (Kim & Choi, 2019) and transgenic *Arabidopsis* carrying *SPO11-1* hypomorphic alleles have fewer DSBs compared to WT, resulting in fewer COs and an altered distribution of COs (Xue et al., 2018). Notably, significantly fewer COs formed in pericentromeric regions of the transgenic plants, emphasizing the potential in future studies to control SPO11 DSB locations. Although yeast has very little heterochromatin, stimulation of meiotic COs at naturally low recombination regions was demonstrated by tethering SPO11 protein to zinc fingers, transcription activator-like effectors or dCas9 (Sarno et al., 2017). Gene promoters and coding sequences previously identified as cold regions exhibited a 2.3-6.3-fold increase in COs, but no recombination was observed in gene terminators or centromeres (Sarno et al., 2017). Tethering genome editing reagents to SPO11 may prove to be a broadly applicable strategy for varying the location of DSBs and associated CO sites and should be explored in crop species for increasing the genetic diversity of gametes (Figure 3.1).

Gene conversions (GCs) are another product of meiotic recombination and can help break up linkage groups (Wang & Copenhaver, 2018; Gardiner et al., 2019). Unlike COs, GCs are the result of the nonreciprocal transfer of short DNA segments between loci and they have been repeatedly observed across centromeric regions (Shi et al., 2010; Gardiner et al., 2019). The wheat *RECQ7* gene, a recently identified homolog of the Werner syndrome helicase, appears to be a pro-recombination factor that controls GCs (Gardiner et al., 2019). Overexpression of *RECQ7* or the expression of RECQ7 tethered to a genome editing reagent might be used to generate GCs across or at targeted locations anywhere within wheat chromosomes, overcoming the bias for recombination toward the chromosome arms to create useful new haplotypes (Gardiner et al., 2019). Identification and manipulation of homologs of RECQ7, or other genes that control GCs, may offer breeders and geneticists a way to enhance allelic diversity in heterochromatic regions in other crop species.

Epigenetic modifications to control recombination

Epigenomics is an emerging field offering further opportunities for plant breeders to manipulate trait variation. Changes to epigenetic marks can influence the rate and location of COs, but impact of epigenetic marks on pericentromeric meiotic recombination is not fully understood. In *Arabidopsis*, mutation of the genome-wide chromatin remodeler *DECREASE IN DNA METHYLATION1* (*ddm1*), increases recombination in euchromatic regions but fails to induce recombination in pericentric regions despite demethylation (Melamed-Bessudo & Levy, 2012). The heterochromatin that surrounds the centromere in plants is epigenetically silenced by histone 3 lysine 9 dimethylation (H3K9me2) and CG and non-CG DNA methylation (Underwood et al., 2018). *Arabidopsis* mutants of maintenance DNA methyltransferase, *met1*, have no CG methylation and limited H3K9me2 across heterochromatic regions, but the DNA remains tightly packed and inaccessible (Tariq et al., 2003). However, recently, combined loss of H3K9me2 and non-CG DNA methylation (with retention of CG-methylation) in *Arabidopsis* was shown to induce meiotic recombination near centromeres (Underwood et al., 2018). The authors note that pericentromeric COs in H3K9me2/non-CG mutants were present in inbred and hybrid plants and may represent both Class I and Class II COs. Why pericentromeric COs occur in H3K9me2/non-CG mutants but not *ddm1* or *met1* lines is not clear but ostensibly relates to the difference in CG methylation (Underwood et al., 2018).

Despite the recent success of genome-wide epigenetic modification to enhance recombination in heterochromatic regions in *Arabidopsis*, whether this approach will translate to crops remains to be determined. Sensitivity to genome-wide modifications of methylation patterns varies across plant species. Use of CRISPR-Cas9 to knock out two orthologs of *DDM1* in tomato resulted in severely altered vegetative and reproductive development (Corem et al., 2018). Extensive hypomethylation and developmental abnormalities were similarly observed in double mutants of the two *DDM1* orthologs in maize and rice (Corem et al., 2018). Additionally, while in many eukaryotes' sites of recombination have been associated with features of accessible chromatin features such as H3K4me3 sites, maize recombination hotspots do not match this pattern and poorly associate with H3K4me3 sites (Shilo et al., 2015; Yelina et al., 2015; He et al., 2017). DNA sequence and shape have been reported to be predictive of COs throughout the plant kingdom using machine learning models, but which features are most predictive varies (Demirci et al., 2018). Taken together, current research suggests that genome-wide manipulation of epigenetic marks holds promise as a means to expand CO position and frequency in plant breeding, yet further research is needed to determine what modification strategies will be effective, and the answer may differ across species.

Genome editing tools tethered to epigenetic modifiers to target activity to specific loci offer a potential alternative to genome-wide perturbation. In particular, a recent success in site-specific epigenome editing via dCas9 fusion to DNA and histone modifiers such as acetyltransferase or methyltransferase (Hilton et al., 2015; Vojta et al., 2016), hold promise for plant breeders to control recombination in a locus specific fashion and to improve CO frequencies across heterochromatin.

CRISPR-Cas based approaches for understanding COs

In addition to their use in genome and epigenome editing, tools based on CRISPR-Cas are being developed that will help decode the basic molecular mechanisms of meiosis and meiotic recombination in plants and identify stages for effective manipulation of CO formation. For example, live-cell CRISPR imaging with dCas9 and fluorescence-labelled proteins facilitated visualization of telomere movements in tobacco leaf cells (Dreissig et al., 2017). Further application of live-cell CRISPR imaging has the potential to guide future studies for more informed targeting and conversion of DSBs to COs. Live-cell CRISPR imaging could also be used for visualizing how epigenetic tags affect meiosis.

Practical considerations for controlled recombination in crop breeding

Efficient induction of, or selection for, recombination events at critical genomic positions would greatly enable breeders and geneticists. Simulations using data from multiple bi-parental crosses of self and cross-pollinated crops, including soybean, barley, wheat, pea and maize, have identified optimal recombination points from estimated genome-wide marker effects that would maximize genetic gain (Figure 3.2) (Bernardo, 2017; Brandariz & Bernardo, 2019; Ru & Bernardo, 2019). Across all traits and populations, predicted genetic gain was greater for two controlled recombinations rather than one per chromosome. However, on a linkage group level, one controlled recombination outperformed two recombinations in 27% of the cases, and a parental genotype outperformed one or two controlled recombinations in about a third of all cases (Ru & Bernardo, 2019). The above simulation studies have yet to be tested *ex silico*, but in theory, the breeder could induce a desirable mitotic HR in tissue culture, screen plant cells for the ideal genotype, and induce doubled-haploids from regenerated plantlets.

Breeders are often concerned with identifying sources of quantitative variation of a trait rather than knocking out genes (Rodríguez-Leal et al., 2017). The CRISPR-Cas system has been widely used for knockouts in readily transformable species and genotypes, but whether it will be adopted for breaking up linkage drag and generating novel allelic variation in a breeding context remains unknown. It is important to note that genetic transformation and tissue culture methods are not yet efficient or feasible in many crop species (Altpeter et al., 2016). Even within a crop species, such as wheat, tissue culture methodology can be genotype specific and restrict the introduction of controlled recombination across a breeding program's germplasm (Hayta et al., 2019). It is yet to be determined whether the time and cost of developing transformation protocols and expertise, and applying them to a given crop species will restrict the breeder's ability to adopt a controlled recombination approach. The challenges may outweigh the advantages, relative to the traditional cost and effort of the needed population size and generations to map genes and breed elite lines.

Another consideration is that altering CO frequency and distribution may increase the frequency of disrupting existing beneficial gene combinations, though it is yet to be seen if that will outweigh the advantage of controlling COs. A challenge in applying controlled recombination to breeding programs is the effect of ascertainment bias on estimating a locus effect, but this bias may be lessened with tools like CRISPEY that can identify thousands of variants with a fitness impact (Sharon et al., 2018; Ramstein et al., 2019). Additionally, it remains to be determined at which stage in the breeding program controlled recombination is the most beneficial, as well as its long term effects on genetic gain [E. Tourrette *et al.* bioRxiv 704544 unpublished]. For example, SPO11-dCas9 may be effective only as a pre-breeding tool to develop recombinant inbred populations using fewer lines, or to increase the COs in doubled-haploid populations, but be too variable for elite line production. Studies manipulating pro and anti-CO gene expression report a modest (roughly 3-fold) increase in CO frequency, and traditional screening for natural recombinants in a large segregating population may be just as efficient or affordable for a breeding program.

Apart from the technical challenges to overcome and the improvements to efficiency that might be needed for controlled recombination to become broadly applicable in plant breeding, a potential bottleneck for adoption of the technology will be government regulation of gene edited crops. It is critical that such regulation reflects the fundamental difference from conventionally genetically modified organisms, primarily that no foreign DNA is present in the final product. When the editing reagents are introduced via a transgenic construct, this can be achieved by obtaining a null segregant (Zhang et al., 2016; Yubing He et al., 2018; Aliaga-Franco et al., 2019). Furthermore, non-transgenic methods for reagent delivery are increasingly being developed and applied (Zhang et al., 2016). A “product” rather than “process” based policy will allow for any potential risks of gene editing to be evaluated alongside the benefits of the technology (Scheben & Edwards, 2018).

Concluding Remarks and Future Perspectives

Stimulation of recombination by perturbation of pro and anti-CO genes and epigenetic modifiers and by targeting DSBs and epigenetic changes using genome editing tools is a novel and promising set of approaches that should be explored for plant breeding. The direct manipulation of recombination frequency for high-resolution genetic maps demonstrated in yeast and heterochromatic COs achieved in *Arabidopsis* are inspiring proof of principle advances. The potential to dramatically decrease the time and cost required to identify causal variants, break undesired linkages of traits and select preferred haplotypes are compelling motivations to pursue further work in this area. Future research should focus on both improving our basic understanding of recombination in plants and on translating the knowledge from model species to economically important crops. It will be particularly important to determine how controlled recombination can most efficiently increase genetic gain during the plant breeding cycle. By accelerating fundamental understanding and practical advances in plant breeding, continued development and adoption of controlled recombination will lay a strong foundation for improving food security and human health.

Acknowledgements

We thank Wojtek Pawlowski, Shantel A. Martinez, Daniel W. Sweeney and Evan R. Rees for critical reading of the manuscript. ET was supported by a grant from the USDA-NIFA (no. 2017-67007-25939) and Hatch Project 149-447.

REFERENCES

Aliaga-Franco, N., Zhang, C., Presa, S., Srivastava, A.K., Granell, A., Alabadí, D., Sadanandom, A., Blázquez, M.A., & Minguet, E.G. (2019). Identification of Transgene-Free CRISPR-Edited Plants of Rice, Tomato, and Arabidopsis by Monitoring DsRED Fluorescence in Dry Seeds. *Frontiers in Plant Science*, *10*, 1–9. https://doi.org/10.3389/fpls.2019.01150

Altpeter, F., Springer, N.M., Bartley, L.E., Blechl, A.E., Brutnell, T.P., Citovsky, V., Conrad, L.J., Gelvin, S.B., Jackson, D.P., Kausch, A.P., Lemaux, P.G., Medford, J.I., Orozco-Cárdenas, M.L., Tricoli, D.M., Van Eck, J., Voytas, D.F., Walbot, V., Wang, K., Zhang, Z.J., Stewart, C.N., & Jr. (2016). Advancing Crop Transformation in the Era of Genome Editing.. *The Plant cell*, *28*, 1510–1520. https://doi.org/10.1105/tpc.16.00196

Badaeva, E., Dedkova, O., Gay, G., Pukhalskyi, V., Zelenin, A., Bernard, S., Bernard, M., Pukhalskyi Vavilov, V.N., & Zelenin Engelhardt, A. (2007). Chromosomal rearrangements in wheat: their types and distribution. *Genome*, *50*, 907–926. https://doi.org/10.1139/G07-072

Bauer, E., Falque, M., Walter, H., Bauland, C., Camisan, C., Campo, L., Meyer, N., Ranc, N., Rincent, R., Schipprack, W., Altmann, T., Flament, P., Melchinger, A.E., Menz, M., Moreno-González, J., Ouzunova, M., Revilla, P., Charcosset, A., Martin, O.C., & Schön, C.-C. (2013). Intraspecific variation of recombination rate in maize. *Genome Biology*, *14*, R103. https://doi.org/10.1186/gb-2013-14-9-r103

Bernardo, R. (2017). Prospective targeted recombination and genetic gains for quantitative traits in maize. *The Plant Genome*, *10*, 1–9. https://doi.org/10.3835/plantgenome2016.11.0118

Blary, A., & Jenczewski, E. (2019). Manipulation of crossover frequency and distribution for plant breeding. *Theoretical and Applied Genetics*, *132*, 575–592. https://doi.org/10.1007/s00122-018-3240-1

Bogdanove, A.J., & Voytas, D.F. (2011). TAL Effectors: Customizable Proteins for DNA Targeting. *Science (New York, N.Y.)*, *333*, 1843–1846. https://doi.org/10.1126/science.1204094

Brandariz, S.P., & Bernardo, R. (2019). Predicted genetic gains from targeted recombination in elite biparental maize populations. *The Plant Genome*, *12*, 180062. https://doi.org/10.3835/plantgenome2018.08.0062

Brown, M.S., & Bishop, D.K. (2014). DNA strand exchange and RecA homologs in meiosis.. *Cold Spring Harbor perspectives in biology*, *7*, 1–29. https://doi.org/10.1101/cshperspect.a016659

Chelysheva, L., Grandont, L., Vrielynck, N., Le Guin, S., Mercier, R., & Grelon, M. (2010). An Easy Protocol for Studying Chromatin and Recombination Protein Dynamics during Arabidopsis thaliana Meiosis: Immunodetection of Cohesins, Histones and MLH1. *Cytogenet Genome Res*, *129*, 143–153. https://doi.org/10.1159/000314096

Choi, K., Zhao, X., Tock, A.J., Lambing, C., Underwood, C.J., Hardcastle, T.J., Serra, H., Kim, J., Cho, H.S., Kim, J., Ziolkowski, P.A., Yelina, N.E., Hwang, I., Martienssen, R.A., & Henderson, I.R. (2018). Nucleosomes and DNA methylation shape meiotic DSB frequency in Arabidopsis thaliana transposons and gene regulatory regions. *Genome research*, *28*, 532–546. https://doi.org/10.1101/gr.225599.117

Consortium, T.I.B.G.S. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature*, *491*, 711–716. https://doi.org/10.1038/nature11543

Copenhaver, G.P., Housworth, E.A., & Stahl, F.W. (2002). Crossover Interference in Arabidopsis. *Genetics*, *160*, 1631–1639

Corem, S., Doron-Faigenboim, A., Jouffroy, O., Maumus, F., Arazi, T., & Bouchéc, N. (2018). Redistribution of CHH methylation and small interfering RNAs across the genome of tomato ddm1 mutants. *The Plant Cell*, *30*, 1628–1644. https://doi.org/10.1105/tpc.18.00167

Crismani, W., Girard, C., Froger, N., Pradillo, M., Santos, J.L., Chelysheva, L., Copenhaver, G.P., Horlow, C., & Mercier, R. (2012). FANCM limits meiotic crossovers.. *Science (New York, N.Y.)*, *336*, 1588–1590. https://doi.org/10.1126/science.1220381

Darrier, B., Rimbert, H., Balfourier, F., Pingault, L., Josselin, A.-A., Servin, B., Navarro, J., Choulet, F., Paux, E., & Sourdille, P. (2017). High-Resolution Mapping of Crossover Events in the Hexaploid Wheat Genome Suggests a Universal Recombination Mechanism. *Genetics*, *206*, 1373–1388. https://doi.org/10.1534/genetics.116.196014

Demirci, S., Peters, S.A., de Ridder, D., & van Dijk, A.D.J. (2018). DNA sequence and shape are predictive for meiotic crossovers throughout the plant kingdom. *The Plant Journal*, *95*, 686–699. https://doi.org/10.1111/tpj.13979

Dreissig, S., Schiml, S., Schindele, P., Weiss, O., Rutten, T., Schubert, V., Gladilin, E., Mette, M.F., Puchta, H., & Houben, A. (2017). Live-cell CRISPR imaging in plants reveals dynamic telomere movements. *The Plant Journal*, *91*, 565–573. https://doi.org/10.1111/tpj.13601

Faris, J.D., Xu, S.S., Cai, X., Friesen, T.L., & Jin, Y. (2008). Molecular and cytogenetic characterization of a durum wheat–Aegilops speltoides chromosome translocation conferring resistance to stem rust. *Chromosome Research*, *16*, 1097–1105. https://doi.org/10.1007/s10577-008-1261-3

Fasching, C.L., Cejka, P., Kowalczykowski, S.C., & Heyer, W.D. (2015). Top3-Rmi1 dissolve Rad51-mediated D loops by a topoisomerase-based mechanism. *Molecular Cell*, *57*, 595–606. https://doi.org/10.1016/j.molcel.2015.01.022

Fernandes, J.B., Duhamel, M., Seguéla-Arnaud, M., Froger, N., Girard, C., Choinard, S., Solier, V., De Winne, N., De Jaeger, G., Gevaert, K., Andrey, P., Grelon, M., Guerois, R., Kumar, R., & Mercier, R. (2018)(a). FIGL1 and its novel partner FLIP form a conserved complex that regulates homologous recombination. *PLoS Genetics*, *14*, e1007317. https://doi.org/10.1371/journal.pgen.1007317

Fernandes, J.B., Séguéla-Arnaud, M., Larchevêque, C., Lloyd, A.H., & Mercier, R. (2018)(b). Unleashing meiotic crossovers in hybrid plants. *Proceedings of the National Academy of Sciences*, *115*, 2431–2436. https://doi.org/10.1073/PNAS.1713078114

Fernandes, J.B., Wlodzimierz, P., Henderson, I.R., & Kelly, S. (2019). Meiotic recombination within plant centromeres. *Current Opinion in Plant Biology*, *48*, 26–35. https://doi.org/10.1016/j.pbi.2019.02.008

Gardiner, L.-J., Wingen, L.U., Bailey, P., Joynson, R., Brabbs, T., Wright, J., Higgins, J.D., Hall, N., Griffiths, S., Clavijo, B.J., & Hall, A. (2019). Analysis of the recombination landscape of hexaploid bread wheat reveals genes controlling recombination and gene conversion frequency. *Genome Biology*, *20*, 69. https://doi.org/10.1186/s13059-019-1675-6

Girard, C., Chelysheva, L., Choinard, S., Froger, N., Macaisne, N., Lehmemdi, A., Mazel, J., Crismani, W., & Mercier, R. (2015). AAA-ATPase FIDGETIN-LIKE 1 and Helicase FANCM Antagonize Meiotic Crossovers by Distinct Mechanisms. *PLOS Genetics*, *11*, e1005369. https://doi.org/10.1371/journal.pgen.1005369

Girard, C., Crismani, W., Froger, N., Mazel, J., Lemhemdi, A., Horlow, C., & Mercier, R. (2014). FANCM-associated proteins MHF1 and MHF2, but not the other Fanconi anemia factors, limit meiotic crossovers. *Nucleic Acids Research*, *42*, 9087–9095. https://doi.org/10.1093/nar/gku614

Glover, N.M., Redestig, H., & Dessimoz, C. (2016). Homoeologs: What Are They and How Do We Infer Them?. *Trends in plant science*, *21*, 609–621. https://doi.org/10.1016/j.tplants.2016.02.005

Gonen, S., Battagin, M., Johnston, S.E., Gorjanc, G., & Hickey, J.M. (2017). The potential of shifting recombination hotspots to increase genetic gain in livestock breeding. *Genetics Selection Evolution*, *49*, 55. https://doi.org/10.1186/s12711-017-0330-5

Hartung, F., & Puchta, H. (2006). The RecQ gene family in plants. *Journal of Plant Physiology*, *163*, 287–296. https://doi.org/10.1016/j.jplph.2005.10.013

Hartung, F., Suer, S., & Puchta, H. (2007). Two closely related RecQ helicases have antagonistic roles in homologous recombination and DNA repair in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *104*, 18836–18841

Hayta, S., Smedley, M.A., Demir, S.U., Blundell, R., Hinchliffe, A., Atkinson, N., & Harwood, W.A. (2019). An efficient and reproducible Agrobacterium-mediated transformation method for hexaploid wheat (Triticum aestivum L.). *Plant Methods*, *121*, 1–15

Hayut, S.F., Melamed Bessudo, C., & Levy, A.A. (2017). Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nature Communications*, *8*, 15605. https://doi.org/10.1038/ncomms15605

He, Y., Wang, M., Dukowic-Schulze, S., Zhou, A., Tiang, C.-L., Shilo, S., Sidhu, G.K., Eichten, S., Bradbury, P., Springer, N.M., Buckler, E.S., Levy, A.A., Sun, Q., Pillardy, J., Kianian, P.M.A., Kianian, S.F., Chen, C., & Pawlowski, W.P. (2017). Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. *Proceedings of the National Academy of Sciences*, *114*, 12231–12236. https://doi.org/10.1073/pnas.1713225114

Higgins, J.D., Osman, K., Jones, G.H., Chris, F., & Franklin, H. (2014). Factors Underlying Restricted Crossover Localization in Barley Meiosis. *Annual Review of Genetics*, *48*, 29–47. https://doi.org/10.1146/annurev-genet-120213-092509

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., & Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology*, *33*, 510–517. https://doi.org/10.1038/nbt.3199

Hu, Q., Li, Y., Wang, H., Shen, Y., Zhang, C., Du, G., Tang, D., & Cheng, Z. (2017). Meiotic Chromosome Association 1 Interacts with TOP3α and Regulates Meiotic Recombination in Rice.. *The Plant cell*, *29*, 1697–1708. https://doi.org/10.1105/tpc.17.00241

Jiang, W., & Marraffini, L.A. (2015). CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. *Annu. Rev. Microbiol*, *69*, 209–228. https://doi.org/10.1146/annurev-micro-091014-104441

Keeney, S., Giroux, C.N., & Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, *88*, 375–84

Kim, J., & Choi, K. (2019). Signaling-mediated meiotic recombination in plants. *Current Opinion in Plant Biology*, *51*, 44–50. https://doi.org/10.1016/j.pbi.2019.04.001

Klug, A. (2010). The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annual Review of Biochemistry*, *79*, 213–231. https://doi.org/10.1146/annurev-biochem-010909-095056

Lande, R., & Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, *124*, 743–756. https://doi.org/10.1046/j.1365-2540.1998.00308.x

Lao, J.P., Cloud, V., Huang, C.-C., Grubb, J., Thacker, D., Lee, C.-Y., Dresser, M.E., Hunter, N., & Bishop, D.K. (2013). Meiotic Crossover Control by Concerted Action of Rad51-Dmc1 in Homolog Template Bias and Robust Homeostatic Regulation. *PLoS Genetics*, *9*, e1003978. https://doi.org/10.1371/journal.pgen.1003978

De Los Santos, T., Hunter, N., Lee, C., Larkin, B., Loidl, J., & Hollingsworth, N.M. (2003). The Mus81/Mms4 Endonuclease Acts Independently of Double-Holliday Junction Resolution to Promote a Distinct Subset of Crossovers During Meiosis in Budding Yeast. *Genetics*, *164*, 81–94

Lowder, L.G., Paul, J.W., & Qi, Y. (2017). Multiplexed Transcriptional Activation or Repression in Plants Using CRISPR-dCas9-Based Systems. *Plant Gene Regulatory Networks* (pp. 167–184). Humana Press, New York, NY.

De Maagd, R.A., Loonen, A., Chouaref, J., Pel, A., Meijer-Dekens, F., Fransz, P., & Bai, Y. (2019). CRISPR/Cas inactivation of RECQ4 increases homeologous crossovers in an interspecific tomato hybrid. *Plant Biotechnology Journal*, 1–9. https://doi.org/10.1111/pbi.13248

Mammadov, J., Buyyarapu, R., Guttikonda, S.K., Parliament, K., Abdurakhmonov, I.Y., & Kumpatla, S.P. (2018). Wild Relatives of Maize, Rice, Cotton, and Soybean: Treasure Troves for Tolerance to Biotic and Abiotic Stresses. *Frontiers in Plant Science*, *9*, 886. https://doi.org/10.3389/fpls.2018.00886

Melamed-Bessudo, C., & Levy, A.A. (2012). Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. *Proceedings of the National Academy of Sciences*, *109*, E981–E988. https://doi.org/10.1073/pnas.1120742109

Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N., & Grelon, M. (2015). The Molecular Biology of Meiosis in Plants. *Annual Review of Plant Biology*, *66*, 297–327. https://doi.org/10.1146/annurev-arplant-050213-035923

Mieulet, D., Aubert, G., Bres, C., Klein, A., Droc, G., Vieille, E., Rond-Coissieux, C., Sanchez, M., Dalmais, M., Mauxion, J.-P., Rothan, C., Guiderdoni, E., & Mercier, R. (2018). Unleashing meiotic crossovers in crops. *Nature Plants*, *4*, 1010–1016. https://doi.org/10.1038/s41477-018-0311-x

Presting, G.G. (2018). Centromeric retrotransposons and centromere function. *Current Opinion in Genetics and Development*, *49*, 79–84. https://doi.org/10.1016/j.gde.2018.03.004

Puchta, H. (1999). Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. *Genetics*, *152*, 1173–1181

Ramstein, G.P., Jensen, S.E., & Buckler, E.S. (2019). Breaking the curse of dimensionality to identify causal variants in Breeding 4. *Theoretical and Applied Genetics*, *132*, 559–567. https://doi.org/10.1007/s00122-018-3267-3

Ren, Z., Tang, Z., Fu, S., Yan, B., Tan, F., Ren, T., & Li, Z. (2017). Molecular Cytogenetic Characterization of Novel Wheat-rye T1RS.1BL Translocation Lines with High Resistance to Diseases and Great Agronomic Traits. *Frontiers in Plant Science*, *8*, 799. https://doi.org/10.3389/fpls.2017.00799

Rey, M.-D., Martín, A.C., Smedley, M., Hayta, S., Harwood, W., Shaw, P., & Moore, G. (2018). Magnesium increases homoeologous crossover frequency during meiosis in ZIP4 (Ph1 Gene) mutant wheat-wild relative hybrids. *Frontiers in Plant Science*, *9*, 509. https://doi.org/10.3389/fpls.2018.00509

Rodgers-Melnick, E., Elshire, R.J., Li, Y., Bradbury, P.J., Mitchell, S.E., Li, C., Glaubitz, J.C., Buckler, E.S., & Acharya, C.B. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences*, *112*, 3823–3828. https://doi.org/10.1073/pnas.1413864112

Rodríguez-Leal, D., Lemmon, Z.H., Man, J., Bartlett, M.E., & Lippman, Z.B. (2017). Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell*, *171*, 470–480. https://doi.org/10.1016/J.CELL.2017.08.030

Ru, S., & Bernardo, R. (2019). Targeted recombination to increase genetic gain in self-pollinated species. *Theoretical and Applied Genetics*, *132*, 289–300. https://doi.org/10.1007/s00122-018-3216-1

Rutkoski, J.E. (2019). A practical guide to genetic gain. *Advances in Agronomy*, *157*, 217–249. https://doi.org/10.1016/bs.agron.2019.05.001

Sadhu, M.J., Bloom, J.S., Day, L., & Kruglyak, L. (2016). CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science (New York, N.Y.)*, *352*, 1113–1116. https://doi.org/10.1126/science.aaf5124

Sarno, R., Vicq, Y., Uematsu, N., Luka, M., Lapierre, C., Carroll, D., Bastianelli, G., Serero, A., & Nicolas, A. (2017). Programming sites of meiotic crossovers using Spo11 fusion proteins. *Nucleic Acids Research*, *45*, e164. https://doi.org/10.1093/nar/gkx739

Scheben, A., & Edwards, D. (2018). Bottlenecks for genome-edited crops on the road from lab to farm. *Genome Biology*, *19*, 178. https://doi.org/10.1186/s13059-018-1555-5

Séguéla-Arnaud, M., Crismani, W., Larchevêque, C., Mazel, J., Froger, N., Choinard, S., Lemhemdi, A., Macaisne, N., Van Leene, J., Gevaert, K., De Jaeger, G., Chelysheva, L., & Mercier, R. (2015). Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM.. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 4713–4718. https://doi.org/10.1073/pnas.1423107112

Serra, H., Lambing, C., Griffin, C.H., Topp, S.D., Nageswaran, D.C., Underwood, C.J., Ziolkowski, P.A., Séguéla-Arnaud, M., Fernandes, J.B., Mercier, R., & Henderson, I.R. (2018). Massive crossover elevation via combination of HEI10 and recq4a recq4b during Arabidopsis meiosis.. *Proceedings of the National Academy of Sciences*, *115*, 2437–2442. https://doi.org/10.1073/pnas.1713071115

Sharon, E., Chen, S.-A.A., Khosla, N.M., Smith, J.D., Pritchard, J.K., & Fraser, H.B. (2018). Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell*, *175*, 544–557. https://doi.org/10.1016/j.cell.2018.08.057

Shi, J., Wolf, S.E., Burke, J.M., Presting, G.G., Ross-Ibarra, J., & Dawe, R.K. (2010). Widespread Gene Conversion in Centromere Cores. *PLoS Biology*, *8*, e1000327. https://doi.org/10.1371/journal.pbio.1000327

Shilo, S., Melamed-Bessudo, C., Dorone, Y., Barkai, N., & Levy, A.A. (2015). DNA Crossover Motifs Associated with Epigenetic Modifications Delineate Open Chromatin Regions in Arabidopsis. *The Plant Cell*, *27*, 2427–2436. https://doi.org/10.1105/tpc.15.00391

Singh, G., Da Ines, O., Eugenia Gallego, M., & White, C.I. (2017). Analysis of the impact of the absence of RAD51 strand exchange activity in Arabidopsis meiosis. *PLoS ONE*, *12*, e0183006. https://doi.org/10.1371/journal.pone.0183006

Tam, S.M., Hays, J.B., & Chetelat, R.T. (2011). Effects of suppressing the DNA mismatch repair system on homeologous recombination in tomato. *Theoretical and Applied Genetics*, *123*, 1445–1458. https://doi.org/10.1007/s00122-011-1679-4

Tariq, M., Saze, H., Probst, A. V, Lichota, J., Habu, Y., & Paszkowski, J. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. *Proceedings of the National Academy of Sciences*, *100*, 8823–8827

Tock, A.J., & Henderson, I.R. (2018). Hotspots for Initiation of Meiotic Recombination. *Frontiers in Genetics*, *9*, 521. https://doi.org/10.3389/fgene.2018.00521

Underwood, C.J., Choi, K., Lambing, C., Zhao, X., Serra, H., Borges, F., Simorowski, J., Ernst, E., Jacob, Y., Henderson, I.R., & Martienssen, R.A. (2018). Epigenetic activation of meiotic recombination near Arabidopsis thaliana centromeres via loss of H3K9me2 and non-CG DNA methylation.. *Genome research*, *28*, 519–531. https://doi.org/10.1101/gr.227116.117

Vojta, A., Dobriní, P., Tadí, V., Bočkor, L., Bočkor, B., Korá, P., Julg, B., Klasí, M., Zoldoš, V., & Zoldoš, Z. (2016). Repurposing the CRISPR-Cas9 system for targeted

DNA methylation. *Nucleic Acids Research*, *44*, 5615–5628. https://doi.org/10.1093/nar/gkw159

Wang, Y., & Copenhaver, G.P. (2018). Meiotic Recombination: Mixing It Up in Plants. *Annual Review of Plant Biology*, *69*, 577–609

Wijnker, E., Van Dun, K., De Snoo, C.B., Lelivelt, C.L.C., Keurentjes, J.J.B., Naharudin, N.S., Ravi, M., Chan, S.W.L., De Jong, H., & Dirks, R. (2012). Reverse breeding in Arabidopsis thaliana generates homozygous parental lines from a heterozygous plant. *Nature Genetics*, *44*, 467–470. https://doi.org/10.1038/ng.2203

Wolde, G.M., Trautewig, C., Mascher, M., & Schnurbusch, T. (2019). Genetic insights into morphometric inflorescence traits of wheat. *Theoretical and Applied Genetics*, *132*, 1661–1676. https://doi.org/10.1007/s00122-019-03305-4

Xue, M., Wang, J., Jiang, L., Wang, M., Wolfe, S., Pawlowski, W.P., Wang, Y., & He, Y. (2018). The Number of Meiotic Double-Strand Breaks Influences Crossover Distribution in Arabidopsis.. *The Plant cell*, *30*, 2628–2638. https://doi.org/10.1105/tpc.18.00531

Yelina, N.E., Lambing, C., Hardcastle, T.J., Zhao, X., Santos, B., & Henderson, I.R. (2015). DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis.. *Genes & development*, *29*, 2183–2202. https://doi.org/10.1101/gad.270876.115

Yubing He, Min Zhu, Lihao Wang, Junhua Wu, Qiaoyan Wang, Rongchen Wang, & Yunde Zhao. (2018). Programmed Self-Elimination of theCRISPR/Cas9 Construct Greatly Accelerates the Isolation of Edited and Transgene-Free Rice Plants. *Molecular Plant*, *11*, 1210–1213. https://doi.org/10.1016/j.molp.2018.05.005

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J.L., & Gao, C. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms12617

Zhu, C., Gore, M., Buckler, E.S., & Yu, J. (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome Journal*, *1*, 5–20. https://doi.org/10.3835/plantgenome2008.02.0089

Ziolkowski, P.A., Underwood, C.J., Lambing, C., Martinez-Garcia, M., Lawrence, E.J., Ziolkowska, L., Griffin, C., Choi, K., Franklin, F.C.H., Martienssen, R.A., & Henderson, I.R. (2017). Natural variation and dosage of the HEI10 meiotic E3 ligase control Arabidopsis crossover recombination. *Genes and Development*, *31*, 306–317. https://doi.org/10.1101/gad.295501.116

CHAPTER 4

IF IT AIN'T BROKE, DON'T FIX IT: EVALUATING THE EFFECT OF INCREASED RECOMBINATION ON RESPONSE TO SELECTION FOR WHEAT BREEDING [3]

Abstract

Meiotic recombination is a source of allelic diversity, but the low frequency and biased distribution of crossovers that occur during meiosis limits the genetic variation available to plant breeders. Simulation studies have previously identified that increased recombination frequency can retain more genetic variation and drive greater genetic gains than wildtype recombination. Our study was motivated by the need to define desirable recombination intervals in regions of the genome where we have historically detected very few crossovers. We hypothesized that deleterious variants, which can negatively impact phenotypes and occur at higher frequencies in low recombining regions where they are linked in repulsion with favorable loci, may offer a signal for evaluation of shifting recombination distributions. Genomic selection breeding simulation models with empirical wheat data were developed to evaluate increased recombination frequency, changing recombination distribution, and QTL variant annotation on response to selection. Comparing high and low values for a range of simulation parameters identified that few combinations retained greater genetic variation and fewer still achieved higher genetic gain than the wild type. More recombination was associated with loss of genomic prediction accuracy, which often outweighed the benefits of disrupting repulsion linkages. Irrespective of

recombination frequency or distribution and QTL annotation, enhanced response to selection under increased recombination largely depended on polygenic trait architecture, high heritability, more repulsion than coupling linkages, and greater than six cycles of genomic selection. Altogether, the outcomes of this research discourage a controlled recombination approach to genomic selection in wheat as a more efficient path to retaining genetic variation and increasing genetic gains compared to existing breeding methods.

[3] Project conducted under the supervision of Dr. Jean-Luc Jannink and intended for journal submission May 2022.

Ella Taagen contributions: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing-original draft; Writing-review & editing.

Introduction

Plant breeders rely on natural recombination of genetic material during meiotic segregation to generate novel allelic combinations and select favorable haplotypes. A single crossover (CO) between each homologous chromosome pair during meiosis is obligate for proper chromosome segregation. Recombination rates above this minimum can vary but are limited in most species (Henderson & Bomblies, 2021). Additionally, the distribution of COs in many eukaryotes, including plants, are skewed away from the pericentromere and toward subtelomeric regions. The low frequency and uneven distribution of COs along chromosomes limits the genetic variation accessible to plant breeders, which requires working with large populations over many cycles of selection to identify new cultivars.

There are evolutionary advantages as well as costs associated with variation in recombination. The benefits of recombination are largely described by two models in population and evolutionary genetics: the Hill-Robertson effect and Muller's Ratchet (Hill & Robertson, 1966; Muller, 1964). Recombination can disrupt repulsion linkages between favorable and deleterious loci, aiding in the efficiency of selection (i.e., The Hill-Robertson effect). In the absence of recombination, a population's deleterious load will steadily increase and can never fall below the lowest load in the original population (i.e., Muller's Ratchet). In regions of low recombination like the pericentromere, deleterious mutations may persist and thus are likely to become linked in repulsion with positive loci (Rodgers-Melnick et al., 2015; Jordan et al., 2018). Significantly increased recombination can come at a cost though, leading to decreased fitness by breaking up beneficial linkages and reduced fertility (Charlesworth & Barton, 1996; Mieulet et al., 2018).

New mutations continuously arise in populations due to errors in DNA replication. Their fitness effects can range from lethal to beneficial. The rate of new mutations in eukaryotes is estimated to be at least 1 x 10^{-8} / base pair / meiosis, and it is predicted that new mutations in coding regions will be deleterious in some of the environments the species inhabits (Ohta, 1972, 1992; Baer et al., 2007). The accumulation of deleterious mutations may also be faster in allopolyploids, compared to diploids, due to the masking effect of unaltered gene copies (Conover & Wendel, 2022). Artificial selection and population improvement (i.e., increased fitness) often result in inbreeding, which reduces the effective recombination rate (Moyers et al., 2018). The "cost of domestication" hypothesis suggests that the process of artificial selection has increased the proportion of deleterious variants in domesticated genomes compared to their wild progenitor (Lu et al., 2006; Moyers et al., 2018). This cost reduces the efficiency of selection, for example as the likelihood of deleterious variants hitchhiking via linkage disequilibrium (LD) is increased.

Overcoming the low rate and biased positioning of COs to reduce deleterious load and increase the genetic variation accessible to breeders is a longstanding goal. Manipulation of pro- and anti-CO factors, and the use of genome editing reagents that induce double-stranded DNA breaks or modify the epigenome at desired sites of recombination (e.g., CRISPR/Cas system), offer novel approaches for modifying CO frequency and distribution. (Hayut et al., 2017; Corem et al., 2018; Fernandes et al., 2018; Mieulet et al., 2018; Serra et al., 2018; Underwood et al., 2018; Taagen et al., 2020). These approaches are collectively referred to as 'controlled recombination' and transitioning them from a lab setting using model species to a breeder's field with diverse applications may more efficiently harness the power of selection for plant breeding (Taagen et al., 2020). For example, mutation of anti-CO factor *recq4* orthologs in rice (*Oryza sativa* L.), pea (*Pisum sativum* L.), and tomato (*Solanum lycopersicum* L.), have been shown to increase CO frequency by threefold compared to wild type (WT) (Mieulet et al., 2018). In Arabidopsis (*Arabidopsis thaliana*), increased copy-number of pro-CO factor HEI10 and mutation of *recq4a* and *reqc4b* additively led to five-fold and 1.5-fold increased meiotic recombination in the chromosome arms and pericentromeric heterochromatin, respectively (Serra et al., 2018). Modified recombination distributions and frequencies have also been achieved by generating allotriploid hybrids in turnip (*Brassica rapa* L.), which reported a remarkable twenty-fold increased recombination rate in the pericentromere (Pelé et al., 2017). All together, these methods may open the door to exciting applications for breeding, including rapid fine-mapping of genes, facilitating reintroduction of genetic variance at sites of selective sweeps, introgression of diverse alleles from wild crop relatives, and maintenance of genetic variation during genomic selection (Tam et al., 2011; Sadhu et al., 2016; Presting, 2018; Rey et al., 2018; Taagen et al., 2020).

While there are a variety of methods under development to introduce controlled recombination to a crop species, determining where in the genome to implement controlled recombination is essential to enhancing breeding efficiency and genetic gain. At present, controlled recombination for inbred and hybrid crops, as well as livestock breeding pipelines has only been tested with simulation models (Battagin et al., 2016; Bernardo, 2017; Gonen et al., 2017; Brandariz & Bernardo, 2019; Johnsson et al., 2019; Ru & Bernardo, 2019, 2020; Tourrette et al., 2019; Oyetunde & Bernardo, 2020). The consensus from simulations across different species and traits is that applying controlled recombination could, at minimum, double genetic gains. However, it is notable that few simulations to date have factored in a cost of novel technology adoption and feasibility of multiplex genome editing. For example, under independent segregation of chromosomes into gametes, the likelihood of generating an individual with two desired site-specific COs on all 21 wheat chromosomes is on the magnitude of $1 / 2^{21}$. In addition, many of these simulations rely on RR-BLUP (Endelman, 2011) to estimate marker effects and predict ideal CO intervals that produce the greatest marker effect sum, which limits the analysis to identifying regions where historical recombination has made quantitative trait loci (QTL) effects apparent (Bernardo, 2017; Brandariz & Bernardo, 2019; Ru & Bernardo, 2019, 2020; Tourrette et al., 2019; Oyetunde & Bernardo, 2020). In regions of low recombination where deleterious mutations are more likely to become linked in repulsion, their potentially deleterious effects are hidden from RR-BLUP's estimation of marker effects. Using approaches to identify deleterious alleles, such as variant effect prediction in coding regions or evolutionary conservation, offers distinct strengths for identifying potentially advantageous controlled recombination targets in plant breeding (Rodgers-Melnick et al., 2015; Kono et al., 2018, 2019; Johnsson et al., 2019).

Many putative deleterious alleles have been identified in pericentromeric areas and shifting the distribution of recombination toward the pericentromere could unlock novel haplotypes and ultimately inform better targets for site specific controlled recombination approaches. While these technologies remain under research and development, initial reports on them prompted us to test the value of recombination in regions of the genome where we suspect it has long been suppressed. We hypothesized that deleterious variants, which can occur more frequently in low recombining regions and have significant negative effects on phenotype, may offer a signal for evaluation of shifting recombination distributions.

We used empirical genotype data and deleterious variant annotations from wheat (*Triticum aestivum* L) to simulate a genomic selection breeding program and identified the parameter space in which increased recombination maintained genetic diversity and increased genetic gain. Previous evaluations of the recombination landscape in this diverse wheat population reported that over 75% of the recombination events fell within 10% of the distal end of chromosomes, and a higher density of putative deleterious variants were detected in the pericentromere versus distal arms of chromosomes (Jordan et al., 2018). We evaluated the impact of several simulation parameters on population improvement, including the number of QTL per chromosome, heritability, the recombination frequency and distribution, whether the QTL were annotated as deleterious variants, coupling versus repulsion between QTL, and finally the relationship matrix used for estimated marker effects. Altogether, our findings highlight the challenging path to the realizing the benefits of increased recombination for a genomic selection wheat breeding program and discuss the feasibility of technology adoption.

Materials and Methods

We used the programming language R and breeding program simulation tool AlphaSimR to evaluate the effects of increased frequency and shifting distribution of recombination on a range of population improvement scenarios following a basic scheme (Figure 4.1) (R Core Team, 2020; Gaynor et al., 2021). The raw data, scripts for the simulations, and supplementary results are available on https://github.com/etaagen/dissertation_chapter_4.

Founder SNP, linkage map data, and population structure

We used a published dataset of 29 genetically and geographically diverse wheat accessions for our simulations, previously selected to develop a nested association mapping population (Jordan et al., 2018). The high-density SNP data for the founder lines was generated using the wheat exome capture assay and mapped to the W7984 reference genome and genetic map (Chapman et al., 2015). We retrieved SNP and InDel data (Table S2 and S3) from http://wheagenomics.plantpath.ksu.edu/nam and removed variants with more than 60% missing genotypes and annotations of unknown chromosomes. The remaining missing genotype calls were imputed with R/qtl2 imputation sim_geno() and monomorphic SNPs were removed (Broman et al., 2019).

The 29 founders were fully inbred, and we designated genotypes as '1' and '0' for the major and minor allele, respectively. A principal component (PC) analysis of the founders identified that the first PC explained 8% of the total variance and was associated with the minor allele count (Supplemental figure_S2.1.pdf). There were three outlier lines (*Cltr_15134, Cltr_11223,* and *PI_366716*) that drove the very high observed rate of coupling among minor alleles, and we removed them from the founder population (see *Simulation parameter: coupling and repulsion*). After this final data filtering step and re-evaluating major and minor alleles, there were 26 founder lines, 352,804 total SNPs, 134,803 SNPs on the A genome, 181,012 SNPs on the B genome, and 36,989 SNPs on the D genome.

We also considered a biparental simulation scheme, where we selected the two most divergent parents from the 26 founders, *PI_220431* and *PI_185715*, whose genotypes were designated as '1' and '0' respectively. The biparental population had 130,127 total SNPs, 61,387 on the A genome, 53,471 on the B genome, and 15,269 on the D genome. The full set of 26 lines and the biparental population served as founders for their respective simulation replicate draws.

We use "SNP" to refer to any variant in the founder population, "marker" to refer to variants assigned to the SnpChip (i.e., these generate observed genotypes in the AlphaSimR simulation), and "QTL" to refer to variants assigned as causal loci. We use the term recombination to refer to the meiotic process where a double-stranded DNA beak is repaired via homologous recombination, resulting in a CO.

Simulation parameter: SnpChip

Each replicate of the simulation (see Figure 4.1) began by randomly sampling SNPs from each cM bin to serve as the markers on the SnpChip. When markers shared the same genetic map position, we jittered them by 0.00001 M. The chromosome length and marker density averaged 1.22 M and 1200 SNPs per chromosome for both the full founder and biparental population approaches.

Simulation parameter: QTL

The QTL effects in AlphaSimR inform the genetic value, which we used to simulate a trait phenotype. All effects were additive (no dominance or epistasis). Two approaches for assigning QTL were considered, either randomly (R) or if we categorized the QTL as potentially harboring a deleterious variant (DV) based on the SnpEff annotation (see *Simulation parameter: Deleterious variant annotations*). We simulated either 2 (oligogenic trait) or 200 (polygenic trait) QTL per chromosome, and there was no overlap between the SnpChip and QTL. We considered traits with heritability values of 0.2 and 0.8. We also included a proof-of-concept simulation using the causal variant relationship matrix. This approach did not require a SnpChip as it used the QTL directly to estimate marker effects and is reflective of perfect LD between the markers and QTL (de Los Campos et al., 2013).

Simulation parameter: deleterious variant annotations

Annotation of deleterious variants in the founder population from exome capture (coding sequence) data were previously published using the SNPeffect program (De Baets et al., 2012; Jordan et al., 2018). We considered SNPs with the 'high' SnpEff putative impact criteria as well as the 'non-synonymous coding' effect to be potentially deleterious variants for DV QTL (see http://pcingola.github.io/SnpEff/). These variants were found to be more frequent in lower recombining regions, which we defined as a 0.2 M bin spanning the centromere on each chromosome (Figure 4.2) (Jordan et al., 2018). Under the biparental population approach there was a 25% loss in SNP density and 59% loss in potentially deleterious SnpEff variants across the lower recombining regions, compared to the full 26 founder set. Given this reduction in polymorphic sites, the impact of increased recombination on DV QTL was only evaluated in the full founder population.

We interpreted the SnpEff annotation as an imperfect approximation for impact on yield. This decision was motivated by yield being a highly complex trait that is influenced by genetic variants on every chromosome and at every stage of plant growth. While the SnpEff annotation is based on predicted impact on a protein, that protein may not impact the phenotype per se, and we varied the pool of potentially deleterious variants for DV QTL each replicate of the simulation. Each replicate sampled 90% of SNPs with the 'high' SnpEff putative impact criteria and 25% of SNPs with the 'non-synonymous coding' effect (Lu et al., 2006). This resulted in an average pool per chromosome of 613 potentially causal SNPs for the DV QTL approach. Note that the SnpEff annotations were lowest on the group D chromosomes, which reduced the number of polygenic trait QTL on the D genome. To simplify simulation comparisons, we set the number of polygenic trait QTL in the R QTL simulations to match the DV QTL number, which averaged 200 QTL per A and B genome chromosomes, and 60 QTL per D genome chromosome.

Simulation parameter: coupling and repulsion

Recombination can be beneficial for genetic improvement if it breaks up QTL in repulsion, but detrimental if it breaks up QTL in coupling. We estimated the amount of repulsion in the full founder population by randomly selecting 200 QTL per chromosome (R and DV approach) one hundred times and measuring the percent of neighboring QTL where minor alleles were in repulsion.

In AlphaSimR the additive effects of each QTL are sampled from a standard normal distribution and the magnitude of the effects is scaled to achieve a user specified genetic variance. As we set the full founder genotypes to ‘1’ and ‘0’ for the major and minor allele, respectively, we could use the QTL effect sign, positive or negative, to introduce different levels of coupling and repulsion between neighboring QTL. We set different levels of coupling versus repulsion linkages among the founders because we were interested in the importance of starting conditions versus ongoing selection effects on the impact of increased recombination. Intuitively, we would assume the minor allele to be deleterious (because it would have been selected against during evolution). On the other hand, domestication and the transition to new agricultural environments may cause changes in the fitness effects of alleles such that an allele historically driven to low frequency by evolution may become favorable. To test the effect of increased recombination on population improvement we tested five different QTL effect sign distribution conditions:

1. Additive effect signs are positive for all QTL (major allele favorable for all QTL)
2. Random 2/3 of additive effect signs are positive and 1/3 are negative for QTL

3. Random 1/2 of additive effect signs are positive and 1/2 are negative for QTL
4. 1/2 of additive effect signs are positive and 1/2 are negative for QTL, alternating positive or negative each QTL
5. Random 1/3 of additive effect signs are positive and 2/3 are negative for QTL

The founder and burned-in population (see *Breeding and selection scheme*) genetic variance were always standardized to one by dividing each QTL effect size by the square root of the population's additive genetic variance. The first condition presents an extreme scenario of coupling between every QTL, which resulted in selection against all minor alleles. The fourth condition presents another extreme scenario of repulsion between every QTL. The third scenario is random and assumes random coupling or repulsion. The remaining conditions present moderate ratios of coupling and repulsion, which test different proportions of selecting against the minor allele. In the biparental population approach we only applied conditions three and four.

Simulation parameter: recombination

After the population simulation parameters were set, and the burn-in population was generated, we initiated a genomic selection scheme (see *Breeding and selection scheme*). At the start of genomic selection, we introduced a variable that scaled the genetic map size by two or twenty-fold, either across the entire chromosome or only in in lower recombining regions compared to the WT genetic map. These map types were respectively named Chromosome, Pericentromere, and WT (Figure 4.2). The size of the genetic map is proportional to the amount of recombination, and we accounted for crossover interference with the Kosambi mapping function.

Breeding and selection scheme

Once the simulation parameters were assigned for the full founder population, we conducted ten cycles of phenotypic selection to burn-in the population. Each cycle consisted of 400 random crosses among 80 doubled haploid (DH) parents. Each cross produced one F1 progeny used to create one DH. Phenotypic selection was used to advance 80 DHs (20% selection intensity) to the next cycle. In the tenth generation all 400 DHs were advanced and assigned to the training population (TP). For the biparental population founders, after the simulation parameters were assigned the burn-in was only one cycle and all 400 unselected DHs were used to generate a TP.

At this point for the full founder or biparental TP, the WT genetic map was used, or the increased recombination parameter was applied. The TP also served as the first genomic selection candidate population. We conducted ten cycles genomic selection, beginning with 400 random crosses that produced one F_1 progeny used to create one DH each, for a total of 400 DHs. The AlphaSimR function *RRBLUP()* was used to estimate breeding values in the TP, set estimated breeding values of the DHs, and select 20 DHs (5% selection intensity) to advance to the next cycle. 160 DHs (40% selection intensity) with the greatest estimated breeding value were phenotyped and added to the existing TP each cycle, and we dropped the bottom 20% of the updated TP each cycle (i.e., the cycle 10 TP consisted of 1200 DHs).

Simulation replicates and statistical analysis

For each cycle of genomic selection we measured the additive genetic variance (AlphaSimR function *varA()*) and calculated the genetic gain (*genomic selection cycle$_x$ population mean genetic value – genomic selection cycle$_1$ population mean genetic value*) (Gaynor et al., 2021). We evaluated the prediction accuracy of the estimated breeding values for the DH population at each cycle of selection by measuring their correlation with the true genetic values. We measured the impact of the Bulmer effect on genetic variance in the DHs using the AlphaSimR functions *varA()* and *genicVarA()*. The *genicVarA()* function calculates the expected variance under Hardy-Weinberg equilibrium (HWE) for each QTL and reports the sum. It thus removes the impact of linkage disequilibrium on genetic variance. As such, we calculated Bulmer effect = *varA() / genicVarA() / 2* (the factor of 2 is because the DH population is fully inbred, and *genicVarA()* assumes HWE). A stronger Bulmer effect (lower values) indicates conditions of greater repulsion linkage between QTL, as shown by (Bulmer, 1971). We also measured the QTL fixation ratio for small (bottom third), medium (middle third), and large effect QTL (top third), and the change in QTL allele frequency.

The complete list of settings included the founder population (all 26 lines or biparental), number of QTL per chromosome (2 or 200), heritability (0.2 or 0.8), recombination frequency (2X or 20X scale), genetic map type (WT, Pericentromere, or Chromosome), QTL type (R or DV), coupling versus repulsion conditions (1, 2, 3, 4, or 5; only 3 or 4 in biparental), and the relationship matrix (GW or CV), resulting in 672 unique simulations (Supplemental /results_S2.1/). To account for the stochastic variation among simulations we ran 100 replicates for each setting. We fit univariate linear models with the R/lme4 package to the following response variables: genetic gain, additive genetic variance, Bulmer effect, prediction accuracy, QTL fixation, and QTL allele frequency (Bates et al., 2015). Separate linear models were fit at breeding cycles 6 and 10, and for the full founder and biparental simulations (Supplemental model_S2.1.md, model_S2.2.md). We chose cycle 6 as it is most comparable to a traditional breeding program duration for cultivar release, and cycle 10 for comparison over time. The models fit fixed effects to all main effects and first-order interactions between number of QTL, heritability, recombination frequency, genetic map type, QTL type, coupling versus repulsion conditions, relationship matrix, and allele (only for QTL fixation and QTL allele frequency: small, medium, large effect). The replicate simulations were fit as random effects. Analysis of variance (ANOVA) of each model and marginal means for the primary effects were evaluated. In addition post hoc comparison of least-squares means for all significant pairwise contrasts were performed with Tukey multiple-test correction, within each model using the *R/car* and *R/emmeans* packages (Supplemental /results_S2.2/) (Fox & Weisberg, 2019; Lenth et al., 2019). In Figures 4.3 & 4.4 we reported the mean of each measurement and did not show the standard error because it was smaller than the size of the symbols on each plot.

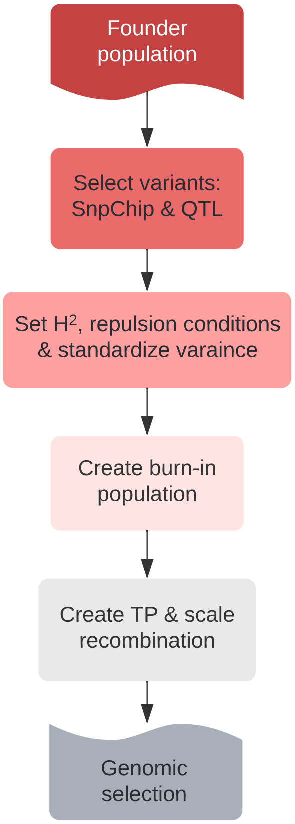
Founder population
Select variants: SnpChip & QTL
Set H², repulsion conditions & standardize varaince
Create burn-in population
Create TP & scale recombination
Genomic selection

Figure 4.1: Each simulation replicate begins with a founder population, either the full set of 26 wheat accessions or PI_220431 and PI_185715 for a biparental population. From observed variants among these founders, marker and QTL variants for simulation are selected. Simulation conditions are set (heritability and linkage disequilibrium), and the variance is standardized to one. The population is burned in using cycles of phenotypic selection and the variance is standardized. A training population (TP) is selected. Finally, the recombination is left unchanged or scaled up by two or twenty-fold across the low recombining region, or across the entire chromosome. Genomic selection is then simulated.

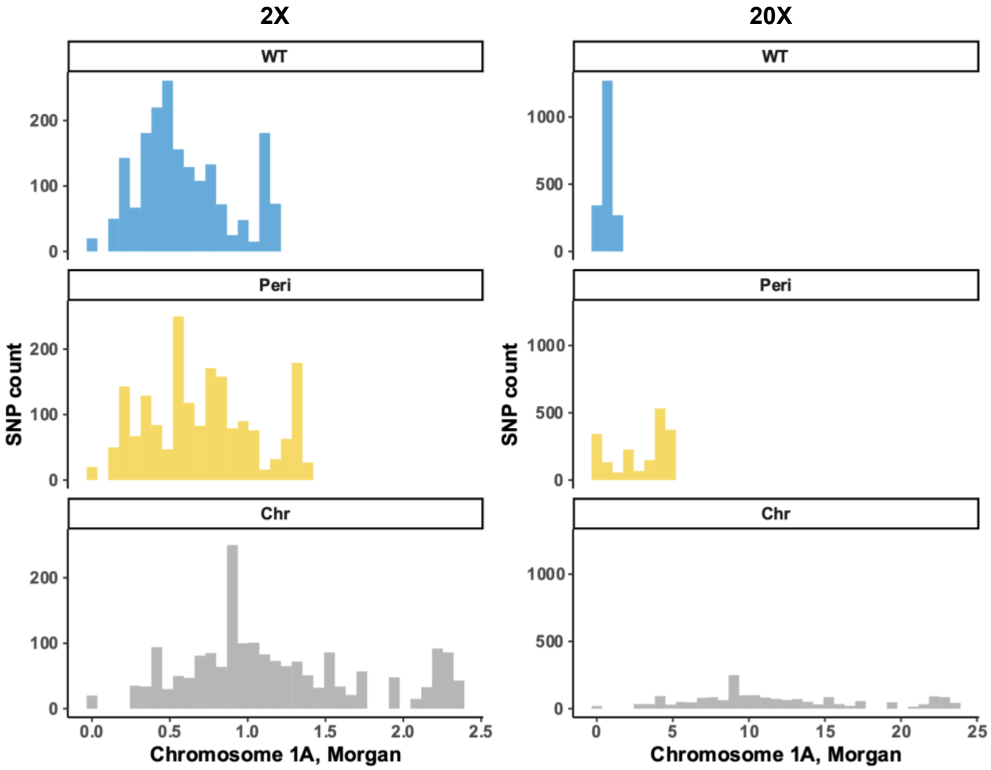

Figure 4.2: Wildtype (WT) genetic map comparison with Pericentromere (Peri) and Chromosome (Chr) genetic map types with 2X (left column) and 20X (right column) increased recombination. The X-axis is the position along Chromosome 1A in Morgans, and the Y-axis is the SNP count. The red bars represent the 0.2 M low recombination regions spanning the centromere on each chromosome that was designated for increased recombination in the Pericentromere map type. See Supplemental figure_S2.2.pdf for genetic maps of all 21 chromosomes.

Results

Comparing high and low values for a range of simulation parameters (number of QTL, heritability, recombination frequency, genetic map type, QTL type, coupling vs. repulsion) helped show which variables had the greatest effect on response to selection. Using values more extreme than realistic helped us identify mechanisms that led to change in the response variables. Recombination can break genetic linkages and change the genetic variance available, but it can also diminish LD between markers and QTL. Linkage between markers and QTL is necessary to detect QTL. The CV relationship matrix is not affected by LD between markers and QTL, and comparisons with the GW relationship matrix results helped us disentangle the effect of LD from other variables on increased recombination. Based on evolutionary and population genetics models, and existing simulation studies, we expected that elevated recombination across the simulation conditions would slow the loss of genetic diversity and increase genetic gains over subsequent cycles of genomic selection (Muller, 1964; Hill & Robertson, 1966; Gonen et al., 2017; Tourrette et al., 2019). Ultimately, we found that there was only a narrow parameter space in which increased recombination significantly preserved genetic diversity and increased genetic gains compared to WT recombination. Unless otherwise stated, the parameters considered in the following sections are full founder population, 200 QTL per chromosome, and heritability of 0.8.

Few simulation settings retain genetic variation and fewer still achieve higher genetic gain

We evaluated genetic variation and genetic gain responses to simulation parameters at genomic selection cycle 6 (typical breeding program) and 10 for comparison over time. In the ANOVA table summaries of the linear models for both response variables and cycles, the largest effects and interactions involved coupling versus repulsion, the number of QTL, heritability, and the relationship matrix (Table

4.1, Supplemental model_S2.1.md, model_S2.2.md). The smallest effects and interactions involved recombination frequency, genetic map type, and QTL type.

In the initial F1 and DH lines there were one or two COs per chromosome per meiosis in our WT genetic map, two or three COs in our 2X larger maps, and fourteen or fifteen COs in our 20X larger maps (Supplemental script_S2.1.R). Evaluation of 2X and 20X recombination rates compared to WT identified that 2X change in recombination rarely retained greater genetic variation, which did not increase genetic gains (Figures 4.3 & 4.4). For example, at cycle 6 both the Chromosome and Pericentromere map type marginal mean genetic variances were at most 12% greater than WT. This trend remained at cycle 10. However, this did not translate to significantly different genetic gains compared to WT. We only observed 2X recombination generate a difference in genetic variation and greater genetic gains compared to WT at cycle 10 when QTL were initially in repulsion (conditions 3, 4, and 5) and using the CV relationship matrix (Figure 4.3 & 4.4). The same trend for 2X recombination was present in the biparental population (Supplemental plots_S2.2.md).

Under 20X greater recombination we observed multiple simulation parameters that retained more genetic variance and had higher genetic gains than WT. Similar to the previous parameters, however, increased genetic variance did not always translate to greater genetic gains. For example, at cycle 6 both the Chromosome and Pericentromere map type with 20X increased recombination had marginal mean genetic variances that were at least 30% greater than WT. This trend was even stronger at cycle 10. Yet the marginal mean genetic gains at cycle 10 were nearly equivalent for WT and Pericentromere map type, and 2.5% less for the Chromosome map type. Comparison across all marginal mean genetic gains at cycle 10 with 20X recombination identified that the only simulation settings where both the Chromosome and Pericentromere map type outperformed WT recombination were when the CV relationship matrix was used. The GW and CV relationship matrix comparisons highlighted that gain under higher recombination (i.e., > 20 M) suffered from LD decay between markers and QTL leading to loss of prediction accuracy (see *Changing recombination frequency or distribution is more efficient when QTL locations are known*).

The most efficient response to selection compared to WT recombination was under the CV relationship matrix and 20X recombination for high repulsion (condition 4). For these parameter settings at cycle 10 the Chromosome and Pericentromere map types both retained at least 37% more genetic variance than WT. This translated to 8.3% and 6.5% greater genetic gains in the Chromosome and Pericentromere map type, respectively, compared to WT recombination. A similar trend was observed in the biparental population condition 4, but total genetic gains and percent differences from WT were smaller (Supplemental plots_S2.2.md).

Increased recombination is beneficial under repulsion, and has a marginal impact under coupling

Thus far we have primarily considered high repulsion simulations (condition 4), as this condition had the greatest response to selection. We also evaluated moderate coupling and repulsion (condition 2 and 5, respectively), random (condition 3), and high coupling (condition 1). Given Chromosome or Pericentromere map type, condition 4 had the greatest overall response to selection and difference from WT recombination compared to conditions 1, 2, 3, and 5 (Figures 4.3 & 4.4). High coupling (condition 1) performed the worst, retaining the least variance and no consistent or significant difference in genetic gain compared to WT recombination. Focusing on the CV relationship matrix results at cycle 10 under moderate coupling, the marginal mean genetic gain compared to WT was at most 2% greater for Chromosome and Pericentromere map types. Under moderate repulsion the marginal mean genetic gain compared to WT was not significantly different for the Pericentromere map type, and 1.6% greater for the Chromosome map type. Under low heritability, 20X increased recombination had a negative impact on genetic gain for all conditions except high repulsion (see *Genetic gain from increased recombination is less efficient for low heritability traits*).

We also evaluated the Bulmer effect across genomic selection cycles (Supplemental plots_2.1.md). Of the five coupling and repulsion scenarios at each cycle, condition 1 had the weakest Bulmer effect, condition 2, 3, and 5 were not consistently different, and condition 4 had the strongest Bulmer effect as it had the most to gain from recombining away from repulsion linkage. Marginal mean comparisons between Chromosome and Pericentromere map types were not significantly different for their Bulmer effect, suggesting both recombination frequency and distribution approaches are beneficial for breaking up repulsion linkages. However, the DV QTL had the strongest Bulmer effects compared to R QTL, as variants in regions of low recombination may have more repulsion linkages.

In addition, we measured the overall change of QTL allele frequency for large, medium, and small effect QTL (Supplemental plots_2.1.md). Large effect QTL had the greatest change in allele frequency while small effects experienced little change. Under high repulsion we observed the small effect QTL negative allele frequency could increase. This phenomenon of increased deleterious variant allele frequency is known as genetic hitchhiking and can occur due to LD with targets of selection where the combined effect is net positive (Moyers et al., 2018). While there was no significant difference due to recombination frequency, distribution, or QTL type on the allele frequency change in the full founder population, there was a difference in the biparental population. Most notably, for condition 4 in the full founder population only small effect QTL negative alleles hitchhiked but in the biparental population both small and medium effect QTL negative alleles hitchhiked.

Increased recombination is not beneficial for oligogenic traits

We tested simulations with 2 and 200 QTL per chromosome to compare how oligogenic and polygenic traits respond to changes in recombination, respectively. With only 2 QTL per chromosome the QTL became fixed sooner. This often occurred before genomic selection cycle 6, and genetic variance and gains plateaued. In fact, the rate of fixation of positive and negative QTL alleles showed no response to changes in heritability, recombination frequency, map type, QTL type, or relationship matrix, and only showed a response to the number of QTL per chromosome (Supplemental plots_2.1.md).

Genetic gain from increased recombination is less efficient for low heritability traits

In our simulations we modeled additive traits with high (0.8) and low (0.2) heritability. More genetic variation was retained for 2X and 20X recombination under low heritability, but the relative genetic gain was significantly less than those under high heritability simulations (Supplemental plots_2.1.md). For example, at cycle 10 across all simulation conditions the marginal mean genetic variances were 0.025 for low heritability and 0.015 for high heritability. And the marginal mean genetic gains were 3.6 and 5.7, respectively. Under low heritability the repulsion scenarios (conditions 3, 4, and 5) did not have significantly different genetic gains, and coupling conditions performed significantly worse than WT. The low heritability simulation settings had a twofold decrease in marginal mean prediction accuracy compared to high heritability. Increased recombination and low heritability may retain more genetic variation, but this did not translate to significant genetic gains compared to WT.

Changing recombination frequency or distribution is more efficient when QTL are well annotated

The Chromosome map type increased the WT recombination rate without changing the distribution, whereas the Pericentromere map type increased the WT recombination rate only in regions previously identified to have suppressed recombination (Figure 4.2) (Jordan et al., 2018). We expected that irrespective of QTL type, the Chromosome map type would perform better than Pericentromere map type because the overall genetic map is larger. We also expected the Pericentromere map type to perform better under the DV approach, compared to the R approach because the DV QTL are primarily found in the pericentromere.

Across all simulation conditions, comparison of the marginal mean genetic gain at cycle 10 for WT, Pericentromere, and Chromosome map type revealed no significant difference (4.66, 4.68, and 4.63, respectively). Under our comparison of QTL type, the R QTL had a 6.6% higher marginal mean genetic gain than DV QTL (4.8 and 4.5, respectively). As previously noted, a significant difference between map types under 20X recombination was associated with polygenic traits, high heritability, and high repulsion. Narrowing our focus to analyzing the effect of map type under these conditions at cycle 10, we recognized that the relationship matrix played significant role in the genetic gain irrespective of map or QTL type. Across these simulation settings the marginal mean genetic gain for the GW relationship matrix was 19% less than the CV relationship matrix genetic gain.

Taking these observations into account, we compared the effect of map type and QTL type for 20X recombination at cycle 10 with a polygenic trait, high heritability, and high repulsion. Under the GW relationship matrix, map types performed equivalently with R QTL. The Pericentromere map type performed the best with DV QTL, while Chromosome map type performed the same as WT (Figure 4.4). Under the CV relationship matrix, Chromosome map type always performed slightly better than Pericentromere map type, and 12.2% better than WT with R QTL and 19.2% better than WT with DV QTL. While the Chromosome map type was largest, the LD decay between SNP and QTL under the GW relationship matrix neutralized the benefits of increased recombination breaking repulsion linkages. The DV QTL marginal mean genetic gains were generally lower than R QTL, but under the CV relationship matrix they were higher than R QTL raw values and significantly better than WT for both Pericentromere and Chromosome map types. If the precise locations of QTL are known, the genetic map size and the distribution of QTL may have comparable response to selection. Note that we only evaluated DV QTL in the full founder population because the biparental population had too few SNPs and SnpEff annotations in the pericentromere.

Prediction accuracy is lower under increased recombination

Consider the GW relationship matrix for comparing the prediction accuracy across simulation parameters, as the CV relationship matrix is reflective of perfect LD between the markers and QTL (Supplemental plots_2.1.md). The marginal mean prediction accuracy at cycle 6 was 0.45 for WT recombination, 0.43 for Pericentromere map type, and 0.41 for Chromosome map type. A similar trend was observed at cycle 10. As noted previously, LD diminished between the markers and QTL for larger genetic maps, leading to lower prediction accuracy.

Effect	Df	Chi Square GV, 6	GV, 6	Chi Square GV, 10	GV, 10	Chi Square GG, 6	GG, 6	Chi Square GG, 10	GG, 10
Map type	2	696.67	***	1617.36	***	89.44	***	24.57	***
Recombination	1	499.15	***	939.62	***	125.42	***	26.09	***
QTL	1	30191.17	***	41806.41	***	13597.37	***	69683.01	***
H2	1	7659.54	***	4447.56	***	30875.10	***	47470.85	***
Repulsion	4	5681.27	***	5513.01	***	45217.37	***	53692.47	***
Matrix	1	6580.72	***	3268.55	***	8567.85	***	6438.86	***
QTL type	1	476.03	***	738.47	***	195.12	***	705.99	***
Map type : Recombination	2	210.07	***	467.67	***	54.95	***	30.75	***
Map type : QTL	2	257.03	***	999.28	***	51.16	***	21.96	***
Map type : H2	2	11.63	**	98.18	***	17.06	***	77.59	***
Map type : Repulsion	8	140.68	***	202.90	***	19.21	*	47.72	***
Map type : Matrix	2	59.46	***	111.13	***	90.79	***	173.35	***
Map type : QTL type	2	23.84	***	52.98	***	10.67	**	14.16	***
Recombination : QTL	1	143.58	***	441.54	***	31.97	***	2.65	NS
Recombination : H2	1	0.22	NS	29.91	***	4.99	*	29.01	***
Recombination : Repulsion	4	100.82	***	115.42	***	29.54	***	50.74	***
Recombination : Matrix	1	46.94	***	68.24	***	69.02	***	114.26	***
Recombination : QTL type	1	8.94	**	8.69	**	1.26	NS	8.83	**
QTL : H2	1	42.57	***	389.92	***	13363.78	***	33221.71	***
QTL : Repulsion	4	3966.33	***	4264.55	***	20914.58	***	20284.46	***
QTL : Matrix	1	884.27	***	37.86	***	328.34	***	116.74	***
QTL : QTL type	1	296.17	***	635.82	***	1.43	NS	185.36	***
H2 : Repulsion	4	667.76	***	314.96	***	921.27	***	1214.00	***
H2 : Matrix	1	861.27	***	417.54	***	77.42	***	36.11	***
H2 : QTL type	1	25.15	***	11.99	***	26.13	***	79.61	***
Repulsion : Matrix	4	636.97	***	155.26	***	1579.54	***	1117.09	***
Repulsion : QTL type	4	2.93	NS	11.90	*	135.04	***	161.25	***
Matrix : QTL type	1	20.54	***	22.32	***	12.88	***	26.49	***

Table 4.1: ANOVA table results for genetic variance (GV) and genetic gain (GG) linear models at cycles 6 and 10. Type II Wald chi-square tests significance codes: P-value < 0.0001 '***', P-value < 0.001 '**', P-value < 0.01 '*', P-value 'NS' not significant. Degrees of freedom (DF).

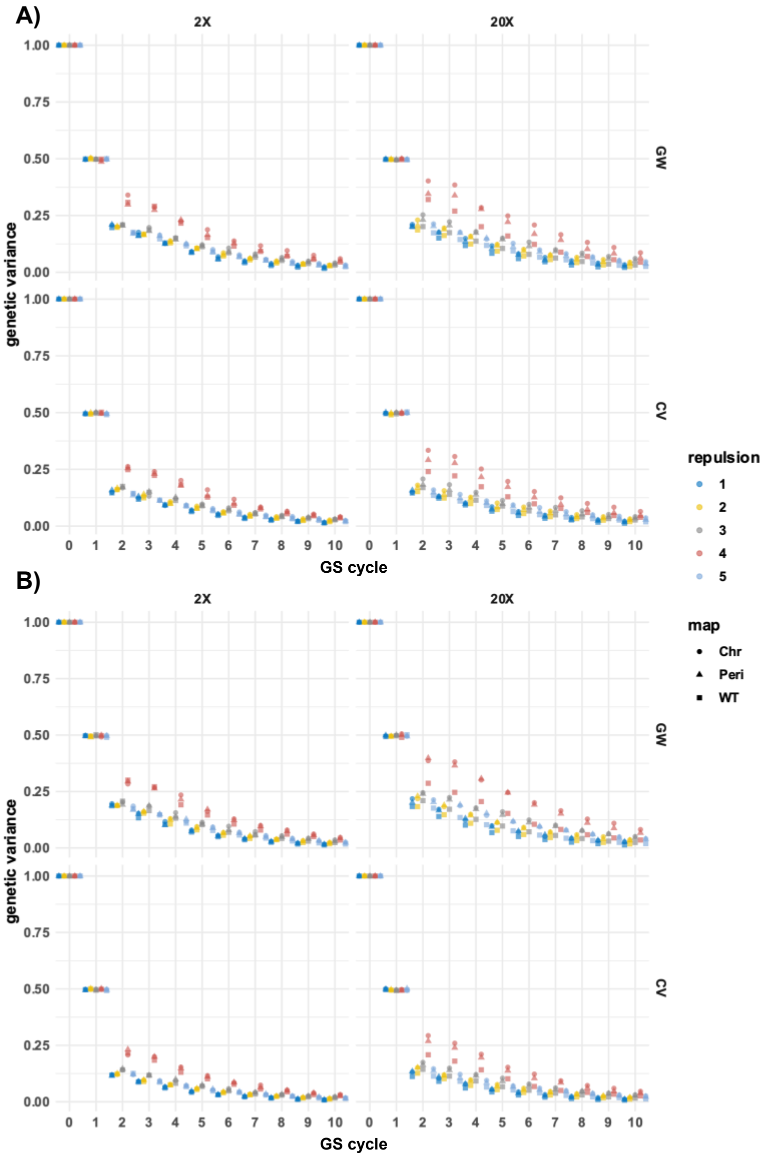

Figure 4.3: Genetic variance for the full founder population, 200 QTL per chromosome, $H^2 = 0.8$. The response is measured at the end of burn-in (0) and each cycle of genomic selection (GS) under wildtype (WT) recombination landscape map (square), pericentromere (triangle), and full chromosome (circle), 2X (left hand columns) and 20X (right hand columns) recombination frequency. The rows compare whether a genomewide (GW) or causal variant relationship matrix (CV) was used to estimate marker effects. The colors represent the five levels of coupling and repulsion (see Methods); dark blue 1: high coupling; yellow 2: moderate coupling; grey 3: random coupling or repulsion; red 4: high repulsion; 5 light blue: moderate repulsion. Each point represents the mean of 100 simulation replicates and there are no standard

error bars because they were smaller than each symbol. The symbols are jittered horizontally by the five conditions for easier visualization. A) QTL are assigned at random, B) QTL are assigned to deleterious variant annotations.

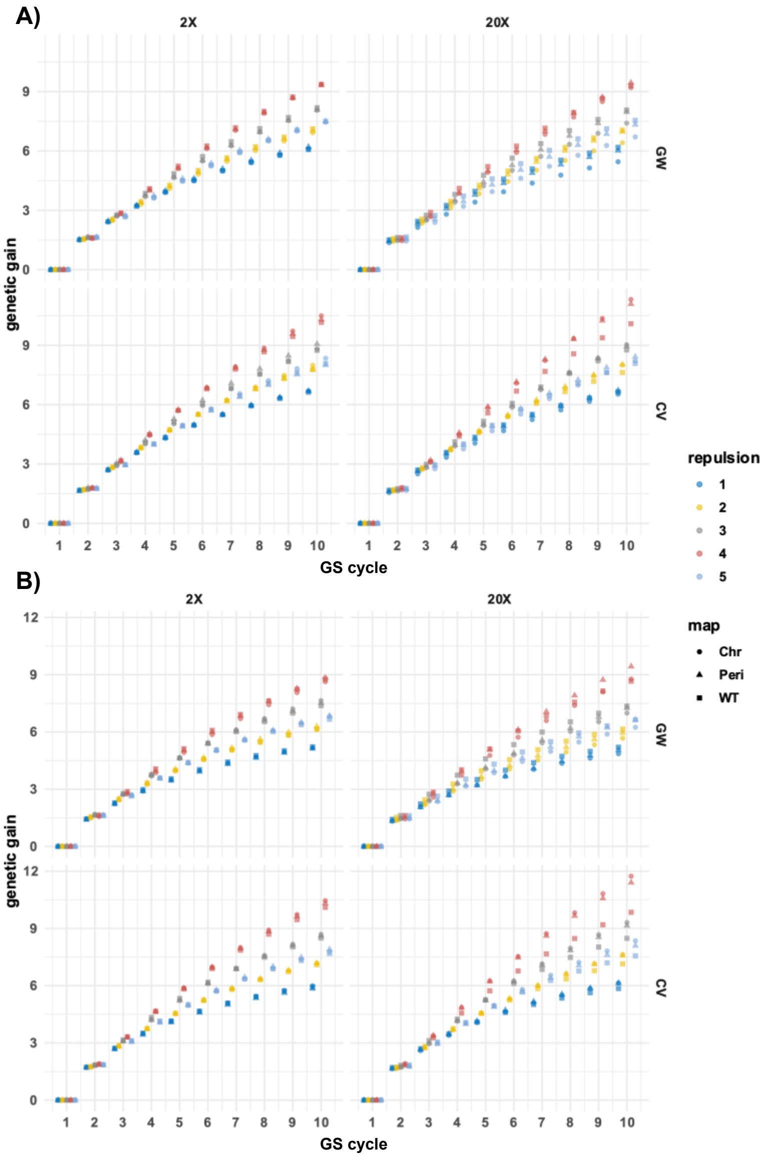

Figure 4.4: Genetic gain for the full founder population, 200 QTL per chromosome, $H^2 = 0.8$. The response (mean genetic value difference between cycle X and cycle 1) is measured at each cycle of genomic selection (GS) under wildtype (WT) recombination landscape map (square), pericentromere (triangle), and full chromosome (circle), 2X (left hand columns) and 20X (right hand columns) recombination frequency. The rows

compare whether a genomewide (GW) or causal variant relationship matrix (CV) was used to estimate marker effects. The colors represent the five levels of coupling and repulsion (see Methods); dark blue 1: high coupling; yellow 2: moderate coupling; grey 3: random coupling or repulsion; red 4: high repulsion; 5 light blue: moderate repulsion. Each point represents the mean of 100 simulation replicates and there are no standard error bars because they were smaller than each symbol. The symbols are jittered horizontally by the five conditions for easier visualization. A) QTL are assigned at random, B) QTL are assigned to deleterious variant annotations.

Discussion

For the majority of simulation parameter scenarios studied, increased recombination did not have a significant impact on genetic gain compared to WT recombination after ten cycles of genomic selection. Testing high and low values of simulation parameters was important for dissecting which combinations did have an impact on translating increased recombination to greater genetic gains.

The variance generated from increased recombination may not increase genetic gain

We observed tradeoffs in genetic gain under increased recombination due to the genetic variation generated from breaking repulsion linkages, and the variation retained from reduced prediction accuracy. Repulsion versus coupling, the relationship matrix, the number of QTL, and heritability had some of the largest effects on our measurements of genetic variance and genetic gain under increased recombination.

The effect of repulsion was most apparent in the 2X increased recombination simulations, where only condition 4 (alternating positive and negative effect alleles along the chromosome) outperformed WT recombination. This represents an extreme level of repulsion unlikely to arise in an elite breeding program but demonstrated that populations with repulsion stand to gain the most from increased recombination. It also highlights that for controlled recombination technologies to be adopted by a breeding program, they will need to achieve greater than 2X increased recombination. While this has been accomplished in Arabidopsis and some crop species, the mechanisms and response (i.e., plant viability and recombination frequency achieved) are not always conserved (Mieulet et al., 2018).

Under 20X increased recombination the coupling scenarios occasionally performed worse than WT, and likely suffered from disrupted blocks of positive alleles. Condition 4 still performed the best at cycle 6 and 10 under 20X recombination, which suggests that the benefits of increased recombination may depend on initial population conditions more so than conditions that arise inherently due to selection. Similar observations about the positive and negative response to increased recombination in repulsion and coupling conditions, respectively, have been observed but methods to evaluate repulsion linkages in a breeding population are slow and not common practice in modern breeding (Inks, 1981; Tourrette et al., 2019).

When we turned our attention to 20X increased recombination, two competing effects became clear. First, increased recombination can break repulsion or coupling linkages, changing the genetic variance available and driving genetic gain, and second, it can diminish the LD between markers and QTL necessary for genomic prediction. For example, the relationship between increased recombination on moderate repulsion, random, and coupling conditions and higher genetic gain was primarily evident with the CV relationship matrix. With the GW relationship matrix, the Chromosome map type generally performed worse than WT irrespective of repulsion conditions. This observation was associated with LD decay between the markers and QTL (data not shown), which led to lower prediction accuracy than WT. The poor performance of Chromosome map type under the GW relationship matrix was not consistent with previous studies, perhaps due to a difference in methodology (Battagin et al., 2016; Gonen et al., 2017; Tourrette et al., 2019). To realize gains from higher recombination frequencies, i.e., 20X larger genetic maps, knowledge of the locations of QTL would be beneficial, as would increasing the marker density beyond the 1200 markers per wheat chromosome used here.

Our designation of the major and minor allele as ‘1’ and ‘0’, respectively, may have contributed to plateauing genetic gains in the coupling scenarios as the initial positive QTL allele frequencies were greater than 0.5. For conditions 3, 4, and 5 under the DV QTL approach, at least half of the causal variants’ negative allele was the major allele, which is likely an overestimate of the population’s genetic load. Using an evolutionary conservation approach such as a Genomic Evolutionary Rate Profiling score to designate the preferred versus deleterious allele for the SnpEff variants would be more accurate for characterizing the population’s genetic load, but it is not clear if this would have lessened the baseline coupling in the founder population.

We learned that increased recombination simulations with polygenic traits (200 QTL per chromosome) performed better than oligogenic traits (2 QTL per chromosome) due to slower rates of QTL fixation. There was no difference in response variables from WT recombination across the oligogenic trait simulation parameters. For the polygenic trait simulations, allele fixation occurred sooner for WT compared to increased recombination only for the biparental founder population, likely due to higher initial LD. The trend that genetic gain increased with more QTL had previously been observed by other simulation studies that increased recombination frequency (Battagin et al., 2016; Tourrette et al., 2019). When there are many QTL the loss of genetic variance is associated with the Bulmer effect, and when there are few QTL, the loss is associated with allele frequencies more quickly approaching fixation. Consequently, generally increasing the frequency or shifting the distribution of recombination may not be appropriate for oligogenic trait architecture. For example, controlled recombination simulations that use genome editing to target COs to specific intervals along each chromosome, rather than modifying the overall CO rate and distribution across many generations, may be suitable for oligogenic traits (Bernardo, 2017; Brandariz & Bernardo, 2019; Ru & Bernardo, 2019, 2020).

In our study heritability also mediated the translation of increased recombination to greater genetic gains. The polygenic trait and low heritability simulation parameters are comparable to the genetic architecture of yield. The marginal mean genetic gain of a polygenic trait at cycle 10 with low heritability for the Chromosome and Pericentromere map type were actually less than WT. Our results suggest that increased recombination is not a promising strategy for low heritability quantitative traits, e.g., yield improvement. This is different from previous controlled recombination simulation findings in rice and turnip, where the response to increased recombination under 0.8 and 0.2 heritability was very similar (Tourrette et al., 2019). Differences in our results for low heritability traits from existing studies may be explained by variations in methodology.

The efficiency of increased recombination may depend on knowledge of QTL locations

Previous simulations have shown that significantly more recombination in low recombining regions increased the efficiency of selection (Gonen et al., 2017; Tourrette et al., 2019). However, when marker effects are less sensitive to repulsion linkages in regions of low recombination, how accurately can simulations evaluate the effect of increased recombination? We developed the Chromosome and Pericentromere map types, in combination with the R and DV QTL approach to test the value of recombination in regions of the genome where historical recombination had not made QTL effects apparent. Note that this approach required a large and diverse founder population (i.e., compared to the biparental) to have sufficient marker density and SnpEff annotations to designate DV QTL.

Contrary to our original expectations, we found that the map type and QTL type had some of the smallest effects on genetic variation and genetic gain under increased recombination. Generally, the R QTL performed better than the DV QTL because they are more evenly distributed. However, when the QTL positions are known (CV relationship matrix) the DV QTL performed better irrespective of the increased map type. Indeed, enhancing recombination in regions where causal variants are concentrated is more likely to aid in selection (Gonen et al., 2017).

While revealing currently inaccessible genetic diversity with more recombination in the pericentromere is an exciting prospect, our simulations with data from real wheat populations indicate it is most beneficial to know the location of causal variants. This information may not be available for many traits in a typical breeding program today, but with advances in genome editing, sequencing, and deleterious variant annotation, identification of variants that underlie quantitative traits in the coming decades may increase.

Feasibility of increased recombination for wheat genomic selection breeding programs

A primary assumption that we made during our simulations was that increasing the frequency and shifting the distribution of recombination in wheat is biologically feasible. Assuming a controlled recombination method is successfully applied to wheat, our results are markedly less promising than previous simulations with comparable parameters, which have reported upwards of 34% greater gains compared to WT recombination (Tourrette et al., 2019). While we are not sure why our increased recombination simulation study had a lower response to selection, Tourrette et al., (2019) did detect a difference in the efficiency of modifying recombination across species (turnip versus rice). Working with wheat, our specific founder population structure, as well as R packages used to generate controlled recombination simulation parameters, could be responsible for the deviation in our results.

Many of the simulation parameters tested in this study (e.g., the five coupling and repulsion conditions) were included because they helped to diagnose the reason for limited response to selection under increased recombination. One persistent limitation was the accuracy of genomic selection under increased recombination, which required retraining our model at each cycle of selection. We recognize that this would be an added cost for a breeding program. In addition, a differential response in genetic gain was not always detected by genomic selection cycle 6, hence we ran our simulations through cycle 10. However, the results at cycle 6 are more representative of the time to deliver a cultivar given traditional breeding methods, and the rate of technology advancement. The time to introduce the technology to a breeding program (e.g., tissue culture methods) is another cycle of generations to consider.

The cost and resources required for adopting a genome-editing mediated technology may be prohibitive for wheat breeding programs. Our simulations showed that the benefits of increased recombination were only realized if the recombination rate was 20-fold greater than WT for many generations. However too much recombination can cause segregation problems and decrease fertility (Pelé et al., 2017; Mieulet et al., 2018). Additionally, requiring many generations of selection may not be appropriate for all trait architectures or outweigh the benefits of traditional plant breeding methods.

Conclusion

Given the potential applications of controlled recombination to future breeding programs, this study was motivated by the need to define desirable recombination intervals in regions of the genome where historically very few COs have been detected. Ultimately our comparison of the recombination frequency and distribution, as well as the QTL annotation to better understand the impact of increased recombination had little impact on response to selection. The initial conditions of the breeding population, especially repulsion linkages, polygenic trait architecture, and heritability had a greater influence on selection under increased recombination. We also identified that the increased genetic variation generated from more recombination may be associated with loss of genomic prediction accuracy (GW relationship matrix) rather than broken repulsion linkages (CV relationship matrix), which narrows the conditions in which controlled recombination may produce greater genetic gains. Collectively, the outcomes of this research challenge whether a controlled recombination application to genomic selection in wheat offers a more efficient path to retaining genetic variation and increasing genetic gains compared to existing breeding methods.

Acknowledgements

We thank Dr. Eduard Akhunov's lab for providing access to data from (Jordan et al., 2018). We are grateful for the contributions of Dr. Wojtek Pawlowski, Ruth Epstein, and Elise Tomaszewski for constructive comments and discussion. Funding for this project was provided by USDA National Institute of Food and Agriculture grant 2017-67007-25939 and 2022-68013-36439 (Wheat CAP).

Supplemental material

All the data referenced in this manuscript, supplementary files, custom functions, and scripts for reproducing results are publicly available at: https://github.com/etaagen/dissertation_chapter_4.

REFERENCES

Baer, C.F., Miyamoto, M.M., & Denver, D.R. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics 2007 8:8*, *8*, 619–631. https://doi.org/10.1038/NRG2158

de Baets, G., van Durme, J., Reumers, J., Maurer-Stroh, S., Vanhee, P., Dopazo, J., Schymkowitz, J., & Rousseau, F. (2012). SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Research*, *40*, D935. https://doi.org/10.1093/NAR/GKR996

Bates, D., Mächler, M., Bolker, B.M., & Walker, S.C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, . https://doi.org/10.18637/jss.v067.i01

Battagin, M., Gorjanc, G., Faux, A.-M., Johnston, S.E., & Hickey, J.M. (2016). Effect of manipulating recombination rates on response to selection in livestock breeding programs. *Genetics Selection Evolution*, *48*, 44. https://doi.org/10.1186/s12711-016-0221-1

Bernardo, R. (2017). Prospective targeted recombination and genetic gains for quantitative traits in maize. *The Plant Genome*, *10*, 1–9. https://doi.org/10.3835/plantgenome2016.11.0118

Brandariz, S.P., & Bernardo, R. (2019). Predicted genetic gains from targeted recombination in elite biparental maize populations. *The Plant Genome*, *12*, 180062. https://doi.org/10.3835/plantgenome2018.08.0062

Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś., Yandell, B.S., & Churchill, G.A. (2019). R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. *Genetics*, *211*, 495. https://doi.org/10.1534/GENETICS.118.301595

Bulmer, M. (1971). The effect of selection on genetic variability. *Am Nat.* , *105*, 201–211

Chapman, J.A., Mascher, M., Buluç, A., Barry, K., Georganas, E., Session, A., Strnadova, V., Jenkins, J., Sehgal, S., Oliker, L., Schmutz, J., Yelick, K.A., Scholz, U., Waugh, R., Poland, J.A., Muehlbauer, G.J., Stein, N., & Rokhsar, D.S. (2015). A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology*, *16*, 26. https://doi.org/10.1186/s13059-015-0582-8

Charlesworth, B., & Barton, N.H. (1996). Recombination load associated with selection for increased recombination. *Genetical research*, *67*, 27–41. https://doi.org/10.1017/S0016672300033450

Conover, J.L., & Wendel, J.F. (2022). Deleterious Mutations Accumulate Faster in Allopolyploid Than Diploid Cotton (Gossypium) and Unequally between Subgenomes. *Molecular Biology and Evolution*, *39*. https://doi.org/10.1093/MOLBEV/MSAC024

Corem, S., Doron-Faigenboim, A., Jouffroy, O., Maumus, F., Arazi, T., & Bouchéc, N. (2018). Redistribution of CHH methylation and small interfering RNAs across the genome of tomato ddm1 mutants. *The Plant Cell*, *30*, 1628–1644. https://doi.org/10.1105/tpc.18.00167

Endelman, J.B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome*, . https://doi.org/10.3835/plantgenome2011.08.0024

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, *78*, 737–756

Fernandes, J.B., Séguéla-Arnaud, M., Larchevêque, C., Lloyd, A.H., & Mercier, R. (2018). Unleashing meiotic crossovers in hybrid plants. *Proceedings of the National Academy of Sciences*, *115*, 2431–2436. https://doi.org/10.1073/PNAS.1713078114

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression*. Third. Sage, Thousand Oaks, CA.

Gaynor, C.R., Gorjanc, G., & Hickey, J.M. (2021). AlphaSimR: An R package for breeding program simulations. *G3: Genes, Genomes, Genetics*, *11*. https://doi.org/10.1093/G3JOURNAL/JKAA017

Gonen, S., Battagin, M., Johnston, S.E., Gorjanc, G., & Hickey, J.M. (2017). The potential of shifting recombination hotspots to increase genetic gain in livestock breeding. *Genetics Selection Evolution*, *49*, 55. https://doi.org/10.1186/s12711-017-0330-5

Hayut, S.F., Melamed Bessudo, C., & Levy, A.A. (2017). Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nature Communications*, *8*, 15605. https://doi.org/10.1038/ncomms15605

Henderson, I.R., & Bomblies, K. (2021). Evolution and Plasticity of Genome-Wide Meiotic Recombination Rates. https://doi.org/10.1146/annurev-genet-021721

Hill, W.G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research*, *8*, 269–294. https://doi.org/10.1017/S0016672300010156

Inks, J.L.J. (1981). The genetic framework of plant breeding. *Trans. R. Soc. Lond. B*

Johnsson, M., Chris Gaynor, R., Jenko, J., Gorjanc, G., de Koning, D.-J., & Hickey, J.M. (2019). Removal of alleles by genome editing (RAGE) against deleterious load. *Genet Sel Evol*, *51*, 14. https://doi.org/10.1186/s12711-019-0456-8

Jordan, K.W., Wang, S., He, F., Chao, S., Lun, Y., Paux, E., Sourdille, P., Sherman, J., Akhunova, A., Blake, N.K., Pumphrey, M.O., Glover, K., Dubcovsky, J., Talbert, L., & Akhunov, E.D. (2018). The genetic architecture of genome-wide recombination rate variation in allopolyploid wheat revealed by nested association mapping. *The Plant Journal*, *95*, 1039–1054. https://doi.org/10.1111/tpj.14009

Kono, T.J.Y., Lei, L., Shih, C.H., Hoffman, P.J., Morrell, P.L., & Fay, J.C. (2018). Comparative genomics approaches accurately predict deleterious variants in plants. *G3: Genes, Genomes, Genetics*, *8*, 3321–3329. https://doi.org/10.1534/g3.118.200563

Kono, T.J.Y., Liu, C., Vonderharr, E.E., Koenig, D., Fay, J.C., Smith, K.P., & Morrell, P.L. (2019). The fate of deleterious variants in a barley genomic prediction population. *Genetics*, *213*, 1531–1544. https://doi.org/10.1534/genetics.119.302733

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package “emmeans.” https://doi.org/10.1080/00031305.1980.10483031 de Los Campos, G., Vazquez, A.I., Fernando, R., Klimentidis, Y.C., Sorensen, D., & Goddard, M.E. (2013). Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. https://doi.org/10.1371/journal.pgen.1003608

Lu, J., Tang, T., Tang, H., Huang, J., Shi, S., & Wu, C.-I. (2006). The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends in Genetics*, *22*, 126–131. https://doi.org/10.1016/j.tig.2006.01.004

Mieulet, D., Aubert, G., Bres, C., Klein, A., Droc, G., Vieille, E., Rond-Coissieux, C., Sanchez, M., Dalmais, M., Mauxion, J.-P., Rothan, C., Guiderdoni, E., & Mercier, R. (2018). Unleashing meiotic crossovers in crops. *Nature Plants*, *4*, 1010–1016. https://doi.org/10.1038/s41477-018-0311-x

Moyers, B.T., Morrell, P.L., & McKay, J.K. (2018). Genetic costs of domestication and improvement. *Journal of Heredity*, *109*, 103–116. https://doi.org/10.1093/jhered/esx069

Muller, H.J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *1*, 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Ohta, T. (1972). Population size and rate of evolution. *Journal of Molecular Evolution 1972 1:4*, *1*, 305–314. https://doi.org/10.1007/BF01653959

Ohta, T. (1992). The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics*, *23*, 263–286. https://doi.org/10.1146/ANNUREV.ES.23.110192.001403

Oyetunde, T., & Bernardo, R. (2020). Linear, funnel, and multiple funnel schemes for stacking chromosomes that carry targeted recombinations in plants. *Theoretical and Applied Genetics*, *133*, 3177–3186. https://doi.org/10.1007/s00122-020-03663-4

Pelé, A., Falque, M., Trotoux, G., Eber, F., Nègre, S., Gilet, M., Huteau, V., Lodé, M., Jousseaume, T., Dechaumet, S., Morice, J., Poncet, C., Coriton, O., Martin, O.C., Rousseau-Gueutin, M., & Chèvre, A.M. (2017). Amplifying recombination genome-wide and reshaping crossover landscapes in Brassicas. *PLOS Genetics*, *13*, e1006794. https://doi.org/10.1371/JOURNAL.PGEN.1006794

Presting, G.G. (2018). Centromeric retrotransposons and centromere function. *Current Opinion in Genetics and Development*, *49*, 79–84. https://doi.org/10.1016/j.gde.2018.03.004

R Core Team. (2020). R: A Language and Environment for Statistical Computing

Rey, M.-D., Martín, A.C., Smedley, M., Hayta, S., Harwood, W., Shaw, P., & Moore, G. (2018). Magnesium increases homoeologous crossover frequency during meiosis in ZIP4 (Ph1 Gene) mutant wheat-wild relative hybrids. *Frontiers in Plant Science*, *9*, 509. https://doi.org/10.3389/fpls.2018.00509

Rodgers-Melnick, E., Elshire, R.J., Li, Y., Bradbury, P.J., Mitchell, S.E., Li, C., Glaubitz, J.C., Buckler, E.S., & Acharya, C.B. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences*, *112*, 3823–3828. https://doi.org/10.1073/pnas.1413864112

Ru, S., & Bernardo, R. (2019). Targeted recombination to increase genetic gain in self-pollinated species. *Theoretical and Applied Genetics*, *132*, 289–300. https://doi.org/10.1007/s00122-018-3216-1

Ru, S., & Bernardo, R. (2020). Predicted genetic gains from introgressing chromosome segments from exotic germplasm into an elite soybean cultivar *133*, 605–614. https://doi.org/10.1007/s00122-019-03490-2

Sadhu, M.J., Bloom, J.S., Day, L., & Kruglyak, L. (2016). CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science (New York, N.Y.)*, *352*, 1113–1116. https://doi.org/10.1126/science.aaf5124

Serra, H., Lambing, C., Griffin, C.H., Topp, S.D., Nageswaran, D.C., Underwood, C.J., Ziolkowski, P.A., Séguéla-Arnaud, M., Fernandes, J.B., Mercier, R., & Henderson, I.R. (2018). Massive crossover elevation via combination of HEI10 and recq4a recq4b during Arabidopsis meiosis.. *Proceedings of the National Academy of Sciences*, *115*, 2437–2442. https://doi.org/10.1073/pnas.1713071115

Taagen, E., Bogdanove, A.J., & Sorrells, M.E. (2020). Counting on Crossovers: Controlled Recombination for Plant Breeding. *Trends in Plant Science*, . https://doi.org/10.1016/j.tplants.2019.12.017

Tam, S.M., Hays, J.B., & Chetelat, R.T. (2011). Effects of suppressing the DNA mismatch repair system on homeologous recombination in tomato. *Theoretical and Applied Genetics*, *123*, 1445–1458. https://doi.org/10.1007/s00122-011-1679-4

Tourrette, E., Bernardo, R., Falque, M., & Martin, O.C. (2019). Assessing by Modeling the Consequences of Increased Recombination in Recurrent Selection of Oryza sativa and Brassica rapa. https://doi.org/10.1534/g3.119.400545

Underwood, C.J., Choi, K., Lambing, C., Zhao, X., Serra, H., Borges, F., Simorowski, J., Ernst, E., Jacob, Y., Henderson, I.R., & Martienssen, R.A. (2018). Epigenetic activation of meiotic recombination near Arabidopsis thaliana centromeres via loss of H3K9me2 and non-CG DNA methylation.. *Genome research*, *28*, 519–531. https://doi.org/10.1101/gr.227116.11