

SUBGENOME DOMINANCE AND GENOME EVOLUTION IN ALLOPOLYPLOIDS

By

Kevin Andrew Bird

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Horticulture – Doctor of Philosophy
Ecology, Evolutionary Biology, and Behavior – Dual Major

2022

ABSTRACT

SUBGENOME DOMINANCE AND GENOME EVOLUTION IN ALLOPOLYPLOIDS

By

Kevin Andrew Bird

The merger of divergent genomes, via hybridization or allopolyploidization, frequently results in a 'genomic shock' that induces a series of rapid genetic and epigenetic modifications as a result of conflicts between parental genomes. This conflict among the subgenomes routinely leads one subgenome to become dominant over the other subgenome(s), resulting in subgenome biases in gene content and expression. Recent advances in methods to analyze hybrid and polyploid genomes with comparisons to extant parental progenitors have allowed for major strides in understanding the mechanistic basis for subgenome dominance. In particular, our understanding of the role that homoeologous exchange might play in subgenome dominance and genome evolution is quickly growing. Here I present novel work in several polyploid species investigating the biological and evolution impact of polyploidy and the evolution of these polyploid species. The first chapter introduces concepts like whole-genome duplication and describes advances in genomic sequencing technology that have accelerated the study of polyploid genomes. The second chapter reviews subgenome dominance and recent breakthroughs in understanding its causes and implications for genome evolution. The third chapter explores the repeatability of subgenome dominance in independently resynthesized *Brassica napus*. The fourth chapter investigates the extent to which genomic rearrangements from chromosomal duplications and deletions and homoeologous exchange can bias the analysis of subgenome expression dominance from RNAseq data. The fifth chapter explores the prevalence and impact of homoeologous exchange on independently resynthesized *Brassica napus*, providing novel evidence that gene dosage changes from homoeologous exchange are constrained by the need to maintain dosage balance of gene products. The sixth chapter explores the origins and admixture of wild octoploid strawberries *Fragaria virginiana* and *Fragaria chiloensis* with newly generated genomic resources applied to global collections.

ACKNOWLEDGMENTS

This dissertation can hardly be considered solely my work; it is the result of all the communities and people who helped and supported me throughout my life and have allowed me to be where I am today. It's an honor and blessing to be able to recognize them here and engage in an ethic of gratitude, acknowledgement and reciprocity. This list is inevitably incomplete. If you feel you have been omitted, you are correct, and I promise this is due to poor memory or space constraints and not a personal slight.

I am first and foremost thankful to my mother, Amy Bird, who made countless personal and financial sacrifices to make my life better and my father, David Bird, who has never failed to let me know how proud he is of me. I am forever thankful to my pediatric cardiologist, Dr. Zuhdi Lababidi, whose innovative medical research and diligent care gave me the life I have today while also instilling the sense of appreciation and wonder toward research that set me on my path to pursue science as a career. I am also thankful for my current cardiologist, Dr. Timothy Cotts, whose care has helped me navigate the latest transition in my health.

Many people throughout my early education contributed to my growth and development as a student and as a person and I am thankful for each of them. I want to specifically thank Shelley Creed for the systems and approaches she brought to the Camdenton school's gifted program that helped a hyperactive monster of a kid become a reasonably well-adjusted, self-motivated learner. I also want to thank Chris Reeves for his mentorship, the work he put in to the Science Research program at Camdenton High School that first exposed me to formal research, and for his continued friendship and support.

I want to thank J. Chris Pires for his invaluable mentorship at Mizzou. He not only helped me learn to navigate the hidden curriculum of academia but also taught me so much about being a good scientist, mentor, and person. I have benefitted greatly from his support and friendship. Thank you also to the Pires lab alums (R. Shawn Abarhams, McKenzie Mabry, Michelle Tang, Wade Dismukes, Dustin Mayfield) during my time for their friendship and contribution to my growth as a scientist. I also want to thank Tim Parshall, Linda Blockus, and Michael Cohen for the incredible work they have done in Mizzou's

Fellowship and Undergraduate Research offices and for their individual help on the fellowship applications that made my graduate work possible.

I want to thank my friends and colleagues at MSU (Elizabeth Alger, Jordan Brock, Julia Brose, Marivi Colle, Anna Haber, Charity Goekeritz, MacKenzie Jacobs, McKenna Lipham, Serena Lotreck, Rose Marks, Jeremy Pardo, Kathleen Rhoades, Scott Teresi, Alan Yocca) for support, distractions, and commiserations throughout graduate school. Thank you to everyone I met and engaged with in the Graduate Employees Union, especially my fellow executive board members (Alex Aaring, Acacia Ackles, Darren Incorvaia, Stephie Kang, Nick Rowe, Shawna Rowe, McKayla Sluga), valued leaders (Malu Castro, AY Odedeyi), and staff (Jordan Lindsay, Shelby Krohn, and Jolyse Race).

I also want to thank my friends outside of MSU. Thank you Kevin Carr, Steve Dawson, Dan Leo, Jason Clark, and Zack Newman for our continued friendship, mutual support, and dedication to several bits over the years despite our paths taking us across the country. Thank you also to James McGuire, Dallas Duncan, Brian Lanigan, and Ryan Saunders for bringing me in to your weekly creative outlets and distractions from whatever stresses are afflicting us. Thank you to the Twitter group DM that has been an unexpected place of support, fun, and decompression. I also want to thank my various haters, who have consistently been my motivators.

Finally thank you to my PhD advisors, Patrick Edger and Bob VanBuren, who have been incredible sources of mentorship and support throughout the last five years. I couldn't have asked for better role-models of how to do good science and run a lab with compassion and integrity. Thank you both for encouraging, or at least tolerating, my independence and various side projects/distractions and for assuaging my numerous anxieties about my future in academia. Thank you also to my committee members, Malia Gehan, Ning Jiang, Emily Josephs, and Chad Niederhuth for their support, encouragement, and occasional collaborations. Thank you to the office staff I've interactive with at MSU (Barbara Bloemers, Melissa Del Rio, Morgan Fowler, Meghan Hill, Sherry Mulvaney, Greta Orzula) and all other workers at MSU whose labor was essential to a smooth experience. I also thank the many coauthors on works I've published for their contributions.

TABLE OF CONTENTS

LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
CHAPTER 1	1
Introduction.....	2
REFERENCES	6
CHAPTER 2	10
The Causes and Consequences of Subgenome Dominance in Hybrids and Recent Polyploids..	11
Abstract	11
Summary	12
CHAPTER 3	13
Replaying the Evolutionary Tape to Investigate Subgenome Dominance in Allopolyploid <i>Brassica napus</i>	14
Abstract	14
Summary	15
CHAPTER 4	16
The Role of Genomic Rearrangements in Biasing Analysis of Subgenome Dominance.....	17
Abstract	17
Introduction	17
Methods	19
Sequencing data.....	19
Genomic rearrangement analysis	19
Homoelog expression bias	19
Results	20
Genomic rearrangements in these resynthesized lines are highly variable and do not show signs of subgenome bias	20
Impact of dosage changes on homoeologous expression bias	20
Discussion	24
REFERENCES	28
CHAPTER 5	32
Gene Dosage Constraints Affect the Transcriptional Response to Allopolyploidy and Homoeologous Exchange in Resynthesized <i>Brassica napus</i>	33
Abstract	33
Introduction	33
Methods	36
Sequencing data	36
Dosage response to polyploidy.....	37
Dosage sensitivity assignment	37
Polyploid response variance	37
Homoeologous exchange response variance.....	38
Results	38
Assessing early gene expression response to dosage changes from allopolyploidy.....	38
Expression changes from homoeologous exchanges appear to behave according to the gene-balance hypothesis	42
Expression changes from homoeologous exchanges are distinct from the effect of polyploidy	45

Discussion	48
Evolutionary dynamics of early expression response to allopolyploidy	49
Homoeologous exchange and early polyploid genome evolution	52
REFERENCES	55
CHAPTER 6	61
Diversification, Spread, and Admixture of Octoploid Strawberry in the Western Hemisphere.....	62
Abstract	62
Summary	63

LIST OF TABLES

Table 4.1 Chi-squared test for bias in direction of gene dosage changes.....	22
Table 4.2 Homoeolog Expression Bias with and without Genomic Rearrangement (Grs) chi-squared table.....	25
Table 5.1 Kruskal-Wallis test exploring the difference in expression coefficient of variation from homoeologous exchange and allopolyploidy induced dosage changes broken down by dosage sensitivity and subgenome bias.....	47
Table 5.2 Kruskal-Wallis test exploring the difference in expression coefficient of variation from homoeologous exchange and allopolyploidy induced dosage changes broken down by dosage sensitivity and generation.....	48

LIST OF FIGURES

Figure 4.1 Variability of gene dosage changes and hotspots in resynthesized <i>B. napus</i>	21
Figure 4.2 Impact of homoeologous exchange on subgenome dominance.....	24
Figure 5.1 Expression response to polyploid induced dosage changes.....	40
Figure 5.2 Expression changes from allopolyploidy reflect predictions from the dosage balance hypothesis.....	41
Figure 5.3 Expression response to non-reciprocal homoeologous exchange induced dosage changes...	43
Figure 5.4 Expression changes from non-reciprocal homoeologous exchange reflect predictions from the dosage balance hypothesis.....	44
Figure 5.5 Expression responses from allopolyploidy and homoeologous exchange appear to be distinct.....	46

CHAPTER 1

Introduction

The completion of the first reference genomes for complex eukaryotic species at beginning of the 21st century marked a turning point in the biological sciences (Goffeau et al. 1996; The C. elegans Genome Consortium, 1998; Adams, 2000; Kaul et al. , 2000; Lander et al. 2001; Venter et al. 2001; Waterson et al., 2002). While early optimism that with the completion of the human genome “*the genetic messages encoded within our DNA molecules will provide the ultimate answers to the chemical underpinnings of human existence*” (Watson et al. 1990) and that it would herald a new age of medical cures was largely misplaced (Lewontin, 1992), it is beyond a doubt that a wealth of evolutionary knowledge was gained within- and across kingdoms of life from subsequent genomic analyses. In plants, the *Arabidopsis thaliana* genome cemented the presence and prevalence of ancient whole-genome duplications (WGD) along the angiosperm phylogeny (Blanc et al. 2000; Kaul et al., 2000; Paterson et al. 2000; Blanc, Hokamp, and Wolfe, 2003; Bowers et al. 2003). Subsequent plant genomes over the next decade would reveal more about genome evolution in plants; the role of WGDs in producing genomic, regulatory, and phenotypic complexity; and facilitate the transfer of functional genomic information to non-model plant species (De Bodt, Maere, and Van de Peer, 2004; Blanc and Wolfe, 2004; Freeling and Thomas, 2006; Thomas, Pedersen, and Freeling,, 2006; Paterson et al. 2010; Jiao et al. 2011).

However, this early ‘golden age’ of plant genome sequencing (Paterson et al. 2010) was hindered by the technological, logistic, and financial barriers of the time. Reference quality genomes like *Arabidopsis* and rice (*Oryza sativa*) required expensive and complex approaches that took years for completion, while later ‘draft assemblies’ from more tractable short-read sequencing and assembly technologies produced largely incomplete and fragmented genomes (Michael and VanBuren, 2015). These early short-read approaches particularly struggled with large, highly heterozygous, or polyploid genomes (Michael and VanBuren, 2015). Only within the last decade were polyploid genomes feasible with cultivated Cotton (*Gossypium arboreum*; Li et al. 2014), *Brassica napus* (Chalhoub et al. 2014) and wheat (*Triticum aestivum*; International Wheat Genome Sequencing Consortium, 2014) all published in 2014 after grueling multi-year projects. In 2015, VanBuren et al. (2015) published the first genome assembled only from long-read technology from PacBio, marking a new-age in plant genomics (Michael and VanBuren, 2020). Over the last three years, nearly 75% of all existing plant genomes were

sequenced and a demonstrated a 32-fold improvement in contig N50 (Marks et al. 2021). The affordances of long-read technology and their rapidly improving quality and lowering cost has also mean that it is for the first time possible to produce chromosome-scale genomes for complex polyploid species at a fraction of the cost and time of the first polyploid genome projects (Zhang et al. 2018; Colle et al. 2019; Edger et al. 2019; Chen et al. 2020; VanBuren et al. 2020; Marks et al. 2021). This expanded capacity of genomic technology has also ushered in pan-genomics as a new paradigm, moving beyond a single reference-genome representation of a species and centering the study of genomic structural variation within a species (Bayer et al. 2020; Danilevicz et al. 2020; Golicz et al. 2020).

Research about the prevalence and impact of whole-genome duplication benefitted immensely from the wealth of genomic data over the last two decades. These results have generated theories about the relationship between whole-genome duplication and evolution of novel traits (Edger et al., 2015; Van de Peer et al., 2017; Qi et al., 2021) and with species diversification (Schranz et al., 2012; Landis et al., 2018). The intersection of these two phenomena represented by Edger et al.'s (2015) genomic investigation of the coevolutionary arms race between plants of the order *Brassicales* and butterflies of the subfamily *Pierinae*. Edger et al. (2015) showed the role of genome duplications in elaborating chemical herbivory defenses of plants in the mustard family (*Brassicaceae*), the subsequent increase species diversification in response to improved defenses, and the concomitant burst of gene duplications in cabbage butterflies that allowed them to defuse the plant's defense and produced its own increase in species diversification rates.

The realization of the widespread prevalence of polyploidy across higher eukaryotes spurred new areas of research on genome evolution, including into how polyploids repeatedly return to a diploid-like state via various processes, collectively called 'diploidization' (Conant et al., 2014). Over the past two decades, numerous studies of the diploidization process in diverse polyploids yielded valuable insights into the underlying mechanisms of duplicate gene retention and the functional divergence of duplicate genes. Chief among these findings are how evolutionary constraints on gene dosage balance (the preservation of stoichiometry among interacting regulatory gene products) and subgenome dominance (the asymmetric expression and regulation of subgenomes in allopolyploids) produce biases in gene retention following WGD. Studies on polyploid genome evolution found that genes that are sensitive to

changes in relative dosage (dosage-sensitive genes) are preferentially retained as duplicates after whole-genome duplication over long evolutionary timescales, leading to increases in copy number and expansions of gene families related to complex regulatory activity (Birchler & Veitia, 2012). Other work demonstrated one of the parental subgenomes generally retains significantly more genes compared with other subgenome(s) (Thomas, Pederson, and Freeling, 2006; Emery et al., 2018). Thomas, Pederson, and Freeling (2006) gave the first detailed account of this phenomenon, a process they termed 'biased fractionation'. The vast majority of duplicated regions retained from the most recent polyploid event in the *Arabidopsis* genome preferentially retained genes in one subgenome (i.e. dominant subgenome) compared with the other 'recessive' subgenome. Subsequent work connected these patterns of biased fractionation with differences in expression of gene copies on different subgenomes (Schnable et al. 2011). Due to technical constraints much of this work deals with ancient polyploid events and diploidized genomes, but analysis of the early impacts of these phenomena in newly formed polyploid genomes is now more feasible than ever.

It is against this backdrop that the research presented in this dissertation was carried out. Chapter two reviews the latest findings in genome evolution in polyploid species, focusing specifically on 'subgenome dominance'. Subgenome dominance describes systematic asymmetry in gene retention, expression, and epigenomic regulation between the distinct subgenomes of polyploid. Until now, the majority of research on subgenome dominance focused on the remnants of polyploid subgenomes from ancient whole-genome duplications identified in extant diploid genomes (Thomas, Pederson, and Freeling, 2006; Schnable et al. 2011). This review focuses on newly formed polyploids, where the wide availability of polyploid genomes has lead to a trove of new findings. It expands on previous models of subgenome dominance that identify differences in transposable element density as the driving force of subgenome dominance, and also integrates genomic restructuring from recombination among subgenomes into the causes and consequences of subgenome dominance.

The chapters following explore various aspects of polyploid genome evolution and evolution of polyploid species. Chapter three, four, and five use a unique population of resynthesized *Brassica napus* plants to explore early genome evolution of polyploids. These lines were independently generated, started completely genetically identical, and were self-fertilized for ten generations, allowing for the analysis of

genomic, transcriptomic, and epigenomic changes over time across truly independent polyploid origins. Chapter three combines genome resequencing, RNA-seq, and Methylseq data to test if the same subgenome is dominant across independent origins and if systematic differences in methylation patterns near genes is associated with biased expression of homoeologous genes, as predicted of the TE density model of subgenome dominance. Chapter four tests whether changes in gene dosage from genomic rearrangements from chromosomal duplications and deletions and from recombination among subgenomes can bias the identification of a dominant subgenome if they are not accounted for. Then chapter five explores whether copy number changes from homoeologous recombination are constrained by the same need to maintain gene dosage balance that constrains the evolution of duplicate genes, as described by the Gene Balance Hypothesis. This work was also able to assess how subgenome dominance interacts with gene dosage constraint and the temporal dynamics of dosage constraint over the first ten generations. Finally, chapter six leverages newly developed genomic resources from the publishing of the octoploid strawberry genome (*Fragaria x ananasa*) to study the population genomics and phylogeography of a global sample of wild octoploid strawberry species (*Fragaria chiloensis* and *Fragaria virginiana*), establishing the monophyly of these species and indicating gene flow between these strawberry subspecies and species.

REFERENCES

REFERENCES

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., ... & Saunders, R. D. (2000). The genome sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185-2195.
- Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., & Edwards, D. (2020). Plant pan-genomes are the new reference. *Nature plants*, 6(8), 914-920.
- Birchler, J. A., & Veitia, R. A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences*, 109(37), 14746-14753.
- Blanc, G., Barakat, A., Guyot, R., Cooke, R., & Delseny, M. (2000). Extensive duplication and reshuffling in the *Arabidopsis* genome. *The Plant Cell*, 12(7), 1093-1101.
- Blanc, G., Hokamp, K., & Wolfe, K. H. (2003). A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome research*, 13(2), 137-144.
- Blanc, G., & Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *The Plant Cell*, 16(7), 1679-1691.
- Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, 422(6930), 433-438.
- C. elegans Sequencing Consortium*. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, 282(5396), 2012-2018.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., ... & Wincker, P. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *science*, 345(6199), 950-953.
- Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., ... & Schmutz, J. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nature genetics*, 52(5), 525-533.
- Conant, G. C., Birchler, J. A., & Pires, J. C. (2014). Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Current opinion in plant biology*, 19, 91-98.
- Colle, M., Leisner, C. P., Wai, C. M., Ou, S., Bird, K. A., Wang, J., ... & Edger, P. P. (2019). Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. *GigaScience*, 8(3), giz012.
- Danilevicz, M. F., Fernandez, C. G. T., Marsh, J. I., Bayer, P. E., & Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. *Current Opinion in Plant Biology*, 54, 18-25.
- De Bodt, S., Maere, S., & Van de Peer, Y. (2005). Genome duplication and the origin of angiosperms. *Trends in ecology & evolution*, 20(11), 591-597.
- Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., ... & Knapp, S. J. (2019). Origin and evolution of the octoploid strawberry genome. *Nature genetics*, 51(3), 541-547.

- Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., ... & Wheat, C. W. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences*, 112(27), 8362-8366.
- Emery, M., Willis, M. M. S., Hao, Y., Barry, K., Oakgrove, K., Peng, Y., ... & Conant, G. C. (2018). Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. *PLoS genetics*, 14(3), e1007267.
- Freeling, M., & Thomas, B. C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome research*, 16(7), 805-814.
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., ... & Oliver, S. G. (1996). Life with 6000 genes. *Science*, 274(5287), 546-567.
- Golicz, A. A., Bayer, P. E., Bhalla, P. L., Batley, J., & Edwards, D. (2020). Pangenomics comes of age: from bacteria to plant and animal applications. *Trends in Genetics*, 36(2), 132-145.
- International Wheat Genome Sequencing Consortium (IWGSC), Mayer, K. F., Rogers, J., Doležal, J., Pozniak, C., Eversole, K., ... & Praud, S. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, 345(6194), 1251788.
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., ... & Depamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, 473(7345), 97-100.
- Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., ... & Somerville, M. C. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *nature*, 408(6814), 796-815.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., ... & Proctor, M. J. (2001). Initial sequencing and analysis of the human genome.
- Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C., & Soltis, P. S. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *American journal of botany*, 105(3), 348-363.
- Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., ... & Yu, S. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature genetics*, 46(6), 567-572.
- Lewontin, R. (1992). The dream of the human genome: doubts about the Human Genome Project. *The New York review of books*, 39(10), 31-40.
- Marks, R. A., Hotaling, S., Frandsen, P. B., & VanBuren, R. (2021). Representation and participation across 20 years of plant genome sequencing. *Nature plants*, 7(12), 1571-1578.
- Michael, T. P., & VanBuren, R. (2015). Progress, challenges and the future of crop genomes. *Current opinion in plant biology*, 24, 71-81.
- Michael, T. P., & VanBuren, R. (2020). Building near-complete plant genomes. *Current Opinion in Plant Biology*, 54, 26-33.
- Paterson, A. H., Bowers, J. E., Burow, M. D., Draye, X., Elisk, C. G., Jiang, C. X., ... & Wright, R. J. (2000). Comparative genomics of plant chromosomes. *The Plant Cell*, 12(9), 1523-1539.
- Paterson, A. H., Freeling, M., Tang, H., & Wang, X. (2010). Insights from the comparison of plant genome sequences. *Annual review of plant biology*, 61, 349-372.

- Qi, X., An, H., Hall, T. E., Di, C., Blischak, P. D., McKibben, M. T., ... & Barker, M. S. (2021). Genes derived from ancient polyploidy have higher genetic diversity and are associated with domestication in *Brassica rapa*. *New Phytologist*, 230(1), 372-386.
- Schnable, J. C., Springer, N. M., & Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences*, 108(10), 4069-4074.
- Schranz, M. E., Mohammadin, S., & Edger, P. P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Current opinion in plant biology*, 15(2), 147-153.
- Thomas, B. C., Pedersen, B., & Freeling, M. (2006). Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome research*, 16(7), 934-946.
- VanBuren, R., Bryant, D., Edger, P. P., Tang, H., Burgess, D., Challabathula, D., ... & Mockler, T. C. (2015). Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature*, 527(7579), 508-511.
- VanBuren, R., Man Wai, C., Wang, X., Pardo, J., Yocca, A. E., Wang, H., ... & Michael, T. P. (2020). Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nature communications*, 11(1), 1-11.
- Van de Peer, Y., Mizrahi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics*, 18(7), 411-424.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... & Kalush, F. (2001). The sequence of the human genome. *science*, 291(5507), 1304-1351.
- Waterston, R. H., & Pachter, L. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), 520-562.
- Watson, J. D. (1990). The human genome project: past, present, and future. *Science*, 248(4951), 44-49.
- Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., ... & Ming, R. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature genetics*, 50(11), 1565-1573.

CHAPTER 2

The work presented in this chapter is part of the final publication

Bird, K.A., VanBuren, R., Puzey, J.R. and Edger, P.P. (2018), The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol*, 220: 87-93.

The Causes and Consequences of Subgenome Dominance in Hybrids and Recent Polyploids

Abstract

The merger of divergent genomes, via hybridization or allopolyploidization, frequently results in a 'genomic shock' that induces a series of rapid genetic and epigenetic modifications as a result of conflicts between parental genomes. This conflict among the subgenomes routinely leads one subgenome to become dominant over the other subgenome(s), resulting in subgenome biases in gene content and expression. Recent advances in methods to analyze hybrid and polyploid genomes with comparisons to extant parental progenitors have allowed for major strides in understanding the mechanistic basis for subgenome dominance. In particular, our understanding of the role that homoeologous exchange might play in subgenome dominance and genome evolution is quickly growing. Here we describe recent discoveries uncovering the underlying mechanisms and provide a framework to predict subgenome dominance in hybrids and allopolyploids with far-reaching implications for agricultural, ecological, and evolutionary research.

Summary

This chapter reviews the history of subgenome dominance and covers recent data and theories for the underlying mechanisms that cause subgenome dominance to occur in hybrid and polyploid genomes with a specific focus on newly formed hybrid and polyploid species. After synthesizing this information it proposes a framework to predict patterns subgenome dominance in hybrids and allopolyploids and discusses the important and implications of subgenome dominance for plant breeding and research in the fields of ecology and evolution. For this chapter, I did the primary literature review, wrote the initial draft of this manuscript and handled incorporating all comments and edits from coauthors.

CHAPTER 3

The work presented in this chapter is part of the final publication

Bird, K.A., Niederhuth, C.E., Ou, S., Gehan, M., Pires, J.C., Xiong, Z., VanBuren, R. and Edger, P.P. (2021), Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *New Phytol*, 230: 354-371.

Replaying the Evolutionary Tape to Investigate Subgenome Dominance in Allopolyploid *Brassica napus*

Abstract

Allopolyploidisation merges evolutionarily distinct parental genomes (subgenomes) into a single nucleus. A frequent observation is that one subgenome is 'dominant' over the other subgenome, often being more highly expressed. Here, we 'replayed the evolutionary tape' with six isogenic resynthesised *Brassica napus* allopolyploid lines and investigated subgenome dominance patterns over the first 10 generations postpolyploidisation. We found that the same subgenome was consistently more dominantly expressed in all lines and generations and that >70% of biased gene pairs showed the same dominance patterns across all lines and an *in silico* hybrid of the parents. Gene network analyses indicated an enrichment for network interactions and several biological functions for the *Brassica oleracea* subgenome biased pairs, but no enrichment was identified for *Brassica rapa* subgenome biased pairs. Furthermore, DNA methylation differences between subgenomes mirrored the observed gene expression bias towards the dominant subgenome in all lines and generations. Many of these differences in gene expression and methylation were also found when comparing the progenitor genomes, suggesting that subgenome dominance is partly related to parental genome differences rather than just a byproduct of allopolyploidisation. These findings demonstrate that 'replaying the evolutionary tape' in an allopolyploid results in largely repeatable and predictable subgenome expression dominance patterns.

Summary

This chapter investigates the repeatability of subgenome dominance by studying six lines of the plant species *Brassica napus* that were generated from a cross between the progenitor species *Brassica rapa* and *Brassica oleracea*. These lines were used to see if the same subgenome was dominantly expressed in each of these lines and across the ten generations for which these lines were sampled. The results showed that the same subgenome had a significantly higher number of genes with higher expression compared to the other subgenomes, and this was found in all lines and generations. Additionally, over 70% of gene pairs that were biased toward the dominant subgenome showed the same biased expression in all lines and in a comparison of the parents' gene expression, indicating a consistent and repeatable pattern of expression bias. Next, these genes were analyzed in the context of a protein-protein interaction network to look for enrichment for network interactions and biological functions. This analysis showed network and functional enrichment for the dominant subgenome biased pairs, but not for the other subgenome. This functional enrichment includes primary metabolism and the organelles which are maternally inherited and may support cyto-nuclear interactions as a contributor to subgenome dominance. Finally, DNA methylation was compared between subgenomes and showed methylation differences in all lines and at all generations that mirrored the differences in gene expression that favored the dominant subgenome. Many of these differences in gene expression and methylation were also found when comparing the progenitor genomes. By using this unique population of genetically identical and independently made polyploids, I could avoid confounding from genetic variation and showed that subgenome dominance is substantially related to parental genome differences rather than just a byproduct of hybridization and genome duplication. For this chapter I planned and designed the experiments and data analysis, performed the data analysis on the whole-genome sequencing and RNAseq, and wrote the first draft of the manuscript and handled incorporated edits from coauthors and journal referees.

CHAPTER 4

The Role of Genomic Rearrangements in Biasing Analysis of Subgenome Dominance

Abstract

Allopolyploidy species, which experience the hybridization of two evolutionary diverged species and the doubling of genomic material, frequently exhibit genomic rearrangements that recombine, duplicate, or delete homoeologous regions of the newly formed genome. We used genomic and transcriptomic data for six independently resynthesized, isogenic *Brassica napus* lines in the first, fifth, and tenth generation to identify genomic rearrangements and assess their impact on the distribution of homoeolog expression bias. We show that genomic rearrangements can quantitatively affect the estimation of homoeolog expression bias, but fail to fully obscure which subgenome is dominantly expressed.

Introduction

Allopolyploid species are those that experience a whole-genome duplication and hybridization of two evolutionarily diverged genomes. Upon the merger of these genomes, epigenetic markers like DNA methylation are frequently remodeled over early generations (Madlung et al., 2001; Edger et al., 2017; Bird et al., 2021) which can lead to major alterations in gene regulation (Chen, 2007) and activation of transposable elements (Vicent and Casacuberta, 2012). Polyploid genomes also must accommodate inherited and novel expression differences in homoeologous genes. A major area of research concerning allopolyploid genome evolution is subgenome dominance, a term used to describe patterns of biased gene expression and regulation between progenitor genomes and the long-term asymmetric retention of duplicate genes (Bird et al. 2018;2021, Wendel et al. 2018).

While the majority of investigations of subgenome dominance have relied upon the detection of ancient subgenomes in species that have rediploidized genomes (Thomas and Freeling, 2006; Schnable et al. 2011; Woodhouse et al. 2014), advances in genome sequencing technologies have produced a surge of studies using natural species with true polyploid genome structure or lab generated resynthesized polyploids. However, the analysis of subgenome dominance in these genomes involves unique challenges due to the often dynamic genome evolution in newly formed polyploids. From the first meiosis in new polyploid genomes, substantial genomic rearrangement from homoeologous recombination, partial or complete chromosomal duplications, and deletions can occur (Szadowski et al.

2010; Xiong et al. 2011; Nicolas et al. 2012; Mason and Wendel 2020). Rearrangements can continue to accumulate over time, producing extensive genomic and phenotypic diversity in early polyploids (Xiong et al. 2011; Mason and Wendel, 2020; Pires et al. 2004; Gaeta et al. 2007; Wu et al. 2021). While natural polyploids often exhibit less extensive rearrangements than resynthesized plants, likely due to selection to maintain genomic stability (Gaeta and Pires, 2010; Pele et al. 2018; Xiong et al. 2020; Gonzalo et al. 2019; Gaebelein et al. 2019; Ferreira de Carvalho et al. 2021), the genomic rearrangements that do exist often underlie important gene presence/absence variation and agronomically valuable quantitative trait loci in species like *Brassica napus* (Stein et al. 2017; Samans et al. 2017; Hurgobin et al. 2017; Bayer et al. 2021). Homoeologous exchanges (non-reciprocal recombination events that swap genomic regions among subgenomes) have also been associated with the generation of novel, chimeric transcripts in multiple polyploid species (Zhang et al, 2020).

An unexplored concern in studying subgenome dominance in polyploid genomes is that genomic rearrangements can alter the global transcriptome and expression levels of homoeologous gene pairs and potentially bias patterns of subgenome dominance. A study using natural *B. napus* demonstrated that homoeologous exchanges caused dosage-dependent gene expression changes (Lloyd et al. 2018). Bird et al. (2018) and Edger et al. (2019) have hypothesized from this observation that the expression changes from homoeologous exchange could alter the global transcriptome in a way that obscures or exaggerates the extent of subgenome dominance. Studies often do not have paired whole-genome sequencing (WGS) and RNAseq data to identify genomic rearrangements and subgenome dominance at the same time. Bird et al. (2021) analyzed subgenome expression dominance in resynthesized *B. napus* but only investigated gene pairs identified as having a 2:2 dosage ratio using WGS data. Looking only at 2:2 dosage regions means the effect of homoeologous exchange on subgenome dominance inference could not assess. These predictions from Bird et al. (2018) and Edger et al. (2019) have yet to be tested.

This study utilized a previously generated population of independently resynthesized *B. napus* lines, produced by hybridizing *B. oleracea* acc. TO1000DH and *B. rapa* acc. IMB-218DH. Importantly, because these lines were created from two doubled haploid parental lines all individuals started completely isogenic. An individual plant from six resynthesized lines was sequenced at the first (S1), fifth (S5), and tenth selfing generation (S10) and analyzed for changes in gene (homoeolog) dosage due to

genomic rearrangement and potential bias toward a particular subgenome. Shifts in the ratio of WGS read depth coverage between these gene pairs allowed us to pinpoint changes in gene dosage from genomic rearrangements across each of the six lines and over the ten generations. These rearrangements may represent homoeologous exchanges, where non-reciprocal homoeologous recombination between syntenic regions of the parental subgenomes replaces one homoeolog with another, chromosomal deletions and duplications, or gene conversion events. Paired RNAseq data was used to determine the effect these genomic rearrangements have on the analysis of subgenome expression dominance.

Methods

Sequencing data

We downloaded the data and files for previously identified genomic rearrangements and transcript quantification based on the filtering and analysis of Bird et al. (2021) from the dryad repository <https://doi.org/10.5061/dryad.h18931zjr>.

Genomic rearrangement analysis

Using identified genomic rearrangements from Bird et al. (2021), a Chi-squared test was used to test for bias in the direction of gene dosage changes. Observed dosage changes were compared against an expected 50/50 ratio with an equal number of events that increase the copy of C-subgenome (BnC) regions increasing and events that increase copy of A-subgenome (BnA) regions. Significant deviations were considered to be biases in genomic rearrangements, either favoring more BnA than BnC copies or vice-versa.

Homoeolog expression bias

We used the published expression quantification data from Bird et al. (2021) to assign homoeolog expression bias designations based on a threshold of $\log_2$ fold change $> |3.5|$. We calculated the number of biased homoeolog pairs for syntenic homoeolog pairs, including those by genomic rearrangements that altered gene dosage, and on a data set including only homoeologous pairs with a 2:2 dosage ratio. We used a Chi-squared test to see if genomic rearrangements significantly altered the proportion of biased homoeologs compared to observed proportions when only analyzing homoeologous pairs with 2:2 dosage. Data were also reanalyzed using a threshold of $\log_2$ fold change $> |2|$.

Results

Genomic rearrangements in these resynthesized lines are highly variable and do not show signs of subgenome bias

Previous studies of this resynthesized *B. napus* population using a handful of DNA or cytogenetic markers identified extensive chromosomal duplications and deletions and homoeologous exchanges that resulted in immense phenotypic variation in both plant height and pollen count (Xiong et al. 2011). However, the previous set of markers had limited resolution and small-scale exchanges were not identifiable. We used a whole-genome resequencing approach to identify at higher resolution genomic rearrangements that altered the relative dosage of homoeologs among individuals across this population.

The direction of dosage changes and proportion of gene pairs with changed dosage varied greatly between lines and generations with no consistent pattern significantly favoring the A subgenome (BnA) or C subgenome (BnC) (Figure 4.1). The number of homoeologous gene pairs affected by genomic rearrangement in individual lines ranged from 114 to 10,231 , representing 0.4% to 39.2% of identified syntenic gene pairs, respectively. Additionally, the number of genes affected by genomic rearrangements consistently increased over time. (Table 4.1). Overall, 9 of 18 plants had significantly more genomic rearrangements increasing BnC copy number than expected, while 8 out of 18 had significantly more rearrangements increasing BnA copy number than expected. Only two lines, EL-300 and EL-1100 showed the same direction of bias in genomic rearrangements for each generation, while the other four lines showed a change in the direction of bias across generations.

Impact of dosage changes on homoeologous expression bias

Next, we took advantage of paired genomic and transcriptomic sequencing data to compare homoeologous expression bias when only analyzing genes inferred as in 2:2 dosage and when including genes involved in genomic rearrangements that altered gene dosage. For this analysis, we used the same definition for homoeolog expression bias as Bird et al. (2021), based on a threshold of $\log_2$ fold change $> |3.5|$.

Figure 4.1 Variability of gene dosage changes and hotspots in resynthesized *B. napus*

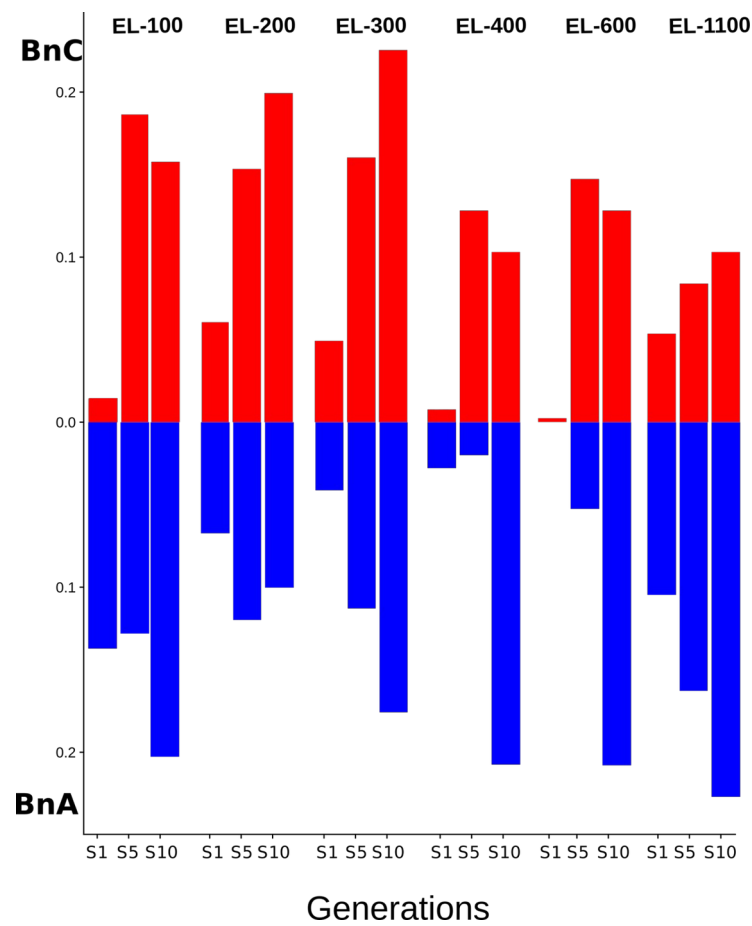


Table 4.1 Chi-squared test for bias in direction of gene dosage changes

<i>Sample</i>	<i>BnC Genes Observed</i>	<i>BnC. Genes Expected</i>	<i>BnA. Genes Observed</i>	<i>BnA. Genes Expected</i>	<i>Chi. Squared</i>	<i>P.value</i>
EL-100S1	368	1931	3494	1931	2530.26	0
EL-100S5	4749	4024.5	3300	4024.5	260.85	1.12e-58
EL-100S10	4007	4573	5139	4573	140.11	2.52e-32
EL-200S1	1535	1620.5	1706	1620.5	9.02	0.003
EL-200S5	3875	3459	3043	3459	100.06	1.48e-23
EL-200S10	5049	3803.5	2558	3803.5	815.71	2.08e-179
EL-300S1	1255	1156	1057	1156	16.96	3.82e-05
EL-300S5	4082	3473	2864	3473	213.58	2.27e-48
EL-300S10	5725	5107.5	4490	5107.5	149.31	2.45e-34
EL-400S1	201	452.5	704	452.5	279.57	9.33e-63
EL-400S5	3207	1855	503	1855	1970.79	0
EL-400S10	2633	3953.5	5274	3953.5	882.11	7.58e-194
EL-600S1	53	26.5	0	26.5	53	3.34e-13
EL-600S5	3748	2542.5	1337	2542.5	1143.15	1.38e-250
EL-600S10	3267	4294	5321	4294	491.26	7.59e-109
EL-1100S1	1366	2019	2672	2019	422.4	7.34e-94

Table 4.1 (cont'd)

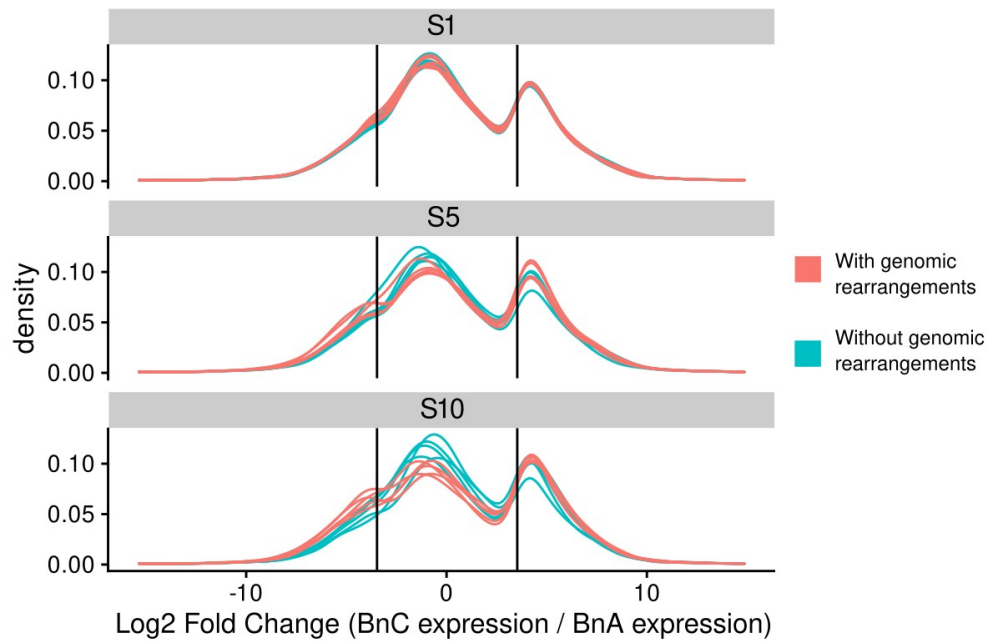
EL1100S5	2133	3171.5	4210	3171.5	680.11	6.33e-150
EL-1100S10	2590	4197	5804	4197	1230.62	1.35e-269

In the first generation, before most gene dosage changes occur, the distribution of the $\log_2$ expression ratio of homoeologous gene pairs when including and excluding gene dosage alterations broadly overlap (Figure 4.2) and the ratio of BnC to BnA biased gene pairs is not significantly different from 1:1 for 4 of 6 lines (χ^2 -test, $p > 0.05$). In the fifth and tenth generations, after more genomic rearrangements accumulate, the distributions visibly begin to diverge (Figure 4.2). Only one of ten of these individuals have ratios of BnC and BnA biased homoeolog pairs that are not significantly different between analyses that exclude gene pairs affected by genomic rearrangements and those that include them (Table 4.2). In 6 of 10 cases, the gene dosage cases reduced the proportion of BnC biased gene pairs and increased the proportion of BnA biased gene pairs. The other 4 of 10 cases showed an increased proportion of BnC biased gene pairs and decreased proportion of BnA biased gene pairs (Table 4.2).

These results demonstrate that gene dosage changes from genomic rearrangement do alter the distribution of homoeolog expression bias and the ratio of biased gene pairs in statistically significant ways. Importantly, however, gene dosage changes never completely reversed the dominance relationship of the subgenomes. In other words, gene dosage events never led to the non-dominant BnA subgenome becoming the dominantly expressed subgenome by having more biased homoeolog pairs compared to the BnC subgenome. Because gene dosage changes in this study were not biased with respect to subgenome, it is unclear if it would be possible to completely reverse subgenome expression dominance relationships if dosage changes occurred in a biased fashion. However, among the 6 lines, there was variation in HE bias. Some lines, like EL1100, tended to have more HEs that increased BnA dosage, and lines like EL300 tended to have more HEs that increased BnC dosage (Figure 4.1). Even in line EL1100 there was never a case where HEs resulted in BnA being the dominant subgenome (Table 4.2).

We also analyzed the impact of genomic rearrangements using a common threshold for homoeolog expression bias of $\log_2$ fold change $> |2|$ (Schnable et al. 2011; Woodhouse et al. 2014; Cheng et al. 2016). With this lower threshold, a greater change in the proportion of biased homoeologs was detected, in several cases, a difference of over 10% was observed; however, BnC still remained the dominant subgenome in all cases.

Figure 4.2 Impact of homoeologous exchange on subgenome dominance



Discussion

Subgenome dominance has become a major focus of genomic studies of polyploids, but the ways genomic rearrangements alter gene expression patterns (Lloyd et al. 2017; Hou et al. 2018) have led to concerns that failing to account for genomic rearrangements in polyploid genomes may lead to biased assessment of subgenome expression dominance (Bird et al. 2018; Edger et al. 2019). Our analysis of genomic rearrangements and homoeologous exchanges in resynthesized *B. napus* confirmed at higher resolution the extensive genomic rearrangement in these lines (Gaeta et al. 2007; Xiong et al. 2011). Leveraging paired RNAseq data, our results suggest that even extensive genomic rearrangement found in resynthesized polyploid lines result in only quantitative changes to the results of subgenome

Table 4.2 Homoeolog Expression Bias with and without Genomic Rearrangement (GRs) chi-squared table

<i>Sample</i>	<i>BnC biased pairs with GRs</i>	<i>BnC biased pairs without GRs</i>	<i>BnA biased pairs with GRs</i>	<i>BnA biased pairs without GRs</i>	<i>Chi. Squared</i>	<i>P.value</i>
RS-100S1	3423 (0.67)	3581.29 (0.70)	1698 (0.33)	1539.71 (0.30)	23.27	1.41e-06
RS-100S5	3355 (0.60)	3886.38 (0.69)	2278 (0.40)	1746.62 (0.31)	234.32	6.81e-53
RS-100S10	4108 (0.66)	4528.96 (0.72)	2162 (0.34)	1741.04 (0.28)	140.91	1.68e-32
RS-200S1	3571 (0.68)	3698.61 (0.70)	1692 (0.32)	1564.39 (0.30)	14.81	1.19e-04
RS-200S5	3879 (0.67)	3866.63 (0.67)	1868 (0.33)	1880.37 (0.33)	0.12	0.73
RS-200S10	4234 (0.68)	3955.51 (0.64)	1984 (0.32)	2262.49 (0.36)	53.89	2.12e-13
RS-300S1	3421 (0.67)	3442.93 (0.68)	1672 (0.33)	1650.07 (0.32)	0.43	0.51
RS-300S10	3979 (0.62)	4317.57 (0.67)	2479 (0.38)	2140.43 (0.33)	80.1	3.55e-19
RS-400S1	3555 (0.67)	3616.74 (0.68)	1739 (0.33)	1677.26 (0.32)	3.33	0.068
RS-400S5	3987 (0.69)	3808.79 (0.66)	1803 (0.31)	1981.21 (0.34)	24.37	7.96e-07

Table 4.2 (cont'd)

RS-600S1	3625 (0.68)	3588.89 (0.68)	1685 (0.32)	1721.11 (0.32)	1.12	0.290
RS-600S5	3528 (0.66)	3246.27 (0.60)	1851 (0.34)	2132.73 (0.40)	61.67	4.07e-15
RS-600S10	4327 (0.65)	4571.01 (0.69)	2280 (0.35)	2035.99 (0.31)	42.27	7.96e-11
RS-1100S1	3581 (0.68)	3626.07 (0.68)	1717 (0.32)	1671.93 (0.32)	1.78	0.183
RS-1100S5	3362 (0.60)	3931.43 (0.70)	2256 (0.40)	1686.57 (0.30)	274.73	1.06e-61
RS-1100S10	4067 (0.63)	4761.25 (0.73)	2424 (0.37)	1729.75 (0.27)	379.87	1.32e-84

expression dominance analysis. Comparing analysis of subgenome dominance that excluded or included genomic rearrangements showed that although the precise proportion of biased homoeologs substantially changed the qualitative direction of the bias did not, even when a line showed strong subgenome bias in direction of homoeologous exchange. These results were found even with a more permissive of $\log_2$ fold change threshold of $|2|$, although a greater shift in the proportion of dominant homoeologs was observed. These results may support the use of a more conservative $\log_2$ fold change threshold when not directly accounting for genomic rearrangement.

One potential contributor to the results of this study is the unbiased nature of the direction of genomic rearrangements. In *B. napus*, this is at odds with observed biases favoring replacing BnC segments with BnA segments in the reference accession Darmor-bzh (Chalhoub et al. 2014) and a population of field-grown natural and synthetic *B. napus* (Samans et al. 2017). A likely explanation is that although the mechanism for homoeologous recombination is largely a random process of meiosis there

are fitness costs in natural environments of the field that select against homoeologous exchanges in a certain direction. Gaebelein et al. (2019) noted reduced fertility when C-genome regions replaced A-genome regions in a *Brassica* allohexaploid (AABBCC), supporting this idea. This population of resynthesized lines was grown in the more hospitable greenhouse and growth chamber conditions and hand-pollinated, which likely offsets the fitness costs identified by other studies and prevented the formation of systematic bias in homoeologous exchange. However, among individual plants there were significant biases in the direction of genomic rearrangements and even in these most extreme cases where BnA segments replaced BnC segments there was still no situation where BnC was not identified as the dominant subgenome.

Although not accounting for genomic rearrangement may lead to imprecise estimates of subgenome dominance dynamics, analyses will likely still provide a reliable estimate of the overall direction of subgenome dominance. Considering these resynthesized lines accumulate more genomic rearrangements than natural *B. napus* it could be even less likely that subgenome dominance estimates are severely biased. Based on these results it is likely that past analyses of subgenome dominance without accounting for possible genomic rearrangement events are reliable and future studies likely do not need to account for genomic rearrangements if the goal is simply to identify the more subgenome with more highly expressed homoeologs.

REFERENCES

REFERENCES

- Bayer, P. E., Scheben, A., Golicz, A. A., Yuan, Y., Faure, S., Lee, H., ... & Edwards, D. (2021). Modelling of gene loss propensity in the pangenomes of three Brassica species suggests different mechanisms between polyploids and diploids. *Plant Biotechnology Journal*.
- Bird, K. A., Niederhuth, C. E., Ou, S., Gehan, M., Pires, J. C., Xiong, Z., ... & Edger, P. P. (2021). Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid Brassica napus. *New Phytologist*, 230(1), 354-371.
- Bird, K. A., VanBuren, R., Puzey, J. R., & Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist*, 220(1), 87-93.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., ... & Wincker, P. (2014). Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *science*, 345(6199), 950-953.
- Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.*, 58, 377-406.
- Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., ... & Knapp, S. J. (2019). Origin and evolution of the octoploid strawberry genome. *Nature genetics*, 51(3), 541-547.
- Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., ... & Puzey, J. R. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *The Plant Cell*, 29(9), 2150-2167.
- Ferreira de Carvalho, J., Stoeckel, S., Eber, F., Lodé-Taburel, M., Gilet, M. M., Trotoux, G., ... & Rousseau-Gueutin, M. (2021). Untangling structural factors driving genome stabilization in nascent Brassica napus allopolyploids. *New Phytologist*, 230(5), 2072-2084.
- Freeling, M., & Thomas, B. C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome research*, 16(7), 805-814.
- Gaeta, R. T., Pires, J. C., Iniguez-Luy, F., Leon, E., & Osborn, T. C. (2007). Genomic changes in resynthesized Brassica napus and their effect on gene expression and phenotype. *The Plant Cell*, 19(11), 3403-3417.
- Gaeta, R. T., & Pires, J. C. (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytologist*, 186(1), 18-28.
- Gaebelin, R., Schiessl, S. V., Samans, B., Batley, J., & Mason, A. S. (2019). Inherited allelic variants and novel karyotype changes influence fertility and genome stability in Brassica allohexaploids. *New Phytologist*, 223(2), 965-978.
- Gonzalo, A., Lucas, M. O., Charpentier, C., Sandmann, G., Lloyd, A., & Jenczewski, E. (2019). Reducing MSH4 copy number prevents meiotic crossovers between non-homologous chromosomes in Brassica napus. *Nature communications*, 10(1), 1-9.

- Hou, J., Shi, X., Chen, C., Islam, M. S., Johnson, A. F., Kanno, T., ... & Birchler, J. A. (2018). Global impacts of chromosomal imbalance on gene expression in Arabidopsis and other taxa. *Proceedings of the National Academy of Sciences*, 115(48), E11321-E11330.
- Hurgobin, B., Golicz, A. A., Bayer, P. E., Chan, C. K. K., Tirnaz, S., Dolatabadian, A., ... & Edwards, D. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus. *Plant biotechnology journal*, 16(7), 1265-1274.
- Lloyd, A., Blary, A., Charif, D., Charpentier, C., Tran, J., Balzergue, S., ... & Jenczewski, E. (2018). Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *New Phytologist*, 217(1), 367-377.
- Madlung, A., Tyagi, A. P., Watson, B., Jiang, H., Kagochi, T., Doerge, R. W., ... & Comai, L. (2005). Genomic changes in synthetic Arabidopsis polyploids. *The Plant Journal*, 41(2), 221-230.
- Mason, A. S., & Wendel, J. F. (2020). Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Frontiers in Genetics*, 11, 1014.
- Pelé, A., Rousseau-Gueutin, M., & Chèvre, A. M. (2018). Speciation success of polyploid plants closely relates to the regulation of meiotic recombination. *Frontiers in plant science*, 9, 907.
- Pires, J. C., Zhao, J., Schranz, M. E., Leon, E. J., Quijada, P. A., Lukens, L. N., & Osborn, T. C. (2004). Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae). *Biological Journal of the Linnean Society*, 82(4), 675-688.
- Samans, B., Chalhou, B., & Snowdon, R. J. (2017). Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species Brassica napus. *Plant Genome*, 10(3), 1-15.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Stein, A., Coriton, O., Rousseau-Gueutin, M., Samans, B., Schiessl, S. V., Obermeier, C., ... & Snowdon, R. J. (2017). Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in Brassica napus. *Plant biotechnology journal*, 15(11), 1478-1489.
- Szadkowski, E., Eber, F., Huteau, V., Lode, M., Huneau, C., Belcram, H., ... & Chèvre, A. M. (2010). The first meiosis of resynthesized Brassica napus, a genome blender. *New Phytologist*, 186(1), 102-112.
- Vicent, C. M., & Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Annals of Botany*.
- Wendel, J. F., Lisch, D., Hu, G., & Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Current opinion in genetics & development*, 49, 1-7.
- Woodhouse, M. R., Cheng, F., Pires, J. C., Lisch, D., Freeling, M., & Wang, X. (2014). Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proceedings of the National Academy of Sciences*, 111(14), 5283-5288.
- Wu, Y., Lin, F., Zhou, Y., Wang, J., Sun, S., Wang, B., ... & Liu, B. (2021). Genomic mosaicism due to homoeologous exchange generates extensive phenotypic diversity in nascent allopolyploids. *National Science Review*, 8(5), nwaa277.

Xiong, Z., Gaeta, R. T., & Pires, J. C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proceedings of the National Academy of Sciences*, 108(19), 7908-7913.

Zhang, Z., Gou, X., Xun, H., Bian, Y., Ma, X., Li, J., ... & Levy, A. A. (2020). Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proceedings of the National Academy of Sciences*, 117(25), 14561-14571.

CHAPTER 5

Gene Dosage Constraints Affect the Transcriptional Response to Allopolyploidy and Homoeologous Exchange in Resynthesized *Brassica napus*

Abstract

Allopolyploidy involves the hybridization of two evolutionary diverged species and the doubling of genomic material. Allopolyploids also exhibit homoeologous exchange that recombines, duplicates, or deletes homoeologous regions of the newly formed genome. These kinds of changes to gene dosage are hypothesized to be constrained by selection to maintain balanced gene dosage. The dynamics of this constraint immediately after allopolyploidy and in response to homoeologous exchange is unknown. We used genomic and transcriptomic data for six independently resynthesized, isogenic *Brassica napus* lines in the first, fifth, and tenth generation to identify genomic rearrangements and assess their impact on gene expression dynamics related to gene dosage constraint. Dosage-sensitive genes show a more coordinated expression response to polyploidy, consistent with selective constraint for balanced gene dosage. We also find that the expression response systematically differs for dosage-sensitive genes depending on whether homoeolog expression is biased toward the dominant or non-dominant subgenome. Expression coordination appears to change over early generations, possibly suggesting a weakening of dosage constraint. Dosage-sensitive genes also exhibit the same kind of coordinated expression response to homoeologous exchanges as they do to genome duplication. Constraint on gene dosage acts on gene expression for newly formed allopolyploids as it does for autopolyploids and exerts a detectable effect on homoeologous exchanges. These findings connect patterns of long- and short-term gene retention in polyploids and suggest novel patterns for the evolution of homoeologous exchanges.

Introduction

Changes in gene dosage are known to be a powerful and important driver of gene expression abundance, quantitative trait variation, and the evolution of genomes (Birchler and Veitia 2007, 2010, 2012). The observation that imbalanced gene dosage changes can have a large phenotypic impact and can be highly deleterious for certain classes of genes, especially those involved in highly connected regulatory networks and multimeric protein complexes lead to the formulation of Gene Balance Hypothesis (Birchler and Newton, 1981; Birchler et al., 2001; Makino and McLysaght, 2010; Birchler and Veitia, 2012). The core of the GBH argues that changing the stoichiometry of members of networks and

protein complexes involved in multicomponent interactions affects their kinetics, assembly, and function of the whole, which causes negative fitness consequences (Birchler et al., 2005; Birchler and Veitia, 2007, 2010, 2012). The need to maintain the stoichiometric balance of gene products in the face of changes in gene dosage from both small-scale and whole-genome duplication influences genome evolution in important and predictable ways. Comparative genomic studies have supported predictions from the GBH, showing that patterns of duplicate gene retention differ based on the function of a gene and whether a gene is duplicated by whole-genome duplication or by small-scale duplications. Specifically, it has been repeatedly observed that genes related to signaling pathways, regulatory pathways, and multi-component protein complexes tend to be over retained after whole-genome duplications. (Blanc and Wolfe, 2004; Maere, 2005; Paterson et al. 2006; Thomas and Freeling, 2006; Freeling, 2009; Edger and Pires, 2009; De Smet et al., 2013; Li et al., 2016; Tasdighian et al., 2018). Additionally, genes that tend to be over-retained after whole-genome duplication are frequently under-retained after small-scale duplications, a phenomena called 'reciprocal retention' (Freeling, 2009; Edger and Pires, 2009)

Many of these studies have focused on meso- or paleopolyploids, where genomes have returned to a diploid-like state, leaving the immediate transcriptional impact of large-scale gene dosage changes less well understood. However, several authors have recently investigated the expression responses caused by aneuploidy and polyploidy (Coate et al. 2016; Hou et al. 2018; Song et al. 2020; Shi et al. 2021; Yang et al. 2021). Coate et al. (2016) and Song et al. (2020), in particular, attempt to connect observed patterns of long-term duplicate gene retention to short-term duplicate gene expression responses. They use tenets of the GBH to predict two patterns in short-term expression response. First, genes that are reciprocally retained after whole-genome duplication (e.g. highly connected in gene networks, involved in multi-component protein complexes, etc.) should experience a change in gene expression in response to genome duplication. Second, these changes should be similar for all genes in the network, what they call a "coordinated response". Coate et al. (2016) address this question using natural soybean (*Glycine* L.) allopolyploids with an origin ~500,000 years ago and known diploid progenitors, while Song et al. (2020) use three *Arabidopsis thaliana* autopolyploid/diploid pairs. Both studies determined that genes that are highly reciprocally retained post-WGD showed a more coordinated gene expression response to polyploidy (Coate et al. 2016; Song et al. 2020).

These investigations have been greatly informative but were unable to address the extent that the immediate transcriptional response differs between a whole-genome duplication involving the hybridization of distinct progenitor genomes (allopolyploidy) and when a whole-genome duplication involves duplication of genetically similar chromosomes (autopolyploidy). While both result in a duplication of the genome, allopolyploidy also involves the merger of evolutionarily diverged genomes, which frequently results in remodeling of epigenetic markers (Madlung et al., 2001; Edger et al., 2017; Bird et al., 2021), alterations in gene regulation (Chen, 2007), and activation of transposable elements (Vicent and Casacuberta, 2012). Polyploid genomes also must accommodate inherited and novel expression differences in homoeologous genes which often results in subgenome dominance, where expression is biased in favor of homoeologs from one progenitor genome over others. (Alger et al. 2021; Bird et al. 2018,2021; Wendel et al. 2018). Studies in resynthesized allopolyploids have shown that from the first meiosis in new polyploid genomes, major reorganizations occur in the form of homoeologous recombination, partial or complete chromosomal duplications, and deletions (Szadowski et al. 2010; Xiong et al. 2011; Nicolas et al. 2012; Mason and Wendel 2020). These rearrangements continue to accumulate over time, generating genomic diversity in early polyploids (Xiong et al. 2011; Mason and Wendel, 2020).

These genomic rearrangements are often destructive to the organism and meiotic stability is more frequently observed in natural polyploids compared to resynthesized (Gaeta and Pires, 2010; Pele et al. 2018; Xiong et al. 2020). It is likely that meiotic stability is under strong selection in natural polyploid populations (Gaeta and Pires, 2010; Pele et al. 2018; Xiong et al. 2020; Gonzalo et al. 2019; Gaebelein et al. 2019; Ferreira de Carvalho et al. 2021). At the same time, genomic rearrangements generate phenotypic novelty in resynthesized polyploids (Pires et al. 2004; Gaeta et al. 2007; Wu et al. 2021) and are frequently observed in natural polyploids (Chalhoub et al. 2014; Lloyd et al. 2018; Edger et al. 2019; He et al. 2017). Additionally, homoeologous exchanges often underlie gene presence-absence variation and agronomically valuable quantitative trait loci in *Brassica napus* (Stein et al. 2017; Samans et al. 2017; Hurgobin et al. 2017; Bayer et al. 2021) and generate novel, chimeric transcripts in multiple polyploid species (Zhang et al, 2020).

Unlike aneuploidy and polyploidy, the impact of gene dosage constraint on gene expression changes from homoeologous exchanges is largely unexplored. There are reasons to believe homoeologous exchange can alter the dosage balance of gene products. Early studies in multiple resynthesized *Brassica napus* lines identified changes in the transcriptome caused by non-reciprocal homoeologous recombination, arguing these transcriptional changes produced phenotypic diversity among the lines (Gaeta et al. 2007). Furthermore, homoeologous exchanges (HEs) have been shown to alter expression in a dosage-dependent manner (Lloyd et al. 2017) that greatly resemble the gene dosage effects seen in aneuploid and polyploid organisms (Birchler and Newton, 1981). Finally, because the main effect of subgenome dominance is an unequal expression of homoeologous copies, altering the ratio of dominant and submissive homoeologs by homoeologous exchange has the potential to change the balance of gene products from the 2:2 tetraploid state. It is unknown whether there are also dosage compensation responses to HEs in other regions of the genome and if the gene expression response to homoeologous exchange follows predictions from the Gene Balance Hypothesis.

We analyzed paired WGS and RNASeq data for six independently resynthesized and isogenic *Brassica napus* (CCAA) lines, which are known to accumulate large amounts of genomic rearrangement (Xiong et al. 2011), at three generations to determine if the immediate gene expression responses to allopolyploidy are consistent with the Gene Balance Hypothesis. Using plants from first, fifth, and tenth generations, we further tested if the gene expression response to both polyploidy and homoeologous exchange changes over time and if it differs based on subgenome dominance of a homoeologous gene pair. We further identified homoeologous exchange events to test if changes in gene expression from homoeologous exchanges exhibit patterns of dosage constrain consistent with the Gene Balance Hypothesis. Our findings provide novel insights into the alteration of global expression by homoeologous exchanges and extend our understanding of how the Gene Balance Hypothesis constrains gene expression and genome evolution across various modes of gene dosage changes.

Methods

Sequencing data

We downloaded the data and files for previously identified genomic rearrangements and transcript quantification from Bird et al. (2021) from the dryad repository <https://doi.org/10.5061/dryad.h18931zjr>

Dosage response to polyploidy

When investigating the dosage response to polyploidy, we limited our analysis to the syntenic homoeologous genes identified as being in a 2:2 dosage ratio. We combined data from all polyploid lines together and calculated expression response to polyploidy for each gene pair, defined as the fold change of polyploid expression for a 2:2 syntenic homoeolog pair and the mid-parent expression of the progenitor

ortholog pair ($\frac{Exp_{B.oleracea} + Exp_{B.rapa}}{2}$). For both polyploid and diploid progenitor samples, Bird et al.

(2021) mapped to the *in silico* polyploid reference genome and transcripts were quantified in the same way so normalization was consistent. The distribution of polyploid dosage response for all sampled gene pairs in all lines was plotted as a histogram, along with the median of the distribution, using ggplot (Wickham) in R v 3.6.3 (R core team, 2020).

Dosage sensitivity assignment

To leverage the well-curated gene annotations of *Arabidopsis thaliana*, and the close phylogenetic relationship between *A. thaliana* and the *Brassica* genus, we assigned our *Brassica* gene pairs to the GO category of their *A. thaliana* ortholog. Orthologs between *A. thaliana* and *Brassica oleracea* were identified with Synmap (Lyons et al. 2008) on CoGe (Lyons and Freeling, 2008) and the *A. thaliana* GO annotations were directly assigned to the *B. oleracea* orthologs and from *B. oleracea* to the *B. rapa* syntelogs. Next, we used the GO category dosage response assignments (dosage-insensitive and dosage-sensitive) from Song et al.'s (2020) analysis of gene retention patterns of *A. thaliana* genes to classify our syntenic homoeologs as belonging to dosage-sensitive and dosage-insensitive GO categories.

Polyploid response variance

We applied the same approach as Coate et al. (2016) and Song et al. (2020) and focused on the

coefficient of variation of expression response ($\frac{\sigma_{exp}}{\mu_{exp}}$), which we similarly termed the polyploid response

variance (PRV). We calculated PRV only for GO terms that contained more than 20 genes. Statistical analysis was done with a Kruskal-Wallis test applied by the function `stat_compare_means()` in the R

package ggpubr v.0.04.0 (R core team, 2020; Kassambara, 2020). When analyzing the response to polyploidy among different homoeolog expression biases, the expression bias of progenitor orthologs was used. Previous analysis showed that for over 70% of homoeologs, all six resynthesized *B. napus* lines shared the same homoeolog expression bias as the parents (Bird et al. 2021).

Homoeologous exchange response variance

We included only syntenic homoeolog pairs that diverged from 2:2 dosage ratio (e.g. gene pairs with read-depth ratio less than 0.4 or greater than 0.6) to investigate the effects of gene dosage changes. To eliminate confounding effects from aneuploidy, we excluded chromosomes where we observed skewed read-depth ratios that spanned the entirety or majority of a chromosome. This resulted in the removal of syntenic homoeologs from chromosomes A1/C1, A2/C2, and A10 from all lines, and chromosome C4 only for line EL-1100 at generation 10. We defined the expression response to

homoeologous exchange as $\frac{Ex\ p_{BnC} + Ex\ p_{BnA}}{Ex\ p_{B.oleracea} + Ex\ p_{B.rapa}}$ which is the fold change of the summed

expression for a homoeologous pair in the polyploids and the summed expression of the progenitor orthologs when mapped to the *in silico* polyploid genome. We calculated the coefficient of variation of this expression response and termed it the homoeologous exchange response variance (HERV). The Kruskal-Wallis implementation from ggpubr (Kassambara, 2020) was used again for statistical analysis. As for the previous analysis, we only included GO terms with 20 or more genes and defined homoeolog expression bias in terms of expression bias in parental orthologs.

Results

Assessing early gene expression response to dosage changes from allopolyploidy

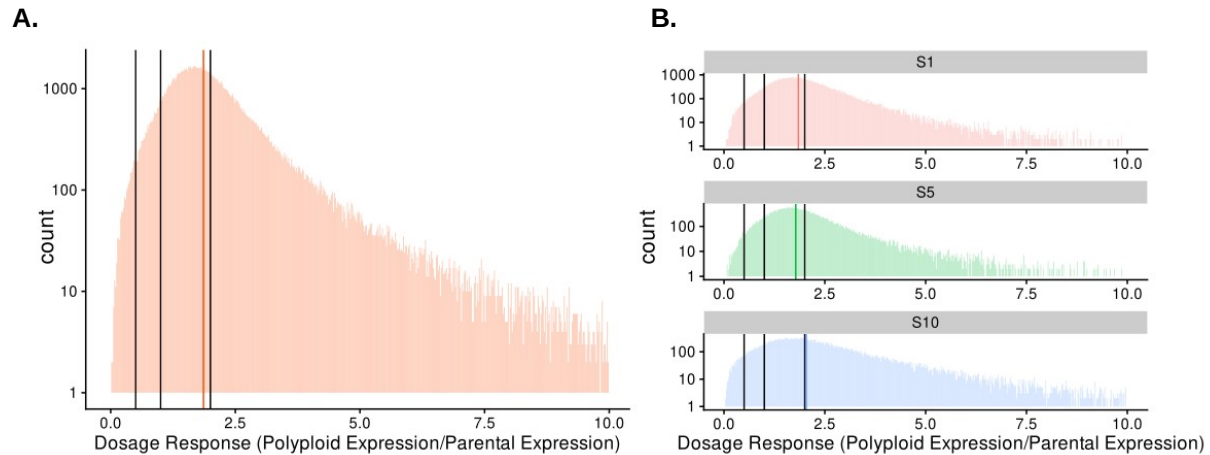
We investigated the relative gene expression change for individual homoeologous gene pairs in 2:2 dosage by taking the fold change of the summed transcript count for homoeologous gene pairs in the allopolyploid individuals and mid-parent value of the progenitors. It should be noted, this approach did not normalize RNA with exogenous spike-in as other studies have, meaning values reported are relative gene expression levels and their response to genome doubling rather than the absolute expression response. While this will introduce some biases to our measures because the increase in transcriptome size of

polyploids does not scale perfectly with the increase in genome size, our ability to detect broad patterns consistent with the Gene Balance Hypothesis should still remain. For this study, a ratio of 1 represents dosage compensation, resulting in no change in expression between polyploid and progenitor genomes, and a ratio of 2 represents a 1:1 expression response to dosage change e.g. doubled expression. Looking at all 16 individuals together, we observed high levels of variation in expression response to polyploidy (Figure 5.1). The median relative expression response to allopolyploidy was 1.86, just below a 1:1 expression response (Figure 5.1a). However, extreme values ranging from a very strong negative dosage response of 0.02 (essentially silenced) to 147 fold increase in expression in response to allopolyploidy were observed. Many genes also exhibited patterns consistent with dosage compensation, with ~8.8% of gene pairs less than or equal to a ratio of 1. These results mirror observed gene expression changes in autotetraploid/diploid maize comparisons (Shi et al. 2021).

When broken down by generation, we observed a progressive change in dosage response. Earlier generations (one and five), show median relative dosage responses of 1.84 and 1.78, respectively. Ten generations after polyploidy, however, the median relative dosage response rises to 2.04 (Figure 5.1b). This change in the median is largely driven by increased variance in expression dosage response. In generations one and five, there are 8.8% and 7.6% of gene pairs with a dosage response less than or equal to 1, respectively, while generation ten showed 11% of gene pairs less than or equal to 1. Likewise, 41.2% and 37.2% of gene pairs had dosage responses greater than 2 in generations one and five, while 51.5% of gene pairs show such a dosage response in generation 10. This increased spread of dosage response in the higher and lower ranges in the tenth generation may suggest that dosage constraint progressively weakens over time in these resynthesized lines. However, it should be noted that our design makes it difficult to distinguish the isolated effects of time against changes in inter-individual variation and concomitant trans-effects, which increase over time due to accumulating genomic rearrangement. As such the increase in the variance of expression response over time may be due to a change in dosage constraint itself, which allows more variance in expression, or a result of accumulating individual variation and trans-effects that increase variance. In either case, the results reveal a greater tolerance to expression variance than suggested by a single generation analysis. It is likely that dosage

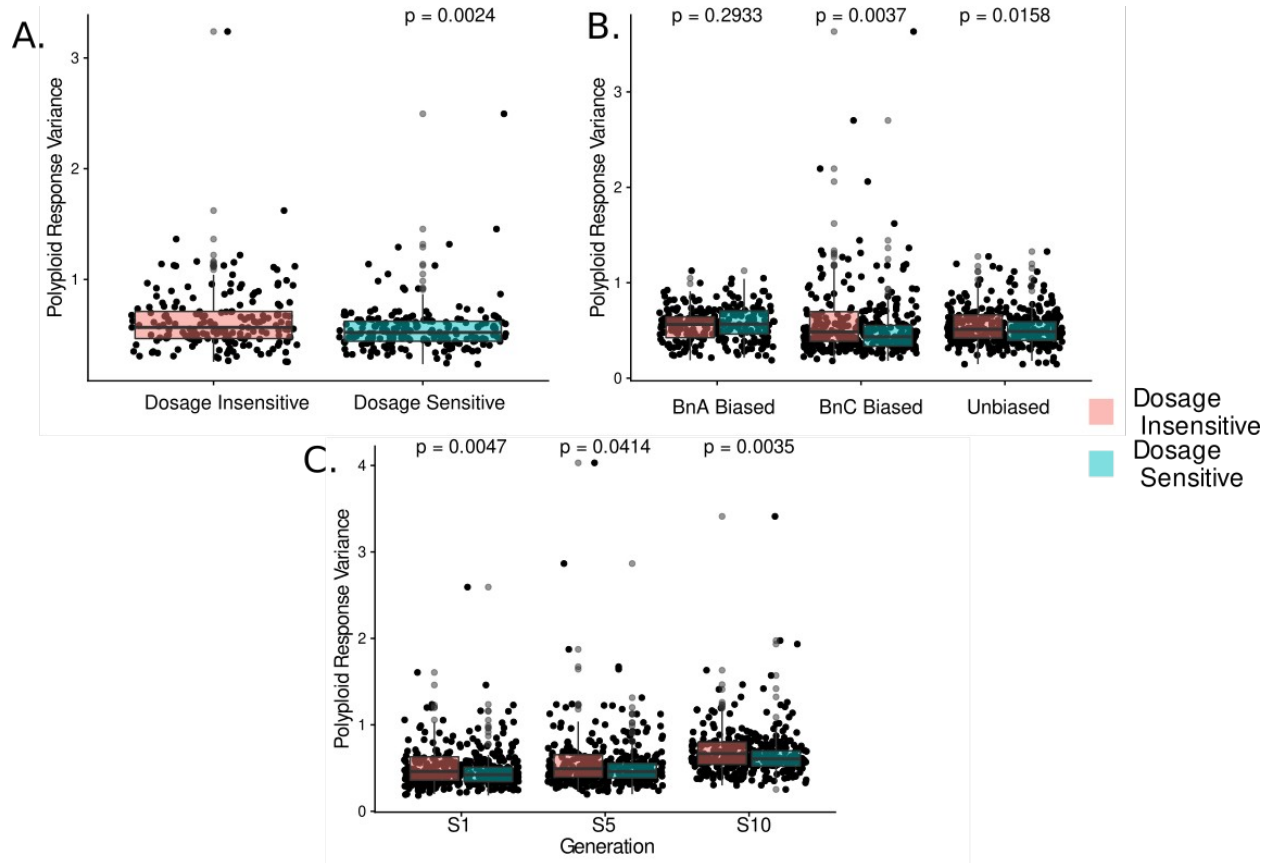
constraint exists on a spectrum, where the weakening of constraint is most prominent for dosage-insensitive genes while dosage-sensitive genes remain relatively unchanged over time.

Figure 5.1 Expression response to polyploid induced dosage changes



To further assess how the dosage sensitivity of genes affects their response to gene dosage changes from allopolyploidy, we used the dosage-balance-sensitivity gene class assignments for *Arabidopsis thaliana* from Song et al. (2020). As per Song et al. (2020), Class I Gene Ontology (GO) categories are putatively dosage-insensitive and Class II are putatively dosage-sensitive based on the observed reciprocal retention of genes from the investigated GO categories following polyploidy across the Angiosperms. To leverage the superior annotation quality of *A. thaliana*, *B. rapa* and *B. oleracea* orthologs were assigned to dosage-sensitivity GO classes based on their ortholog in *Arabidopsis*. These dosage-sensitivity assignments were used to assess how dosage response differs between classes in the resynthesized allopolyploids. We also used the polyploid response variance (PRV measure from Song et al. (2020) and Coate et al. (2016), defined as) the coefficient of variation of the relative expression response, to assess how coordinated the expression response to polyploidy is in the different gene classes.

Figure 5.2 Expression changes from allopolyploidy reflect predictions from the dosage balance hypothesis



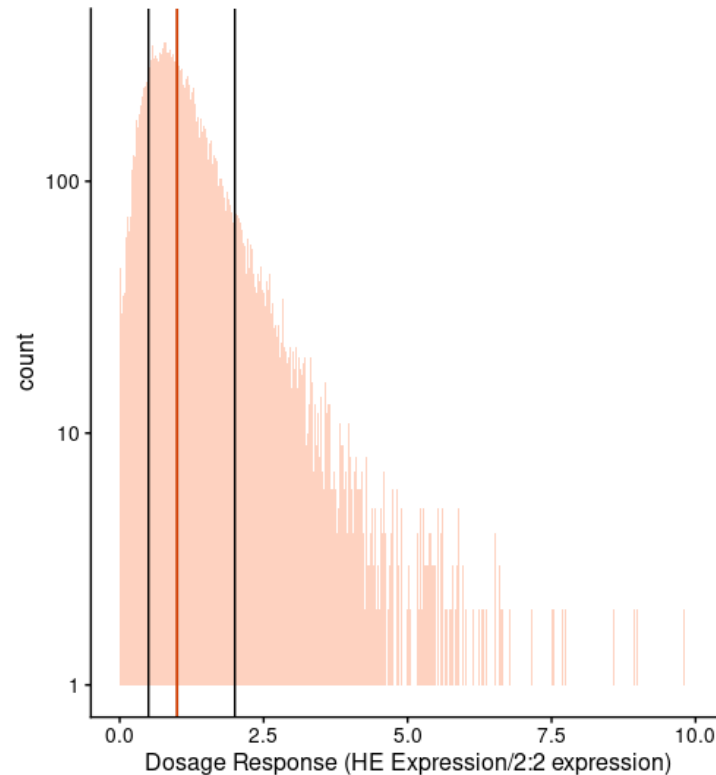
As observed previously in resynthesized autopolyploids and natural *Glycine* allopolyploids, the polyploid response variance was significantly lower (i.e. the expression response was more coordinated) in genes from GO categories in the dosage-sensitive class compared to the dosage-insensitive class (Kruskal-Wallis test, $p=0.0024$; Figure 5.2; Figure 5.2a). Using an allopolyploid gave us the opportunity to observe if gene pairs with different homoeolog expression biases respond differently to whole-genome duplication. We compared the dosage-sensitive and dosage-insensitive GO categories broken down by homoeolog expression bias of the gene pair and found that pairs with expression biased toward the *B. napus* C subgenome (BnC) biased and pairs with unbiased expression show the same significant difference between PRV of dosage-sensitive and dosage-insensitive GO categories as above (Kruskal-

Wallis test, $p=0.0037$; 0.0158). However, gene pairs biased toward the *B. napus* A subgenome (BnA) showed no significant difference in PRV between dosage-sensitive and insensitive GO classes (Kruskal-Wallis test, $p=0.2933$; Figure 5.2b). This result suggests that constraint on the gene dosage response manifests differently depending on homoeolog expression bias. When broken down by generation, we observe the an increase in the coefficient of variation over time, with both dosage-sensitive and dosage-insensitive showing higher PRV in generation ten than in the first generation (Fig 5.2c). Notably, in generation ten the dosage-sensitive GO categories show higher mean polyploidy response variance than dosage-insensitive GO categories in the first generation.

Expression changes from homoeologous exchanges appear to behave according to the gene-balance hypothesis

The extensive genomic rearrangements observed in this population of resynthesized lines (Xiong et al. 2011; Bird et al. 2021) provide an opportunity to test for the first time whether gene expression changes from homoeologous exchange events experience dosage balance constraints as predicted by the gene balance hypothesis. Using the published results from Bird et al. (2021), we focused on genomic regions identified as not be in 2:2 dosage, representing homoeologous exchanges with 0:4, 1:3, 3:1, and 4:0 dosage ratios (BnC:BnA). To avoid the inclusion of likely aneuploidy events, genes on chromosomes that frequently showed dosage changes for the entirety or majority of the chromosome were excluded. With this dataset of likely gene pairs affected by homoeologous exchange events, we compared their expression to the summed expression of the gene pair in the progenitor genomes. Plotting the expression response to homoeologous exchange shows a skewed distribution with a median of 0.99, almost equivalent to 1, which represents compensated expression. However, the distribution shows high variability in expression responses (Figure 5.3). Since each gene pair will have different expression fold change differences between homoeologs, it is impossible to know precisely which ratio represents a proportional dosage increase. Still, over 25% of homoeologous exchange gene pairs are either twice as expressed or half as expressed as when in a 2:2 dosage state (Figure 5.3). As before, our design prevents fully distinguishing the isolated effects of HEs from the impact of novel trans- regulation in the hybrid and allopolyploid genome.

Figure 5.3 Expression response to non-reciprocal homoeologous exchange induced dosage changes

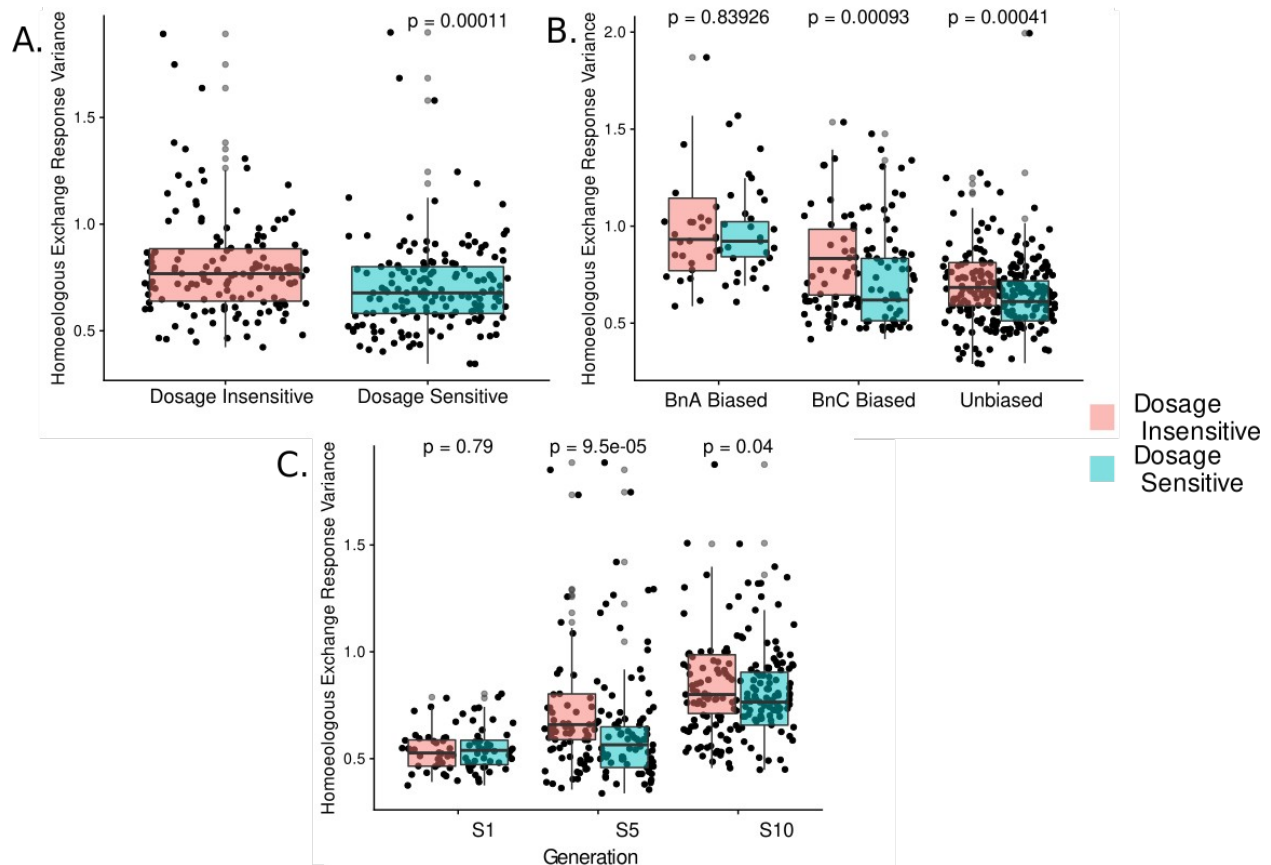


Next, we investigated the extent that expression responses from homoeologous exchanges systematically differ among the identified dosage-sensitive and dosage-insensitive GO categories (Fig 5.4). We again used the coefficient of variation, this time termed Homoeologous Exchange Response Variance (HERV), to assess how coordinated the expression response was for genes from dosage-sensitive and insensitive GO categories. Across all lines, genes belonging to putatively dosage-sensitive GO categories again showed significantly lower HERV, indicating a more coordinated expression response than that for genes from putatively dosage-insensitive GO categories (Figure 5.4a, Kruskal-Wallis test, $p=0.00011$). When broken down by direction of homoeolog expression bias we again see that homoeologous gene pairs with expression biased toward the dominant BnC subgenome (Kruskal-Wallis test, $p=0.00093$) and unbiased gene pairs (Kruskal-Wallis test, $p=0.00041$) show significantly lower HERV in dosage-sensitive GO terms than dosage-insensitive GO terms (Figure 5.4b). Again we see that

homoeologous gene pairs with expression biased toward the submissive BnA subgenome do not show a difference in homoeologous exchange response variance between dosage-sensitive and insensitive GO terms (Figure 5.4b, Kruskal-Wallis test, $p=0.83926$).

Furthermore, we found that there was not a significant difference in HERV between dosage-sensitive and dosage-insensitive GO terms in the first generation (Figure 5.4c, Kruskal-Wallis test, $p=0.79$), but dosage-sensitive and insensitive GO terms did show different HERV in the fifth and tenth generations (Figure 5.4c, Kruskal-Wallis test, $p=9.5 \times 10^{-5}$, $p=0.04$). We also found that homoeologous exchange response variance increased over time with dosage-sensitive and dosage-insensitive GO terms showing mean HERV of 0.547 and 0.540, respectively, in generation one and increasing to 0.789 and 0.860, respectively, in generation ten.

Figure 5.4 Expression changes from non-reciprocal homoeologous exchange reflect predictions from the dosage balance hypothesis



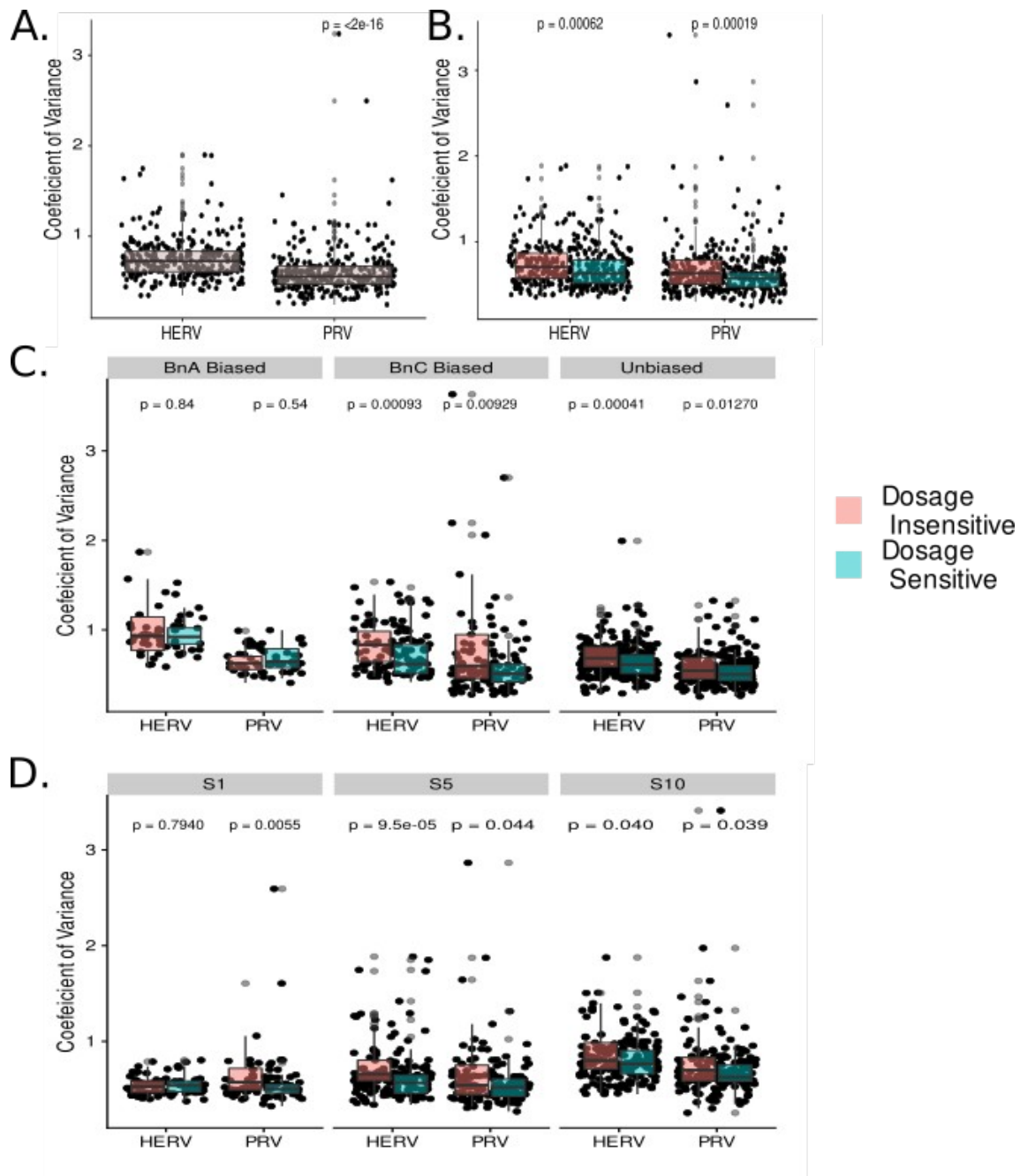
Expression changes from homoeologous exchanges are distinct from the effect of polyploidy

While our findings suggest that dosage changes caused by homoeologous exchanges increase the copy number of one homoeolog over the other, it is possible these results are an artifact of our analysis also picking up the effects of dosage changes caused by allopolyploidy or aneuploidy. To determine if the results obtained for homoeologous exchanges are distinct from the effect of polyploidy, we directly compared the coefficient of variation for the expression response of the two dosage change conditions (Figure 5.5).

First, we compared the proportion of gene pairs belonging to dosage-sensitive and dosage-insensitive GO terms in all 16 individuals for the polyploidy and homoeologous exchange analysis. For the polyploid analysis, the mean proportion of genes belonging to dosage-insensitive GO terms is 0.554, while it is 0.541 for the homoeologous exchange analysis, a significant difference (t-test, $p=0.021$). However, a greater proportion of gene pairs having dosage-insensitive GO terms would be predicted to result in a higher coefficient of variation. Instead, we found a significantly higher coefficient of variation from homoeologous exchanges (Figure 5.5a, Kruskal-Wallis test, $p<2\times 10^{-16}$), which had a lower proportion of genes belonging to dosage-insensitive GO categories. Both allopolyploidy and homoeologous exchange dosage changes produced significantly different expression responses from genes belonging to dosage-sensitive and insensitive GO categories (Figure 7b), and we determined that the coefficient of variation was significantly different between polyploidy and homoeologous exchange dosage changes for gene pairs from both dosage-sensitive (Kruskal-Wallis test, $p = 3.56\times 10^{-14}$) and dosage insensitive (Kruskal-Wallis test, $p=1.153\times 10^{-12}$) GO categories.

Likewise, for both homoeologous exchange and polyploidy induced dosage changes, the difference in expression response between genes belonging to dosage-sensitive and insensitive GO terms was significantly different for BnC biased and unbiased homoeologous pairs, but not for BnA biased pairs (Figure 5.5c). Our results also showed that the coefficient of variation from homoeologous exchange induced dosage changes was significantly higher than for polyploidy induced dosage changes for gene pairs belonging to both dosage-sensitive and insensitive for all homoeolog expression bias relationships (Table 5.1).

Figure 5.5 Expression responses from allopolyploidy and homoeologous exchange appear to be distinct



In generational comparisons, homoeologous exchange and polyploidy induced dosage changes showed the same patterns for differences in coefficient of variation in generations five and ten, but not generation one where the coefficient of variation did not significantly differ by dosage sensitivity for

homoeologous exchange induced dosage changes (Figure 5.5d). We also found that the coefficient of variation for homoeologous exchange induced dosage changes was significantly higher than for dosage changes induced by polyploidy for both dosage-sensitive and insensitive GO terms, but only for generations five and ten (Table 5.2).

That the expression response to homoeologous exchanges and polyploidy induced dosage changes are significantly different overall, and among several comparisons is strong evidence that the patterns observed for homoeologous exchange induced dosage changes are distinct from the effects of polyploidy induced dosage change. Furthermore, it is likely that dosage constraint is weaker for dosage changes from homoeologous exchange, leading to a less coordinated expression response compared to polyploidy. This is because the coefficient of variation for the expression response to homoeologous exchange dosage changes was higher than that for polyploidy induced dosage changes for both dosage-sensitive and dosage-insensitive GO terms.

Table 5.1 Kruskal-Wallis test exploring the difference in expression coefficient of variation from homoeologous exchange and allopolyploidy induced dosage changes broken down by dosage sensitivity and subgenome bias

GO Class	Subgenome Bias	HERV mean (SD)	PRV mean (SD)	X2	df	p-value
Dosage Insensitive	BnC Biased	0.846 (0.240)	0.792 (0.585)	7.428	1	0.0064
Dosage Insensitive	BnA Biased	0.997 (0.313)	0.656 (0.141)	22.948	1	9.90x10⁻⁷
Dosage Insensitive	Unbiased	0.708 (0.183)	0.585 (0.183)	26.173	1	3.12x10⁻⁷
Dosage Sensitive	BnC Biased	0.721 (0.269)	0.569 (0.331)	17.342	1	3.122x10⁻⁵
Dosage Sensitive	BnA Biased	0.930 (0.142)	0.681 (0.141)	22.69	1	1.90x10⁻⁶
Dosage Sensitive	Unbiased	0.634 (0.193)	0.525 (0.150)	34.658	1	3.93x10⁻⁹

Table 5.2 Kruskal-Wallis test exploring the difference in expression coefficient of variation from homoeologous exchange and allopolyploidy induced dosage changes broken down by dosage sensitivity and generation

GO Class	Generation	HERV mean (SD)	PRV mean (SD)	X2	df	p-value
Dosage Insensitive	S1	0.540 (0.0989)	0.629 (0.225)	2.9305	1	0.086
Dosage Insensitive	S5	0.747 (0.298)	0.634 (0.282)	8.6133	1	0.0033
Dosage Insensitive	S10	0.860 (0.231)	0.766 (0.381)	14.394	1	0.0015
Dosage Sensitive	S1	0.547 (0.0985)	0.551 (0.326)	2.6211	1	0.105
Dosage Sensitive	S5	0.615 (0.259)	0.555 (0.297)	5.4126	1	0.0199
Dosage Sensitive	S10	0.789 (0.214)	0.666 (0.198)	25.114	1	5.4x10⁻⁷

Discussion

The gene balance hypothesis has garnered extensive empirical support and has guided understanding of many aspects of genome evolution, such as biased retention of duplicate genes from particular functional categories (Maere et al. 2005; Paterson et al. 2006; Freeling, 2009; Tasdhigian et al. 2018). Two recent investigations have helped demonstrate the connection between gene expression responses to dosage changes and dosage sensitivity (Coate et al. 2016; Song et al. 2020). These authors showed in synthetic *Arabidopsis* autopolyploids and natural *Glycine* allopolyploids that the expression response to WGD in dosage-sensitive genes was more coordinated than for dosage-insensitive genes. They concluded that dosage constraints produce a coordinated expression for dosage-sensitive genes and that this provides a proximal mechanism by which dosage constraint can impact long-term gene retention.

By leveraging our population of resynthesized allopolyploid *B. napus* lines, this study directly tested how similar auto- and allopolyploids immediately respond to WGD. The unique aspects of *B. napus* also allowed for a novel investigation of how subgenome dominance interacts with dosage balance constraints and how dosage changes from homoeologous exchanges are constrained to maintain dosage balance. However, there are some key limitations to this study that warrant future follow-up. There are

several trans-effects on expression, both from hybridization and aneuploidy experienced in these lines that could not be controlled for when assessing expression changes. As such, the expression responses we detect are an unknown combination of responses to WGD and homoeologous exchange in addition to these trans-effects. However, previous analysis of gene expression in these resynthesized lines over ten generations showed that over 70% of genes showed the same biased expression toward the dominant subgenome and over 50% showed the same biased expression toward the non-dominant subgenome across all six lines and between the progenitor genomes (Bird et al. 2021). This suggests that trans-effects from hybridization and unshared genomic rearrangements should not entirely alter expression in a way that invalidates comparisons of progenitor genomes and resynthesized allopolyploids.

Additionally, due to the small number of genes generally affected by homoeologous recombination we combined all dosage combinations (AAAA, AAAC, ACCC, CCCC), which makes it difficult to ascertain the specific direction of expression changes or to isolate particular kinds of homoeologous exchanges. As genomic rearrangements accumulate and diversity over time, merging these factors will increase inter-individual variation. This means the comparisons across generations will be confounded by changing inter-individual variation and interpretation is not straightforward. If there were ways to generate or introduce homoeologous exchanges of a specific dosage in a controlled genetic background a more precise investigation of the effect of these dosage changes would be possible. Despite these shortcomings, this study provides new insight into the role of dosage constraint and gene balance in affecting gene expression changes from genomic rearrangements and opens up avenues for future investigation.

Evolutionary dynamics of early expression response to allopolyploidy

Our analysis of the relative expression response to allopolyploidy reinforces the idea that a general response to dosage changes is expression changing in a variety of ways ranging from compensation to dosage-dependent, as previously observed in an *Arabidopsis* aneuploid series (Hou et al. 2017), *Arabidopsis* autopolyploids (Song et al. 2020), and an *Arabidopsis* allopolyploid dosage series (Shi et al. 2015). We further identified similar patterns of more coordinated expression responses among putatively dosage-sensitive genes, similar to the reports from synthetic autopolyploid *Arabidopsis* (Song et al. 2020) and wild allopolyploid *Glycine* that originated ~500,000 years ago (Coate et al. 2016). Overall,

these results suggest that the effect of dosage constraint on the global expression response to polyploidy is similar between newly formed auto- and allopolyploids, as expected if dosage constraint was a general evolutionary force acting on all polyploid genomes immediately upon duplication.

Dosage constraint and selection on relative gene dosage is not the only evolutionary force that leads to biases in gene loss and retention following WGD. Subgenome dominance also drives the biased retention of genes from one subgenome in allopolyploid genomes. This biased retention is hypothesized to be caused by higher expression of homoeologs from the dominant subgenome (Schnable et al. 2011; Woodhouse et al. 2014; Renny-Byfield et al. 2015; Renny-Byfield et al. 2017). Importantly, because subgenome dominance only occurs in allopolyploid species, previous work on resynthesized autopolyploids (e.g. Song et al. 2020) could not investigate the interplay of dosage constraint and subgenome dominance. Our results suggest novel interaction between subgenome dominance and dosage constraint such that dosage-sensitive genes show more coordinated expression when homoeolog expression is unbiased or biased toward the dominant subgenome, but not when biased toward the non-dominant subgenome.

Over the long term, this would be predicted to preserve more dosage-sensitive genes from the dominant subgenome than the non-dominant. In line with this, Schnable et al. (2012) observed that biased retention of dosage sensitive genes broke down over time, with only 50% of genes retained from one genome duplication event being retained in duplicate after a subsequent duplication event. They further observed that the lower expressed copy was more likely to be lost and proposed the lower expressed copies contribute less to overall gene product dosage, and so experience less purifying selection and weaker dosage constraint (Schnable et al. 2012). Similarly, when subgenome dominance was first described in *Arabidopsis*, the dominant subgenome was also associated with the production of clusters of dosage-sensitive genes (Thomas et al. 2006).

Our results provide a unified account for short-term and long-term interactions of subgenome dominance and dosage constraint. Upon duplication, a more coordinated expression response for homoeologs biased toward the dominant subgenome will produce greater retention of dosage-sensitive genes from the dominant genome and concomitant under-retention from the non-dominant subgenome. Additionally, previous analysis of these resynthesized lines showed that homoeologous pairs biased

toward the dominant subgenome were highly connected in a protein-protein interaction network, while pairs with expression biased toward the non-dominant subgenome showed no such connectivity (Bird et al. 2021). This lack of connectivity may explain why putatively dosage-sensitive genes with biased expression toward the non-dominant subgenome do not show coordinated expression; without high connectivity in gene networks, they do not experience strong dosage constraints.

Selective constraints due to dosage sensitivity act immediately on duplicate genes and previous work suggests dosage constraint remains for long evolutionary periods, though is not permanent (Conant et al., 2014; Schnable et al., 2012). Although previous analysis of synthetic and natural *Arabidopsis* autopolyploids did not show marked differences in coordination of gene expression (Song et al. 2020), we observed a general increase in polyploid response variance for both dosage-sensitive and -insensitive genes over the ten generations observed, suggesting a decrease in coordination over a short period of time. Indeed, by the tenth generation, the dosage-sensitive genes showed less expression coordination than the dosage-insensitive genes in the first generation. This potentially suggests that the strength of dosage constraint starts to change earlier in polyploid evolution than previously thought. Alternatively, it is known that dosage changes induce trans-expression effects on chromosomes that did not have their dosage altered. In our plants, several genomic rearrangements occurred simultaneously with lines exhibiting aneuploidy and homoeologous exchanges and rearrangements occurring on multiple chromosomes. Later generations also accumulated more genomic rearrangements than earlier ones. We were unable to control or measure these kinds of trans dosage effects and they could potentially create inter-individual variation and drive these observed changes in expression coordination between earlier and later generations.

Previous analysis of duplicate gene retention across angiosperms described three broad groups of genes: those with a strong preference for single copy, those with duplicates retained in most or all species, and those that are retained as duplicates for a prolonged period of time and then return to single copy (Li et al. 2016). It is possible our results reflect the start of dosage constraint loosening on some of these intermediately retained genes. However, if our results were driven by inter-individual variation from trans effects, instead of showing a loosening of dosage constraint, we would be revealing a greater

tolerance for uncoordinated expression responses than one would infer from the levels of coordination in the first generation.

Homoeologous exchange and early polyploid genome evolution

Homoeologous exchanges have long been recognized as an engine of phenotypic diversity and novelty in newly formed polyploids (Pires et al. 2004; Gaeta et al. 2007). Our analysis of genomic rearrangements and homoeologous exchanges in resynthesized *B. napus* confirmed at higher resolution the extensive rearrangements in these lines (Gaeta et al. 2007; Xiong et al. 2011). Investigations of genome imbalance and dosage sensitivity have predominately focused on polyploidy and aneuploidy as the sources of gene dosage alteration (Hou et al. 2018; Yang et al. 2021; Shi et al. 2021). These studies have greatly increased our understanding of how changes in dosage affect cis- and trans-gene expression, and subsequent analysis has connected these kinds of expression changes to long-term evolutionary patterns of gene retention (Song et al. 2020). However, homoeologous exchanges, which alter the ratio of parental chromosomes, have also been shown to produce dosage-dependent expression changes (Lloyd et al. 2017). These dosage changes from homoeologous exchanges have not been investigated for dosage constraints or more general patterns of expression response expected from the gene balance hypothesis.

Our results show that expression response to homoeologous exchanges exhibits a variety of behavior with expression sometimes staying equal to the 2:2 expression level but other times increasing or decreasing far beyond that baseline. Because these HE events represent multiple dosage changes and directions, and the homoeolog specific expression levels change between gene pairs it's not clear what proportion is changing in a dosage-dependent or independent manner or being dosage compensated. Previous results from an *Arabidopsis* allopolyploid dosage (AAAA, AAAT, AATT, ATTT, TTTT) series showed that the majority of genes (54%) changed expression in a dosage-dependent manner for both homoeologs (Shi et al. 2015). However, our results suggest a more varied response to homoeologous exchange than Lloyd et al. (2017), who determined over 95% of expression changes from homoeologous exchanges were dosage-dependent. Overall, the variation in expression response from homoeologous exchanges appears to be broadly similar to the response to polyploidy. \

We further find that dosage changes resulting from homoeologous exchanges produce the same patterns of more coordinated expression responses from dosage sensitive genes. We also saw similar patterns of lower expression coordination in later generations and a lack of differences in expression coordination from homoeolog pairs biased toward the non-dominant subgenome that we observed when investigating expression response to polyploidy. Such results have not been reported before, to our knowledge, and suggest that homoeologous exchanges also experience selective constraint for balanced gene dosage in the same way as genes affected by polyploidy or aneuploidy.

If homoeologous exchanges evolve in ways predicted by the gene balance hypothesis then we might expect selection to disfavor homoeologous exchanges containing dosage-sensitive genes, producing biases in gene functions surviving homoeologous exchanges to be similar to small-scale duplications. Following these predictions, Hurgobin et al. (2017) and Bayer et al. (2021) identified a significant degree of gene presence-absence variation in *B. napus* arising from homoeologous exchanges, and these genes were associated with membership in the protein-protein interaction network (Bayer et al. 2021) and GO terms related to plant defense and stress pathways (Hurgobin et al. 2017). They also observed several homoeologous exchanges generating presence-absence variation in paralogs of the large gene family *FLC*, which regulates flowering time. Analysis of expression dynamics of *FLC* paralogs in *B. napus* showed that while *FLC* paralogs are dosage-sensitive, dosage constraints act on overall *FLC* gene family expression allowing compensatory drift (Thompson et al. 2016) and expression divergence (Calderwood et al. 2020). This *FLC* example shows that the interplay of homoeologous exchange and dosage constraint may be highly dynamic depending on the gene family in question. Homoeologous exchange may also drive systematic subgenome biases in the direction of homoeologous exchange. For example, Edger et al. (2019) proposed that constraints on stoichiometric balance and altered gene dosage explained the overwhelming bias in direction of homoeologous exchange, favoring the dominant subgenome, in the octoploid strawberry genome.

Our comparison of homoeologous exchange and polyploidy response variance showed that overall gene expression was less coordinated in response to homoeologous exchange compared to polyploidy. This may mean that genes affected by homoeologous exchange experience weaker dosage constraints, although it may also simply be due to high levels of inter-individual variation among lines.

While the patterns observed for homoeologous exchanges could be an artifact of the effect of polyploidy, the fact that the patterns for response to homoeologous exchange are significantly different than the polyploidy response suggests this is a distinct phenomenon. This could be a promising avenue for future comparative and evolutionary genomic studies to investigate.

REFERENCES

REFERENCES

- Alger, E. I., & Edger, P. P. (2020). One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Current opinion in plant biology*, 54, 108-113.
- Bayer, P. E., Scheben, A., Golicz, A. A., Yuan, Y., Faure, S., Lee, H., ... & Edwards, D. (2021). Modelling of gene loss propensity in the pangenomes of three Brassica species suggests different mechanisms between polyploids and diploids. *Plant Biotechnology Journal*.
- Birchler, J. A., Bhadra, U., Bhadra, M. P., & Auger, D. L. (2001). Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Developmental biology*, 234(2), 275-288.
- Birchler, J. A., & Newton, K. J. (1981). Modulation of protein levels in chromosomal dosage series of maize: the biochemical basis of aneuploid syndromes. *Genetics*, 99(2), 247-266.
- Birchler, J. A., & Veitia, R. A. (2007). The gene balance hypothesis: from classical genetics to modern genomics. *The Plant Cell*, 19(2), 395-402.
- Birchler, J. A., & Veitia, R. A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytologist*, 186(1), 54-62.
- Birchler, J. A., & Veitia, R. A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences*, 109(37), 14746-14753.
- Bird, K. A., Niederhuth, C. E., Ou, S., Gehan, M., Pires, J. C., Xiong, Z., ... & Edger, P. P. (2021). Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid Brassica napus. *New Phytologist*, 230(1), 354-371.
- Bird, K. A., VanBuren, R., Puzey, J. R., & Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist*, 220(1), 87-93.
- Blanc, G., & Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *The Plant Cell*, 16(7), 1679-1691.
- Calderwood, A., Lloyd, A., Hepworth, J., Tudor, E. H., Jones, D. M., Woodhouse, S., ... & Morris, R. J. (2021). Total FLC transcript dynamics from divergent paralogue expression explains flowering diversity in Brassica napus. *New Phytologist*, 229(6), 3534-3548.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., ... & Wincker, P. (2014). Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *science*, 345(6199), 950-953.
- Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.*, 58, 377-406.
- Coate, J. E., Song, M. J., Bombarely, A., & Doyle, J. J. (2016). Expression-level support for gene dosage sensitivity in three Glycine subgenus Glycine polyploids and their diploid progenitors. *New Phytologist*, 212(4), 1083-1093.

Conant, G. C., Birchler, J. A., & Pires, J. C. (2014). Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Current opinion in plant biology*, 19, 91-98.

De Smet, R., Adams, K. L., Vandepoele, K., Van Montagu, M. C., Maere, S., & Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences*, 110(8), 2898-2903.

Edger, P. P., & Pires, J. C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research*, 17(5), 699-717.

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., ... & Knapp, S. J. (2019). Origin and evolution of the octoploid strawberry genome. *Nature genetics*, 51(3), 541-547.

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., ... & Puzey, J. R. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *The Plant Cell*, 29(9), 2150-2167.

Ferreira de Carvalho, J., Stoeckel, S., Eber, F., Lodé-Taburel, M., Gilet, M. M., Trotoux, G., ... & Rousseau-Gueutin, M. (2021). Untangling structural factors driving genome stabilization in nascent *Brassica napus* allopolyploids. *New Phytologist*, 230(5), 2072-2084.

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual review of plant biology*, 60, 433-453.

Freeling, M., & Thomas, B. C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome research*, 16(7), 805-814.

Gaeta, R. T., Pires, J. C., Iniguez-Luy, F., Leon, E., & Osborn, T. C. (2007). Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *The Plant Cell*, 19(11), 3403-3417.

Gaeta, R. T., & Pires, J. C. (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytologist*, 186(1), 18-28.

Gaebelein, R., Schiessl, S. V., Samans, B., Batley, J., & Mason, A. S. (2019). Inherited allelic variants and novel karyotype changes influence fertility and genome stability in *Brassica* allohexaploids. *New Phytologist*, 223(2), 965-978.

Gonzalo, A., Lucas, M. O., Charpentier, C., Sandmann, G., Lloyd, A., & Jenczewski, E. (2019). Reducing MSH4 copy number prevents meiotic crossovers between non-homologous chromosomes in *Brassica napus*. *Nature communications*, 10(1), 1-9.

He, Z., Wang, L., Harper, A. L., Havlickova, L., Pradhan, A. K., Parkin, I. A., & Bancroft, I. (2017). Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNA seq-based visualization. *Plant biotechnology journal*, 15(5), 594-604.

Hou, J., Shi, X., Chen, C., Islam, M. S., Johnson, A. F., Kanno, T., ... & Birchler, J. A. (2018). Global impacts of chromosomal imbalance on gene expression in *Arabidopsis* and other taxa. *Proceedings of the National Academy of Sciences*, 115(48), E11321-E11330.

Hurgobin, B., Golicz, A. A., Bayer, P. E., Chan, C. K. K., Tirnaz, S., Dolatabadian, A., ... & Edwards, D. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant biotechnology journal*, 16(7), 1265-1274.

Kassambara, Alboukadel (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. <https://CRAN.R-project.org/package=ggpubr>

Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y., & De Smet, R. (2016). Gene duplicability of core genes is highly consistent across all angiosperms. *The Plant Cell*, 28(2), 326-344.

Lloyd, A., Blary, A., Charif, D., Charpentier, C., Tran, J., Balzergue, S., ... & Jenczewski, E. (2018). Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *New Phytologist*, 217(1), 367-377.

Lyons, E., & Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. *The Plant Journal*, 53(4), 661-673.

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., ... & Freeling, M. (2008). Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant physiology*, 148(4), 1772-1781.

Madlung, A., Tyagi, A. P., Watson, B., Jiang, H., Kagochi, T., Doerge, R. W., ... & Comai, L. (2005). Genomic changes in synthetic Arabidopsis polyploids. *The Plant Journal*, 41(2), 221-230.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., & Van de Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences*, 102(15), 5454-5459.

Makino, T., & McLysaght, A. (2010). Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proceedings of the National Academy of Sciences*, 107(20), 9270-9274.

Mason, A. S., & Wendel, J. F. (2020). Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Frontiers in Genetics*, 11, 1014.

Paterson, A. H., Chapman, B. A., Kissinger, J. C., Bowers, J. E., Feltus, F. A., & Estill, J. C. (2006). Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. *Trends in genetics*, 22(11), 597-602.

Pelé, A., Rousseau-Gueutin, M., & Chèvre, A. M. (2018). Speciation success of polyploid plants closely relates to the regulation of meiotic recombination. *Frontiers in plant science*, 9, 907.

Pires, J. C., Zhao, J., Schranz, M. E., Leon, E. J., Quijada, P. A., Lukens, L. N., & Osborn, T. C. (2004). Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae). *Biological Journal of the Linnean Society*, 82(4), 675-688.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Renny-Byfield, S., Gong, L., Gallagher, J. P., & Wendel, J. F. (2015). Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Molecular biology and evolution*, 32(4), 1063-1071.

Renny-Byfield, S., Rodgers-Melnick, E., & Ross-Ibarra, J. (2017). Gene fractionation and function in the ancient subgenomes of maize. *Molecular Biology and Evolution*, 34(8), 1825-1832.

Samans, B., Chalhoub, B., & Snowdon, R. J. (2017). Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species Brassica napus. *Plant Genome*, 10(3), 1-15.

- Schnable, J. C., Springer, N. M., & Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences*, 108(10), 4069-4074.
- Schnable, J. C., Wang, X., Pires, J. C., & Freeling, M. (2012). Escape from preferential retention following repeated whole genome duplications in plants. *Frontiers in plant science*, 3, 94.
- Shi, X., Zhang, C., Ko, D. K., & Chen, Z. J. (2015). Genome-wide dosage-dependent and independent regulation contributes to gene expression and evolutionary novelty in plant polyploids. *Molecular Biology and Evolution*, 32(9), 2351-2366.
- Shi, X., Yang, H., Chen, C., Hou, J., Hanson, K. M., Albert, P. S., ... & Birchler, J. A. (2021). Genomic imbalance determines positive and negative modulation of gene expression in diploid maize. *The Plant Cell*, 33(4), 917-939.
- Song, M. J., Potter, B. I., Doyle, J. J., & Coate, J. E. (2020). Gene balance predicts transcriptional responses immediately following ploidy change in *Arabidopsis thaliana*. *The Plant Cell*, 32(5), 1434-1448.
- Stein, A., Coriton, O., Rousseau-Gueutin, M., Samans, B., Schiessl, S. V., Obermeier, C., ... & Snowdon, R. J. (2017). Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. *Plant biotechnology journal*, 15(11), 1478-1489.
- Szadkowski, E., Eber, F., Huteau, V., Lode, M., Huneau, C., Belcram, H., ... & Chèvre, A. M. (2010). The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytologist*, 186(1), 102-112.
- Tasdighian, S., Van Bel, M., Li, Z., Van de Peer, Y., Carretero-Paulet, L., & Maere, S. (2017). Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity. *The Plant Cell*, 29(11), 2766-2785.
- Thomas, B. C., Pedersen, B., & Freeling, M. (2006). Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome research*, 16(7), 934-946.
- Thompson, A., Zakon, H. H., & Kirkpatrick, M. (2016). Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes. *Genetics*, 202(2), 765-774.
- Vicient, C. M., & Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Annals of Botany*.
- Wendel, J. F., Lisch, D., Hu, G., & Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Current opinion in genetics & development*, 49, 1-7.
- Woodhouse, M. R., Cheng, F., Pires, J. C., Lisch, D., Freeling, M., & Wang, X. (2014). Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proceedings of the National Academy of Sciences*, 111(14), 5283-5288.
- Wu, Y., Lin, F., Zhou, Y., Wang, J., Sun, S., Wang, B., ... & Liu, B. (2021). Genomic mosaicism due to homoeologous exchange generates extensive phenotypic diversity in nascent allopolyploids. *National Science Review*, 8(5), nwaa277.

Xiong, Z., Gaeta, R. T., & Pires, J. C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proceedings of the National Academy of Sciences*, 108(19), 7908-7913.

Zhang, Z., Gou, X., Xun, H., Bian, Y., Ma, X., Li, J., ... & Levy, A. A. (2020). Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proceedings of the National Academy of Sciences*, 117(25), 14561-14571.

CHAPTER 6

The work presented in this chapter is part of the final publication

Bird, K. A., Hardigan, M. A., Ragsdale, A. P., Knapp, S. J., VanBuren, R., and Edger, P. P.. 2021. Diversification, spread, and admixture of octoploid strawberry in the Western Hemisphere. *American Journal of Botany* 108(11): 2269– 2281.

Diversification, Spread, and Admixture of Octoploid Strawberry in the Western Hemisphere

Abstract

Polyploid species often have complex evolutionary histories that have, until recently, been intractable due to limitations of genomic resources. While recent work has further uncovered the evolutionary history of the octoploid strawberry (*Fragaria* L.), there are still open questions. Much is unknown about the evolutionary relationship of the wild octoploid species, *Fragaria virginiana* and *Fragaria chiloensis*, and gene flow within and among species after the formation of the octoploid genome. We leveraged a collection of wild octoploid ecotypes of strawberry representing the recognized subspecies and ranging from Alaska to southern Chile, and a high-density SNP array to investigate wild octoploid strawberry evolution. Evolutionary relationships were interrogated with phylogenetic analysis and genetic clustering algorithms. Additionally, admixture among and within species is assessed with model-based and tree-based approaches. Phylogenetic analysis revealed that the two octoploid strawberry species are monophyletic sister lineages. The genetic clustering results show substructure between North and South American *F. chiloensis* populations. Additionally, model-based and tree-based methods support gene flow within and among the two octoploid species, including newly identified admixture in the Hawaiian *F. chiloensis* subsp. *sandwicensis* population. *F. virginiana* and *F. chiloensis* are supported as monophyletic and sister lineages. All but one of the subspecies show extensive paraphyly. Furthermore, phylogenetic relationships among *F. chiloensis* populations supports a single population range expansion southward from North America. The inter- and intraspecific relationships of octoploid strawberry are complex and suggest substantial gene flow between sympatric populations among and within species.

Summary

This work sought to characterize the interp-specific and intra-specific phylogenetics, of the two wild octoploid species, *Fragaria virginiana* and *Fragaria chiloensis* using a global collection of *Fragaria* from the USDA in combination with a newly developed genotyping array. It also used population genomic methods to assess gene flow within and among this species pair. Coalescent-based phylogenetic analysis using Black raspberry (*Rubus occidentalis*) as an outgroup demonstrated that the two octoploid strawberry species are monophyletic sister lineages. The topology of the phylogenetic tree and the genetic clustering analyses revealed that the movement from North to South America created substructure between these populations of *F. chiloensis*. Finally, distinct methods to infer admixture and gene flow suggested sizable and detectable events within and among the two octoploid species. This includes a potential admixture event in the Hawaiian *F. chiloensis* subsp. *sandwicensis* population, whose sequence data was analyzed here for the first time. These results phylogenetic support a single population range expansion southward from North America that introduced substructure in these species, and that these species maintain complex inter- and intraspecific relationships involving substantial gene flow between sympatric populations among and within species. For this chapter, I conceptualized the study and analysis, and performed the phylogenetic and population genomic analyses and visualization of results. I also wrote the original draft and incorporated coauthors' and journal referees' edits and comments into subsequent drafts.