Multiple origins, one evolutionary trajectory: gradual genome evolution in the allotetraploid grass *Brachypodium hybridum*

By

Virginia T. Scarlett

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John P. Vogel, Co-Chair
Professor Benjamin K. Blackman, Co-Chair
Professor Michael Freeling
Professor Abby F. Dernburg

Spring 2022

Abstract

Multiple origins, one evolutionary trajectory: gradual genome evolution in the allotetraploid grass *Brachypodium hybridum*

by

Virginia T. Scarlett

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor John P. Vogel, Co-Chair

Professor Benjamin K. Blackman, Co-Chair

Polyploidy, the condition of having more than two sets of chromosomes, is very common in the plant kingdom. However, the role of polyploidy, or whole-genome duplication (WGD), in plant evolution is far from clear. In some cases, polyploidy appears to be an engine of saltational evolution, while in other cases it seems to have little appreciable effect. To understand the role of polyploidy in plant evolution, we must look to the genetics and genomics of the WGD event itself. The progenitors' genomes play a crucial role in establishing the polyploid's genetic stability and ultimate evolutionary trajectory.

We have investigated genome structure and evolution in *Brachypodium hybridum* (2n=4x=30). The purple false brome *B. distachyon* (2n=2x=10) is a well-established model organism, and it is also one of the progenitors of *B. hybridum*, the other being *B. stacei* (2n=2x=20). Despite the availability of many genetic resources for *B. distachyon*, little is known about the allotetraploid *B. hybridum*. Here, we describe several genomic features of two independent natural lineages of *B. hybridum.*

In chapter 1, I describe the transcriptional landscape of the model *B. hybridum* accession ABR113. We found that polyploidy had no appreciable effect on the transcriptome, and differences between progenitor and subgenome were within the level of variation we would expect between accessions of the same diploid species. We used a novel and straightforward analytical approach to cross-species RNA-seq that may be valuable to other researchers. We also demonstrated the importance of including the progenitors in studies of polyploid gene expression, as 'parental legacy' appeared to be the main driver of gene expression. This was part of a larger effort to survey the genome of ABR113, which was published in Gordon et al. (2020).

Chapter 2 constitutes the main project of my doctoral research. This project describes the genome of *B. hybridum* accession Bhyb26. Bhyb26 is an older lineage of *B. hybridum*: the Bhyb26 WGD event is estimated to have occurred 1.4 million years ago, while the ABR113 WGD occurred approximately 140,000 years ago (Gordon et al. 2020). Both lineages possess the expected karyotype, that is, a D subgenome of 10 chromosomes and an S subgenome of 20 chromosomes, with no major rearrangements or sequence loss. However, we found that the

Bhyb26 genome shows subtle genomic changes consistent with relaxed natural selection where the younger ABR113 genome did not. When we searched for biased genome evolution favoring one subgenome in Bhyb26, the results were mixed. We conclude that Bhyb26 is evolving toward the diploid state in a very gradual and largely unbiased manner. This study constitutes a rare glimpse of diploidization "caught in the act" on two evolutionary timescales.

Chapter 3 is a survey of the 3D genome topology of *B. hybridum* ABR113 and its progenitors using Hi-C technology. During the course of our survey, we stumbled onto a mysterious 3D chromatin structure in the *B. hybridum* genome. We find evidence that this 3D chromatin structure, composed of mutually contacting loci akin to the KEEs/IHIs described in *Arabidopsis* (Feng et al. 2014; Grob et al. 2014), was passed down from progenitors to polyploid. The two separate chromatin interaction networks merge to form a single network in the polyploid, which we refer to as the KNOT, in keeping with Grob et al. (2014). Transposable elements (TEs) in the *Brachypodium* KNOT bear certain characteristics that are consistent with proposed functions of the *Arabidopsis* KNOT in TE defense, though the participating loci are not as TE-rich as those hypotheses predict. This study approaches the question of 'parental legacy' from an unusual angle: the inheritance of 3D chromatin structural features.

In summary, we find that *B. hybridum* belongs in a class of polyploids that show a subtle and unbiased, rather than a disruptive, genomic response to WGD. We provide theoretical arguments explaining this observation. *B. hybridum* supports the hypothesis that even an allopolyploid from a very wide cross can undergo gradual, unbiased evolution provided that the progenitor genomes are similar in their TE load.

Dedication

To my parents, who never doubted me and never pressured me,

to John, who applied just the right amount of pressure,

to Cindy, who taught me to question myself,

to Megan, who taught me to believe in myself.

Acknowledgements

I am deeply indebted to more people than I can name here, and those who I do name deserve more thanks than I can convey on this page. First, the work conducted here would not have been possible without my advisor John Vogel. I was a feisty first-year grad student, with a lot of big ideas and high expectations. John nurtured those big ideas, allowing me to take risks but always keeping me focused on deliverables. John acted as my "sanity checker" again and again, helping me to decide which tasks were worth pursuing, which were worth delegating, and which belonged on the shelf. As I came to appreciate the immense amount of labor that goes into a single paper, I also came to appreciate the mountain of work that John and the whole *Brachypodium* community put into building the resources that made possible my project and my education.

On that note, I am also indebted to the *Brachypodium* community, one of the warmest and most collaborative professional spaces in which I've ever had the privilege of travelling. Pilar Catalán has been incredibly supportive of my research and my professional growth. From helping me navigate spreadsheets over Zoom to helping me practice my Spanish over beers, Pilar taught me that collaboration is about much more than grants and papers. Similarly, Robert Hasterok and all the wonderful people I met in his lab, including Dominika Idziak-Helmcke and her family, as well as Natalia Borowska-Żuchowska, Ewa Robaszkiewicz, and many others, overwhelmed me (in a good way) with both their prodigious scientific expertise and their kind generosity. In Katowice, I learned the meaning of the word 'hospitality'.

It takes a village to train a grad student. Countless people at the JGI helped me learn the ropes of science, from Vasanth Singan teaching me about RNA-seq, to Dawn Chiniquy and Ben Cole sharing their grad school do's and don't's, to Hope Hundley and Ronan O'Malley without whose patient, open collaboration chapter 3 of this dissertation would not exist. Amy Cartwright and Sean Gordon took me under their wings when I had no clue what I was doing. I'll never forget the look on Sean's face when I asked him, "What's the command line?" and he realized just how much guidance I would need to get up and running with the supercomputer. Of course, he stuck with me and we got there. Amy taught me more about *Brachypodium* than anybody. She was knowledgeable, hard-working and patient, a role model I still aspire to. She also taught me important lessons about designing my life and appreciating the value of my own labor. In the later years of my PhD, Li Lei was a wonderful mentor and friend. With her love for science and her open attitude, Li embodies what's best about science—passion, collaboration, innovation, and fun. There are many others who contributed to this work and to whom I am incredibly grateful, including Mike Shao, Przemek Tomczyk, and Jacob Espinosa. Their efforts were vital not only to my project but also to my personal growth, and they have my deepest thanks.

The PMB department at UC Berkeley has been a warm and cozy incubator for my intellectual and emotional development. Interactions with my cohort and with folks in the DEB book club never failed to brighten my day. Peggy Lemaux and the CLEAR team helped me see how the world sees science, and more importantly they made me rethink how I see myself. Ben Blackman was a dedicated mentor in all his roles—teacher, quals chair, committee member, and faculty liaison to the (sometimes disgruntled) Bio1B teaching assistants. Abby Dernburg, while not part of the PMB department, played a vital role in supporting my work, and consistently gave me very sound advice at every committee meeting. Finally, Mike Freeling inspired me more than

he knows, with his audacity to do what so few scientists are both willing and able to do—to have fun with ideas.

Finally, I would not have even made it *to* grad school, much less *through* grad school, without the support of my friends and family. "Dad jokes" are a cliché, but it's true that my dad has a way of turning hopelessness into humor, and conjuring the sharp insight that humor brings. Long phone calls with Mom, even the tearful ones, always left me feeling empowered and optimistic for the future. My parents are goofy, crazy, and about as close to perfect as parents get. My siblings Tony, Lisa, Natassia and Sean are my cheer squad, my sounding boards, and my role models. I struggle to appreciate what a privilege it is to be their sister. Rachel, my soul sister, does more than keep me sane—she's somebody I can be crazy with. Heaven knows I needed that over these last six years.

Every time I look up papers on Google Scholar, I see their motto, "Stand on the shoulders of giants." At first, I thought those giants were the big names in the literature, the biologists that all biologists have heard of. To some extent that's true, but those scientific celebrities are not really *my* giants. *My* giants, the people who support me, who lift me up, and who help me grow, are not necessarily famous. Like giants, they may not feel that their actions are big, but to me their actions are huge. Which just goes to show that one does not have to be big to be a giant, and that sometimes small actions can move mountains.

Table of Contents

General Introduction

Plant genomes have a far greater capacity for structural upset than our own mammalian genomes. There are most likely several reasons for this, notably plants' lifelong regeneration of the germline, their production of large quantities of gametes for external fertilization, and the developmental and morphological plasticity that is demanded by their sessile lifestyle (Kejnovsky et al. 2009). A consequence of plants' genomic resilience is that they show astounding diversity in genome size (Leitch and Bennett 2004; Leitch and Bennett 2007). For example, in the genus of carnivorous corkscrew plants *Genlisea*, *G. tuberosa* has a genome of size ~61Mbp, while its close relatives *G. violacea* and *G. lobata* have genomes of ~1700Mbp. (Fleischmann et al. 2014). The largest known eukaryotic genome is the Japanese canopy flower *Paris Japonica*, at an impressive 152Gbp (Pellicer et al. 2010). The genome size variation observed in plants can be attributed in part to massive genome restructuring events that can instantly increase genome size, including interspecific hybridization and whole-genome duplication (WGD). It has been estimated that roughly 35% of vascular plant species are polyploid (Wood et al. 2009), and nearly all plant lineages have at least one WGD event in their evolutionary history (Clark and Donoghue 2018). Considering the prevalence of polyploidy among plants, it is perhaps not so remarkable that plant genomes can be so large, but rather that any of them are small (Wang et al. 2021).

Polyploidy is the condition of having more than two sets of chromosomes. Polyploids usually have an even number of sets of chromosomes, though stable triploids and pentaploids do exist. Polyploidy is common in plants, and not uncommon in amphibians and fish as well. Many of our favorite crops are polyploid including cotton, wheat, bamboo, sugarcane, peanut, strawberry, banana, quinoa, potato, and more. This may not be a coincidence: hybridization can produce unexpectedly vigorous plants (Shull 1914), and while the association between polyploidy and domestication remains a matter of debate, there is evidence that polyploid plants have historically been more likely to be selected for domestication than diploid relatives (Salman-Minkov et al. 2016). Human goals aside, polyploidy appears to be an important force in plant evolution. While the precise relationship between polyploidy and speciation and/or diversification remains elusive (Kellogg 2016), no plant scientist today, it seems, would regard polyploidy as a consistent "evolutionary dead-end" (Stebbins 1950; Soltis et al. 2014). The prevalence of polyploidy in plant lineages (Clark and Donoghue 2018), the fact that redundant sequence is lost non-randomly over evolutionary time (Freeling 2009), the apparent burst of WGD events 66 MYA during a period of global climate change (Van de Peer et al. 2017; Van de Peer et al. 2021), and isolated case studies in which polyploids seem to have a fitness advantage over their progenitors (Ramsey 2011; Godfree et al. 2017; Wei et al. 2019; Lovell et al. 2021), all suggest that, provided the initial founder(s) can overcome the challenges of establishment (Thompson and Lumaret 1992; Ramsey and Schemske 2002), polyploidy provides genetic novelty that may promote ecological success.

Most diploid plants are descended from polyploid ancestors (Clark and Donoghue 2018). Modern diploid plants have undergone a process known as diploidization or genome downsizing to the diploid state (Soltis et al. 2016). This process typically takes tens of millions of years, with DNA losses averaging to 4–70 Mbp per million years in flowering plants (angiosperms) (Wang et al. 2021). Compared to gymnosperms, angiosperms have relatively high rates of WGD but relatively small genomes, suggesting that angiosperms face selective pressure for genome size reduction. Many explanations for this have been speculated, from reduction of nutritional

demands for nitrogen and phosphorous, to promotion of genetic linkage of beneficial alleles (Bowers and Paterson 2021; Wang et al. 2021). Whatever the causes, the fact that flowering plants undergo a 'wondrous cycle' of diploidy, instant polyploidy, and gradual return to diploidy is well-established (Wendel 2015).

Sometimes a distinction is made between genetic diploidization, that is, the loss or inactivation of duplicate genes, and cytological diploidization, or the suppression of aberrant chromosome pairing and/or restoration of disomic inheritance (Feldman et al. 2012; Grandont et al. 2013). For our purposes, diploids are plants that contain homologous chromosomes but not homeologous chromosomes (that is, 'corresponding' chromosomes from distinct progenitor lineages). I use the term 'diploidization' to refer to the holistic process of reaching this diploid state. To refer specifically to the acquisition of bivalent chromosome pairing in nascent polyploids, I will use the term 'cytological diploidization'.

Polyploids can arise through several different pathways. One is somatic genome doubling, that is, through mitotic non-disjunction in the zygote, in the embryo, or in a meristem. However, it is generally accepted that this is not the most common route to polyploidy in nature (Ramsey and Schemske 1998). Most natural polyploids form sexually, through one of several pathways involving unreduced gametes (Tayalé and Parisod 2013). Unreduced gametes may form through any of a number of possible meiotic abnormalities (Pelé et al. 2018). Interestingly, these abnormalities are to some degree under genetic control, and those genetic controls themselves can be sensitive to environmental stress, again provoking speculation that polyploidy can be a path to ecological success, especially during challenging periods (Pelé et al. 2018).

The two most common ways of classifying polyploids are by age (Lewis 1980; Gaut 2002; Blanc and Wolfe 2004) and by the genetic divergence between progenitors (Clausen et al. 1945; Manton 1950; Stebbins 1950; Stebbins 1971; Ramsey and Schemske 1998). The distinction by age uses the terms neopolyploid and paleopolyploid (and occasionally the less common term mesopolyploid). The term neopolyploid is generally reserved for resynthesized polyploids—those that have been recreated in the laboratory through manual crossing—and natural polyploids that formed "recently", i.e. on the order of decades or centuries ago. Paleopolyploid generally refers to diploids whose genomes reveal footprints of an ancient WGD event, namely, homeologous regions, but not homeologous chromosomes. Because a WGD preceded the diversification of angiosperms, all diploid angiosperms are actually paleopolyploids (Jiao et al. 2011).

Classification by genetic divergence uses the terms auto/allopolyploidy. Autopolyploids formed from the union of two similar genomes, usually of the same species, while allopolyploids formed from the union of distinct genomes, usually of different species. (This is not remarkable in angiosperms, where approximately 25% of species do hybridize in nature with another species (Mallet 2005).) In other words, allopolyploids are essentially hybrids and polyploids at the same time. Autopolyploids are less well-studied, but they are probably about as common as allopolyploids in nature (Barker et al. 2016). The terms auto- and allo- are today generally regarded as opposite ends of a spectrum, with most natural polyploids lying somewhere in between, having formed from progenitors that are neither completely identical nor completely unrelated (Mason and Wendel 2020).

While these classification schemes are useful and informative, they have limitations. Polyploidy is too complex to be described as a single axis with two poles. A richer framework for understanding polyploidy would have to consider several crucial variables. The "genomic shock response" is a useful framework for understanding polyploidy. Through it, we see how

characteristics of the progenitor genomes set a polyploid on a particular evolutionary trajectory, while at the same time emphasizing the potential for polyploid genomes to defy expectations.

Rethinking Genome Shock

The concept of genome "shock" was proposed by McClintock (1984), and has since been taken up by the polyploidy community. The term is usually not well-defined by those who employ it, in part because McClintock herself used it rather expansively, if not a little ambiguously. McClintock's research established that large-scale chromosomal rearrangements such as deletions, duplications and translocations are common byproducts of the cell's struggle to maintain genomic and mitotic integrity following chromosome breakage. These rearrangements are not always predictable, especially given that they may activate potentially destructive transposable elements. Toward the end of her illustrious career, McClintock hinted that hybridization with or without WGD may be one more example of a genome challenge that could elicit unpredictable genome restructuring (McClintock 1984).

Today the term "genomic shock" is often used in the polyploidy literature to indicate any sort of dramatic consequence of WGD. For example, Bird et al. (2018) refer to genomic shock as an event "that induces a series of rapid genetic and epigenetic modifications as a result of conflicts between parental genomes" (Bird et al. 2018). This evolution of the term reflects not mere ignorance of McClintock's intention, but an evolving understanding of the nature of the WGD response. New technologies and new data have revealed that the genome responds to WGD at many organizational levels. Furthermore, the mechanistic underpinnings of the most dramatic WGD responses point again and again to a role for progenitor genome divergence. The scientific community's continued interest in the term "genomic shock" speaks to the persistence of McClintock's question: is the genome response to hybridization/WGD predictable or not?

Decades of polyploidy research reveal that the response to WGD is neither completely chaotic nor completely deterministic. While there is still no "formula" to predict a plant's response to WGD, broad trends can be pointed to. In the next two subsections, I will briefly review the trends in plant responses to WGD. We will see that all of these trends have exceptions. Our mini-review will focus on allopolyploidy, on newer polyploids rather than on paleopolyploids, and highlight the importance of progenitor divergence. This sets the stage for our investigation of the allotetraploid *B. hybridum*, which sheds light on the role of parental conflict in the genomic shock response.

Overview of Responses to WGD

A common response to WGD is meiotic instability. The most common type of meiotic aberration in polyploids is the formation of synaptic multivalents—pairwise associations of three or more chromosomes during prophase I (Grandont et al. 2013), which may or may not lead to improper pairing at metaphase (Jenkins 1986). Multivalent formation has been observed in some, but by no means all, resynthesized auto- and allopolyploids (Ramsey and Schemske 2002; Grandont et al. 2013; Li et al. 2021). Natural polyploids, auto and allo, generally show bivalent pairing. Thus the process of cytological diploidization is instantaneous in some polyploids and evolved in others (Yant et al. 2013; Li et al. 2021).

Cytogenetic and genomic studies have demonstrated that multivalent formation can result in aneuploidy and/or crossover events between non-homologous—usually homeologous—

chromosomes (Nicolas et al. 2007; Chester et al. 2012; Zhang et al. 2013). Among allopolyploids, the classic case studies for these phenomena are *Brassica napus* and the natural neoallopolyploid *Tragopogon miscellus*. In both organisms, sequence loss is rampant in the early generations following polyploidy (Gaeta et al. 2007; Szadkowski et al. 2010; Xiong et al. 2011; Buggs et al. 2012; Chester et al. 2012). Some of this sequence loss is due to aneuploidy, and some is due to homeologous exchange, that is, homeologous recombination leading to reciprocal or non-reciprocal translocations between homeologs ('swaps' or duplications/deletions, respectively) (Gaeta and Pires 2010; Soltis et al. 2012). Like its synthetic counterpart, the genome of natural *B. napus*, which formed at most 7500 years ago, also bears signs of extensive chromosomal rearrangement, most of which appears to be due to homeologous recombination (Chalhoub et al. 2014; Samans et al. 2017). Rapid sequence loss has been observed in other polyploids as well, including *Nicotiana tabacum* (Skalická et al. 2005; Renny-Byfield et al. 2012) and certain wheats (Feldman et al. 1997; Liu et al. 1998; Ozkan et al. 2001; Shaked et al. 2001). However, not all polyploids show such dramatic genome reorganization. Some species, such as cotton and coffee, show homeologous exchange but not large-scale structural rearrangement (Brubaker et al. 1999; Liu et al. 2001; Guo et al. 2014; Lashermes et al. 2014). Resynthesized *A. suecica* shows aneuploidy and chromosome breaks, but scant evidence of homeologous recombination (Madlung et al. 2005). Finally, a number of species show no evidence of aneuploidy, homeologous exchange, or large-scale chromosomal rearrangement (Douglas et al. 2015; Griffiths et al. 2019; Hardigan et al. 2020; VanBuren et al. 2020; Burns et al. 2021).

The most obvious explanation for why some neoallopolyploids undergo chromosomal rearrangement, particularly between homeologs, while others do not is that in some allopolyploids, the homeologous chromosomes are sufficiently diverged in karyotype and/or sequence homology that improper pairing is unlikely. Indeed, Xiong et al. (2011) found that the more similar homeologs were, the more more likely they were to undergo chromosomal rearrangement in resynthesized *B. napus* (Xiong et al. 2011). By this rule, allopolyploids with distinct progenitors would be expected to undergo the least chromosomal rearrangement, while autopolyploids and allopolyploids with similar progenitors would undergo the most. While this is true on the whole, a literature review by Ramsey and Schemske (2002) revealed that autopolyploids form bivalents more than would be theoretically expected, and allopolyploids form multivalents more than would be expected (Ramsey and Schemske 2002). Likewise, some allopolyploids show more chromosomal rearrangements than some autopolyploids (Ozkan et al. 2006). Among natural *Nicotiana* polyploids, for example, the only lineage with clear evidence of homeologous exchange is *N. tabacum*, the Nicotiana polyploid with the *most* diverged progenitors (Leitch et al. 2008).

One possible explanation for these exceptions is variation in genetic control of chromosome associations. Meiotic pairing has been shown to be under genetic, typically polygenic, control in wheat, *B. napus*, *A. suecica*, *A. arenosa*, and others (Jenczewski and Alix 2004; Grandont et al. 2013; Hollister 2015). It remains to be seen whether 'pairing control loci' such as *Ph1* in wheat can fully explain the discrepancy between expected and observed rates of post-WGD chromosomal rearrangement/homeologous exchange. Wang et al. (2004) speculate that epigenetic changes associated with WGD may lead to altered expression of canonical DNA recombination/repair genes, and this too, might affect the incidence of chromosomal rearrangements. It has also been observed that the pathway of polyploid formation affects the frequency of aneuploidy and chromosomal rearrangements (Szadkowski et al. 2011; Pelé et al.

2018). Demographic factors, too, may affect the rate of genome stabilization (Le Comber et al. 2010).

Changes to epigenetic features such as histone modifications, microRNA expression, and in particular DNA methylation and small interfering RNA (siRNA) expression, are also commonly observed in new polyploids (Madlung and Wendel 2013; Song and Chen 2015). These epigenetic changes may largely explain the gene expression changes that are often observed in polyploids, a topic I will review in chapter 1. Epigenetic changes can be extensive and can influence phenotypes (Ding and Chen 2018). In the genus *Spartina*, for example, two natural hybrids of independent origin and a related neoallopolyploid were subjected to methylation sensitive amplified polymorphism (MSAP) analysis (Salmon et al. 2005). It was found that about 30% of parental fragments were differently methylated in the hybrids and polyploid relative to their parents. It has been speculated that this epigenetic plasticity may explain the surprising phenotypic diversity of *S. anglica* (Madlung and Wendel 2013). Interestingly, most of the methylation changes were due to hybridization rather than WGD per se (Parisod et al. 2009), similar to the findings of Madlung et al. (2002) in Arabidopsis, though not entirely consistent with the findings of Kashkush et al. (2002) and Shaked et al. (2001) in wheat. Thus, like chromosome restructuring, epigenetic reprogramming following WGD is sometimes observed, occasionally extensive, and is not easily mechanistically explained.

Derepression (transcription and/or transposition) of certain TEs is sometimes observed in polyploids as well (Vicient and Casacuberta 2017). This is perhaps not surprising, given that most TEs are epigenetically silenced, and epigenetic changes following polyploidy are not uncommon (Gill et al. 2021). In some cases, changes in DNA methylation, small RNA expression, or other epigenetic features have been directly linked to loss of TE silencing (Kashkush et al. 2002; Madlung et al. 2005; Yaakov and Kashkush 2012; Ben-David et al. 2013) (but see Ha et al. (2009)). TE derepression following hybridization/WGD typically seems to involve a modest number of TE copies from one or a few TE families (Madlung et al. 2005; Parisod et al. 2009; Yaakov and Kashkush 2012; Sarilar et al. 2013; Gantuz et al. 2021). Actual transposition is generally not observed for most transposons (Kashkush et al. 2003; Kraitshtein et al. 2010; Mestiri et al. 2010; Zhao et al. 2011; Usai et al. 2020). In many cases, the TE derepression is largely limited to one or two generations post-polyploidy, and TE repression is eventually restored (Kraitshtein et al. 2010; Yaakov and Kashkush 2012; An et al. 2014). The case of hybrid sunflower, while not a study of WGD, is often cited as an extraordinary TE burst: the same sublineage of transposable elements was largely responsible for a genome size increase of more than 50% in three independent natural diploid hybrids (Ungerer et al. 2006; Ungerer et al. 2009). However, this was an exceptionally massive TE proliferation, and it is still not known whether the TE activity was due to genetic incompatibilities associated with hybridization, by genetic drift in the early founder populations, and/or by environmental stress. So the sunflower case does not necessarily demonstrate that massive TE bursts are to be expected in hybrids or polyploids. Nevertheless, numerous studies have established that WGD can sometimes instantly induce expression and/or mobilization of certain TEs.

The mechanism of epigenetic reprogramming and possible TE mobilization following WGD remains somewhat obscure, but again progenitor divergence is expected to play a role. Stochastic TE activity in natural polyploids with different combinations of parental genomes (Parisod et al. 2012; Senerchia et al. 2015), suggests that the parental combination may affect the epigenetic impact of WGD. Variation in genetic and epigenetic outcomes in new polyploids is probably due to genetic incompatibilities between parental genomes. The most well-studied

mechanism for this is trans chromosomal methylation/demethylation (Martienssen 2010; Groszmann et al. 2011). TE silencing in plants is largely mediated by siRNAs, particularly 24 nucleotide-long siRNAs, that maintain TE repression through RNA-directed DNA methylation (Gill et al. 2021). If a locus is methylated via a small RNA-directed pathway in one parent but not the other, those small RNAs may target all parental alleles indiscriminately in the hybrid embryo or endosperm, leading to hypermethylation of the formerly unmethylated allele. This process would be expected to be sensitive to parental genome dosage, so each parent's ploidy and origin as maternal or paternal genome donor would affect outcomes (since the maternal parent's genome is overrepresented 2:1 in the triploid endosperm). Studies of resynthesized *Arabidopsis* and *Mimulus* hybrids and polyploids bear out this theory (Josefsson et al. 2006; Greaves et al. 2012; Rigal et al. 2016; Kinser et al. 2021), though Satyaki and Gehring (2019) caution that the full picture is more complex (Satyaki and Gehring 2019). The greater body of work on hybrid incompatibility suggests that epigenetic changes, and subsequent changes in gene and TE expression, would be expected in polyploids whose progenitors dissimilar in their epigenetic constitution.

Plant genomes also respond to hybridization/WGD at the level of spatial arrangement. It is well-established that the "noodle soup model" of chromatin organization in the nucleus is incorrect; rather, chromosomes occupy non-randomly arranged chromosome territories (Cremer and Cremer 2006). The effect of hybridization on chromosome territories has been cytologically investigated in natural and resynthesized wide hybrids, polyploids, and chromosome introgression lines (lines in which a chromosome from another species has been introduced) (Jones and Hegarty 2009; Vimala and Lavania 2021). These studies reveal that in some cases, the distinct subgenomes (or the introgressed chromosome(s) and the background genome), are spatially separated in the interphase hybrid nucleus, a phenomenon sometimes called "genome territories" (Concia et al. 2020; Vimala and Lavania 2021). In these cases, chromosomes from the subgenome with larger chromosomes tend to be located at the nuclear periphery, although in unstable hybrids the genome destined to be eliminated tends to occupy the periphery regardless of chromosome size (Vimala and Lavania 2021). In allopolyploids and hybrids, subgenome positioning probably depends on progenitor differences in chromosome size, gene density, and kinetochore assembly (Heride et al. 2010; Sanei et al. 2011; Marimuthu et al. 2021). It is tempting to speculate that mitotic and interphase chromosome organization in stable hybrids and allopolyploids may affect rates of transcription and DNA repair, and perhaps impact plant success (Cavalli and Misteli 2013; Sunder and Wilson 2019). However, this question remains under-explored.

In summary, different plants respond to WGD in different ways. Some plants show extensive sequence loss, gene expression change, and epigenetic change. Others show more modest responses, such as sequence exchange with minimal loss (Bertioli et al. 2019; Hu et al. 2021), or epigenetic change with minimal sequence change (Parisod et al. 2009; Zhao et al. 2011). Some plants show only subtle changes that probably emerged over evolutionary time (Douglas et al. 2015; Ågren et al. 2016; Baduel et al. 2019). In some cases, such as *A. suecica* (Madlung et al. 2005; Burns et al. 2021), resynthesized polyploids and their natural counterparts are quite different, suggesting that nature may select for polyploids with certain characteristics, or that, as McClintock predicted, the WGD response may involve a certain element of random chance. There is no single genome feature that predicts the WGD response. Progenitor divergence is one critical factor, along with homolog recognition mechanisms and ecological context.

Genome dominance and long-term trajectories

Does a plant's initial response to WGD have anything to do with its long-term fate? This question is critical to understanding the history of modern polyploids and the fundamental principles of polyploid evolution. The evidence is mixed, but there is persuasive evidence that, at least in some cases, the evolutionary trajectory of a polyploid is established early on. The biological factors that dictate whether and how a polyploid's fate is sealed is a matter of much debate.

The discourse on paleopolyploidy centers largely on the important phenomenon of genome dominance. Genome dominance refers to the observation that in some paleopolyploids, genes derived from one ancient progenitor are more likely to be retained and are more highly expressed (Schnable et al. 2011; Woodhouse et al. 2014; Cheng et al. 2018). Genes from the other ancient subgenome(s) are overall lower-expressed, less abundant, and are more degraded or "fractionated" (Langham et al. 2004; Thomas et al. 2006).

It has been proposed that there are two "kinds" of paleopolyploid: those that show genome dominance and those that do not (Garsmeur et al. 2014). Two classic case studies are maize and soybean. Maize shows strong subgenome dominance (Schnable et al. 2009; Woodhouse et al. 2010; Schnable et al. 2011; Woodhouse et al. 2014). Soybean shows genome downsizing, that is, diploidization, but neither subgenome is markedly higher in gene content or gene expression (Zhao et al. 2017). It has been suggested that these two "types" of paleopolyploid exemplified by maize and soybean may correspond to descendants of allopolyploids and autopolyploids, respectively (Garsmeur et al. 2014; Zhao et al. 2017). Genome dominance has been observed in a number of paleopolyploids including *Arabidopsis thaliana* and *Brassica* species (Thomas et al. 2006; Murat et al. 2014; Parkin et al. 2014; Cheng et al. 2016). Other paleopolyploids such as pear and poplar do not show genome dominance (Garsmeur et al. 2014; Li et al. 2019). These studies generally conform to the expectation that allopolyploids show genome dominance while autopolyploids do not, though there are exceptions. Genome dominance is subtle in paleoallotetraploid *Miscanthus*, and absent in paleoallotetraploid *Cucurbita* genomes (Sun et al. 2017; Mitros et al. 2020).

Freeling et al. (2012) and Woodhouse et al. (2014) proposed a mechanism for genome dominance based on the "positional effect" of TE silencing. They suggested that the subgenome that started out with the lower TE load—that is, the genome with lower relative TE content—would be more likely to become the dominant subgenome. The subgenome(s) with a higher TE load would experience more gene silencing, primarily due to the "spread" of repressive DNA methylation marks from TEs to nearby genes, a phenomenon that is not without empirical support (Hollister and Gaut 2009). When two diverged genomes come together, the theory goes, a gene that is near a methylated TE is more likely to be the lower-expressed homeolog; deleterious mutations in the lower-expressed homeolog are more likely to be tolerated; over time, genes from the TE-rich genome acquire deleterious mutations (typically small deletions (Freeling et al. 2015)) and the entire subgenome gradually becomes fractionated. Additionally, preferential retention of genes from one subgenome would have the advantage of preserving the stoichiometry of dosage-sensitive genes, such as those that encode components of protein complexes (Papp et al. 2003; Birchler et al. 2005; Freeling and Thomas 2006; Thomas et al. 2006; Freeling 2009).

Other theories have been proposed. Renny-Byfield et al. (2015) argue that positional effects are not sufficient to explain genome dominance in cotton, and suggest that a combination

of factors related to TE density may promote genome dominance. For example, fractionation might primarily be the consequence of lower recombination rates in TE-rich regions (Renny-Byfield et al. 2015). Steige and Slotte (2016) suggest a role for demography in setting up subgenome dominance. They propose that if one progenitor has more deleterious alleles, for example because it is a selfing species with a small effective population size, the corresponding subgenome may be more prone to fractionation (Steige and Slotte 2016). This theory remains compelling (Douglas et al. 2015; Paape et al. 2018), but has received scant attention compared with the TE theory.

The hypothesis that initial epigenetic differences between the subgenomes determine a polyploid's evolutionary trajectory is difficult to test, because it requires studying the same polyploid on short-term and long-term timescales. In most cases (or arguably all cases), the exact progenitor genotypes of the natural polyploid are unknown. Still, several studies of paleopolyploids have demonstrated a connection between TE load, targeting of TEs by 24-nt siRNAs, and biased expression of homeologous genes (Woodhouse et al. 2014; Cheng et al. 2016; Zhao et al. 2017). A landmark study in favor of the theory was that of Edger et al. (2017). They studied DNA methylation and gene expression in the natural neoallohexaploid *Mimulus peregrinus*, its progenitors *M. guttatus* (a diploid) and *M. luteus* (an allotetraploid), the F1 hybrid and the resynthesized allohexaploid. They found that the *M. luteus*-derived subgenome is consistently higher-expressed, and that this bias increases over the generations. This subgenome is substantially more gene-rich and TE-poor than the *M. guttatus*-derived subgenome, and it also has lower methylation levels near genes. This system is well in-line with the predictions of Freeling et al. (2012). Edger et al. (2017) is not a smoking gun, since it does not prove that the differences in TE content cause the expression bias, nor that expression bias is causing biased fractionation (the natural polyploid is too young to observe this). Still, the best explanation for the expression bias in *M. peregrinus* does seem to be the disparity in TE content and genome features related to that disparity. More data are needed to definitively understand whether disparities in TE load set a polyploid on the path of biased genome evolution.

Mimulus is not the only polyploid system that suggests that genome dominance is established immediately. A number of resynthesized allopolyploids and neoallopolyploids show biased gene expression, biased sequence loss, or both (Alger and Edger 2020). In some cases, the bias is repeatable, that is, multiple independent WGD events result in bias in the same direction (Buggs et al. 2012; Renny-Byfield et al. 2012; Bird et al. 2021; X. Yu et al. 2021). On the other hand, genome dominance is not always present and it is not always clear-cut (Wang et al. 2004; Douglas et al. 2015; Griffiths et al. 2019; VanBuren et al. 2020; Wu et al. 2021). A good example of a complicated case is cotton. In allotetraploid cotton, which formed 1-1.5 MYA, the A subgenome seems to be slightly richer in post-WGD structural rearrangements and degraded genes (Fang et al. 2017; Hu et al. 2021). However, studies examining genome dominance at the level of gene expression have yielded mixed results, with resynthesized lines not recapitulating natural lines, and outcomes depending on the tissue (Yang et al. 2006; Flagel and Wendel 2010; Yoo et al. 2013). Cotton appears to show subtle genome dominance at the structural level and, at best, tissue-dependent dominance at the expression level. The cotton case suggests that polyploids do not necessarily fall into clear-cut categories of genome dominance or no dominance; there may be evidence for dominance at some levels of genome organization but not others.

In summary, in cases where a particular polyploid shows a consistent response to WGD, that response tends to persist over time. That is, polyploids that show immediate genome

dominance generally continue showing genome dominance, and those that are initially unbiased probably remain unbiased. We have noted that some polyploids do not show a consistent WGD response, undermining this trend. Furthermore, species like *Miscanthus sinensis*, which are around the center of the auto/allo spectrum, pose problems as well (Mitros et al. 2020). These are preliminary conclusions drawn from limited data, but they are sufficient to develop a set of hypotheses:

- Autopolyploids should show chromosomal rearrangements (i.e., meiotic instability) in the short-term and they should not show genome dominance.
- Allopolyploids whose subgenomes are substantially different in TE load should not show chromosomal rearrangements, but should show genome dominance.

Based on this reasoning, we may further hypothesize:

- Allopolyploids whose subgenomes are similar in TE load should show neither chromosomal rearrangements nor genome dominance (Fig 0.1).

This third situation teases apart progenitor divergence *per se* from progenitor divergence *in terms of TE load*. If TE load is the decisive factor determining genome dominance, then a polyploid with similarly compact subgenomes should show unbiased evolution, even if it is a very wide cross. *Brachypodium hybridum* is an ideal polyploid with which to test this prediction.

The *Brachypodium hybridum* polyploid complex

The genus *Brachypodium* encompasses 18 grass species native to the circum-Mediterranean region, though today *Brachypodium* is found on all continents except Antarctica (Catalán et al. 2014; Catalán et al. 2016). This genus belongs to the subfamily Pooideae, which includes many economically important grasses such as wheat, oat, barley, and a number of forage grasses. The genus includes a range of grasses that vary in ploidy and life history strategy. *Brachypodium* is a popular model genus for researchers the world over, and the species *Brachypodium distachyon* is a particularly popular model organism (International Brachypodium Initiative 2010; Vogel 2016). *B. distachyon* is a selfing annual plant with a short generation time, small stature (about 20 cm at maturity), non-shattering seeds, and simple growth requirements, making it easier to handle than most, if not all, crop grasses (Draper et al. 2001). The *B. distachyon* genome is one of the smallest high-quality grass genomes sequenced to date (250 Mb) (International Brachypodium Initiative 2010). A robust collection of more than 1800 wild *Brachypodium* accessions has been genotyped (Wilson et al. 2019), and 63 diverse, complete genome sequences are available for *B. distachyon* (Gordon et al. 2017).

B. hybridum is an allotetraploid species formed from the union of the diploid species *B. distachyon* and *B. stacei* (Catalán et al. 2012). All are selfing, and these three species are the only annuals in the genus (Catalán et al. 2016). The two diploids last shared a common ancestor approximately 11.5 MYA (Gordon et al. 2020), making this a wider cross than is typical of model polyploids. Both *B. stacei* and *B. distachyon* have undergone descending dysploidy, primarily via nested chromosomal fusions, since the ancestral grass karyotype of n=12 chromosomes (Lusinska et al. 2018a). *B. stacei* has 2n=2x=20 chromosomes and *B. distachyon* has 2n=2x=10 chromosomes (Catalán et al. 2012), so unlike bread wheat, tobacco, and cotton, *B. hybridum* is formed from progenitors with distinct chromosome number and structure.

It was recently discovered that *B. hybridum* consists of at least two independent lineages with separate origins (Gordon et al. 2020). Plastome sequencing revealed that most lines in our collection had *B. stacei* as the maternal progenitor, while two lines had *B. distachyon* as the maternal progenitor; these were dubbed "S plastotype" and "D plastotype" lines, respectively (Gordon et al. 2020). It was found that the D-plastotype lineage of *B. hybridum* formed 1.4 MYA, while the S-plastotype lineage formed 140 KYA (Gordon et al. 2020).

Phylogenomic and genomic structure analysis supports a scenario in which these two lineages each formed from one WGD event with little to no gene flow between them (Gordon et al. 2020). To further investigate this, I performed reciprocal crosses between the two flagship genotypes of *B. hybridum*: Bhyb26 representing the D-plastotype lineage and ABR113 representing the S-plastotype lineage. I found that F1 offspring from crossing these two genotypes did not produce seed, further corroborating that the two lineages are reproductively isolated (Fig. 0.2).

A resynthesized line of *B. hybridum* was created by Dinh Thi et al. (2017). Through a Herculean effort involving 9,388 crosses between the two diploid species over the course of four years, six F1 amphihaploid interspecific hybrids were produced (Dinh Thi et al. 2016). These were treated with colchicine to induce WGD, and two independent bonafide allotetraploids were produced, one of which was fertile. Karyotype analysis revealed no aneuploidy, and SSR- and gene-derived PCR markers revealed no rearrangements, in the F1 hybrids or in the S1 or S2 generations of the synthetic allotetraploid. In other words, *B. hybridum* is a remarkably stable polyploid formed from a remarkably wide cross.

B. hybridum is well-positioned to shed light on several biological questions. Is the WGD response consistent across lines and across timescales? Is diploidization necessarily biased in allopolyploids? Does genome dominance depend on TE load, genetic divergence, or both? How does genome composition affect chromosome positioning, gene expression, and the pace of evolutionary change? This study will address these questions and more. *B. hybridum*, which consists of two lineages representing two snapshots in evolutionary time, provides insights into the "genomic shock", or lack thereof, imposed by WGD.

Introduction Figures

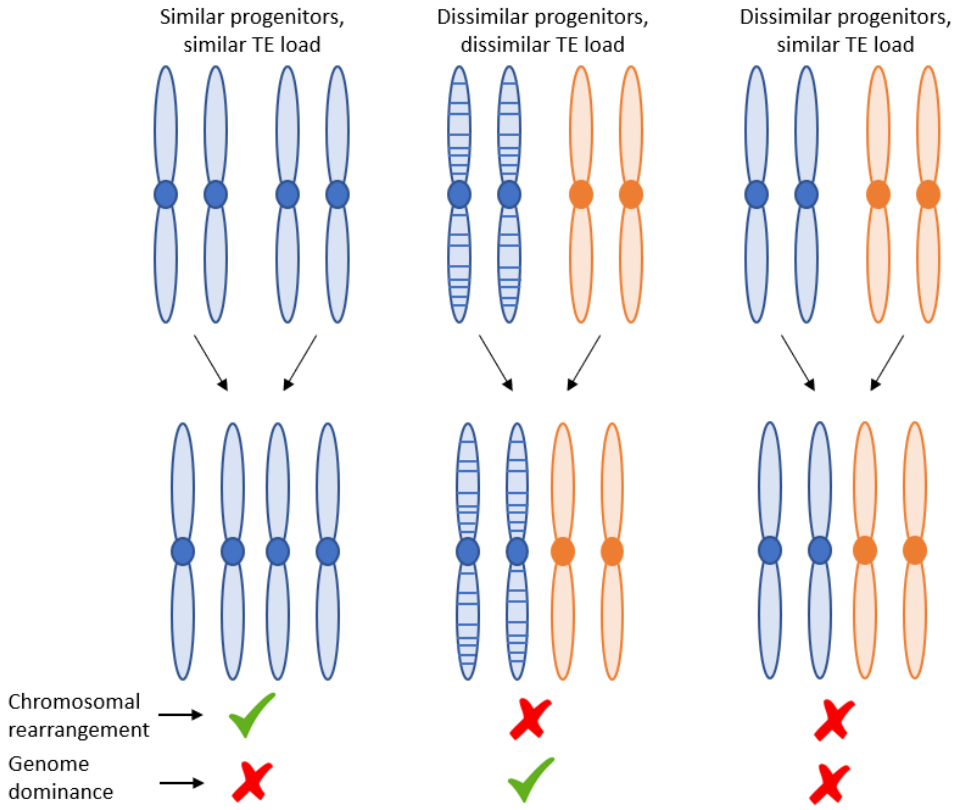

Fig. 0.1. Hypothesis for this study. Progenitor divergence in terms of sequence may predict chromosomal stability, but progenitor divergence in terms of TE load may be the strongest predictor of genome dominance, such that a polyploid with distantly related but similarly compact progenitors may show neither chromosomal rearrangement nor genome dominance. This study tests the third hypothesis.

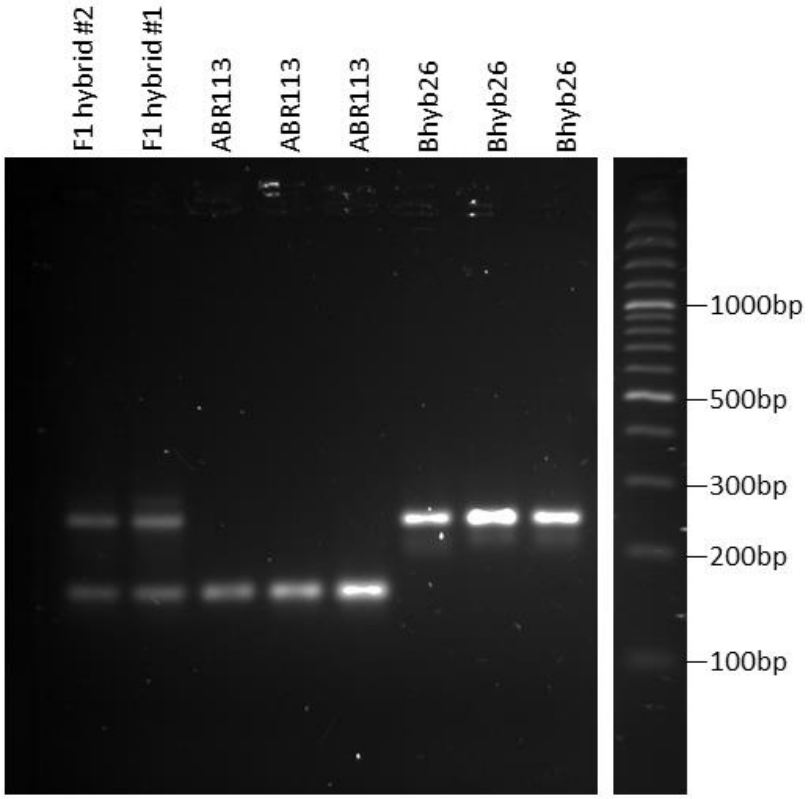

Fig. 0.2. PCR verification of F1 hybrids. A genotyping gel for F1 progeny of *B. hybridum* ABR113 x Bhyb26 crosses. Image shows a representative gel from one of 3 primer pairs from different locations throughout the genome that discriminated between the two parental accessions. With all three primer pairs, the F1 hybrids show both parental band sizes. All F1 hybrids failed to set seed. See Gordon et al. (2020) for more details.

Chapter 1: Parental legacy drives gene expression in the allotetraploid *Brachypodium hybridum*

Abstract

One of the most common responses to polyploidy or whole-genome duplication (WGD) is altered patterns of gene expression. It has been speculated that allopolyploids, which are formed from two or more distinct genomes, might be expected to show biased gene expression, and this bias may be exacerbated over evolutionary time. With this hypothesis in mind, we performed RNA-seq on a natural allotetraploid *Brachypodium hybridum* and its diploid progenitor species. Contrary to expectations, we observed no expression bias at the subgenome or homeolog level. We also observed no convincing evidence of subfunctionalization or pseudogenization. These results suggest that the auto/allo distinction is not sufficient to predict whether a polyploid will show expression bias, and other factors such as TE content may contribute to the post-WGD transcriptional landscape.

Introduction

For plants, gene expression is a fast-acting line of defense against external challenges. Polyploidy represents a kind of internal challenge, in which two or more genomes are suddenly thrust together, and presumably they must coordinate gene expression. For this reason, polyploidy is sometimes called a "transcriptomic shock", a play on McClintock's "genomic shock" (McClintock 1984). However, the precise mechanisms by which polyploid subgenomes adjust gene expression post-WGD are not fully understood.

Expression reprogramming in polyploids may be characterized in terms of two phases: an initial "transcriptomic shock" response phase, and a more gradual but ultimately more impactful adaptation phase in which homeologous genes diverge and develop specialized or novel functions (Flagel et al. 2008; Wang et al. 2016). Studies on numerous resynthesized polyploids have revealed extensive expression changes relative to diploid progenitors immediately following WGD. Studies in *Brassica* (Gaeta et al. 2007; Higgins et al. 2012; Jiang et al. 2013; Zhang et al. 2015; Wu et al. 2018), wheats (Kashkush et al. 2003; Pumphrey et al. 2009; Wang et al. 2016; Jiao et al. 2018; Yuan et al. 2020), and *Arabidopsis* (Madlung et al. 2002; Wang et al. 2006; Shi et al. 2012) provide just a few examples.

The latter "adaptation" phase, in which homeologs either diverge in function or remain redundant and risk being lost, is more difficult to observe. Nevertheless, several examples have been characterized (Liu and Adams 2010; Zhang et al. 2011; Sharma and Kramer 2013). The term for this acquisition of specialized function is subfunctionalization. The rate at which subfunctionalization unfolds, and the likelihood of its occurring at all, depends on the functional and dosage constraints of the gene in question (Cheng et al. 2018).

In some polyploids, a transcriptionally dominant subgenome is clearly distinguishable, as in monkeyflower (Edger et al. 2017) or paleopolyploid maize (Schnable et al. 2011; Woodhouse et al. 2014). In other cases, the absence of dominance is just as clear, as in the paleopolyploids soybean (Zhao et al. 2017), pear (Li et al. 2019), or cucurbits (Sun et al. 2017). In a third set of cases, subgenome dominance is subtle, as in paleoallotetraploid *Miscanthus sinensis* (Mitros et al. 2020), or conditional on tissue, line, and/or developmental stage, as in cotton (Adams et al.

2003; Yoo et al. 2013; Zhang et al. 2015), *Capsella bursa-pastoris* (Kryvokhyzha et al. 2019) or blueberry (Colle et al. 2019). It is not clear whether subgenome dominance sets in during the earlier or later stages of polyploid evolution. Some evidence suggests it may be established immediately and exacerbated over time (Edger et al. 2017).

Strategies for evaluating polyploid gene expression vary according to the technology used and the biological samples available. Thus, there is a plethora of terms for describing polyploid gene expression outcomes (Grover et al. 2012). Grover et al. (2012) worked to standardize this language, outlining two categories of systematic expression bias. The first is homeolog expression bias (HEB). The term HEB is typically applied to a pair of homeologs (or triad for hexaploids, quartet for octaploids, etc.). A particular homeolog set can be said to show HEB if one homeolog is significantly more highly expressed than the other(s) (Smith et al. 2019).

The second category of systematic expression bias outlined by Grover et al. (2012) refers to the total aggregate expression of all homeologs having preferential fidelity to one parent's expression levels. This is to some extent a holdover from microarray technologies, in which individual homeologs were difficult or impossible to distinguish. Newer RNA-seq technologies that allow for easier discernment of homeologous transcripts have instigated a paradigm shift away from aggregating the homeologs and instead evaluating each subgenome's fidelity to parental expression levels separately. We will call this phenomenon—an individual gene's fidelity to parental expression levels—parental legacy (PL) (Buggs et al. 2014).

Here, we investigated gene expression between a natural allotetraploid and its diploid progenitors. *B. hybridum* ABR113 formed approximately 140,000 years ago from a wide cross between diploid species *B. distachyon* and *B. stacei* (Catalán et al. 2012; Gordon et al. 2020). We assessed both HEB and PL to ascertain whether *B. hybridum* ABR113 shows signs of either a transcriptomic shock response or subfunctionalization. Technically, we handled the cross-species RNA-seq like an alternative splicing experiment in which the expected primary transcript length in each sample is known *a priori*, and in this way, we controlled for differences in gene length between diploid and polyploid. Examining two tissues, we found no evidence of transcriptional reprogramming or systematic HEB favoring one subgenome. Genes from both subgenomes adhere strongly to parental expression patterns. This study shows that polyploidy is not necessarily accompanied by massive or systematic gene expression changes, suggesting that something about this polyploid may have set it upon a path of gradual, unbiased genome evolution.

Methods

Sample collection and library preparation

RNA-seq samples were obtained for *B. hybridum* ABR113, *B. stacei* ABR114, and *B. distachyon* Bd21 leaves and spikelets. Leaf samples were from seedlings at the 3–4 leaf stage. Each spikelet sample consisted of the spikelets from one plant, with each spikelet collected separately 3 days after inflorescence emergence. RNA was extracted using TRIzol (Life Technologies) or PureLink (ThermoFisher) kits, DNAse-treated using the DNA-free™ Kit (Life Technologies), and RNA quality was assessed by Nanodrop (ThermoFisher), agarose gel, and BioAnalyzer (Agilent). Strand-specific libraries were prepared using the Illumina TruSeq kit, and library quality was checked by BioAnalyzer (Agilent). Libraries were sequenced using

Illumina technology. The average number of total mapped paired-end reads ranged from ~60 to ~100 million reads; average coverage ranged from ~18-22; average depth per base ranged from ~32-45.

RNA-seq data analysis

Raw RNA-seq reads were filtered and trimmed using BBDuk (v37) from the BBtools package (v. 38.0). Reads were aligned to the complete reference genome (ABR113 v. 1.0) using BBmap (v37) (https://sourceforge.net/projects/bbmap). For subgenome dominance analyses, ABR113 reads were aligned to the complete reference genome (ABR113 v. 1.0). To increase mapping stringency, reads were required to share 90% sequence identity with the target location, and ambiguous reads were discarded. Gene-level counts were obtained using HTSeq (v. 0.9.1) (Anders et al. 2015). Transcripts per million (TPM) values were calculated using a custom Python script (https://github.com/vtartaglio/scripts/blob/master/TPMs/countsToTPMbasicNEW.py).

For cross-species differential expression and homeolog expression bias (HEB) analyses, the same mapping criteria were applied but using a slightly modified pipeline. Using BBsplit (v. 37), reads from each library were simultaneously aligned to the reference nuclear or plastid genomes (Bd21 v. 3.1, ABR114 v. 1.0, or ABR113 v. 1.0) in fastq output mode. This read-sorting procedure yields separate fastq files for each of the polyploid subgenomes. Alignments to the appropriate nuclear genome or subgenome were then performed on the fastq files using BBmap. Counts were then obtained for each of the separate subgenomes' bam files.

DESeq2 (Love et al. 2014) was used to calculate the log fold change in expression between a polyploid gene and its diploid orthologs, or between homeologs. The same approach was used in both cases. Homeologs were determined using Phytozome's Phytomine pipeline, incorporating both homology and synteny (Goodstein et al. 2012). The length of each gene was recorded in a gene lengths matrix, which was incorporated into normalization factor estimation, which is in turn used to estimate each gene's true expression value. This option was designed mainly to account for differences in alternative transcript usage between conditions, but can be used to control for any known systematic bias for a given gene between two samples. We then tested for significant differential expression or HEB as one would test for differential expression between any two treatments, except our "treatments" were the two (sub)genomes. Our models for the likelihood ratio test (using Wilkinson notation) were: full=~library+genome, and reduced=~library, so library was essentially handled as a covariate, and genome of origin was the variable of interest (i.e. progenitor or polyploid). Functional enrichment analysis of differentially expressed genes was performed using GOATOOLS (Klopfenstein et al. 2018), and the significant categories were visually displayed with REVIGO (Supek et al. 2011).

Our candidate pseudogene identification protocol was based on the method of Session et al. (2016). *B. hybridum* subgenomes were compared with their corresponding diploid genomes three genes at a time to identify triplets where the central gene lacked a *B. hybridum* ortholog but the two flanking genes each had a syntenic *B. hybridum* ortholog. The *B. hybridum* genomic region between the two flanking orthologs was extracted using bedtools (v2.27.1) (Quinlan 2014), and the diploid gene was aligned to this sequence using the codon- and intron-aware protein2genome model of exonerate (v. 2.4.0) (Slater and Birney 2005). Alignments were parsed with the Biopython (v. 1.7.0) package ExonerateIO. RNA-seq counts were obtained strictly for the aligned regions using the customizable functionalities of HTSeq, and normalized by the total number of mapped reads for each library. Pairwise nonsynonymous to synonymous substitution

rate ratios were calculated using the yn00 program from PAML (v. 4.9 h) (Yang 2007). Final candidate pseudogenes were required to have a nonsynonymous to synonymous substitution rate ratio (dN/dS) of 0.5 or higher, and the log ratio of diploid to polyploid expression for the aligned region had to be >0.3. Genes were manually inspected for proximity to a transposon using Jbrowse on the CoGe platform (https://www.genomevolution.org/coge/).

Results

mRNA-seq was performed on leaf and spike from the reference accessions of *B. hybridum* (Bh), *B. distachyon* (Bd), and *B. stacei* (Bs). A concern was that the subgenomes might be too similar for reliable read mapping. To obtain a sense of the similarity of the subgenomes, the coding sequences of 10,000 randomly sampled homeologous gene pairs from the reference genome were aligned using BLAST+ (v. 2.6.0) (Camacho et al. 2009), and the number of mismatches was tallied. On average, there were 4.4 nucleotide mismatches per 100 aligned nucleotides. Our RNA-seq reads were between approximately 100 and 150 bp in length, suggesting that our reads would capture this divergence. Indeed, when we required that only uniquely mapped reads be used for downstream analysis, more than 96% of reads were retained in all RNA-seq samples.

We began by comparing the overall expression levels of the two polyploid subgenomes to each other. The distribution of gene expression values was very similar in each subgenome (Fig. 1.1A). A paired t-test and Bartlett test of log-transformed transcripts per million (TPM) values from homeologous genes indicated that the means and variances of the D- and S-subgenome gene expression values were equivalent ($p=0.40$ and $p=0.82$, respectively). We also compared the TPM values from individual homeolog pairs and recorded which homeolog was more highly expressed (Fig. 1.1B). The binomial test has been used on polyploid expression data under the moniker “horse race” to test whether genes from one subgenome are more highly expressed than their homeologs with a higher-than-expected frequency (Woodhouse et al. 2014). Our horse race results indicate that the chances of the D or S homeolog being the more highly expressed one do not significantly deviate from 50/50 ($p=0.60$). Finally, we also compared the total raw counts from each subgenome. Our null expectation for the contribution of raw counts was calculated from the number of genes from each subgenome: there are 37,711 BhD genes and 32,449 BhS genes in *B. hybridum* ABR113, so we expected roughly 46.25% of counts to originate from the BhS subgenome. In all samples, almost exactly 50% of counts came from each subgenome (Fig. 1.1C). This was somewhat surprising, since it implies that the two subgenomes contribute equally to the transcriptome, even though they are not exactly the same size.

To examine HEB formally, we used a likelihood ratio test implemented in DESeq2, explicitly accounting for gene length and variation between libraries (see Methods). Remarkably, in leaves, 2,476 gene pairs showed higher expression of the D homeolog, while 2,473 showed higher expression of the S homeolog. Spike showed a similarly balanced profile: 3,944 homeolog pairs favored D, while 3,977 favored S. 13,184 genes were unbiased in leaf and 10,698 genes were unbiased in spike. Nearly half of all homeologous gene pairs were not biased in either tissue, and about two-thirds of gene pairs that showed any bias showed it only in one tissue (Fig. 1.1D). These results confirm the absence of any systematic preferential expression of genes from one subgenome.

Next, we compared each *B. hybridum* subgenome to its corresponding diploid progenitor to uncover expression changes that may have occurred post-polyploidization. Using the default significance threshold in DESeq2 (Love et al. 2014), a cross-species analysis of syntenic orthologous genes showed that 36% and 25% of testable BhD and BhS genes, respectively, were differentially expressed relative to their diploid ortholog (Fig. 1.2A). The number of up- vs. down-regulated genes was fairly similar in each case: 4,464 up and 4,229 down in Bd vs. BhD, and 2,700 up and 2,904 down in Bs vs. BhS. The higher number of differentially expressed genes (DEGs) between Bd and BhD than between Bs and BhS was in accord with our expectations, since the BhS subgenome is more closely related to the Bs reference line (ABR114) than the BhD subgenome is to the Bd reference line (Bd21) (Gordon et al. 2020). These results suggest that, while a substantial portion of genes are differentially expressed relative to the corresponding gene in the progenitor, the majority conform to parental expression levels.

We performed functional enrichment analysis on the genes that were DE between polyploid and progenitor. More functional categories were significant between Bd and BhD than between Bs and BhS, consistent with the higher number of DEGs. GO categories that were enriched among BhS DEGs were also enriched among BhD DEGs, with the exception of one category that was enriched in Bs-BhS but not Bd-BhD ("cellular aromatic compound metabolic process"). Overall, the GO categories point to a possible trend of nucleic acid-related processes (Fig. 1.2B), which is interesting since we might expect that a key difference between a polyploid and its diploid progenitor would be demands related to maintaining more DNA.

To further assess the extent of PL and possible subfunctionalization, we compared parental and polyploid genes in terms of their expression domain, that is, the combination of tissues in which they are expressed. Genes were considered expressed or non-expressed if they had an average TPM among biological replicates of > 1.0 or < 0.1, respectively. Out of approximately 20,000 ortholog pairs for each subgenome-progenitor comparison, or about 40,000 pairs total, fewer than 1,000 pairs showed that the Bh copy expressed in a different set of tissues than the parental ortholog (Fig.1.3A). If we consider expression domain to be a rudimentary proxy for gene function, these data provide little evidence of subfunctionalization in *B. hybridum*.

Finally, the *B. hybridum* genome was scanned for pseudogenes--non-functional relicts of genes--which are expected to accumulate in polyploids due to the relaxed natural selection that is believed to accompany genome duplication (Schnable et al. 2011). We confined our search to regions of the *B. hybridum* genome where no gene was annotated, but where a gene was expected to exist based on synteny with the diploid (Session et al. 2016). By aligning the diploid coding sequence to the *B. hybridum* candidate region, we recovered unannotated genes and gene fragments. Putative pseudogenes were filtered based on non-synonymous to synonymous substitution rate ratio (dN/dS) and gene expression (see Methods). Even with our relatively generous criteria, only 27 candidate pseudogenes were identified. 4 of these were actually transposons. Surprisingly, 21 of the remaining 23 genes were from the D subgenome. None of the putative pseudogenes appeared to be near a novel transposon insertion, a proposed mechanism of post-polyploidy pseudogenization (Wendel et al. 2018). Only 4 of the 23 candidates were completely silenced in the polyploid, all of them "shell" genes in the *Brachypodium* pan-genome (Gordon et al. 2017) and all lacking any functional annotation. The distribution of pan-genome categories among the 21 BhD candidate pseudogenes matched the distribution of pan-genome categories in the ABR113 genome overall; no category was disproportionately represented (exact multinomial test, p=0.3484). Overall, it is difficult to say

whether these are true pseudogenes or simply a random sampling of low-quality genes from the genome. The overrepresentation of the D subgenome may simply be due to dubious orthology calls, given the greater distance between BhD and Bd than BhS-Bs. The small number of candidate pseudogenes indicates that pseudogenization is not widespread in *B. hybridum*, if there is any pseudogenization at all.

Discussion

Our investigation of HEB and PL in the allotetraploid *B. hybridum* reveals a case of remarkable genome stability following WGD. Each subgenome contributes equally to the entire polyploid transcriptome, at both the whole-genome and homeolog levels. When each subgenome was compared to its polyploid, there were some DEGs, though some of these were DE due to intraspecific variation rather than the effects of polyploidy *per se*. Interestingly, the predicted functions of the DEGs suggest roles in nucleic acid metabolism, which might signal adjustments to the higher DNA content in the polyploid. Very few genes have altered expression in the polyploid in terms of the tissue in which they are expressed, which does not support the possibility of substantial subfunctionalization. Finally, we found little evidence of pseudogenization, suggesting that diploidization has barely begun in *B. hybridum* ABR113.

We explored each subgenome's contribution to the transcriptome via several metrics. We compared the distribution of TPM values and the number of raw counts from each subgenome. Interestingly, in both cases the two transcriptomes seemed to comprise almost exactly 50% of the total, even though the two subgenomes are not exactly the same size. ABR113D is 269Mb, while ABR113S is 240Mb. It would be worthwhile to repeat this experiment with the natural polyploid and the resynthesized polyploid (Dinh Thi et al. 2016) to see whether the contribution of each subgenome is 1:1 immediately following WGD. It is possible that the S subgenome is in fact slightly dominant, in the sense that it is more highly expressed than we would expect given its size. This is not unprecedented: Coate and Doyle (2010) found that the transcriptome size of *Glycine dolichocarpa* may have increased disproportionately (2.4x) with WGD, though they could not discriminate each subgenome's contribution (Coate and Doyle 2010).

We developed a novel, straightforward approach to differential expression analysis for assessment of HEB and PL. The question of cross-species expression analysis is not a straightforward one, and it has led to confusion and to the development of many analytical approaches (Higgins et al. 2012; Buggs et al. 2014). Our DESeq2-based (Love et al. 2014) method for HEB is similar to the method of Smith et al. (2019). Both methods model gene-level counts on a negative binomial distribution, accounting for library size and the global mean-variance relationship, and then test for differential expression between the homeologs via a likelihood ratio test followed by Benjamini-Hochberg correction. The main differences between our method and that of Smith et al. (2019) are the use of a Bayesian shrinkage estimation approach rather than maximum likelihood, and the way gene length was subsequently incorporated into those gene expression estimates. Our approach carries all the uncertainty that comes with any attempt at cross-species RNA seq. Notably, it disregards genes that do not have a homeolog, and we did not investigate the effect of this decision on the accuracy of estimation of the mean-variance relationship. Nevertheless, our method is very accessible, in contrast to that of Smith et al. (2019), and it delivered results that were consistent with our expectations.

Each *B. hybridum* subgenome carried a strong parental legacy. Most testable genes in each subgenome were not differentially expressed relative to their progenitors, and nearly all

were expressed in the same set of tissues. We consistently found more DEGs between BhD and Bd than between BhS and Bs. While this may reflect greater transcriptional reprogramming in one subgenome than the other, we think it is probably due to the fact that the Bd accession used in this experiment is not as closely related to its corresponding subgenome as the Bs accession is (Gordon et al. 2020). This experiment could have been improved by using Bd1-1 as the Bd reference accession, especially now that a high-quality reference genome for this accession is available (https://phytozome-next.jgi.doe.gov/). However, even with diploid reference lines that are closely related to the true diploid progenitors of ABR113, there would still be a small fraction of gene expression differences due to normal intraspecific variation rather than polyploidy. This experiment could have been further improved by inclusion of many accessions of both *B. distachyon* and *B. stacei*. With those data, the relationship between phylogenetic distance and number of DEGs could potentially be modeled. Such an investigation would give us a more precise answer as to whether the number of DEGs in each polyploid subgenome conforms to the expected variation between diploids of the same species.

B. hybridum is an allopolyploid without much transcriptional reprogramming post-WGD. Like hexaploid bread wheat, we did find a number of differences between subgenomes, and between each subgenome and its progenitor, but we did not find evidence that one subgenome is systematically dominant (Ramírez-González et al. 2018). Rather, as in the wheat case, the clearest bias emerged when comparing subgenomes against progenitors, highlighting the importance of accounting for phylogenetic distance from progenitor to subgenome when studying gene expression in natural polyploids. Unlike many paleopolyploids, in which millions of years of evolution have led to degradation of one or more subgenomes, *B. hybridum* shows little evidence of pseudogenization at all, much less biased retention or utilization of one subgenome. It is more like the case of *Ephedra*, another allopolyploid with substantially diverged progenitors, in which transcriptome evolution was described as “even and slow” (Wu et al. 2021). It may be the case that the progenitors of *B. hybridum* ABR113 are highly epigenetically compatible (Adams et al. 2003; Edger et al. 2017), so that it did not receive a “shock” at the time of WGD, which may have set it on a course of gradual diploidization.

Chapter 1 Figures

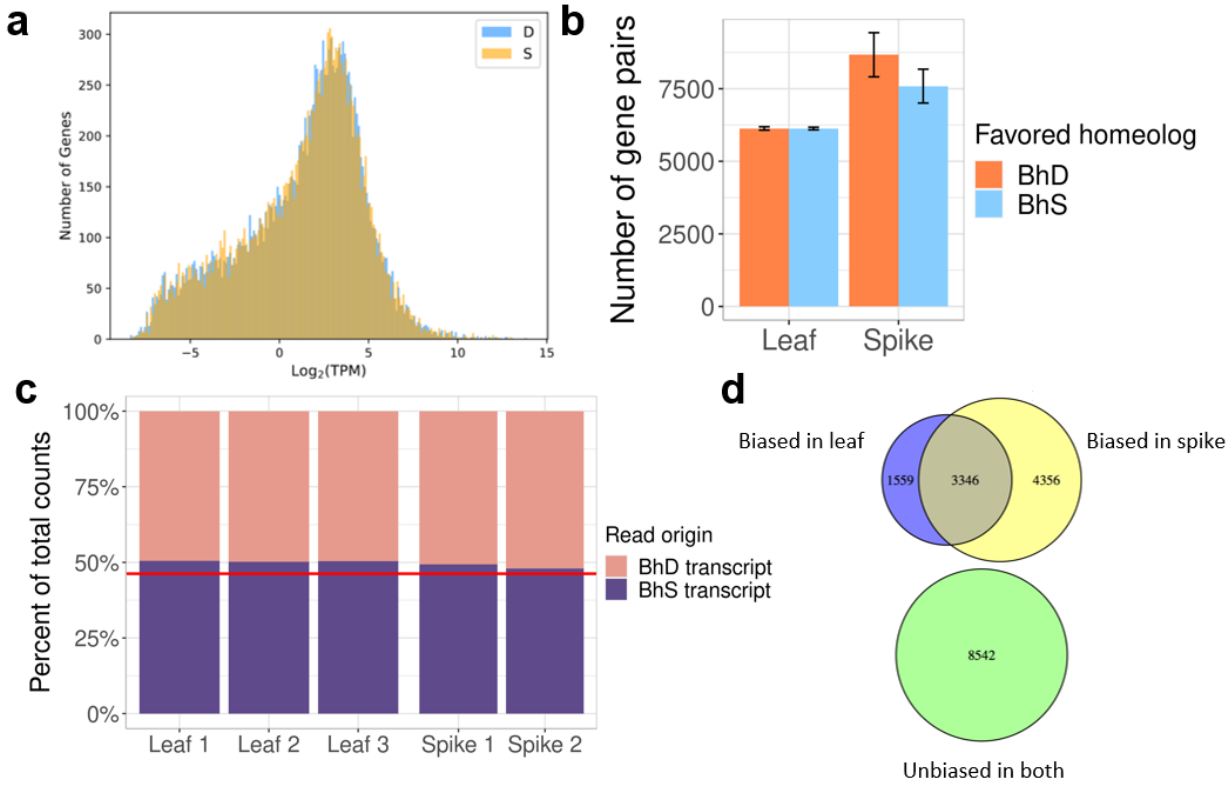

Fig. 1.1. Absence of genome expression dominance. (a) Normalized expression values for homeologous genes of the B. hybridum ABR113 subgenomes D (blue) and S (green) obtained by mapping leaf RNA-seq reads to the ABR113 reference genome. Non-expressed genes were included in statistical tests but removed from graph for visual clarity. (b) Grouped bar chart showing the more highly-expressed homeolog for homeologous gene pairs. Only gene pairs where at least one homeolog had a TPM > 1.0 were considered. (c) Stacked bar chart showing percent of RNA-seq reads mapped to each subgenome. Red line indicates what percent of genes originate from the S subgenome. (d) Results of formal HEB analysis by tissue, for all BhD-BhS homeolog pairs.

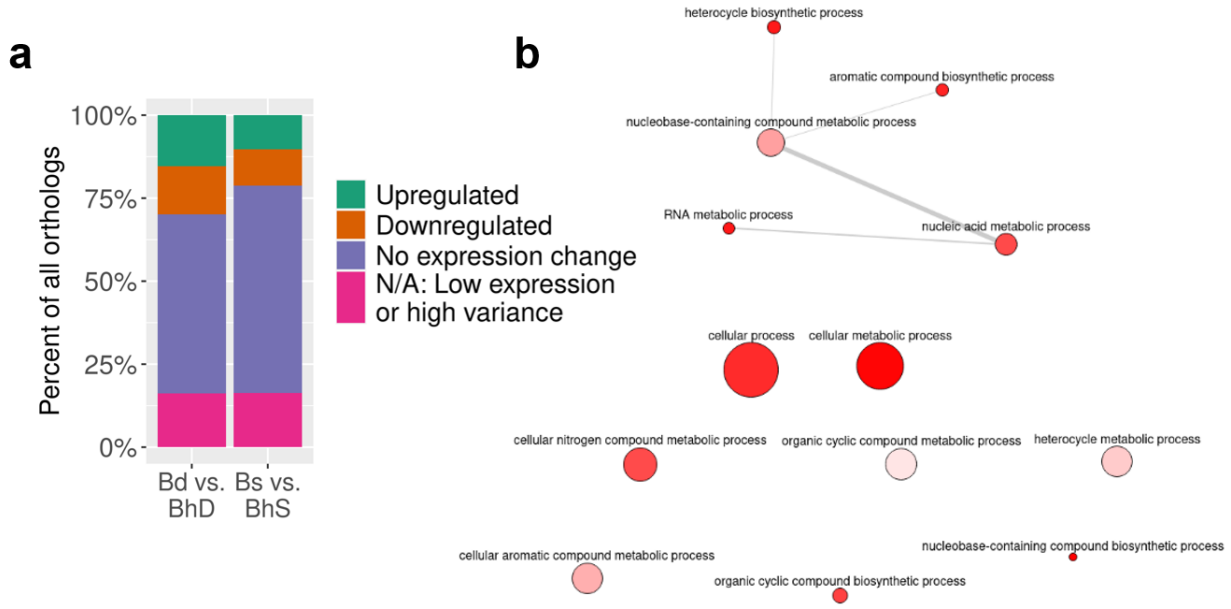

Fig 1.2. Cross species RNA-seq reveals parental legacy. (a) DE analysis from RNA-seq of orthologs between each B. hybridum subgenome and its diploid progenitor (leaf samples). (b) Network of significant GO categories from differentially expressed genes between polyploid and progenitor (leaf samples). For clarity, only BhS-Bs GO terms are shown.

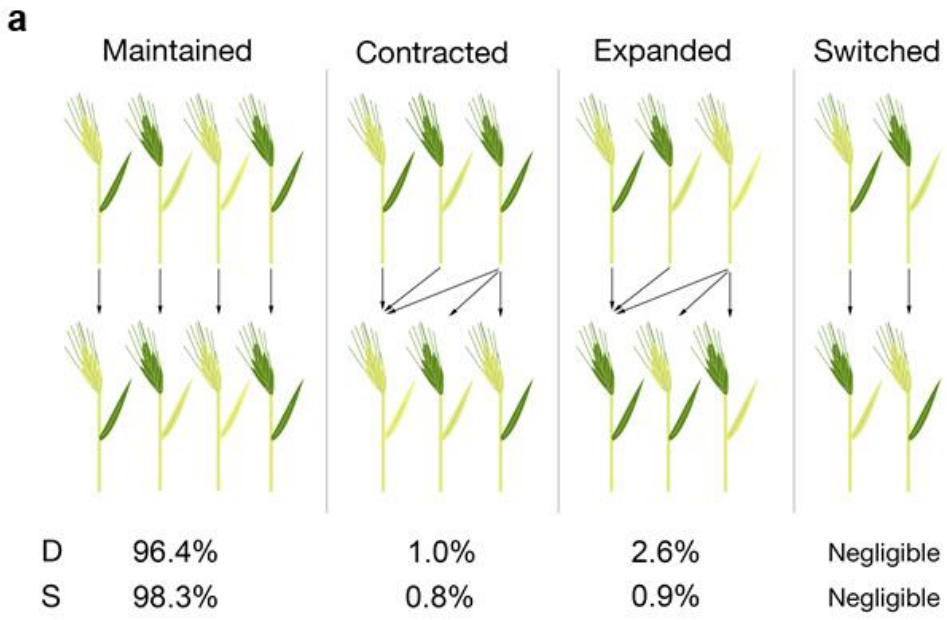

Fig 1.3. Tissue expression pattern differences between diploid and polyploid orthologs. (a) TPM cut-offs were used to determine whether a gene was expressed (TPM >1) or not expressed (TPM < 0.1), for both leaf and spike samples. Expression pattern was recorded for progenitor gene (top row) and its polyploid ortholog in corresponding subgenome (bottom row). Expression pattern changes were grouped into categories. Each plant in the top row corresponds to a tissue expression pattern, and arrows denote possible expression patterns in the polyploid that put that ortholog pair in a particular category. For example, 96.4% of valid BhD genes "maintained" the parental expression configuration.

Chapter 2: Gradual evolution characterizes distinct lineages of allotetraploid *Brachypodium*

Abstract

The 'genomic shock' hypothesis posits that unusual challenges to genome integrity such as whole genome duplication (WGD) may induce chaotic genome restructuring. Decades of research on polyploid genomes have revealed that this is often, but not always the case. While some polyploids show major chromosomal rearrangements and de-repression of transposable elements (TEs) in the immediate aftermath of WGD, others do not, though all polyploids show gradual diploidization over evolutionary time. We produced a high-quality reference genome for the natural allotetraploid grass *Brachypodium hybridum*, accession Bhyb26. This new reference genome allowed us to conduct a detailed comparison between two independent accessions of *B. hybridum* and their deeply diverged diploid progenitors *Brachypodium distachyon* and *Brachypodium stacei*. The two *B. hybridum* lineages provide a natural timecourse in genome evolution because one formed 1.4 million years ago, and the other formed 140 thousand years ago. The newly sequenced genome of the older lineage reveals signs of gradual post-WGD genome evolution including minor gene loss and genome rearrangement that are missing from the younger lineage. In neither *B. hybridum* lineage do we find signs of homeologous recombination or pronounced TE activation, though we find evidence supporting steady post-WGD TE activity in the older lineage. Gene loss in the older lineage was slightly biased toward one subgenome, but genome dominance was not observed at the transcriptomic level. We propose that relaxed selection, rather than an abrupt genomic shock response, drives evolutionary novelty in *B. hybridum*, and that the progenitors' similarity in TE load may account for the subtlety of the observed genome dominance.

Introduction

Nearly all plant lineages have at least one polyploidy event, or whole genome duplication (WGD), in their recent or ancient past (Clark and Donoghue 2018). Today's diploids have undergone a process known as genetic diploidization, in which a polyploid loses genomic sequence over evolutionary time until it becomes diploid again, though some duplicate genes are retained. Polyploidy is an important source of genetic novelty and contributes to adaptive evolution (Van de Peer et al. 2017; Baduel et al. 2018; Van de Peer et al. 2021).

In many cases, WGD is accompanied by rapid genome restructuring, in line with the hypothesis that WGD may represent a kind of 'genomic shock' (McClintock 1984). The most dramatic example of this is chromosomal rearrangement, mostly resulting from recombination between homeologous chromosomes, which may occur in the early generations post-WGD (Ramsey and Schemske 2002). Such homeologous rearrangements have been observed in allopolyploids (those whose progenitors are different species) and autopolyploids (those whose progenitors are from the same species) (Grandont et al. 2013). Homeologous rearrangements are common in resynthesized polyploids (Mason and Wendel 2020), and evidence for them has been

observed in a number of natural polyploids including *Brassica napus* (Chalhoub et al. 2014; Hurgobin et al. 2018), cotton (Guo et al. 2014) domesticated strawberry (Edger et al. 2019), quinoa (Jarvis et al. 2017), peanut (Bertioli et al. 2019), *Perilla frutescens* (Zhang et al. 2021), and the neoallopolyploid *Tragopogon miscellus* (Chester et al. 2012).

Some polyploids exhibit a dominant subgenome, whose genes are expressed at higher levels than their homeolog(s) on the other subgenome(s) (Alger and Edger 2020). It remains unclear to what extent genome dominance is established instantaneously or gradually. The evidence suggests both: expression bias established in the early generations following WGD may be reinforced over evolutionary time, with the dominant subgenome ultimately contributing more genes to the fully diploidized genome (Flagel et al. 2008; Feldman and Levy 2009; Flagel and Wendel 2010; Woodhouse et al. 2014; Edger et al. 2017).

Transposable element (TE) activation (transcription and/or transposition) can also occur following WGD on short and long timescales. Post-WGD epigenetic changes are not uncommon in polyploids (Ha et al. 2009; Parisod et al. 2009; Kenan-Eichler et al. 2011; Yaakov and Kashkush 2012; Yuan et al. 2020; Jiang et al. 2021). In allopolyploids, a single TE family or several families may be activated immediately following WGD, probably due to epigenetic incompatibilities between subgenomes (Madlung et al. 2005; Parisod et al. 2009; Martienssen 2010; Groszmann et al. 2011; Yaakov and Kashkush 2012; Sarilar et al. 2013; Gantuz et al. 2021). TE proliferation can also occur in polyploids over longer timescales due to relaxed selection because duplicate genes allow for a greater tolerance for TE insertions (Ågren et al. 2016; Baduel et al. 2019)

While some polyploids show chromosome rearrangements, expression dominance, and TE activation following WGD, these responses are not universal. Many natural allopolyploids show little to no genome restructuring, including *Arabidopsis suecica* (Burns et al. 2021), *Eragrostis teff* (VanBuren et al. 2020), *Capsella bursa-pastoris* (Douglas et al. 2015), white clover (Griffiths et al. 2019), and the octoploid progenitors of domesticated strawberry (Hardigan et al. 2020). Thus, while WGD is often regarded as a profound genomic shock, a number of species seem to contradict this paradigm. The plant response to WGD is controlled by several complex factors including meiosis-related genes (Grandont et al. 2013), progenitor divergence (Ramsey and Schemske 2002), TE abundance or TE load (Woodhouse et al. 2014; Wendel et al. 2018), and demographic factors (Steige and Slotte 2016). Given the complexity of the plant response to WGD, simple model organisms are needed to reveal how genome characteristics may predispose a polyploid to a particular evolutionary trajectory.

B. hybridum (2n=30) is an annual allotetraploid grass that is native to the Mediterranean region but has spread all over the world, surpassing the range of either of its diploid progenitors, *B. stacei* (2n=20), or the well-known model grass *B. distachyon* (2n=10) (Catalán et al. 2012; Catalán et al. 2016). We know that *B. hybridum* has multiple origins because some lineages have chloroplasts that resemble the chloroplasts in *B. distachyon* (D-plastotype accessions), and other lines have chloroplasts that resemble the chloroplasts of *B. stacei* (S-plastotype accessions) (Gordon et al. 2020). Since chloroplasts are only inherited from the maternal parent, *B. hybridum* must have arisen from more than one cross. This was confirmed by phylogenetics (Gordon et al. 2020). In a previous study (Gordon et al. 2020), we designated the accession Bhyb26 as the model D-plastotype lineage, and ABR113 as the model S-plastotype lineage. Crosses between these two *B. hybridum* accessions resulted in sterile offspring, consistent with the lack of genetic evidence for hybridization between them (Gordon et al. 2020). The compact, naturally inbred genomes of these two polyploid lineages, their reproductive isolation, and the relative simplicity

of the WGD make this system a valuable model for detailed study of polyploid genome evolution.

We previously demonstrated that *B. hybridum* ABR113 shows no sign of genome rearrangement nor of substantial gene loss (Gordon et al. 2020). A resynthesized *B. hybridum* line also bore no evidence of genomic rearrangements, based on a panel of SSR- and gene-derived PCR markers (Dinh Thi et al. 2016). This contrasts with some polyploid plants, such as *B. napus* (Szadkowski et al. 2010), tobacco (Lim et al. 2006), cucumber (X. Yu et al. 2021), and certain wheats (Mirzaghaderi and Mason 2017) in which the first generation following WGD is genetically unstable, and meiosis may (Tian et al. 2010) or may not (Gou et al. 2018) stabilize over the first few generations. All *B. hybridum* lines examined so far show no sign of aneuploidy, homeologous exchange, nor chromosomal rearrangement (Dinh Thi et al. 2016; Gordon et al. 2020).

B. hybridum ABR113 formed roughly 140,000 years ago, making it a relatively "young" polyploid, so it was difficult to draw conclusions about its diploidization. *B. hybridum* Bhyb26, on the other hand, formed 1.4 million years ago, meaning that this lineage has had substantially more time for evolution toward diploidization (Gordon et al. 2020). Here, we present a high-quality PacBio-based reference genome for *B. hybridum* Bhyb26, and we perform an in-depth survey of its structure and TE landscape. The Bhyb26 genome, like the other *B. hybridum* genomes, reveals no convincing evidence of homeologous rearrangement. However, we did find evidence that Bhyb26, unlike the younger lineage, has experienced post-WGD structural change and slightly but significantly biased gene loss. Remnants of these 'lost' genes show signs of pseudogenization. We did not find evidence of increased TE proliferation, nor did we observe increased TE insertion in or near genes, a mechanism by which TEs have been proposed to drive diploidization (Wendel et al. 2018). Therefore, TEs do not seem to be contributing to the observed gene loss. Our study demonstrates that polyploids with multiple origins can be effectively used to study polyploid evolution, serving—with some caveats—as natural replicates of the diploidization experiment.

Methods

Sample preparation

For details on the lines used in this study and preparation of high-molecular weight DNA for Pac-Bio sequencing, see Gordon et al. (2020). PacBio sequencing was performed on a PacBio RSII instrument at the HudsonAlpha Institute.

RNA-Seq

To collect leaf tissue, plants were grown in a growth chamber in short-day conditions (26C 10h light, 18C 14h dark). Leaf tissues were harvested from plants at the 4-5 leaf stage. To collect spikelets, plants were grown in long-day conditions (26C 16h light, 18C 8h dark). Spikelets were harvested three days after inflorescence emergence. For root tissue, plants were grown in plastic sundae cups with lids on sterile MS medium, and roots were harvested at 1-2 weeks. Callus tissue was prepared as described (Bragg et al. 2015). RNA was extracted using TRIzol (Invitrogen) and purified with the Purelink RNA Mini Kit (Invitrogen) including DNA removal with the Purelink DNAse Set (Invitrogen). Stranded RNASeq libraries were created

using the Illumina TruSeq kit and quantified by qPCR. Sequencing was performed using an Illumina NovaSeq S4 instrument yielding 50-175 million 2x151bp reads.

Raw RNA-seq reads were filtered and trimmed using BBDuk from the BBtools package (https://sourceforge.net/projects/bbmap). Reads were aligned to the complete reference genome (Bhyb26 v2.1) using BBmap. To increase mapping stringency, given the redundancy of a polyploid genome, reads were required to share 90% sequence identity with the target location, and ambiguous reads were discarded. Gene-level counts were obtained using HTSeq (Anders et al. 2015). Transcripts per million (TPM) values were calculated using a custom Python script (https://github.com/vtartaglio/scripts/tree/master/TPMs/countsToTPMbasicNEW.py).

Assembly

Assembly of the Bhyb26 genome was performed with MECAT (Xiao et al. 2017) and polished using ARROW (https://github.com/PacificBiosciences/GenomicConsensus).

Misjoins in the assembly were identified using HiC data as part of the JUICER pipeline (Durand, Shamim, et al. 2016). No misjoins were identified in the polished assembly. Scaffolds were then oriented, ordered, and joined together using HiC scaffolding. Significant telomeric sequence was properly oriented in the assembly. HiC reads were then aligned to the joined release. A contact map was generated using JUICER-pre and visualized using JUICEBOX (Durand, Robinson, et al. 2016) as a quality control check on the order/orientation of contigs in the chromosomes. Care was taken to ensure that telomeres were properly oriented in the chromosomes, and the resulting sequence was screened for retained adapter/vector and/or contaminants.

Adjacent alternative haplotypes were identified on the joined contig set. Althap regions were collapsed using the longest common substring between the two haplotypes. A total of 22 adjacent alternative haplotypes were collapsed. Chromosomes were numbered and oriented relative to the previous V1 release. Finally, homozygous SNPs and INDELs were corrected in the release sequence using 40x of Illumina reads.

Annotation

Transcript assemblies were made from Illumina RNA-seq reads using PERTRAN, which conducts genome-guided transcriptome short read assembly via GSNAP (Wu and Nacu 2010) and builds splice alignment graphs after alignment validation, realignment and correction. PacBio Iso-Seq CCSs were corrected and collapsed by a genome guided correction pipeline, which aligns CCS reads to the genome with GMAP (Wu and Nacu 2010) and clusters alignments when all introns are the same or 95% overlap for single exon. Subsequently 625,901 transcript assemblies were constructed using PASA (Haas et al. 2003) from the RNA-seq transcript assemblies. Loci were determined by transcript assembly alignments and/or EXONERATE (Slater and Birney 2005) alignments of proteins from diverse plant species and Swiss-Prot proteomes to the repeat-soft-masked *Brachypodium hybridum* Bhyb26 genome using RepeatMasker (Smit et al. 2013 2015). Gene models were predicted by homology-based predictors, FGENESH+ (Salamov and Solovyev 2000), FGENESH_EST, and EXONERATE, PASA assembly ORFs (in-house homology constrained ORF finder) and from AUGUSTUS (Stanke et al. 2006) trained by the high confidence PASA assembly ORFs and with intron hints from short read alignments. The best scored predictions for each locus are selected using multiple positive factors including EST and protein support, and one negative factor: overlap

with repeats. The selected gene predictions were improved by PASA. Improvement includes adding UTRs, splicing correction, and adding alternative transcripts. PASA-improved transcripts were selected based on Cscore, protein coverage, EST coverage, and their CDS overlapping with repeats. Weak gene models, incomplete gene models, gene models whose protein is more than 30% in Pfam TE domains, low homology supported without fully transcriptome supported gene models, and gene models consisting of a short single exon without protein domain nor good expression gene models were manually filtered out.

BAC-FISH

BAC-FISH was performed on *B. distachyon* Bd21, *B. stacei* ABR114, and *B. hybridum* ABR113 and Bhyb26 with *B. distachyon*- and *B. stacei*-derived Bacterial Artificial Chromosome clones (BACs): BAC ABR1-63-E6 and BAC a0047D12 are cloned from *B. distachyon* genomic DNA. The former is from the ABR1 library of (Hasterok et al. 2006), and the latter is from the BD_ABa library of (Febrer et al. 2010). BAC 08P20 was cloned from *B. stacei* genomic DNA as part of an unpublished *B. stacei* BAC library made by B. Chaloub. Chromosome preparation and fluorescence in situ hybridization (FISH) were performed according to the method of (Jenkins and Hasterok 2007), with modifications described in (Lusinska et al. 2018b). This procedure is described in (Gordon et al. 2020).

Synteny and gene loss

The GENESPACE pipeline (Lovell et al. 2018; Lovell et al. 2022) (https://code.jgi.doe.gov/plant/genespace-r) was run with default parameters to evaluate synteny among *Brachypodium* genomes and rice (*B. distachyon* Bd21 v3.2, proteome id:556; *B. stacei* v1.1, proteome id: 316; *B. hybridum* ABR113 v1.1, proteome id: 463; *B. hybridum* Bhyb26 v.2.1, proteome id: 693; *O. sativa* MSU v0.7, Phytozome, proteome id: 323). All reference genomes were obtained from Phytozome (Goodstein et al. 2012) (https://phytozome-next.jgi.doe.gov/). GENESPACE infers orthology relationships among primary peptide sequences using orthofinder (Emms and Kelly 2019), but limits the search to colinear (syntenic) blocks identified by MCScanX (Wang et al. 2012). GENESPACE output includes syntenic dotplots and riparian plots, which were used to visually assess structural variation, and orthogroups, which were the basis of the gene loss analysis. As a “sanity check”, structural variation was also visually assessed with CoGe’s SynMap tool (Schnable and Lyons 2012) (https://genomevolution.org/coge/).

Our procedure for pseudogene identification is essentially that of (Gordon et al. 2020), except that we started with incomplete orthogroups rather than incomplete gene triplets. The neighborhood of the “missing gene” in Bhyb26 was identified based on orthology relationships of ten genes flanking, or nearly flanking, the diploid gene from the progenitor corresponding to the subgenome with the missing gene. The protocol was as follows: once we had identified the diploid gene corresponding to the missing Bhyb26 gene, we 'walked' outward along the diploid chromosome in both directions, checking whether each nearby gene had a single ortholog in the appropriate Bhyb26 subgenome. If a gene had no orthologs or many orthologs, it was skipped and we proceeded to the next-closest gene. This process was repeated until we had ten informative genes flanking the original diploid gene, five on each side. The syntenic orthogroup was discarded if we had to check more than 25 genes on one side, or if we ran off the chromosome before we had 5 good neighbors. At this point, 588 of our original 664 orthogroups

remained. Next, we required that at least four of the five neighboring genes on either side of the original diploid genes had orthologs in the same 200kb region of the Bhyb26 genome. At this point 534 orthogroups remained. Finally, we recorded the Bhyb26 orthologs of the upstream and downstream neighbors that were closest to the original diploid gene, and extracted the region between and including these two 'anchor' genes. If the region was greater than 20kb, the orthogroup was discarded. Finally, 517 candidate Bhyb26 regions remained.

Once we had identified the Bhyb26 genomic region potentially containing the missing gene, the region was extracted using bedtools (Quinlan 2014). The diploid peptide was then aligned to that region using the codon- and intron-aware protein2genome model of EXONERATE (Slater and Birney 2005). We found that these EXONERATE alignments were of excellent quality, but EXONERATE codon-aware DNA-DNA alignments were of poor quality, especially on long genes containing frameshifts. Therefore we next aligned the diploid coding sequence (from Phytozome) to the inferred Bhyb26 coding sequence (from EXONERATE) using MACSE (Ranwez et al. 2011; Ranwez et al. 2018), and these alignments were used to calculate pairwise nonsynonymous to synonymous substitution rate ratios via the yn00 program from PAML (Yang 2007). The same procedure was applied to fully conserved Bhyb26 genes as a control, with each of 1000 trials consisting of 224 BhD genes and 240 BhS genes (464 total), since this was the final number of aligned 'missing' genes from each subgenome.

TE annotation and analysis

TE annotation was performed with an in-house pipeline. The pipeline was not designed for external use, but the scripts are available at https://github.com/vtartaglio/scripts/tree/master/TEs/TE_pipeline. First, monocot TEs were pulled from the RepeatMasker database, and these were concatenated to the TREP database to create an initial TE library. To discover TEs from the *Brachypodium* genomes that are not in public databases, we ran a suite of TE discovery tools on the following lines: *B. distachyon* (Bd21 v3.0, Bd1-1 v1.1), *B. stacei* (ABR114 v1.1), and *B. hybridum* (ABR113 v1.1, Bhyb26 v2.1). Tools used were LTR-Harvest (Ellinghaus et al. 2008), LTR_retriever (Ou and Jiang 2018), TransposonPSI (http://transposonpsi.sourceforge.net/), MITE-Tracker (Crescente et al. 2018), and RepeatModeler2 (Flynn et al. 2019). These TEs were added to the library and redundancy was removed with CD-HIT (Fu et al. 2012) according to the 80-80-80 rule. (cd-hit-est -c 0.8 -G 0 -aS 0.8 -n 5 -T 0 -d 0 -M 0) Sequences were clustered if they had 80% identity locally, and the alignment had to cover at least 80% of the shorter sequence. Only the longest sequence (the representative sequence) from each cluster was retained. Representative sequences less than 80bp were discarded. Next, ProtExcluder from the MAKER-P pipeline (Campbell et al. 2014) was used to search the TE library against a plant protein database, and TEs with significant hits to genes were removed. The result of this process is a non-redundant library containing TE exemplars from a variety of monocots and a de novo TE exemplars from that *Brachypodium* genome. Each genome had its own separate TE library.

All the genomes listed above were annotated with RepeatMasker (Smit et al. 2013 2015) using the appropriate TE library. Non-contiguous genomic sequences that match the same exemplar were designated fragments of a single TE copy if certain distance and orientation criteria were met, using 'one code to find them all' (Bailly-Bechet et al. 2014) with default parameters. Genomic TE copies that were hits to a particular exemplar were considered to belong to the same TE family. Subgenome-specific TE families were those that had at least 5

members and that had at least 90% of their copies on one of the two subgenomes (this latter criterion comes from (Wicker et al. 2018)).

TEMP2 (T. Yu et al. 2021), was used to identify TE polymorphisms relative to the ABR113 reference genome. Library quality was assessed with FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Short-read libraries were the same as those used in (Gordon et al. 2020). High-quality reference genomes for the parental genotypes of the resynthesized polyploid, *B. distachyon* Bd3-1 and *B. stacei* Bsta5, are not available, so the reference was a concatenation of *B. distachyon* Bd21 and *B. stacei* ABR114. Using this reference, transposon insertion polymorphisms (TIPs) were called for the resynthesized allopolyploid (S4 generation) *and* for the pooled Illumina reads of the true diploid parents. TIPs present in the resynthesized polyploid and absent in the parent pool were considered post-WGD polymorphisms. Only TIPs that were supported by reads on both ends ('1p1') and that had a frequency of 20%--that is, at least 20% of sequenced genome supports the insertion--were considered.

To estimate the insertion times of intact LTR-RTs, we largely followed the method of (Wicker et al. 2018), which itself derives from (SanMiguel et al. 1998). The 3' and 5' LTRs of individual LTR-RTs were aligned to each other with MAFFT (Katoh et al. 2002) (einsi --adjustdirectionaccurately) and trimmed with trimAl (Capella-Gutiérrez et al. 2009) (trimal -gapthreshold 0.8). Then EMBOSS distmat (http://emboss.sourceforge.net/) was run on each alignment with the Kimura 2-parameter correction (distmat -nucmethod 2) to obtain the percent identity between the LTRs. Insertion time was calculated with the equation: $T = D/2t$, where T is the time elapsed since the insertion, D is the estimated LTR divergence and t is the substitution rate, for which we used 1.3*10^-8 substitutions per site per year (Ma and Bennetzen 2004).

Results

Assembly and annotation

We assembled a chromosome-scale reference genome of the naturally inbred allotetraploid *Brachypodium hybridum* accession Bhyb26, which was collected in the wild in Jaen, Spain. In a previous study, we built an Illumina-based genome assembly (Bhyb26 v1.1). The new genome assembly (Bhyb26 v2.1) was constructed de novo using long read and Hi-C technologies. The main assembly was performed with MECAT (Xiao et al. 2017) using 45x PacBio coverage (average read length of 19,692), and the resulting assembly was polished with 40x Illumina reads using Arrow (https://github.com/PacificBiosciences/GenomicConsensus). Hi-C scaffolding was performed using the Juicer pipeline (Durand, Shamim, et al. 2016). A total of 51 joins were applied to the broken assembly to form the final assembly consisting of 15 chromosomes, with a total of 99.69% of the assembled sequence contained in the chromosomes. The main genome consists of 32 scaffolds with an N50 of 31.9Mb. 15 of the 32 scaffolds were chromosome-scale (>20Mb), and the remaining 17 totaled about 1.6Mb of sequence. The final genome size is 528.5 Mb, and contains less than 0.1% gaps.

Annotation was performed with the JGI pipeline (see Methods). Transcript assemblies were made from ~290 million pairs of 2X150 stranded paired-end Illumina RNA-seq reads and 23 million PacBio Iso-Seq circular consensus sequences (CCSs), each generated from four tissues: leaf, spikelet, root, and callus. The annotation (v.2.1) contains 53,864 primary transcripts

with an average of 5.1 exons, a median exon length of 166bp, and a median intron length of 142 bp. The BUSCO v3.0.2 score on Embryophyta odb9 is 99.7% complete.

Synteny and structural variation

An initial survey of Bhyb26 genome structure was conducted using molecular cytogenetics. Fluorescent in situ hybridization (FISH) experiments with BACs containing large genomic DNA (gDNA) inserts as probes (BAC 'landing') (Jenkins and Hasterok 2007) were conducted using clones from previously constructed BAC libraries (Hasterok et al. 2006; Febrer et al. 2010) (B. Chalhoub, unpublished) (Fig. 2.1B). Two BACs containing *B. distachyon* gDNA were found to reliably hybridize with the entire D subgenome, but they did not hybridize with any chromosomes of the S subgenome. A third BAC containing *B. stacei* gDNA hybridized with the entire S subgenome, but not the D subgenome. The three BACs discriminated between subgenomes in both polyploids. Therefore in both Bhyb26 and ABR113, the subgenomes are readily distinguishable at the level of molecular probes, and no evidence of sequence exchange between subgenomes was observed.

Next, we performed a computational survey of Bhyb26 genome structure. Syntenic blocks between each polyploid subgenome and its diploid progenitor were identified using the GENESPACE pipeline (Lovell et al. 2018). 97.9% of the Bhyb26 D subgenome was contained within blocks syntenic to *B. distachyon* (Bd21 v3.2), and 93.4% of the Bhyb26 S subgenome was contained within blocks syntenic to *B. stacei* (ABR114 v1.1). There were 41 Bd-BhD syntenic blocks and on average they were 6.6Mb in length, while the 124 Bs-BhS syntenic blocks were on average 2Mb in length. This lower concordance between the S subgenome and its progenitor species may be attributable to the *B. stacei* reference genome being a lower-quality Illumina assembly, and therefore does not necessarily reflect biological divergence.

The synteny results revealed several inversions in Bhyb26 relative to its diploid progenitors. There is a 2.6Mb inversion on chromosome BhD3, as well as one 5.2 Mb and another 6.3 Mb inversion on chromosome BhD5 (Fig. 2.2A). There is one inversion on the S subgenome, a 4.5 Mb inversion at the top of chromosome BhS8. Because the *B. stacei* reference genome is lower quality than the *B. distachyon* reference genome, we also compared the Bhyb26 genome to Bd28, a *B. hybridum* accession from the younger lineage (S-plastotype) with a Pac-Bio genome. The inversions were still unique to Bhyb26 in this comparison, indicating that the inversions, particularly those on the S-subgenome, are not simply absent from the *B. stacei* reference due to genome quality.

To ascertain whether inversions are common between diploid *Brachypodium* accessions, we re-ran our synteny pipeline on several high-quality *B. distachyon* genomes, one from each of the three major populations of *B. distachyon*: Bd21 representing the Turkish+ clade, Bd30-1 representing the Spanish+ clade, and Bd1-1 representing the extremely delayed flowering+ clade (Gordon et al. 2017). We detected no inversions between these diploid genomes. We also re-ran our synteny pipeline on each *B. hybridum* ABR113 subgenome against the diploid progenitor species, and, as previously reported, found no inversions (Gordon et al. 2020). These results indicate that the Bhyb26 genome contains several inversions that are private to that lineage. While it is possible that these inversions were present in the actual progenitors of Bhyb26 prior to polyploidization, the absence of any similarly dramatic structural variation in the widely sampled natural diversity of *B. distachyon* suggests that these inversions may well have occurred post-polyploidy.

Gene loss

In a previous analysis (Gordon et al. 2020), we ascertained that Bhyb26 had more genetic variation relative to the *B. distachyon* reference genome than did ABR113, but the low-quality assembly did not permit in-depth analysis of this variation. We were particularly interested in degradation or loss of genes, which would be indicative of the early stages of diploidization. However, identifying genes that have been lost in Bhyb26 since WGD is difficult without its *true* progenitors, since gene presence-absence variation would be common among arbitrary accessions of *B. distachyon* and *B. stacei* (Gordon et al. 2017). We therefore searched for losses of highly conserved genes, reasoning that any gene that is conserved within and beyond the genus *Brachypodium* was probably present in the true progenitors of Bhyb26. Using the synteny and homology-based pipeline GENESPACE (Lovell et al. 2018; Lovell et al. 2022), we identified 15,217 orthogroups--groups of orthologous genes--that contained at least one orthologous gene in both subgenomes of both polyploids, each diploid genome, and rice (*Oryza sativa*). In other words, we identified many thousands of genes that are widely conserved across the genus *Brachypodium* and in rice. We then identified orthogroups where all but one genome or subgenome was represented (Fig. 2.3A). Unsurprisingly, orthogroups that had an ortholog in every *Brachypodium* sub/genome but not rice were most common (3,912 orthogroups). More surprisingly, the number of cases where a gene was 'missing' from a single polyploid subgenome was greater than we would expect by summing the progenitors, and this discrepancy was greater for the Bhyb26 subgenomes than for those of ABR113. For example, 299 genes were present in every sub/genome except Bhyb26S; meanwhile, only 108 genes were present in every sub/genome except ABR113S. The high number of conspicuously absent genes in Bhyb26 suggests that at least some of these genes may be true pseudogenes or deletions that occurred post-WGD as a consequence of relaxed selection.

We interrogated these putative 'lost genes' more closely (Fig. 2.3B). Six hundred sixty-four broadly conserved genes were absent in one Bhyb26 subgenome. In 517 of those cases, we were able to definitively identify a region of the Bhyb26 genome where the missing gene "should" be (see Methods). 464 of those 517 regions contained sequence that could be aligned to the peptide sequence of the corresponding diploid gene. These sequences were scattered throughout the Bhyb26 genome (Fig. 2.4B). As a control, we ran the same procedure on a set of 464 randomly selected Bhyb26 genes that were completely conserved, that is, genes from orthogroups in which all genomes are represented (see Methods). We repeated the random sampling and analysis for a total of 1,000 times. These alignments were longer: for fully conserved genes, the alignments were 22.7bp shorter than the peptide on average, while for the putative pseudogenes, they were 224.3bp shorter (Table 2.1). Furthermore, 18.8% of the Bhyb26 putative pseudogenes contained a premature termination codon (PTC), while none of the alignments between a fully conserved Bhyb26 gene and its diploid ortholog contained a PTC. While the difference in mean expression values (TPM) was not significant between the pseudogenes and control genes (two-sided t-test p-value=0.748), this is probably due to a few very highly expressed outliers among the candidate pseudogenes, and in fact removing the single most highly expressed candidate pseudogene from the pool did result in a significant difference (t-test p-value=0.022). Visual inspection of the data reveals that putative pseudogenes were clearly relatively enriched for genes with little to no expression, highly-expressed outliers not withstanding (Fig. 2.4B). Finally, a valid dN/dS ratio, that is, the ratio of non-synonymous to synonymous amino acid substitutions (Yang 2007), could be calculated for 423 of the 464 genes. The average dN/dS for the lost genes was 0.53 as opposed to the conserved genes' 0.36,

consistent with relaxed selection. There were no cases where the average dN/dS from a control trial exceeded the average dN/dS from the lost genes. Interestingly, when we repeated this procedure on the ABR113 lost genes, obtaining 220 putative pseudogenes and 106 dN/dS values (many of the alignments had no substitutions in the polyploid), we did not observe a difference from the conserved genes (Fig. 2.3C). Thus, while the ABR113 lost genes do not show signs of pseudogenization, the Bhyb26 genes do. They went unannotated due to lack of homology, incidence of premature stop codons, and weak transcriptome support (Table 2.1).

We hypothesized that TE insertion into the gene body may have contributed to the inactivation of these putative pseudogenes. 130 of the 464 putative pseudogenes contained a TE somewhere between the start and end of the alignable region. Meanwhile, in the 1,000 control trials, on average 196 of the 464 randomly selected conserved genes contained a TE. This shows that the putative pseudogenes are not more likely to contain a TE than we would expect by random chance, although it is still possible that TE insertions in nearby regulatory regions may have deactivated some of the genes.

Finally, we noticed that both polyploid lineages had apparently lost more genes from the D subgenome than the S subgenome (Fig. 2.3A). We performed a chi-square test to test whether the biased loss was significantly different from a bias we might expect by chance, based on the total number of genes in each subgenome. In Bhyb26, the difference was significant ($p = 0.031$), but not in ABR113 ($p = 0.34$). Together, all these results indicate that (1) a significant portion of the 'missing' genes in Bhyb26 are of dubious functionality, (2) the gene loss is marked by small-scale substitutions and deletions rather than by rampant TE insertions or by deletion of entire genes, and (3) in Bhyb26, the S subgenome is slightly but significantly dominant in terms of gene retention.

Gene expression

Using the Illumina RNA-seq data, we investigated whether one subgenome was systematically more highly expressed than the other in Bhyb26. Two analytical approaches were used: one for homeolog expression bias and one for subgenome expression dominance. Since we did not have biological replicates, we could not conduct a formal HEB analysis. Nevertheless, our experiment should be enough to distinguish a genome-wide trend, since the >50,000 genes provide a sort of replication, as do the four tissues sampled. We used the GENESPACE pipeline (Lovell et al. 2018) to identify 1:1 homeologs between the subgenomes, and then filtered out noisy gene pairs (those where both homeologs had a TPM < 1.0) and recorded whether the BhD homeolog or the BhS homeolog was more highly expressed. The chance that the homeolog from a particular subgenome was more highly expressed was near 50/50 in all tissues (Fig. 2.5A). The most extreme deviation from 50/50 was observed in leaf, in which 49.2% of gene pairs favor the BhD homeolog while 50.8% favor the BhS homeolog. To test whether the deviation from 50/50 was significant in any tissue, we performed a binomial test. Leaf was closest to significance ($p=0.052$, alpha=0.0125 with Bonferroni correction), but in no case was the pattern of homeolog expression bias significantly different from what would be expected by random chance.

Finally, we checked for subgenome expression dominance, that is, evidence that the majority of expressed transcripts are coming from one of the two subgenomes. To control for the fact that one subgenome may simply have more genes, we summed the lengths of the primary transcripts from all genes in each subgenome, and took the number of genes in each subgenome to be our null expectation: 50.7% of counts would be expected to originate from BhD, and 49.3% from BhS. All four tissues were close to this ratio, with floret being the most extreme deviation:

46.41% of counts were from BhD transcripts (Fig. 2.5B). While there may be some subtle subgenome expression dominance in floret, there is no evidence for overall subgenome expression dominance in Bhyb26.

Gradual TE activity post-WGD

We developed a workflow for TE annotation in *Brachypodium* genomes, since disparities in TE content may drive biased genome evolution in polyploids (Woodhouse et al. 2014; Edger et al. 2017). Publicly available repeat sequences and de novo TEs were identified in five *Brachypodium* genomes (see Methods). The TE content of each polyploid subgenome was compared to its progenitor species (Fig. 2.6B, Table 2.2). All the genomes are compact, though the D sub/genomes were slightly more TE-rich than the S sub/genomes. The Bhyb26 D subgenome was most TE-rich at 31% TEs, while the *B. stacei* genome was the most TE-poor at 20% TEs. The latter figure may be an underestimate since *B. stacei* is a short read assembly, though it is close to the Bhyb26 S figure (24%). The Bhyb26 S subgenome was enriched for full-length LTR-retrotransposons (LTR-RTs) relative to the other two S sub/genomes (122 versus 108 and 64), which might be due to the long-read assembly. RLG* and RLC elements (see Note beneath this section) occupied most of the TE space in all genomes. The ratio of RLG to RLC LTR-RT copies ranged from 1.16:1 to 1.30:1 in all genomes. Non-LTR retrotransposons also comprised a substantial portion of the TE space, from 2.6 Mb in *B. stacei* to 5.4 Mb in Bhyb26 D.

ABR113 had a slightly higher proportion of subgenome-specific TEs than Bhyb26 (Fig. 2.6A inner track, Fig. 2.6C). 11.4% and 16.8% of all TE copies in Bhyb26 and ABR113, respectively, were from subgenome-specific TE families. This difference is slight, but the trend was consistent across chromosomes (Fig. 2.6B), and a paired t-test of the chromosome-level percentages was highly significant ($p = 2.722 \times 10E\text{-}8$). This slight but significant difference suggests that either the TE landscapes of the true progenitors of Bhyb26 were more similar to each other than those of ABR113 at the time of WGD, or there has been some small-scale transfer of TEs between subgenomes in Bhyb26 post-WGD. Moderate post-WGD TE activity and exchange between subgenomes is plausible, particularly in light of our other findings, though it cannot be confirmed without the progenitors.

We also looked for evidence of overall increased TE activity in and around genes in Bhyb26. We recorded the number of TEs that overlap a gene in each genome, requiring that the TE and gene be on the same strand, and UTR and intronic TEs were included. We found that 43% and 42% of Bhyb26 and ABR113 genes, respectively, overlap or contain a TE. In Bhyb26, 2.1% of exons overlap a TE, while in ABR113 5.1% of exons overlap a TE. TE overlap with genes remained remarkably similar between the two polyploids when only TEs in either centromeres, pericentromeres, or distal regions were considered. The mean distance from a TE to a gene was similar in both polyploids: 1272 bp in Bhyb26, and 1289 bp in ABR113. Thus, TEs in Bhyb26 show no elevated propensity to insert in or near genes compared to ABR113.

We surveyed TE polymorphisms among *B. hybridum* lines. We used the TE polymorphism detection software TEMP2 (T. Yu et al. 2021) as implemented in the McClintock pipeline (Nelson et al. 2017) to quantify TE polymorphisms in short-read data from 20 *B. hybridum* lines, using ABR113 as our reference genome (Fig. 2.7A). We focused on transposon insertion polymorphisms (TIPs), that is, locations where a TE insertion was present in a resequenced genome but not in the reference. Bhyb26 had by far the greatest number of TIPs relative to ABR113. This increase was not due to sequencing technology since all samples are

short read data from the same experiment. TIP number was not correlated with sequencing depth and only loosely correlated with library quality (total TIPs vs. median per sequence quality score R-squared=0.34), so it is not obviously a technical artifact. Interestingly, 118-5 is of the same plastotype as Bhyb26 but does not have nearly as many TIPs (Fig. 2.7A). Previous phylogenetic analysis (Gordon et al. 2020) strongly suggests that 118-5 and Bhyb26 are of the same origin and neither is admixed with S-plastotype lineages of *B. hybridum*, so the greater number of polymorphisms in Bhyb26 suggests an uptick in TE activity since its divergence from 118-5. Many TE families contribute to TE diversity in *B. hybridum*. The 10 TE families that contribute the greatest number of TIPs are responsible for 52% of all TIPs. The majority of TIPs came from RLG and RLC LTR-RTs, the most active classes of TE among *B. distachyon* accessions as well (Stritt et al. 2018; Stritt et al. 2020) (Fig. 2.7B). No single TE family contributed more than 25% of a genome's total TIPs in any *B. hybridum* accession. This is similar to what was observed in *B. distachyon*, where no single family dominates the TE diversity (Stritt et al. 2018).

Because LTR-RTs are among the most abundant and most active TEs in *B. hybridum* (Table 2.2, Fig. 2.7B), we estimated insertion times for intact LTR-RTs in several *Brachypodium* genomes (Fig. 2.6D). The number of intact LTR-RTs across our dataset appears to be a function of genome size and assembly quality. All full-length ABR113 LTR-RTs pre-date the WGD. The percentage of full-length LTR-RTs less than 1.4 million years old was similar in ABR113 and Bhyb26: 40% and 44%, respectively. That Bhyb26 has slightly more 'young' TEs again hints at the possibility of an uptick in TE activity since the WGD, though the difference is very slight.

*Note: In the spirit of inclusion and respect for all people, we avoid using the common name of this transposon superfamily, which derives from a cultural stereotype (Kim and Belyaeva 1991). Instead, we refer to the three LTR-retrotransposon superfamilies by their three-letter abbreviations: RLG, RLC, and RLX (Wicker et al. 2007).

Discussion

Diploidization is underway in *B. hybridum*, an allotetraploid with multiple origins. In contrast to the more recent *B. hybridum* lineage ABR113, the older Bhyb26 lineage shows several megabase-scale inversions and a greater extent of pseudogenization. In both lineages, gene loss slightly favored retention of the S subgenome, though the difference was significant only in the older line, and was not supported by expression data. Finally, we found evidence for gradual rather than instantaneous post-WGD TE activity. We argue that these genomic changes were most likely made possible by relaxed selection post-WGD. The changes are modest overall, consistent with slow and gradual post-WGD evolution.

The chromosomal rearrangements observed in Bhyb26 are not characteristic of homeologous exchange, a classic genomic shock response. Homeologous recombination can lead to duplications, deletions, and translocations (Mason and Wendel 2020). Inversions, on the other hand, more likely result from ectopic recombination or non-homology directed DNA repair within a single chromosome. The inversions that are unique to Bhyb26 probably did not arise through homeologous exchange, so they could have occurred either pre- or post-WGD. However, we find no megabase-scale inversions between diverse accessions of the well-sampled diploid progenitor species *B. distachyon*. Thus, the available evidence suggests that such large inversions are not typical of intraspecific variation within *Brachypodium* diploids. While it is still

possible that the true progenitors of Bhyb26 each happened to harbor large inversions relative to all well-characterized modern *B. distachyon* and *B. stacei* lines, we think a more likely explanation is that the relaxed selection accompanying WGD allowed inversions to persist in the polyploid. Whether these inversions harbor adaptive alleles, as is sometimes the case (Huang and Rieseberg 2020), will be an interesting area for future study.

Some gene loss or gain between lineages, even within the same species, is expected in the normal course of evolution (Gordon et al. 2017). Indeed, we observed that all our *Brachypodium* reference genomes lack at least several dozen genes that are otherwise widely conserved within and beyond the genus. However, such conspicuously absent genes were more common in the polyploids than in the diploid *Brachypodium* genomes, and they were more common in the older polyploid than the younger one. In Bhyb26, the remnants of these genes were shorter, less-expressed, and contained more premature stop codons and non-synonymous substitutions than would be expected by random chance, suggesting that these were not, or at least not entirely, real genes that were missed due to annotation error. Given that gene loss in ABR113 was greater than the sum of its progenitors, and gene loss in Bhyb26 was greater than in ABR113, gene loss appears to be progressing gradually with time. It is worth noting that our current study uncovered more potential pseudogenes in ABR113 than our previous study, likely due to our more sophisticated methods of calling synteny (Gordon et al. 2020).

Bhyb26 shows some evidence of post-WGD TE activity: it is slightly depleted for subgenome-specific TEs, it has more TIPs than its closest relative, and it is slightly more TE-rich than the other *Brachypodium* lineages. These data are reminiscent of the *Capsella bursa-pastoris* case, in which relaxed selection permitted gradual TE proliferation following WGD (Ågren et al. 2016). However, we cannot exclude the possibility that these genome features were already present in the true progenitors of Bhyb26; for instance, the progenitors may have had many TEs in common at the time of WGD. Furthermore, it is possible that the inversions, gene losses, and slight TE activation are not really due to buffering by duplicate genes, but due to some demographic factor, such as a smaller population and greater genetic drift in Bhyb26 than in ABR113 (or its close relative, Bhyb118-5). Broader sampling of the D-plastotype (older) lineage would allow for greater insight into those polyploids' demographic histories.

It is not unusual for allopolyploids to preferentially retain genes from one dominant subgenome (Garsmeur et al. 2014; Woodhouse et al. 2014; Alger and Edger 2020), and it has been proposed that dominance is established immediately following WGD and increases over time (Edger et al. 2017). *B. hybridum* supports this model in the sense that the biased gene loss does appear to be stronger in the older lineage. However, given that the RNA-seq data do not reveal any genome dominance in either lineage, which is crucial to the proposed mechanism of genome dominance (Freeling et al. 2012), we cannot conclude that *B. hybridum* shows subgenome dominance in the classic sense. *B. hybridum* seems to resemble the paleoallopolyploid *Miscanthus sinensis* (Mitros et al. 2020) or *Cucurbita* (Sun et al. 2017) genomes, as it is an allopolyploid that shows little to no genome dominance. Similar to cotton, our expression data are equivocal, with neither subgenome emerging as dominant across all tissues (Fang et al. 2017).

McClintock's genome shock question remains a matter of much debate today: is the response to WGD chaotic or predictable? This depends, perhaps, on the sophistication of our predictions. The minimal genome restructuring in Bhyb26 is in line with the prediction that allopolyploids should be the more genetically stable class of polyploids (Ramsey and Schemske 2002). However, our genome dominance results qualify the predictions of Garsmeur et al.

(2014). Alger and Edger (2020), Wendel et al. (2018), and a number of papers from the Freeling lab including Freeling et al. (2012) build on Garsmeur et al.'s prediction, emphasizing that the key predictor of genome dominance is not necessarily progenitor divergence *per se*, but progenitor divergence *in terms of TE load*. *B. hybridum* is in line with this refined prediction, not unlike the cases of *Ephedra* (Wu et al. 2021) and teff (VanBuren et al. 2020). Genome evolution in *B. hybridum* is largely subtle and unbiased, even though it formed from a remarkably wide cross (Dinh Thi et al. 2016), perhaps because its progenitors bore a similar TE load. Many genetic, epigenetic, and environmental factors contribute to a polyploid's fate, and there is still much work to be done to determine the how these factors interact with each other. *B. hybridum* has shed some light on this complex question by providing a rare glimpse of diploidization "caught in the act".

Data availability

The Bhyb26 genome and standard annotation files are available on Phytozome (https://phytozome-next.jgi.doe.gov/).
Bhyb26 DNA and RNA reads used for genome assembly and annotation are available on the JGI genome portal at this link:
https://genome.jgi.doe.gov/portal/BrahybStandDraft_11_FD/BrahybStandDraft_11_FD.info.htm h or under the Proposal ID: 503504 and the following Project IDs:

Root, leaf, floret and callus Illumina and Pac-Bio (Iso-seq) data used for annotation: 1229574

High-weight genomic DNA Pac-Bio sequencing: 1229573

Additional Illumina libraries mentioned in this report are the same as were used in Gordon et al. (2020). See that publication for details.

Chapter 2 Figures

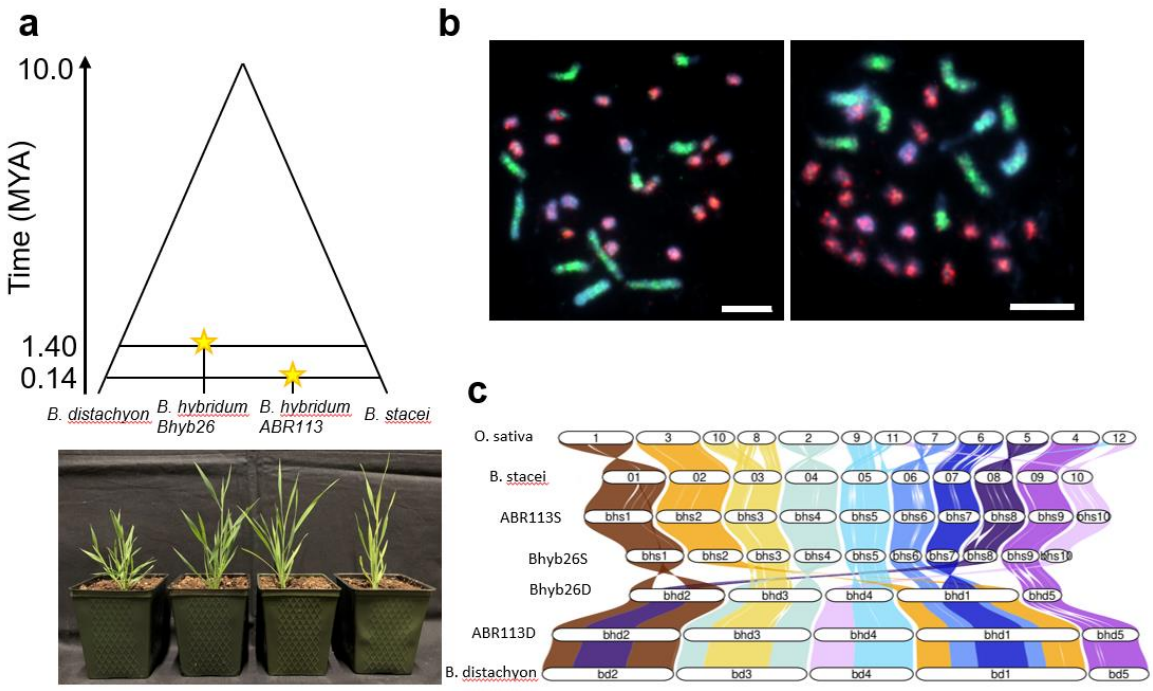

Fig 2.1. Both independent allopolyploid lineages reveal remarkable genome stability. (a) Cladogram illustrating relationships in the *B. hybridum* polyploid complex. (b) BAC-FISH with probes specific to either the S subgenome (8P20) or D subgenome (ABR1-63E-6) indicate considerable sequence heterogeneity between subgenomes. Left, Bhyb26, right, ABR113. Blue fluorescence, DAPI. Bars, 5 µm. (c) Riparian plot showing high collinearity between each subgenome and its progenitor, and low collinearity between the polyploid subgenomes, consistent with the high progenitor divergence.

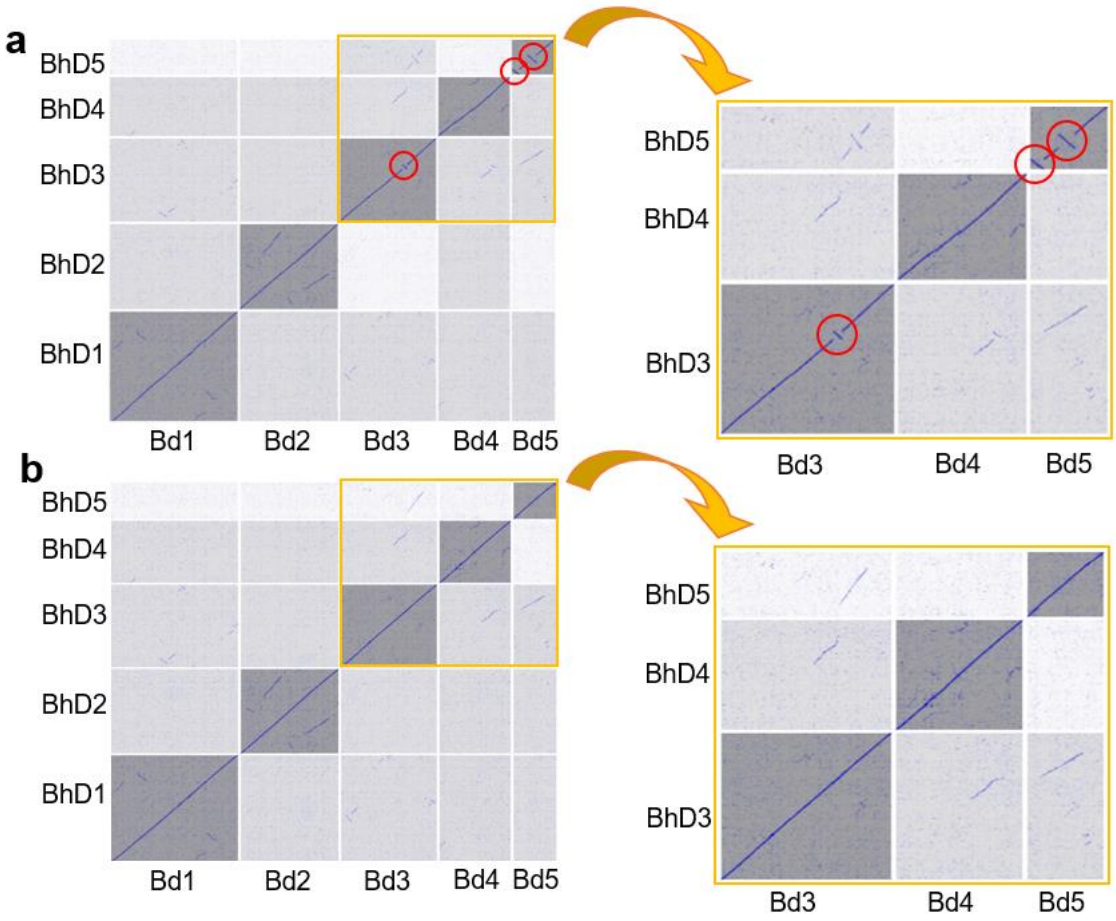

Fig 2.2. Inversions in Bhyb26 suggest possible relaxed selection since WGD. Syntenic dot plots relative to diploid progenitors revealed ~2Mb-6Mb inversions (red circles) on both subgenomes of Bhyb26, but similar structural variation was absent from ABR113. (a) D subgenome of Bhyb26 vs. B. distachyon. (b) D subgenome of ABR113 vs. B. distachyon.

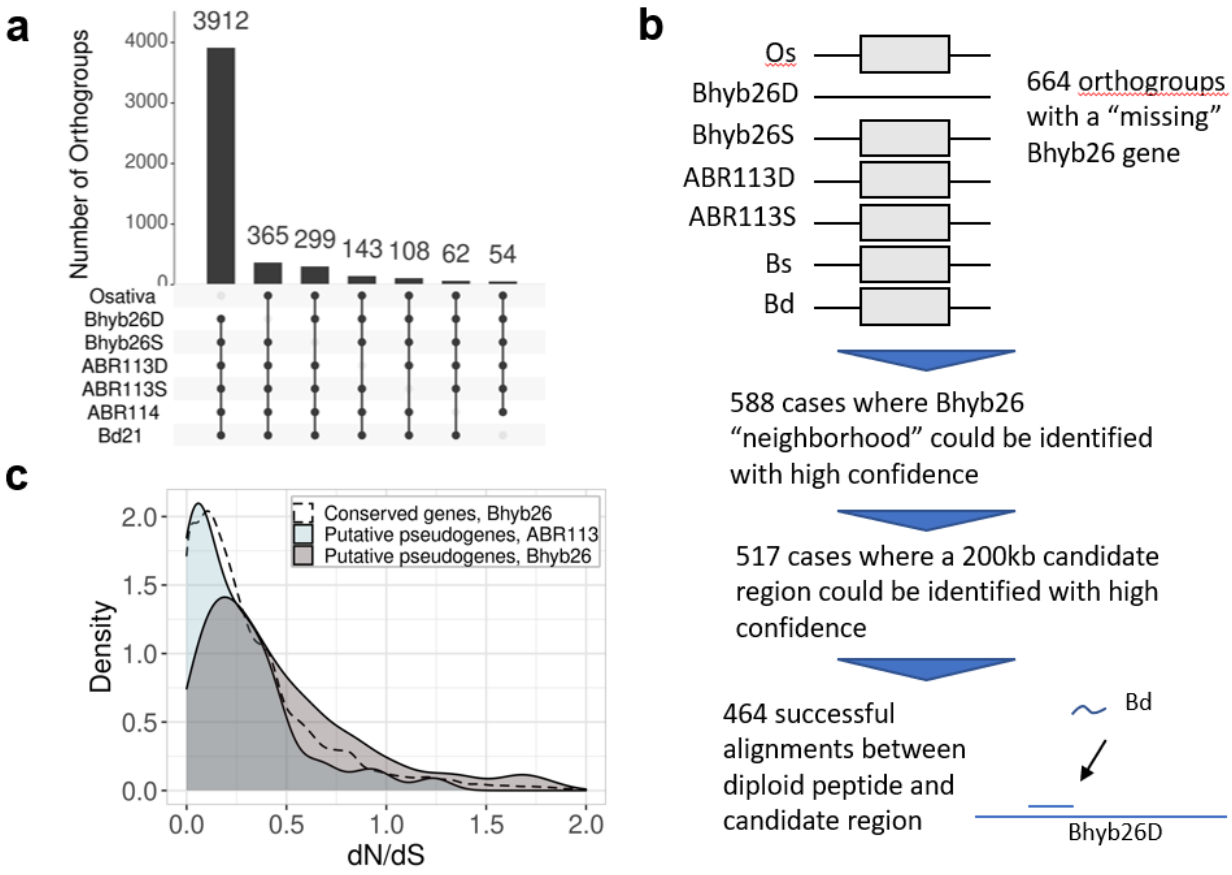

Fig 2.3. Bhyb26 shows more gene loss than ABR113. (a) UpSet plot of orthogroups (groups of orthologous genes) reveals a high number of cases where a single Bhyb26 subgenome lacks an ortholog of an otherwise widely conserved gene. (b) Workflow for identifying putative pseudogenes. (c) Distribution of dN/dS ratios for Bhyb26 and ABR113 'lost genes', and Bhyb26 widely conserved genes. All dN/dS values are relative to the corresponding diploid ortholog.

Metric	Putative pseudogenes	Conserved genes (average of 1000 trials)
Mean length difference between original diploid gene and Bhyb26 alignment	224.3bp	22.7bp
Percent of alignments that contained a premature stop codon in Bhyb26 relative to diploid gene	18.8%	0.0%
Mean dN/dS (aligned to diploid ortholog)	0. 53	0.36
Median transcripts per million (TPM)	0.0	4.14

Table 2.1. Characteristics of Bhyb26 putative pseudogenes vs. Bhyb26 annotated, conserved genes.

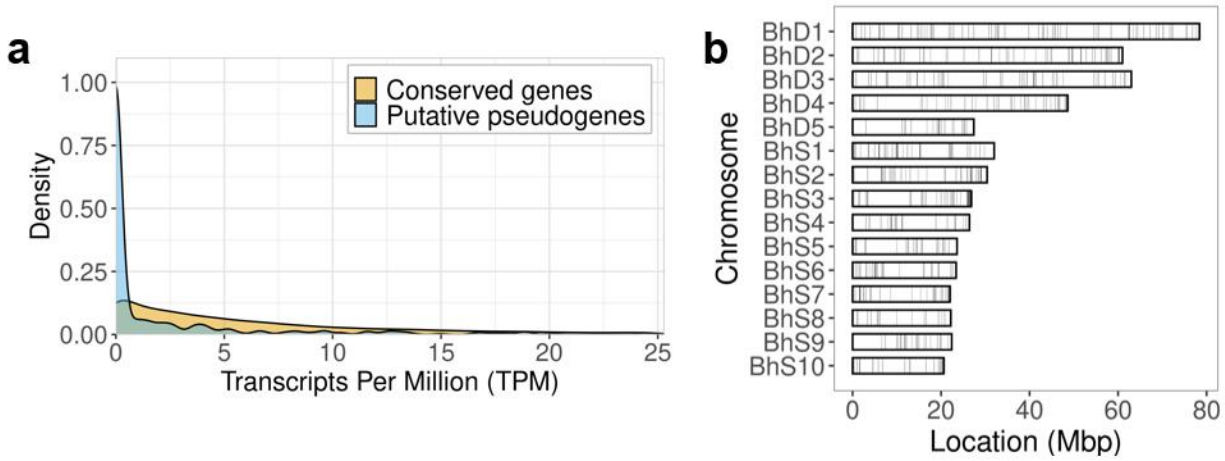

Fig. 2.4. Additional characteristics of Bhyb26 putative pseudogenes. (a) The pool of Bhyb26 candidate pseudogenes is enriched for genes with low expression. Conserved gene data reflect all widely-conserved Bhyb26 genes. (b) Distribution of putative pseudogenes across the Bhyb26 chromosomes.

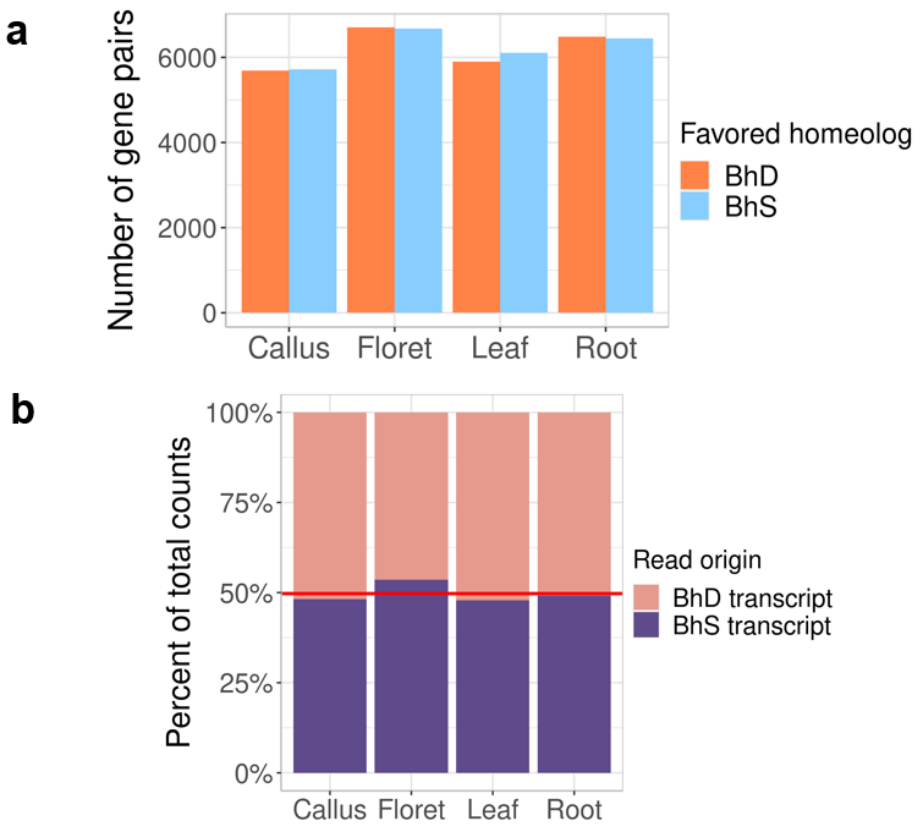

Fig. 2.5. No expression bias in Bhyb26. (a) Grouped bar chart showing the more highly-expressed homeolog in gene pairs from four tissues. Only gene pairs where at least one homeolog had a TPM > 1.0 were considered. (b) Stacked bar chart showing percent of RNA-seq reads mapped to each subgenome. Red line indicates the percent of primary transcriptome base pairs that are from BhS transcripts.

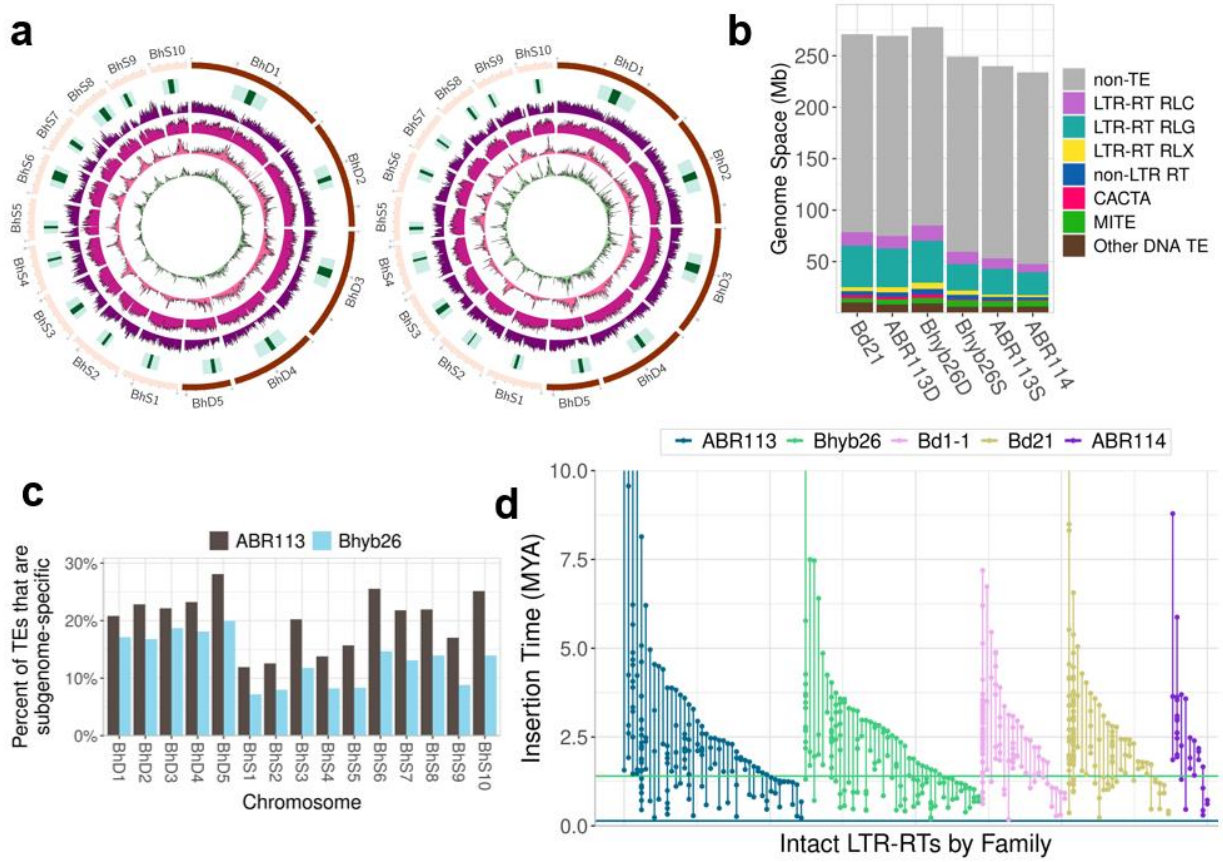

Fig. 2.6. Gradual TE activity in *B. hybridum*. (a) Overview of Bhyb26 (left) and ABR113 (right) genomes. Tracks, outer to inner: pericentromeres and centromeres, gene density, TE diversity, TE density, density of subgenome-specific TEs. (b) TE composition of sub/genomes by TE class. (c) Subgenome-specific TEs, as a percent of total TEs, per chromosome in the two polyploids. (d) Insertion time analysis of intact LTR-retrotransposons. Each vertical line is a TE family and each point is an individual TE copy. Horizontal lines denote WGD events for Bhyb26 (green) and ABR113 (blue).

	Sub/genome size (Mb)	TE space (% of genome)	Genome space occupied by DNA transposons (%)	Genome space occupied by LTR retrotransposons (%)	Total solo LTRs	Total full-length LTR retrotransposons	Average length of full-length LTR retrotransposons
Bd21	275	29.0	6.5	22.5	2244	185	4706
ABR114	234	20.4	5.5	14.9	1259	64	4663
ABR113-D	269	27.8	6.0	20.3	3087	125	4864
ABR113-S	240	22.2	5.2	15.6	1973	108	5636
Bhyb26-D	278	30.6	6.5	22.1	3625	137	4888
Bhyb26-S	249	23.9	5.4	16.8	3250	122	5116

Table 2.2. Survey of TEs in the *B. hybridum* complex.

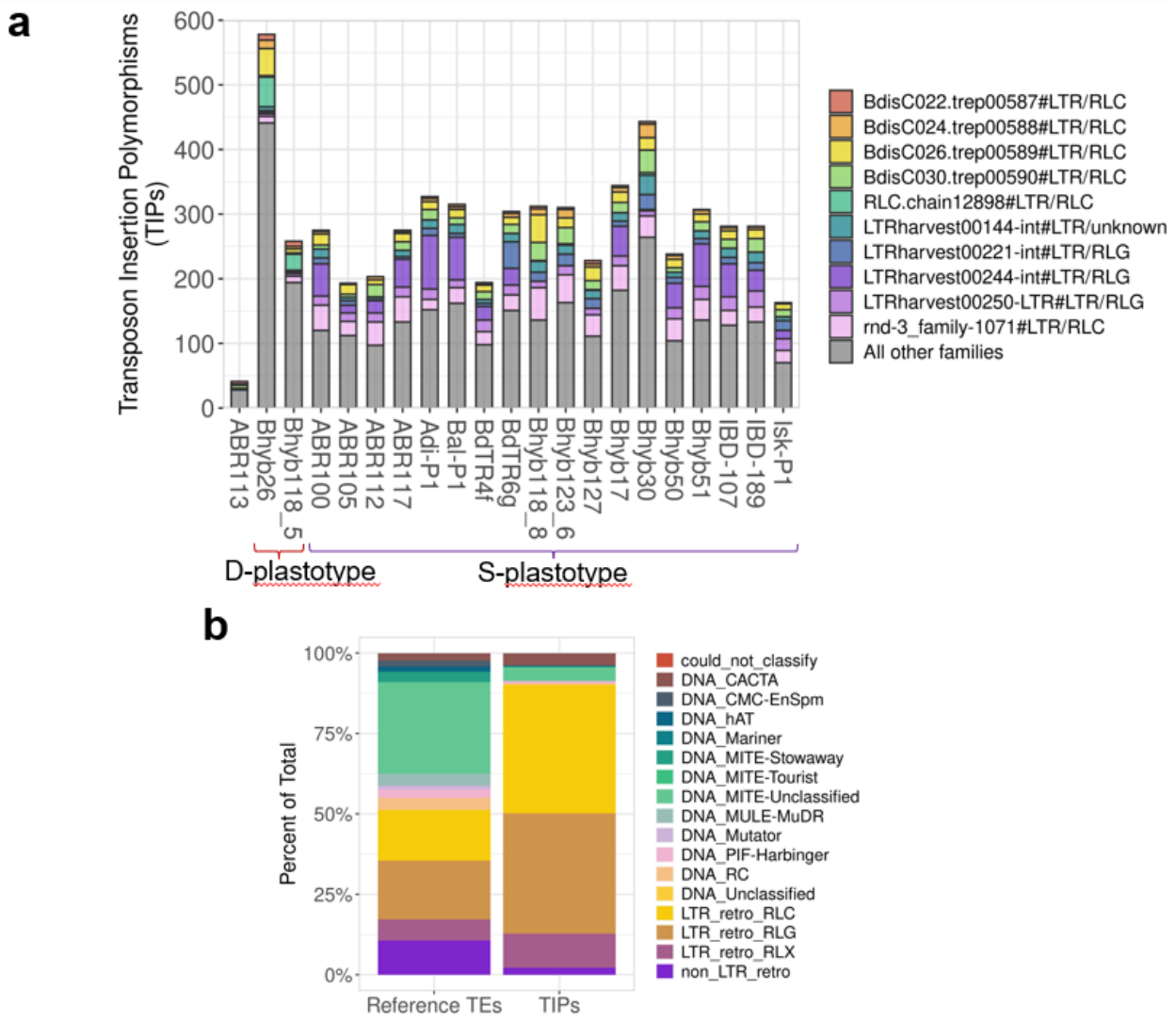

Fig. 2.7. TE diversity *B. hybridum*. (a) Transposon Insertion Polymorphisms (TIPs) in 20 B. hybridum lines relative to the ABR113 reference genome. Colored segments show abundance of top 10 families, by median TIP contribution across lines. Gray segment shows TIPs from all other TE families. Left, ABR113 control. (b) *B. hybridum* TIP contribution disproportionately comes from RLC and RLG retrotransposons. Left, TE composition of ABR113 genome. Right, TIPs from 20 *B. hybridum* lines using ABR113 as reference.

Chapter 3: The *Brachypodium* KNOT is shared between polyploid subgenomes

Abstract

In allopolyploid plants, two diverged genomes come together in one nucleus, possibly initiating reorganization of chromosome territories. Hi-C uses sequencing technology to generate a map of DNA-DNA interaction frequencies for the whole genome, and so it is a promising means to study large-scale and fine-scale chromatin dynamics in polyploids. Here, we performed Hi-C in diploid and polyploid lines of the wild allotetraploid *Brachypodium hybridum* and conducted an exploratory survey of the large-scale chromatin features. Intra-subgenome contacts were not more frequent than inter-subgenome contacts, showing that the subgenomes do not confine themselves to physically separate 'genome territories'. We found that chromosome configurations may differ between the subgenomes. We also discovered a trans-chromosomal interaction network similar to the KNOT/IHIs previously described in *Arabidopsis* and rice. The interactive loci are syntenic, suggesting an ancient origin for this 3D structure and conservation through time. However, analysis of transposable elements within the structure yielded mixed results, with only some features of KEE/IHI TEs being consistent with the "transposon trap" theory for their function.

Introduction

The spatial organization of interphase chromosomes has major implications for gene expression and genome integrity. Interphase chromosomes are not randomly strewn about the nucleus, but are arranged in chromosome territories, or discrete regions within the nucleus (Cremer and Cremer 2010). Chromosomes consist of topological and functional domains that may bear distinctive epigenetic marks, 3D topology, or associations with certain nuclear features (e.g. the lamina or nucleolus) (Bickmore and van Steensel 2013). The arrangement of chromosome territories and of chromatin domains can impact the accessibility of certain regions of the genome for transcription and possibly DNA repair, and affects the probability of long-range DNA-DNA interactions (Heard and Bickmore 2007; Misteli and Soutoglou 2009; Cremer and Cremer 2010; Cavalli and Misteli 2013). Thus spatial genome organization, and biological events that might disrupt it such as cancer, DNA damage, and polyploidy, are of great biological importance (Misteli 2010).

Chromosome conformation capture technologies including 3C, 4C, 5C, and Hi-C allow for the study of 3D genome arrangement with unprecedented ease and resolution (McCord et al. 2020). They are also changing the way we think about the 3D genome, as they reveal new biological phenomena. For example, the concept of 'topologically associated domains' (TADs) has received much attention since the concept was introduced in 2012 (Dixon et al. 2012; Nora et al. 2012). These physically self-interacting chromatin domains can harbor co-regulated genes, and may be evolutionarily well-conserved, although these are topics of ongoing study and debate (Acemel et al. 2017; Lazar et al. 2018 Jun 18; Eres and Gilad 2021).

Hi-C has also revealed the existence of a mysterious 3D chromatin structure in plants. Two side-by-side papers working in the thale cress *Arabidopsis* described a series of 10 loci of approximately 200kb-1.5Mb that physically contact one another with a surprisingly high frequency (Feng et al. 2014; Grob et al. 2014). Because Hi-C is an average of chromatin contacts across a large population of cells, it is not known whether these loci interact one at a time, several at a time, or all at once. One of these studies dubbed these loci Interactive Heterochromatic Islands (IHIs), and the other dubbed the whole structure the KNOT and dubbed the individual loci KNOT-Engaged Elements (KEEs). One group concluded that the interactive loci were heterochromatic, while the other group concluded that epigenetic marks for these loci were somewhat mixed. Grob et al. (2014) conducted follow-up experiments in which they confirmed the DNA-DNA interactions with Fluorescence In-Situ Hybridization (FISH). Grob et al. (2014) also revealed a remarkable discovery: by surveying a database of *Arabidopsis* mutants created with transposon mutagenesis, they found a statistically significant increase in transposon insertions in these regions. In other words, these loci may act as a sort of molecular sponge, soaking up transposons and perhaps protecting other regions of the genome from deleterious insertions. This is very much like the 'transposon trap' model that has been proposed to describe the PIWI-interacting RNA (piRNA) producing locus *flamenco* in the fruit fly *Drosophila* (Coline et al. 2014). piRNAs are a class of small non-coding RNAs in animals that function in transposon defense in the germline. piRNA clusters like *flamenco* are TE-rich loci transcribed and processed into piRNAs that then direct silencing of homologous TE copies elsewhere in the genome (Siomi et al. 2011). Grob et al. (2014) found that the KEE/IHI regions were significantly enriched for smRNAs. They also analyzed *Drosophila* Hi-C data and found that several piRNA clusters physically contacted one another at a significantly elevated frequency, like the KEEs/IHIs do. In short, Hi-C revealed a series of mutually interacting loci in *Arabidopsis* with intriguing similarities to piRNA clusters in *Drosophila*.

Few studies have used Hi-C technology to investigate chromosome positioning in plants, and studies in polyploid plants are even fewer. What is known about genome arrangement in polyploid plants mostly comes from cytogenetic methods, which remain the gold standard for 3D genome research. These methods have shown that polyploidy can induce large-scale changes in chromosome configuration. The most dramatic example occurs in cases of genome elimination, in which a wide cross results in rapid loss of one of the hybrid or polyploid subgenomes (Ishii et al. 2016). In these cases, the subgenome that will ultimately be eliminated typically resides at the nuclear periphery, while the other remains in the center (Vimala and Lavania 2021). Stable wide hybrids may also show distinct 'genome territories'—that is, the subgenomes occupy separate regions of the nucleus—and the relative positions of the subgenomes may depend on chromosome size (Vimala and Lavania 2021). Similarly, studies involving introgression of a distantly related chromosome or chromosome arm have shown that the introgressed chromosome or chromosome arm tends to occupy a well-defined, discrete space separate from the rest of the genome (Abranches et al. 1998; Koláčková et al. 2019).

Do stable, natural polyploids also show systematic trends in chromosome positioning? Apparently some do, with one subgenome preferentially occupying the nuclear periphery and the other occupying the center (Bennett and Bennet 1992; Han et al. 2015). It has also been reported that chromosomes of natural polyploid wheats show centromere associations earlier in the cell cycle than their diploid progenitors, suggesting that ploidy affects chromosome configuration (Martinez-Perez et al. 2000). There are still many unresolved questions regarding interphase chromosome organization in plants. Are separate (sub)genome territories typical of established

polyploids? Are TADs typically conserved from progenitors to polyploid? Do chromosome configurations change upon WGD, and if so, what factors influence this? Do all plants have a KNOT, and if so, do the KEEs/IHIs of polyploid subgenomes mutually interact in a single network?

We conducted an exploratory study of *Brachypodium* Hi-C data to answer some of these questions. We focused on large-scale chromosome positioning, partly due to logistical constraints and partly due to constraints in the resolution of our Hi-C data. In the course of our survey of large-scale chromosome configurations, we discovered KNOT-like loci in diploid and polyploid *Brachypodium*. We leveraged a high-quality TE annotation to characterize TEs in the *B. distachyon* KNOT. We found several features that are consistent with the transposon trap hypothesis. However, given the limitations of this study, whether the *Brachypodium* KNOT is a functional transposon trap remains an open question.

Methods

Hi-C library preparation

Plants were grown in a growth chamber in short-day conditions and harvested at two to three weeks. Hi-C samples were prepared using the Arima Hi-C kit (Cat. no. A510008) according to the manufacturer's instructions. Briefly: fresh tissue was diced and crosslinked using formaldehyde under vacuum pressure. Nuclei were isolated from frozen, cross-linked tissue using the Sigma-Aldrich CelLytic Plant Nuclei Isolation/Extraction Kit according to the manufacturer’s instructions. Cross-linked nuclei were treated with two restriction enzymes with cut sites ^GATC and G^ANTC, where N can be any of the 4 nucleotides. Cut sites were filled in with dNTPs including a biotinylated nucleotide, and the spatially proximal digested ends were ligated with additional kit enzymes. Crosslinking was reversed by incubation at 68C. Libraries were prepared with the Roche KAPA Hyper Prep Kit, with specific modifications described in the Arima Hi-C kit documentation (A160139 v00). Samples were sequenced on the NovaSeqS4 platform to produce paired, stranded, 150bp reads.

Hi-C data analysis

Hi-C reads were trimmed and filtered to remove low-quality and adapter sequences using BBDuk (https://sourceforge.net/projects/bbmap/) (k=23 ktrim=r mink=11 hdist=1 trimq=18 maq=15 minlength=18 tpe tbo). Hi-C data analysis was performed using the HiC-Pro pipeline (Servant et al. 2015) (https://github.com/nservant/HiC-Pro). Contact matrices from HiC-Pro were converted to .hic format using the HiC-Pro utility hicpro2juicebox.sh and were visualized in Juicebox (Durand, Robinson, et al. 2016) (https://github.com/aidenlab/Juicebox). Default HiC-Pro parameters were used, except that the minimum MAPQ score was set to 20 and the minimum distance for cis contacts was set to 1000 bp since we were using multiple restriction enzymes and anticipated many short-range contacts from unligated fragments. We employed no hard and fast rules for determining whether a library passed QC; rather, the cis:trans ratio, mapping rate, overall depth, number of duplicates, and relative orientations of read pairs were considered in light of downstream goals. Visual inspection of the contact map at low resolution was often informative even if the library was of mediocre quality. A more detailed description of

this process can be found on github (https://github.com/vtartaglio/notebook/HiCPro_README.txt).

ABR113 TADs were visually identified in Juicebox and their coordinates were recorded in a bedpe-formatted file. The bedpe file was then layered onto a diploid contact map as an annotation layer, and overlap between the annotation and the putative TAD in the contact map was scored as described in the text.

Whole-chromosome interaction frequencies were calculated from matrices normalized via the ICE method (Imakaev et al. 2012) as implemented in HiC-Pro. We arbitrarily selected a bin size of 150kb. For a given chromosome pair i, j, we calculated the interaction frequency as the number of reads between chromosomes i and j / (lengthi + lengthj) / 1,000,000.

TE and rDNA identification

TE annotation of the *Brachypodium* genomes was performed as described in chapter 2 of this dissertation. LTR-retrotransposon (LTR-RT) insertion times were also calculated as described in chapter 2. TE density was simply the number of TEs per 200kb bin. TE diversity was calculated as the number of families or classes per bin divided by the total number of TEs per bin. Our annotation consisted of 17 possible classes based on the Wicker classification scheme (Wicker et al. 2007). Individual TE copies considered "invading" TEs were those that come from a family with more than 5 members and at least 90% of copies on one subgenome, yet the copy of interest is located on the other subgenome.

Permutation tests were conducted with the regioneR R package (Gel et al. 2016), where the KEEs were Region set A and the rest of the genome was Region set B. When testing for a significant difference in average LTR-RT insertion time between KEE and non-KEE genomic bins, we masked bins that did not contain any full-length LTR-RTs. All permutation tests in this study used 1,000 trials.

To search for rDNA in the KEE/IHI loci, ABR113 Illumina reads (JGI Genome Portal, Project ID 1005164, Proposal ID 277/300658; https://genome.jgi.doe.gov/portal/BrahybStandDraft_FD/BrahybStandDraft_FD.info.html) were aligned to published rDNA sequences from *Brachypodium distachyon* and *Oryza sativa* (Takaiwa et al. 1984; Takaiwa et al. 1985; Takaiwa et al. 1990; Hsiao et al. 1994; Kim and Nahm 1995) using bwa (Li and Durbin 2009). Reads containing rDNA were mapped to the ABR113 reference genome, and read pile-ups were visually inspected with CoGe (https://genomevolution.org/coge/) and also with custom R scripts.

Small RNA-Seq library preparation

Small RNA-seq (smRNA-seq) experiments were performed on old and young flowers from *Brachypodium distachyon* Bd21. smRNA-seq data for *B. distachyon* Bd21 leaves were obtained courtesy of Lifeng Liu. For leaf tissue collection, plants were grown in short-day conditions (26C 10h light, 18C 14h dark). Tissues were harvested at 2 or 3 weeks old (three samples from each were used in this study). For floral tissue collection, plants were grown in a growth chamber in long-day conditions (26C 18h light, 18C 6h dark). We collected the bottom half of an unfertilized floret: bottom of palea and lemma, plus the stigma, anthers, and rachilla. The tip of the floret was snipped off to remove the awn and to enrich for reproductive tissue. For "old flowers", the palea was the same length as the lemma, but the anthers had not yet dehisced. For "young flowers", the palea was shorter than the lemma, about 1/2 or 3/4 the lemma length.

We obtained two high-quality libraries for young flowers and three high-quality libraries for old flowers.

RNA was extracted and purified using the miRNEasy kit (Qiagen). 1 ug of total RNA retaining small RNA was used to make small RNA library using the Truseq Small RNA Library Prep Kit (Illumina). RNA was ligated with 3' RNA adapters followed by 5' RNA adapters. First strand was generated using SSII (Invitrogen) followed by 11 cycles of PCR to generate a double stranded cDNA library. cDNA library was size selected for 118-153 bp using the Pippin Prep or Pippin HT instrument (Sage Science). The prepared libraries were then quantified using KAPA Illumina library quantification kit (Roche) and run on a LightCycler 480 real-time PCR instrument (Roche). Leaf libraries were sequenced on a HiSeq-2500, and flower libraries were sequenced on a NovaSeq S4 sequencer. Reads were paired, stranded, 100 (leaf) or 150 (flower) base pair reads.

Small RNA-Seq data analysis

Raw reads were filtered and trimmed using BBDuk (https://sourceforge.net/projects/bbmap/) (rna=t trimfragadapter=t qtrim=r trimq=6 maxns=1 maq=10 minlen=17 mlf=0 khist=t trimk=23 mink=3 hdist=1 hdist2=1 ktrim=r sketch mito chloro taxlevel=species). Next, monocot rRNA, snoRNA, and tRNA sequences were downloaded from Rfam (Griffiths-Jones et al. 2003) using the "Entry type" search, and monocot sequences were selected from relevant fasta files using a custom script. In a typical smRNA analysis, smRNAs homologous to TEs may be discarded as well, but in our case, these were the smRNAs of interest. We then used BBmap (vslow perfectmode) to map our smRNA-seq reads to rRNA, snoRNA, and tRNA, and mapped reads were discarded. We then mapped the filtered library to the reference genome using ShortStack (Axtell 2013), retaining multi-mapping reads and allowing multi-mapped reads to contribute to the densities that guide read placement (--mmap f, –ranmax 1000, --bowtie_m all, --nohp).

Synteny

Syntenic blocks between the subgenomes of ABR113 were obtained using MCScanX (Wang et al. 2012) as described here: https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version). The significance of overlaps between syntenic blocks containing KEEs/IHIs was evaluated with regioneR (Gel et al. 2016) using numOverlaps as the evaluation function. Region set A was the set of blocks containing a KEE/IHI in the D subgenome, and Region set B was the set of blocks containing a KEE/IHI in the S subgenome. All permutation tests in this study used 1,000 trials.

Possible segmental duplications in B. distachyon Bd21 were investigated using the Plant Genome Duplication Database (Lee et al. 2017) (http://chibba.agtec.uga.edu/duplication/index/dotplot). Intragenomic syntenic blocks were manually examined to determine whether any linked one Bd21 KEE/IHI to another.

Results

We began our study of chromatin topology in *Brachypodium* by performing Hi-C on leaves collected from the *B. hybridum* reference line ABR113. This experiment yielded 95

million valid Hi-C contacts. The cis to trans ratio (that is, the ratio of intrachromosomal to interchromosomal contacts,) was 1.9. Because cis contacts are expected to be more common than trans contacts, the cis to trans ratio should be greater than 1 in a high-quality library (Lajoie et al. 2015). Later Hi-C experiments yielded as few as 4 million contacts to as many 178 million contacts, and cis to trans ratios ranged from less than 1 to more than 4. All Hi-C experiments in this study were performed on leaves. Several attempts were made to perform Hi-C on *B. hybridum* Bhyb26. Oddly, three attempts were made by our team, and one attempt was made at the HudsonAlpha Institute, to obtain high-quality Hi-C data from Bhyb26 leaves, but the QC outcomes were always poor. The Bhyb26 Hi-C data were of sufficient quality to visually assess whole-chromosome interaction patterns, but we did not attempt quantitative analysis of the Bhyb26 contact matrix.

Visualization of the ABR113 contact map in Juicebox revealed several intriguing features of the chromatin landscape. First, we noticed that the S subgenome showed an anti-diagonal pattern that seems to indicate frequent interactions between chromosome arms (Fig. 3.1A). This would be consistent with the possibility that the S subgenome spends more time than the D subgenome in the chromosome configuration known as the Rabl configuration, which is characterized by centromere interactions at one end of the nucleus and telomere interactions at the other. At the very least, the pattern suggests strong interactions between arms of individual chromosomes. To determine whether this phenomenon was reproducible, we repeated the experiment. Though the second experiment yielded a lower-quality library than the first experiment, the same anti-diagonal pattern was visible for the S subgenome but not the D subgenome. To determine whether this chromosome configuration may have been inherited from the respective progenitors, we performed Hi-C on the diploids *B. distachyon* Bd21 and *B. stacei* ABR114. Surprisingly, *B. stacei* ABR114 did not show an anti-diagonal pattern, even though this line is thought to be a close relative of the true progenitor of B. hybridum ABR113 (Fig 3.1B). Despite technical difficulties, *B. hybridum* Bhyb26 yielded data of sufficient quality for this gross level of observation. Surprisingly, Bhyb26 showed the same pattern as ABR113, in spite of the fact that its *B. stacei* progenitor is expected to be quite diverged from that of ABR113 (Gordon et al. 2020). This suggests, intriguingly, that the chromosome configuration of the S subgenome may have changed as a consequence of polyploidy.

Another interesting feature of the *B. hybridum* contact map was the presence of TADs. Given that TADs are not a prominent feature of the *Arabidopsis* 3D genome (Wang et al. 2015), which is a compact plant genome like *Brachypodium*, this result was somewhat surprising, though not unexpected. To scan for large-scale chromatin conformation differences between diploid and polyploid, I visually identified large, obvious TADs in ABR113 and then searched for TADs in the corresponding progenitor region. In general, agreement between each subgenome and its corresponding progenitors was high (Fig. 3.1C), although some regions were more ambiguous than others. This was in part due to errors in the ABR113 and especially the ABR114 assemblies (Fig. 3.1D). I overlayed the 46 TAD annotations from ABR113 onto the progenitor contact maps, using ABR113 genomic coordinates, and scored the overlap between each diploid TAD and the lifted-over region as either 'Good', indicating strong TAD conservation, which included cases where the TAD was somewhat shifted due to slight differences in chromosome size between progenitor and subgenome; 'Mediocre', indicating little TAD-like signal in the region where the TAD was expected; or 'Poor', indicating the clear absence of a TAD in the region where a TAD was expected. None of the TADs scored as 'Poor'. 79% of the *B. distachyon* TADs scored as 'good', while this was 89% in *B. stacei*. This hints at

the possibility of greater Bs-BhS than Bd-BhD TAD conservation, though this experiment was too subjective to support any substantive claims. At any rate, it appears that the majority of large TADs are conserved between polyploid and subgenome.

We investigated whether DNA-DNA interactions within subgenomes were more frequent than those between subgenomes. We calculated interchromosomal interaction frequencies (IFs) for all distinct chromosome combinations in ABR113. The IFs were simply the total number of contacts between two given chromosomes, divided by the total length of those chromosomes (Concia et al. 2020; Jia et al. 2021). Overall, we did not find convincing evidence for reduced contacts between D and S chromosomes, relative to D-D or S-S chromosome pairs (Fig. 3.2A). Chromosome BhD5, the only chromosome of the D subgenome that is comparable in size to those of the S subgenome, did interact with the S subgenome's chromosomes at a lower frequency than any S-S chromosome pair (Fig. 3.2B). This may indeed be due to a "genome territories"-like phenomenon. However, since the remaining inter-subgenome interactions show about as many contacts as we would expect for their size, we are not convinced that ABR113 is systematically depleted for inter-subgenome interactions. Surprisingly, interchromosomal interactions within the D subgenome were significantly more frequent than interchromosomal interactions within the S subgenome (t-test, p=0.0013). Our conclusion that ABR113 subgenomes do not segregate in the nucleus is consistent with the contact maps, which do not show enrichment of intra-subgenome contacts by visual inspection.

A final notable feature of the ABR113 contact map was a series of interaction hotspots that form a single network of mutual DNA-DNA interactions (Fig. 3.3A-C). While the spatio-temporal dynamics of these chromatin interactions remain unclear (Fig. 3.3B), these regions of the genome at least contact one another at an elevated frequency. We identified 13 such hotspots in the *B. hybridum* genome, and they range in size from 200kb-3Mb. Their chromosomal locations varied; those in the S subgenome were often in or near the telomeres, while those in the D subgenome were scattered throughout. The interaction strength among them also varied (Supplementary Data File 1). All had at least one strong interaction with another hotspot. Most hotspots had several strong interactions with several other hotspots, and medium or weak interactions with the rest. Interestingly, several of them occurred as pairs of adjacent hotspots (Fig. 3.3A). Some of them were clear TADs on the cis contact matrices, though many were not (Fig. 3.3C). We suspected that these loci might contain rDNA, since the nucleolus is the most obvious genome feature with the propensity for long-range DNA-DNA contacts. However, after scanning the ABR113 genome for rDNA using published rDNA consensus sequences (see Methods), we found that the KEEs did not contain a significant number of hits to rDNA.

To validate whether these structures were analogous to the previously identified KEEs/IHIs, we investigated one of their signature features: signs of elevated small RNA (smRNA) production. Previously, KEEs/IHIs were found to have high numbers of “smRNA associated regions”, although the authors did not specify whether this was specific to one tissue type (Grob et al. 2014). These “associated regions” are essentially clusters of smRNA-seq reads. We performed smRNA-seq in young flowers and old flowers, and obtained smRNA-seq data from leaves, for *B. distachyon* Bd21. We compared KEEs to the genome average in terms of reads per million, average reads per smRNA cluster, and number of smRNA clusters, and calculated the p-value for randomly sampled genome tiles. While the former two metrics were not significantly different for the KEEs, the latter was significant (p=0.001, permutation test, alpha=0.017 after Bonferroni correction). Interestingly, the number of smRNA clusters was significantly greater for KEEs in old and young flowers (p=0.001 for both), though the difference

was not significant in leaves (Fig. 3.4A). This suggests that in flowers, the clusters present in the KEEs are not necessarily high depth, but there are more clusters than we would expect by random chance. However, it should be noted that read depth is not necessarily particularly informative when mapping small RNAs to the genome, since the reads are spread evenly across all hit sites. Thus, even though we cannot distinguish between highly- and low-targeted sites, or between smRNA template and target, our result is consistent with the "transposon trap" model, in which many templates for smRNA production are in close proximity to one another.

Hi-C in the diploid progenitors *B. distachyon* and *B. stacei* revealed more or less the same interaction pattern among the same loci. One KEE/IHI from the S subgenome was not well-supported in *B. stacei*, but the remaining 12 of the 13 *B. hybridum* KEEs/IHIs were visible in the corresponding location in the diploid progenitor. The KEEs/IHIs seem unlikely to be an assembly artifact, as all three contact maps were based on different reference genomes. The Bhyb26 contact maps did not reveal KEEs/IHIs. Whether this is because Bhyb26 truly lacks KEEs/IHIs or because the libraries were too low-quality to discern them remains unclear.

We conducted a synteny analysis with MCScanX to identify syntenic blocks between the D subgenome/*B. distachyon* and the S subgenome/*B. stacei*. (Because the polyploid bears so little rearrangement, the synteny relationships between the subgenomes and between the diploids are practically identical (Gordon et al. 2020)). Remarkably, all the *B. distachyon*/D subgenome KEEs/IHIs are contained within syntenic blocks that map to *B. stacei*/S subgenome regions that also contain at least one KEE/IHI. In other words, the location of the KEEs appears to be conserved between *B. distachyon* and *B. stacei*, despite 11 million years of evolution since their last common ancestor (Gordon et al. 2020). We ran a permutation test (see Methods), and found that the overlap between syntenic blocks containing a KEE/IHI between the subgenomes was significant ($p=0.002$).

Given the surprising discovery that the locations of the Brachypodium KEEs/IHIs are fairly well-conserved among *Brachypodium* species, we attempted to determine whether the conservation of *Brachypodium* KEE/IHI locations extended to rice. Dong et al. (2018) discovered KEEs/IHIs in rice, although their search criteria were quite different than those of Grob et al. (2014) or Feng et al. (2014)—they relied on epigenetic marks rather than visualization to call the loci, although they argue that their KEEs/IHIs also form hotspots on the contact map (Dong et al. 2018). We obtained the coordinates of their KEEs/IHIs in the *O. sativa* MSU v.7 genome from the authors, but we found that the vast majority of them did not form hotspots on the contact map we obtained from running their raw data through our pipeline. Those *O. sativa* loci that did form hotspots with a few other KEEs/IHIs were clearly not syntenic to the *Brachypodium* KEEs/IHIs.

If the KEEs/IHIs are functional elements of a "transposon trap", the TEs within them may be expected to be distinct in their size, age, or other properties. We found little evidence that this is the case. We split up the *B. distachyon* Bd21 genome into 200kb bins and recorded various properties of the TEs in each bin. We then used regioneR (Gel et al. 2016) to perform permutation tests against random samples of non-KEE genomic bins. Bins within KEEs/IHIs were not significantly different from the genome average in terms of TE density, TE diversity by class, TE length, LTR-RT length, or LTR-RT age (counting only those bins that contain at least one full-length LTR-RT). However, two metrics were significant: TEs from KEE bins were from smaller families than we would expect based on random chance, and they were significantly more diverse than the genome average in terms of TE family ($p=0.001$ for both, $alpha=0.00625$ after Bonferroni correction). The mean diversity by family in the KEE bins is 0.804, while for

non-KEE bins the average is 0.767. The mean family size for KEE bins is 549 members, while for non-KEE bins it is 644 members. That the KEE/IHI TEs are diverse and from small families has intriguing implications for the transposon trap hypothesis.

Comparing the KEEs/IHIs against the whole genome would not necessarily tell us if the KEEs/IHIs are, for example, TE-rich "islands" relative to the surrounding regions. Therefore, we also compared the KEEs to flanking regions, in this case, the 5Mb on either side of each KEE/IHI. Since bins on the centromere side might be expected to have distinct TE characteristics from bins on the telomere side, we examined the KEEs versus each flank separately, with the centromere-facing flank being dubbed “inner” and the telomere-facing flank dubbed “outer” (Fig. 3.4A). Mean family size of KEE bins was significantly different from outer but not inner flanking bins ($p=0.001$). TE family diversity was higher in KEE bins than in either flank (inner: $p=0.005$, outer, $p=0.001$). TE number per bin, that is, TE density, was not significantly different between KEEs/IHIs and either flank. Together, these data suggest that the KEEs are “islands” in terms of TE family diversity, and perhaps also in family size, but they are not “heterochromatic islands” in the sense of being TE-dense.

If the *Brachypodium* KNOT is a functional transposon trap, we might expect that perhaps the KEEs/IHIs in ABR113 may protect each subgenome from “invasion” by TEs of the other subgenome. We recorded “invading” TEs across the ABR113 genome, namely, those that may have spread from one subgenome to the other post-WGD (see Methods). However, we found no evidence that “invading” TEs are more common in KEEs/IHIs than elsewhere in the genome. On the contrary, according to our permutation test, the KEEs were significantly depleted for such TEs, as KEE bins had on average 0.20 “invaders”, while non-KEE bins had an average of 0.41.

Finally, Zanni et al. (2013) found evidence that the *flamenco* locus may actually appear twice in the Drosophila genome due to a segmental duplication. This is intriguing because Grob et al. (2014) noted that some of the KEEs occur in adjacent pairs, and we observed this in *Brachypodium* as well. We searched for evidence of segmental duplications of the KEEs by visually assessing dot plots and intragenomic syntenic blocks created through the Plant Genome Duplication Database (Lee et al. 2017). However, we found no evidence that our KEEs or any substantial portion of them were duplicated in the Bd21 genome.

Discussion

Hi-C is a promising new technology for the investigation of the effects of polyploidization on chromatin and chromosomes. In this exploratory study, we identified some advantages and some challenges of the application of Hi-C to diploid and polyploid plants. We found some evidence that chromosome configuration changed upon polyploidization. On the other hand, large TADs were mostly conserved between corresponding polyploid and progenitor chromosomes. The chromosomes of the subgenomes did not appear to systematically segregate into “genome territories”. Finally, we observed a KNOT-like structure whose location is conserved and which harbors diverse TEs. In the polyploid, the two subgenomic KNOTS form a single interaction network.

Overall, the progenitor and polyploid chromosomes showed similar large-scale chromatin topology and high conservation of large TADs. Our survey was rudimentary, and a far better experiment would have been to repeat the visual analysis with many researchers scoring unlabeled chromosomes, or to use a TAD calling software. However, both these possibilities would have required more time and effort than was available.

The S subgenome of both polyploids shows antidiagonals that are typically interpreted as the Rabl configuration (Dudchenko et al. 2017; Mascher et al. 2017; Jia et al. 2021), while the corresponding progenitor does not. This finding is exciting, though slightly puzzling. The Rabl configuration is thought to be characteristic of large chromosomes and large genomes (Dong and Jiang 1998), so it is somewhat unexpected that the smaller chromosomes of the S subgenome should assume this configuration. It is also surprising that one subgenome should assume the Rabl configuration while the other does not, as we are not familiar with a case where this has been observed in a polyploid plant. Idziak-Helmcke et al. (2015) examined *B. stacei*, *B. distachyon*, and *B. hybridum* root tip cells, and found that only *B. distachyon* took the Rabl arrangement. They also examined *B. distachyon* leaves and found that those cells' chromosomes were non-Rabl. Therefore, the only experiment we have in common is *B. distachyon* leaves, and in that case our results agree. Taken together, all these results suggest that *B. distachyon* is Rabl only in roots, *B. hybridum* is Rabl only in leaves and only for the S subgenome, and *B. stacei* is non-Rabl. This tissue-specific variation in chromosome configuration is not unprecedented (Prieto et al. 2004). It may be due to differences in centromere association, how mitotically active the tissue is, or in timing of endoreduplication (Cowan et al. 2001; Santos and Shaw 2004). It is possible that polyploidization induced changes in one or more of these factors, and thus the previously non-Rabl S chromosomes adopted a greater propensity for the Rabl configuration post-polyploidization. We attempted a pilot Hi-C experiment on ABR113 root tips, but the QC results were poor. An attempt was also made to flow-sort nuclei prior to Hi-C to control for ploidy differences due to endoreduplication, but sufficient material was not obtained and was not likely to be obtained in future attempts.

Two important caveats to our analysis must be noted. First, Hi-C is an aggregate of many tissue types and stages of the cell cycle, unlike the cytogenetic approaches of Idziak-Helmcke et al. (2015). Second, there is no rigorous methodology for distinguishing the Rabl configuration from, e.g., a rosette or some other configuration, with Hi-C data. We do not believe that the Xs are a technical artifact, since we observed them in several polyploid samples (two ABR113 replicates and a Bhyb26 attempt) with independently assembled genomes. However, our interpretation could be flawed, since it is far from clear how chromosome configurations of individual cells are translated into the crude averages delivered by Hi-C.

We found that inter-subgenome interactions are quite common in ABR113. The spatial separation of subgenomes has been observed in a number of unstable polyploids and some established polyploids (Vimala and Lavania 2021). In unstable polyploids, this phenomenon appears to be related to epigenetic differences between the progenitors with consequences for kinetochore assembly (Sanei et al. 2011; Marimuthu et al. 2021). Among plants and animals, it is thought that chromosome positioning depends to some extent on chromosome size, though there are exceptions to this rule (Parada and Misteli 2002). We hypothesized that *B. hybridum* might show 'subgenome territories' because it is a wide cross and because its subgenomes have chromosomes of very different sizes. Thus, the lack of subgenome separation in *B. hybridum* was surprising. Apparently the two progenitors, while genetically distinct, are sufficiently epigenetically similar that the kinetochores assemble with similar efficiency. Meanwhile, our study shows, as others have before, that chromosome size alone does not necessarily predict chromosome positioning.

According to the "transposon trap" hypothesis in animals, piRNA clusters are regions of the genome that are extremely TE-rich and serve as templates for piRNA production. Those piRNAs, complexed with PIWI proteins, then promote post-transcriptional silencing of TEs in

the germline and gonadal somatic cells (Siomi et al. 2011). Grob et al. (2014) went on to speculate that the 3D topology of transposon trap loci may be somehow conducive to TE insertion, which would explain why several *Drosophila* piRNA clusters and their own KNOT formed distinctive interaction patterns in the Hi-C data.

We found that several features of the *Brachypodium* KNOT are consistent with this hypothesis. The *Brachypodium* KNOT is apparently well-conserved in terms of its location, despite substantial karyotypic rearrangement since the divergence of *B. stacei* and *B. distachyon*, which we take as circumstantial evidence that these loci may serve some useful function. It is true that we did not identify a KNOT in *B. hybridum* Bhyb26, but we suspect this is an artifact of low sequencing depth and poor library quality. Indeed, another research team failed to find a *B. distachyon* KNOT with low-depth Hi-C data (Anne Roulin, pers. comm.), and a low-depth replicate of our ABR113 Hi-C showed KEEs/IHIs only in cis, but not in trans. Based on this, we conclude that high sequencing quality and depth are needed to see KEEs/IHIs. The KEEs/IHIs in *B. distachyon* are enriched for smRNA clusters, but only in flowers, while *flamenco* (*flam*) appears to serve its TE defense functions only in reproductive cells (Coline et al. 2014). Finally, TEs in the *Brachypodium* KNOT are diverse and come from smaller-than-average families. Zanni et al. (2013) noted that TEs in *flam* were particularly diverse. While Zanni et al. (2013) did not compare *flam* TEs to genome-wide TEs, the transposon trap hypothesis suggests that TEs in the trap would not come from large families, since proliferation would be halted as soon as a TE lands in the trap.

On the other hand, some of our results were not entirely consistent with the transposon trap hypothesis. For one thing, TE density was not particularly high in the *Brachypodium* KEEs, in contrast to the quantitative and qualitative observations of Feng et al. (2014) and Grob et al. (2014), respectively, in *Arabidopsis*. Also, Zanni et al. (2013) found that *flam* TEs were largely recent insertions. It was difficult to estimate insertion times of TEs in our KEEs since full-length LTR-RTs are few and far between in *Brachypodium* (see chapter 2), but our KEEs were not particularly long, as one would expect recent TE insertions to be. We also failed to find an element that "invaded" an ABR113 subgenome post-WGD and was "trapped" by a KEE, like the *Pifo* element described by Zanni et al. (2013). However, our failure to find such an element may easily be due to two facts: first, there has not been much post-WGD TE activity in ABR113 (see chapter 2), and second, when the entire subgenome was imported into the nucleus during the WGD, that subgenome's TE surveillance machinery was imported, too.

In summary, we uncovered some evidence that the *Brachypodium* KEEs may be a functional transposon trap, though our findings are somewhat mixed, and we did not have the resources to check for a "smoking gun". There is no publicly available, well-characterized mutant collection for *Brachypodium* that was created via transposon mutagenesis, which would reveal an elevated propensity for TE insertions in the KEEs, as it did for Grob et al. (2014) in *Arabidopsis*. Furthermore, many of the findings in this study demand confirmation with molecular cytogenetic approaches. I had intended to use such approaches to validate the *Brachypodium* KEE interactions, but such an effort was ambitious to begin with, and then was disrupted by the COVID-19 pandemic. Cytogenetic approaches remain the "gold standard" for understanding chromatin dynamics in polyploids. Nevertheless, Hi-C is a relatively accessible way to obtain a broad overview of chromatin topology and chromosome interactions. This study contributes to a growing body of work that explores how this new type of data may be harnessed to reliably and reproducibly reveal polyploid chromatin dynamics.

Data availability

ABR113 Hi-C data are available on the JGI genome portal under Proposal ID: 503504 and Sequencing Project ID: 1264506 or at this link: https://genome.jgi.doe.gov/portal/BrahybAABR113HiC_FD/BrahybAABR113HiC_FD.info.html

Unfortunately, the highest quality ABR113 Hi-C library, CHCBS, is not available on the JGI genome portal, nor are the Bhyb26 Hi-C data. Interested researchers may contact Virginia T. Scarlett or John P. Vogel for these data.

Flower smRNA-seq data are available on the JGI genome portal under Proposal ID: 503504 and Sequencing Project ID: 1297379 or via this link: https://genome.jgi.doe.gov/portal/BrahybStandDraft_11_FD/BrahybStandDraft_11_FD.info.html

Leaf smRNA-seq data are available on the JGI genome portal under Proposal ID: 2856 and Sequencing Project ID: 1105443 or via this link: https://genome.jgi.doe.gov/portal/BraEncnscriptome_FD/BraEncnscriptome_FD.info.html

The smRNA libraries from the above *Brachypodium* ENCODE transcriptome project that were used in this study are BCNUX 0d_cold-1, BCNUZ 0d_cold-2, BCNUN 0d_cold-3, BCNUS 7dNonCold-1, BCNTU 7dNonCold-2, and BCNTX 7dNonCold-3.

Chapter 3 Figures

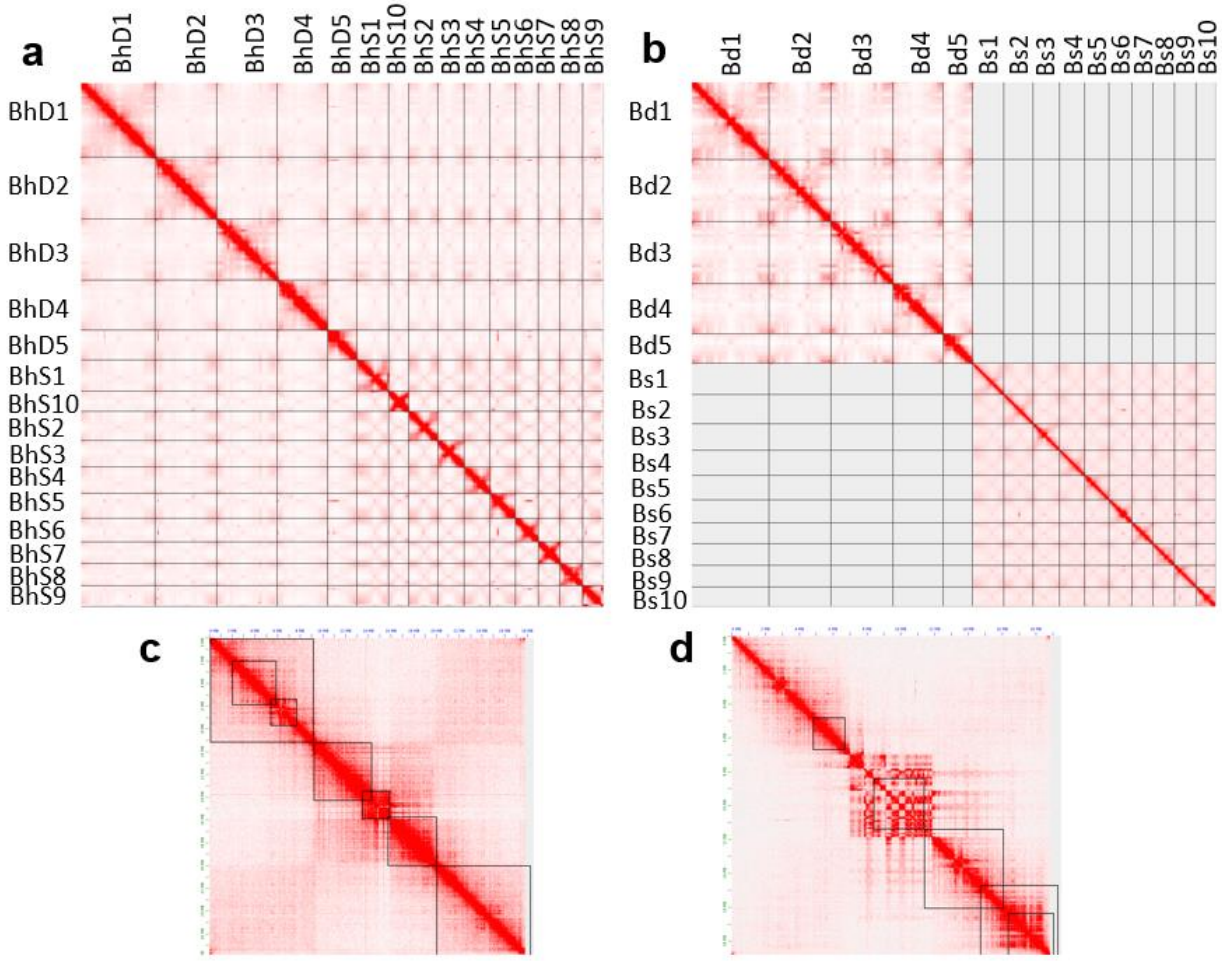

Fig. 3.1. Contact maps and chromosome conformation. (a) Hi-C contact map for ABR113 leaves. Note the X's characteristic of the Rabl configuration in the S subgenome. (b) Contact maps for Bd21 and ABR114 leaves, created separately but visualized together. (c) TADs from ABR113 BhS2 (annotated boxes) overlaid on ABR114 Bs2 show reasonable TAD conservation, despite slight differences in chromosome size. (d) ABR113 BhS10 TADs lifted over to Bs10 are difficult to compare due assembly quality.

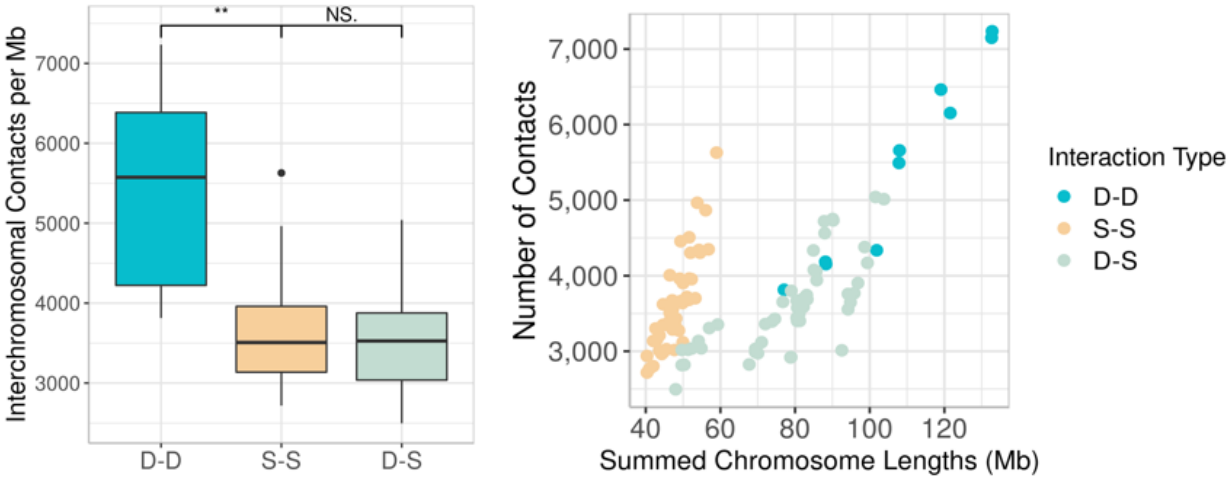

Fig. 3.2. Whole-chromosome interaction frequencies. (a) Whole-chromosome interaction frequencies for all possible combinations of ABR113 chromosomes. Interactions between BhD chromosomes are more frequent than between BhS chromosomes, and interactions between subgenomes are not uncommon. (b) Whole-chromosome interaction frequencies as a function of chromosome length. Each point represents a chromosome pair.

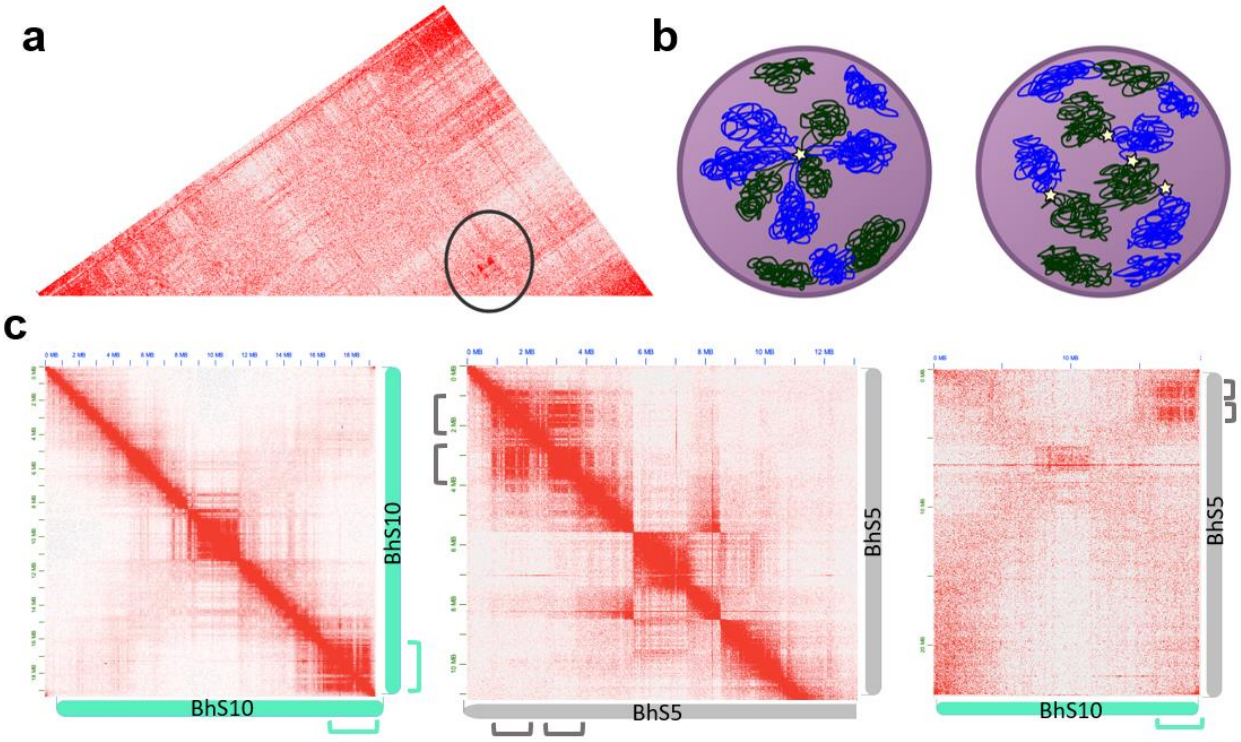

Fig 3.3. KNOT-like interactions in *Brachypodium*. (a) Interchromosomal Hi-C contact map for a pair of Bd21 chromosomes showing KEE/IHI-like interaction hotspots. (b) Hypotheses for putative KEE interaction patterns, showing that mutual interactions may occur simultaneously, pairwise, or in some other combination. Blue and green squiggles indicate chromosomes, and stars indicate a KEE-KEE interaction. If this were a *B. hybridum* nucleus, we may imagine that the blue chromosomes are BhD chromosomes and the green are BhS, for example. (c) Some but not all putative KEEs are themselves TADs. Left and middle, intra chromosomal contact maps for three putative KEEs: one on ABR113 BhS10 (green), and two on ABR113 BhS5 (gray) (marked with brackets at margins). Right, interchromosomal contact map of the two chromosomes.

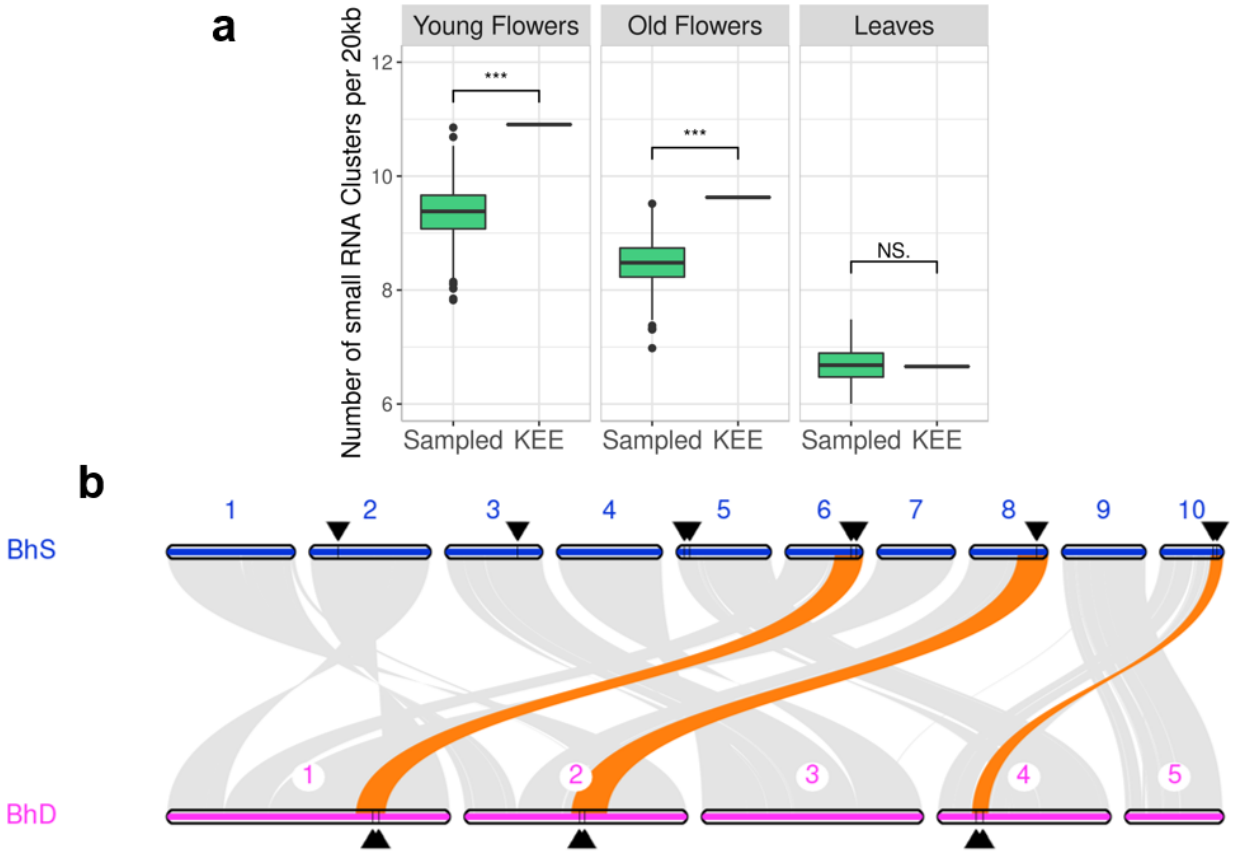

Fig 3.4. Additional features of the *Brachypodium* KNOT. (a) smRNA-seq in Bd21 reveals that putative KEEs have a greater number of smRNA clusters than randomly sampled genome tiles of the same size. However, this trend in smRNA pattern is tissue-specific. (b) Riparian plot showing syntenic blocks between ABR113 subgenomes reveals that many putative KEEs are syntenic between subgenomes. Black triangles indicate KEE locations. Orange highlights indicate syntenic blocks that contain a KEE in both subgenomes.

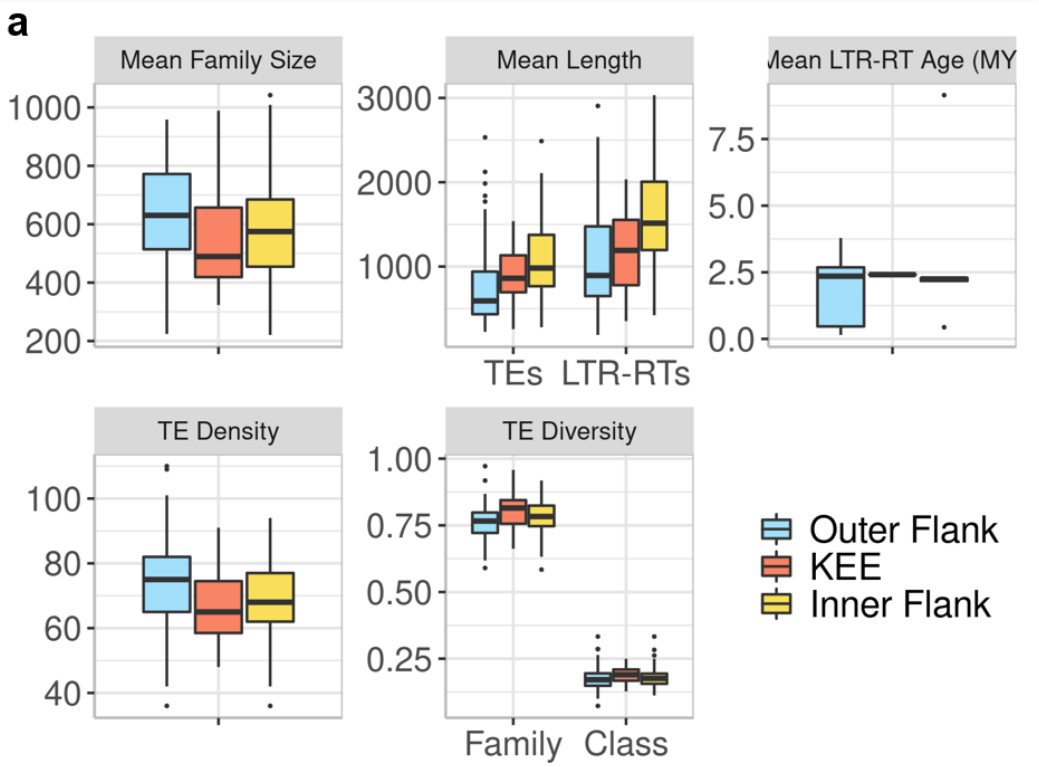

Fig 3.4. TE characteristics of the *Brachypodium* KNOT. (a) Various TE characteristics for KEEs and the 5Mb flanking the KEEs on either side. "Outer flank" refers to the 5Mb on the telomere-facing side, while "inner flank" refers to the 5Mb on the centromere-facing side. Non-KEE regions between side-by-side KEEs were excluded.

Bibliography

Abranches R, Beven AF, Aragón-Alcaide L, Shaw PJ. 1998. Transcription Sites Are Not Correlated with Chromosome Territories in Wheat Nuclei. J Cell Biol. 143(1):5–12.

Acemel RD, Maeso I, Gómez-Skarmeta JL. 2017. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. Wiley Interdiscip Rev Dev Biol. 6(3). doi:10.1002/wdev.265.

Adams KL, Cronn R, Percifield R, Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. PNAS. 100(8):4649–4654. doi:10.1073/pnas.0630618100.

Ågren JA, Huang H-R, Wright SI. 2016. Transposable element evolution in the allotetraploid Capsella bursa-pastoris. Am J Bot. 103(7):1197–1202. doi:10.3732/ajb.1600103.

Alger EI, Edger PP. 2020. One subgenome to rule them all: underlying mechanisms of subgenome dominance. Current Opinion in Plant Biology. 54:108–113. doi:10.1016/j.pbi.2020.03.004.

An Z, Tang Z, Ma B, Mason AS, Guo Y, Yin J, Gao C, Wei L, Li J, Fu D. 2014. Transposon variation by order during allopolyploidisation between Brassica oleracea and Brassica rapa. Plant Biol (Stuttg). 16(4):825–835. doi:10.1111/plb.12121.

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 31(2):166–169. doi:10.1093/bioinformatics/btu638.

Axtell MJ. 2013. ShortStack: comprehensive annotation and quantification of small RNA genes. RNA. 19(6):740–751. doi:10.1261/rna.035279.112.

Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. 2018. The “Polyploid Hop”: Shifting Challenges and Opportunities Over the Evolutionary Lifespan of Genome Duplications. Frontiers in Ecology and Evolution. 6:117. doi:10.3389/fevo.2018.00117.

Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V. 2019. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. Nat Commun. 10(1):5818. doi:10.1038/s41467-019-13730-0.

Bailly-Bechet M, Haudry A, Lerat E. 2014. “One code to find them all”: a perl tool to conveniently parse RepeatMasker output files. Mobile DNA. 5(1):13. doi:10.1186/1759-8753-5-13.

Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. 2016. On the relative abundance of autopolyploids and allopolyploids. New Phytologist. 210(2):391–398. doi:10.1111/nph.13698.

Ben-David S, Yaakov B, Kashkush K. 2013. Genome-wide analysis of short interspersed nuclear elements SINES revealed high sequence conservation, gene association and retrotranspositional activity in wheat. The Plant Journal. 76(2):201–210. doi:10.1111/tpj.12285.

Bennett ST, Bennet MD. 1992. Spatial Separation of Ancestral Genomes in the Wild Grass Milium montianum Parl. Annals of Botany. 70(2):111–118. doi:10.1093/oxfordjournals.aob.a088446.

Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, Leal-Bertioli SCM, Ren L, Farmer AD, Pandey MK, et al. 2019. The genome sequence of segmental allotetraploid peanut Arachis hypogaea. Nat Genet. 51(5):877–884. doi:10.1038/s41588-019-0405-z.

Bickmore WA, van Steensel B. 2013. Genome architecture: domain organization of interphase chromosomes. Cell. 152(6):1270–1284. doi:10.1016/j.cell.2013.02.001.

Birchler JA, Riddle NC, Auger DL, Veitia RA. 2005. Dosage balance in gene regulation: biological implications. Trends Genet. 21(4):219–226. doi:10.1016/j.tig.2005.02.010.

Bird KA, Niederhuth CE, Ou S, Gehan M, Pires JC, Xiong Z, VanBuren R, Edger PP. 2021. Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid Brassica napus. New Phytologist. 230(1):354–371. doi:10.1111/nph.17137.

Bird KA, VanBuren R, Puzey JR, Edger PP. 2018. The causes and consequences of subgenome dominance in hybrids and recent polyploids. New Phytologist. 220(1):87–93. doi:10.1111/nph.15256.

Blanc G, Wolfe KH. 2004. Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes[W]. The Plant Cell. 16(7):1667–1678. doi:10.1105/tpc.021345.

Bowers JE, Paterson AH. 2021. Chromosome number is key to longevity of polyploid lineages. New Phytologist. 231(1):19–28. doi:10.1111/nph.17361.

Bragg JN, Anderton A, Nieu R, Vogel JP. 2015. Brachypodium distachyon. In: Wang K, editor. Agrobacterium Protocols. Vol. 1223. New York, NY: Springer New York. (Methods in Molecular Biology). p. 17–33. [accessed 2022 Mar 31]. http://link.springer.com/10.1007/978-1-4939-1695-5_2.

Brubaker CL, Paterson AH, Wendel JF. 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome. 42(2):184–203. doi:10.1139/g98-118.

Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. 2012. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. Curr Biol. 22(3):248–252. doi:10.1016/j.cub.2011.12.027.

Buggs RJA, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE. 2014. The legacy of diploid progenitors in allopolyploid gene expression patterns. Philosophical Transactions of the Royal Society B: Biological Sciences. 369(1648):20130354. doi:10.1098/rstb.2013.0354.

Burns R, Mandáková T, Gunis J, Soto-Jiménez LM, Liu C, Lysak MA, Novikova PY, Nordborg M. 2021. Gradual evolution of allopolyploidy in Arabidopsis suecica. Nat Ecol Evol. 5(10):1367–1381. doi:10.1038/s41559-021-01525-w.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421. doi:10.1186/1471-2105-10-421.

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, et al. 2014. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. Plant Physiology. 164(2):513–524. doi:10.1104/pp.113.230144.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25(15):1972–1973. doi:10.1093/bioinformatics/btp348.

Catalán P, Chalhoub B, Chochois V, Garvin DF, Hasterok R, Manzaneda AJ, Mur LAJ, Pecchioni N, Rasmussen SK, Vogel JP, et al. 2014. Update on the genomics and basic biology of Brachypodium: International Brachypodium Initiative (IBI). Trends Plant Sci. 19(7):414–418. doi:10.1016/j.tplants.2014.05.002.

Catalán P, López-Álvarez D, Díaz-Pérez A, Sancho R, López-Herránz ML. 2016. Phylogeny and Evolution of the Genus Brachypodium. In: Vogel JP, editor. Genetics and Genomics of Brachypodium. Cham: Springer International Publishing. (Plant Genetics and Genomics: Crops and Models). p. 9–38. [accessed 2022 Jan 20]. https://doi.org/10.1007/7397_2015_17.

Catalán P, Müller J, Hasterok R, Jenkins G, Mur LAJ, Langdon T, Betekhtin A, Siwinska D, Pimentel M, López-Alvarez D. 2012. Evolution and taxonomic split of the model grass Brachypodium distachyon. Ann Bot. 109(2):385–405. doi:10.1093/aob/mcr294.

Cavalli G, Misteli T. 2013. Functional implications of genome topology. Nat Struct Mol Biol. 20(3):290–299. doi:10.1038/nsmb.2474.

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. 2014. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science. 345(6199):950–953. doi:10.1126/science.1253435.

Cheng F, Sun C, Wu J, Schnable J, Woodhouse MR, Liang J, Cai C, Freeling M, Wang X. 2016. Epigenetic regulation of subgenome dominance following whole genome triplication in Brassica rapa. New Phytol. 211(1):288–299. doi:10.1111/nph.13884.

Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X. 2018. Gene retention, fractionation and subgenome differences in polyploid plants. Nature Plants. 4(5):258–268. doi:10.1038/s41477-018-0136-7.

Chester M, Gallagher JP, Symonds VV, Silva AVC da, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid

species, Tragopogon miscellus (Asteraceae). PNAS. 109(4):1176–1181. doi:10.1073/pnas.1112041109.

Clark JW, Donoghue PCJ. 2018. Whole-Genome Duplication and Plant Macroevolution. Trends in Plant Science. 23(10):933–945. doi:10.1016/j.tplants.2018.07.006.

Clausen JL, Keck DD, Hiesey WM. 1945. Experimental Studies on the Nature of Species. II. Plant Evolution through Amphiploidy and Autoploidy, with Examples from the Madiinae. Carnegie Institution of Washington. 564.

Coate JE, Doyle JJ. 2010. Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. Genome Biol Evol. 2:534–546. doi:10.1093/gbe/evq038.

Coline G, Théron E, Brasset E, Vaury C. 2014. History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in Drosophila melanogaster. Frontiers in Genetics. 5. [accessed 2022 Mar 18]. https://www.frontiersin.org/article/10.3389/fgene.2014.00257.

Colle M, Leisner CP, Wai CM, Ou S, Bird KA, Wang J, Wisecaver JH, Yocca AE, Alger EI, Tang H, et al. 2019. Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. GigaScience. 8(3):giz012. doi:10.1093/gigascience/giz012.

Concia L, Veluchamy A, Ramirez-Prado JS, Martin-Ramirez A, Huang Y, Perez M, Domenichini S, Rodriguez Granados NY, Kim S, Blein T, et al. 2020. Wheat chromatin architecture is organized in genome territories and transcription factories. Genome Biology. 21(1):104. doi:10.1186/s13059-020-01998-1.

Cowan CR, Carlton PM, Cande WZ. 2001. The Polar Arrangement of Telomeres in Interphase and Meiosis. Rabl Organization and the Bouquet1. Plant Physiology. 125(2):532–538. doi:10.1104/pp.125.2.532.

Cremer T, Cremer C. 2006. Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. Eur J Histochem. 50(4):223–272.

Cremer T, Cremer M. 2010. Chromosome Territories. Cold Spring Harb Perspect Biol. 2(3):a003889. doi:10.1101/cshperspect.a003889.

Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. BMC Bioinformatics. 19(1):348. doi:10.1186/s12859-018-2376-y.

Ding M, Chen ZJ. 2018. Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. Current Opinion in Plant Biology. 42:37–48. doi:10.1016/j.pbi.2018.02.003.

Dinh Thi VH, Coriton O, Clainche IL, Arnaud D, Gordon SP, Linc G, Catalan P, Hasterok R, Vogel JP, Jahier J, et al. 2016. Recreating Stable Brachypodium hybridum Allotetraploids by

Uniting the Divergent Genomes of B. distachyon and B. stacei. PLOS ONE. 11(12):e0167171. doi:10.1371/journal.pone.0167171.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 485(7398):376–380. doi:10.1038/nature11082.

Dong F, Jiang J. 1998. Non-Rabl patterns of centromere and telomere distribution in the interphase nuclei of plant cells. Chromosome Res. 6(7):551–558. doi:10.1023/a:1009280425125.

Dong Q, Li N, Li X, Yuan Z, Xie D, Wang X, Li J, Yu Y, Wang J, Ding B, et al. 2018. Genome-wide Hi-C analysis reveals extensive hierarchical chromatin interactions in rice. The Plant Journal. 94(6):1141–1156. doi:10.1111/tpj.13925.

Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R, Ågren JA, Hazzouri KM, Wang W, et al. 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid Capsella bursa-pastoris. PNAS. 112(9):2806–2811. doi:10.1073/pnas.1412277112.

Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge APM. 2001. Brachypodium distachyon. A New Model System for Functional Genomics in Grasses. Plant Physiology. 127(4):1539–1555. doi:10.1104/pp.010196.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 356(6333):92–95. doi:10.1126/science.aal3327.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. 3(1):99–101. doi:10.1016/j.cels.2015.07.012.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3(1):95–98. doi:10.1016/j.cels.2016.07.002.

Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, Smith RD, Teresi SJ, Nelson ADL, Wai CM, et al. 2019. Origin and evolution of the octoploid strawberry genome. Nat Genet. 51(3):541–547. doi:10.1038/s41588-019-0356-4.

Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, Yuan Y, Bewick AJ, Ji L, Platts AE, Bowman MJ, et al. 2017. Subgenome Dominance in an Interspecific Hybrid, Synthetic Allopolyploid, and a 140-Year-Old Naturally Established Neo-Allopolyploid Monkeyflower. Plant Cell. 29(9):2150–2167. doi:10.1105/tpc.17.00010.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 9(1):18. doi:10.1186/1471-2105-9-18.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 20(1):238. doi:10.1186/s13059-019-1832-y.

Eres IE, Gilad Y. 2021. A TAD Skeptic: Is 3D Genome Topology Conserved? Trends in Genetics. 37(3):216–223. doi:10.1016/j.tig.2020.10.009.

Fang L, Guan X, Zhang T. 2017. Asymmetric evolution and domestication in allotetraploid cotton (Gossypium hirsutum L.). The Crop Journal. 5(2):159–165. doi:10.1016/j.cj.2016.07.001.

Febrer M, Goicoechea JL, Wright J, McKenzie N, Song X, Lin J, Collura K, Wissotski M, Yu Y, Ammiraju JSS, et al. 2010. An Integrated Physical, Genetic and Cytogenetic Map of Brachypodium distachyon, a Model System for Grass Research. PLoS One. 5(10):e13461. doi:10.1371/journal.pone.0013461.

Feldman M, Levy AA. 2009. Genome evolution in allopolyploid wheat—a revolutionary reprogramming followed by gradual changes. Journal of Genetics and Genomics. 36(9):511–518. doi:10.1016/S1673-8527(08)60142-3.

Feldman M, Levy AA, Fahima T, Korol A. 2012. Genomic asymmetry in allopolyploid plants: wheat as a model. Journal of Experimental Botany. 63(14):5045–5059. doi:10.1093/jxb/ers192.

Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM. 1997. Rapid Elimination of Low-Copy DNA Sequences in Polyploid Wheat: A Possible Mechanism for Differentiation of Homoeologous Chromosomes. Genetics. 147(3):1381–1387. doi:10.1093/genetics/147.3.1381.

Feng S, Cokus SJ, Schubert V, Zhai J, Pellegrini M, Jacobsen SE. 2014. Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. Mol Cell. 55(5):694–707. doi:10.1016/j.molcel.2014.07.008.

Flagel L, Udall J, Nettleton D, Wendel J. 2008. Duplicate gene expression in allopolyploid Gossypiumreveals two temporally distinct phases of expression evolution. BMC Biology. 6(1):16. doi:10.1186/1741-7007-6-16.

Flagel LE, Wendel JF. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol. 186(1):184–193. doi:10.1111/j.1469-8137.2009.03107.x.

Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Temsch EM, Greilhuber J, Müller KF, Heubl G. 2014. Evolution of genome size and chromosome number in the carnivorous plant genus Genlisea (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. Annals of Botany. 114(8):1651–1663. doi:10.1093/aob/mcu189.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2019. RepeatModeler2: automated genomic discovery of transposable element families. Genomics. [accessed 2021 Nov 9]. http://biorxiv.org/lookup/doi/10.1101/856591.

Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol. 60:433–453. doi:10.1146/annurev.arplant.043008.092122.

Freeling M, Scanlon MJ, Fowler JE. 2015. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. Curr Opin Genet Dev. 35:110–118. doi:10.1016/j.gde.2015.11.002.

Freeling M, Thomas BC. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. 16(7):805–814. doi:10.1101/gr.3681406.

Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC. 2012. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Current Opinion in Plant Biology. 15(2):131–139. doi:10.1016/j.pbi.2012.01.015.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 28(23):3150–3152. doi:10.1093/bioinformatics/bts565.

Gaeta RT, Pires JC. 2010. Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytologist. 186(1):18–28. doi:10.1111/j.1469-8137.2009.03089.x.

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007. Genomic Changes in Resynthesized Brassica napus and Their Effect on Gene Expression and Phenotype. The Plant Cell. 19(11):3403–3417. doi:10.1105/tpc.107.054346.

Gantuz M, Morales A, Bertoldi MV, Ibañez VN, Duarte PF, Marfil CF, Masuelli RW. 2021. Hybridization and polyploidization effects on LTR-retrotransposon activation in potato genome. J Plant Res. doi:10.1007/s10265-021-01354-9. [accessed 2022 Jan 10]. https://doi.org/10.1007/s10265-021-01354-9.

Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M. 2014. Two evolutionarily distinct classes of paleopolyploidy. Mol Biol Evol. 31(2):448–454. doi:10.1093/molbev/mst230.

Gaut BS. 2002. Evolutionary dynamics of grass genomes. New Phytologist. 154(1):15–28. doi:10.1046/j.1469-8137.2002.00352.x.

Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. 2016. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics. 32(2):289–291. doi:10.1093/bioinformatics/btv562.

Gill RA, Scossa F, King GJ, Golicz AA, Tong C, Snowdon RJ, Fernie AR, Liu S. 2021. On the Role of Transposable Elements in the Regulation of Gene Expression and Subgenomic Interactions in Crop Genomes. Critical Reviews in Plant Sciences. 40(2):157–189. doi:10.1080/07352689.2021.1920731.

Godfree RC, Marshall DJ, Young AG, Miller CH, Mathews S. 2017. Empirical evidence of fixed and homeostatic patterns of polyploid advantage in a keystone grass exposed to drought and heat stress. R Soc open sci. 4(11):170934. doi:10.1098/rsos.170934.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40(Database issue):D1178-1186. doi:10.1093/nar/gkr944.

Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, Session A, Martin J, Cartwright A, Katz A, et al. 2020. Gradual polyploid genome evolution revealed by pan-genomic analysis of Brachypodium hybridum and its diploid progenitors. Nat Commun. 11(1):3670. doi:10.1038/s41467-020-17302-5.

Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin AC, Schackwitz W, Tyler L, et al. 2017. Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. Nat Commun. 8(1):2184. doi:10.1038/s41467-017-02292-8.

Gou X, Bian Y, Zhang A, Zhang H, Wang B, Lv R, Li J, Zhu B, Gong L, Liu B. 2018. Transgenerationally Precipitated Meiotic Chromosome Instability Fuels Rapid Karyotypic Evolution and Phenotypic Diversity in an Artificially Constructed Allotetraploid Wheat (AADD). Mol Biol Evol. 35(5):1078–1091. doi:10.1093/molbev/msy009.

Grandont L, Jenczewski E, Lloyd A. 2013. Meiosis and Its Deviations in Polyploid Plants. CGR. 140(2–4):171–184. doi:10.1159/000351730.

Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES. 2012. Trans Chromosomal Methylation in Arabidopsis hybrids. PNAS. 109(9):3570–3575. doi:10.1073/pnas.1201043109.

Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby R, Nagy I, Khan A, Larking A, et al. 2019. Breaking Free: The Genomics of Allopolyploidy-Facilitated Niche Expansion in White Clover. Plant Cell. 31(7):1466–1487. doi:10.1105/tpc.18.00606.

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. Nucleic Acids Res. 31(1):439–441.

Grob S, Schmid MW, Grossniklaus U. 2014. Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Mol Cell. 55(5):678–693. doi:10.1016/j.molcel.2014.07.009.

Groszmann M, Greaves IK, Albert N, Fujimoto R, Helliwell CA, Dennis ES, Peacock WJ. 2011. Epigenetics in plants—vernalisation and hybrid vigour. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 1809(8):427–437. doi:10.1016/j.bbagrm.2011.03.006.

Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. New Phytologist. 196(4):966–971. doi:10.1111/j.1469-8137.2012.04365.x.

Guo H, Wang X, Gundlach H, Mayer KFX, Peterson DG, Scheffler BE, Chee PW, Paterson AH. 2014. Extensive and Biased Intergenomic Nonreciprocal DNA Exchanges Shaped a Nascent Polyploid Genome, Gossypium (Cotton). Genetics. 197(4):1153–1163. doi:10.1534/genetics.114.166124.

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang X-J, Chen ZJ. 2009. Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. PNAS. 106(42):17835–17840. doi:10.1073/pnas.0907003106.

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31(19):5654–5666. doi:10.1093/nar/gkg770.

Han J, Zhou B, Shan W, Yu L, Wu W, Wang K. 2015. A and D genomes spatial separation at somatic metaphase in tetraploid cotton: evidence for genomic disposition in a polyploid plant. Plant J. 84(6):1167–1177. doi:10.1111/tpj.13074.

Hardigan MA, Feldmann MJ, Lorant A, Bird KA, Famula R, Acharya C, Cole G, Edger PP, Knapp SJ. 2020. Genome Synteny Has Been Conserved Among the Octoploid Progenitors of Cultivated Strawberry Over Millions of Years of Evolution. Frontiers in Plant Science. 10:1789. doi:10.3389/fpls.2019.01789.

Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, Wolny E, Idziak D, Draper J, Jenkins G. 2006. Alignment of the Genomes of Brachypodium distachyon and Temperate Cereals and Grasses Using Bacterial Artificial Chromosome Landing With Fluorescence in Situ Hybridization. Genetics. 173(1):349–362. doi:10.1534/genetics.105.049726.

Heard E, Bickmore W. 2007. The ins and outs of gene regulation and chromosome territory organisation. Curr Opin Cell Biol. 19(3):311–316. doi:10.1016/j.ceb.2007.04.016.

Heride C, Ricoul M, Kiêu K, von Hase J, Guillemot V, Cremer C, Dubrana K, Sabatier L. 2010. Distance between homologous chromosomes results from chromosome positioning constraints. Journal of Cell Science. 123(23):4063–4075. doi:10.1242/jcs.066498.

Higgins J, Magusin A, Trick M, Fraser F, Bancroft I. 2012. Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species Brassica napus. BMC Genomics. 13(1):247. doi:10.1186/1471-2164-13-247.

Hollister JD. 2015. Polyploidy: adaptation to the genomic environment. New Phytologist. 205(3):1034–1039. doi:10.1111/nph.12939.

Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 19(8):1419–1428. doi:10.1101/gr.091678.109.

Hsiao C, Chatterton NJ, Asay KH, Jensen KB. 1994. Phylogenetic relationships of 10 grass species: an assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. Genome. 37(1):112–120. doi:10.1139/g94-014.

Hu G, Grover CE, Yuan D, Dong Y, Miller E, Conover JL, Wendel JF. 2021. Evolution and Diversity of the Cotton Genome. In: Rahman M-, Zafar Y, Zhang T, editors. Cotton Precision Breeding. Cham: Springer International Publishing. p. 25–78. [accessed 2022 Jan 20]. https://doi.org/10.1007/978-3-030-64504-5_2.

Huang K, Rieseberg LH. 2020. Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants. Frontiers in Plant Science. 11:296. doi:10.3389/fpls.2020.00296.

Hurgobin B, Golicz AA, Bayer PE, Chan C-KK, Tirnaz S, Dolatabadian A, Schiessl SV, Samans B, Montenegro JD, Parkin IAP, et al. 2018. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus. Plant Biotechnol J. 16(7):1265–1274. doi:10.1111/pbi.12867.

Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J, Mirny LA. 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods. 9(10):999–1003. doi:10.1038/nmeth.2148.

International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature. 463(7282):763–768. doi:10.1038/nature08747.

Ishii T, Karimi-Ashtiyani R, Houben A. 2016. Haploidization via Chromosome Elimination: Means and Mechanisms. Annual Review of Plant Biology. 67(1):421–438. doi:10.1146/annurev-arplant-043014-114714.

Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Michell CT, Saber N, et al. 2017. The genome of Chenopodium quinoa. Nature. 542(7641):307–312. doi:10.1038/nature21370.

Jenczewski E, Alix K. 2004. From Diploids to Allopolyploids: The Emergence of Efficient Pairing Control Genes in Plants. Critical Reviews in Plant Sciences. 23(1):21–45. doi:10.1080/07352680490273239.

Jenkins G. 1986. Synaptonemal complex formation in hybrids of Lolium temulentum • Lolium perenne (L.). Chromosoma. 93:413–419.

Jenkins G, Hasterok R. 2007. BAC “landing” on chromosomes of Brachypodium distachyon for comparative genome alignment. Nat Protoc. 2(1):88–98. doi:10.1038/nprot.2006.490.

Jia J, Xie Y, Cheng J, Kong C, Wang M, Gao L, Zhao F, Guo J, Wang K, Li G, et al. 2021. Homology-mediated inter-chromosomal interactions in hexaploid wheat lead to specific subgenome territories following polyploidization and introgression. Genome Biol. 22(1):26. doi:10.1186/s13059-020-02225-7.

Jiang J, Shao Y, Du K, Ran L, Fang X, Wang Y. 2013. Use of digital gene expression to discriminate gene expression differences in early generations of resynthesized Brassica napus and its diploid progenitors. BMC Genomics. 14(1):72. doi:10.1186/1471-2164-14-72.

Jiang X, Song Q, Ye W, Chen ZJ. 2021. Concerted genomic and epigenomic changes accompany stabilization of Arabidopsis allopolyploids. Nat Ecol Evol.:1–12. doi:10.1038/s41559-021-01523-y.

Jiao W, Yuan J, Jiang S, Liu Y, Wang L, Liu M, Zheng D, Ye W, Wang X, Chen ZJ. 2018. Asymmetrical changes of gene expression, small RNAs and chromatin in two resynthesized wheat allotetraploids. The Plant Journal. 93(5):828–842. doi:10.1111/tpj.13805.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature. 473(7345):97–100. doi:10.1038/nature09916.

Jones RN, Hegarty M. 2009. Order out of chaos in the hybrid plant nucleus. Cytogenet Genome Res. 126(4):376–389. doi:10.1159/000266171.

Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during interspecies hybridization. Curr Biol. 16(13):1322–1328. doi:10.1016/j.cub.2006.05.045.

Kashkush K, Feldman M, Levy AA. 2002. Gene Loss, Silencing and Activation in a Newly Synthesized Wheat Allotetraploid. Genetics. 160(4):1651–1659. doi:10.1093/genetics/160.4.1651.

Kashkush K, Feldman M, Levy AA. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet. 33(1):102–106. doi:10.1038/ng1063.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research. 30(14):3059–3066. doi:10.1093/nar/gkf436.

Kejnovsky E, Leitch IJ, Leitch AR. 2009. Contrasting evolutionary dynamics between angiosperm and mammalian genomes. Trends in Ecology & Evolution. 24(10):572–582. doi:10.1016/j.tree.2009.04.010.

Kellogg EA. 2016. Has the connection between polyploidy and diversification actually been tested? Current Opinion in Plant Biology. 30:25–32. doi:10.1016/j.pbi.2016.01.002.

Kenan-Eichler M, Leshkowitz D, Tal L, Noor E, Melamed-Bessudo C, Feldman M, Levy AA. 2011. Wheat Hybridization and Polyploidization Results in Deregulation of Small RNAs. Genetics. 188(2):263–272. doi:10.1534/genetics.111.128348.

Kim AI, Belyaeva ES. 1991. Transposition of mobile elements gypsy (mdg4) and hobo in germ-line and somatic cells of a genetically unstable mutator strain of Drosophila melanogaster. Molec Gen Genet. 229(3):437–444. doi:10.1007/BF00267467.

Kim J-K, Nahm BH. 1995. Rice 5S Ribosomal RNA and Its Binding Protein Genes. Mol Cells. 5(4):381–387.

Kinser TJ, Smith RD, Lawrence AH, Cooley AM, Vallejo-Marín M, Conradi Smith GD, Puzey JR. 2021. Endosperm-based incompatibilities in hybrid monkeyflowers. The Plant Cell. 33(7):2235–2257. doi:10.1093/plcell/koab117.

Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, Naldi A, Mungall CJ, Yunes JM, Botvinnik O, Weigel M, et al. 2018. GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep. 8(1):10872. doi:10.1038/s41598-018-28948-z.

Koláčková V, Perničková K, Vrána J, Duchoslav M, Jenkins G, Phillips D, Turkosi E, Šamajová O, Sedlářová M, Šamaj J, et al. 2019. Nuclear Disposition of Alien Chromosome Introgressions into Wheat and Rye Using 3D-FISH. Int J Mol Sci. 20(17):4143. doi:10.3390/ijms20174143.

Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K. 2010. Genetic and Epigenetic Dynamics of a Retrotransposon After Allopolyploidization of Wheat. Genetics. 186(3):801–812. doi:10.1534/genetics.110.120790.

Kryvokhyzha D, Milesi P, Duan T, Orsucci M, Wright SI, Glémin S, Lascoux M. 2019. Towards the new normal: Transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (Capsella bursa-pastoris). PLOS Genetics. 15(5):e1008131. doi:10.1371/journal.pgen.1008131.

Lajoie BR, Dekker J, Kaplan N. 2015. The Hitchhiker's Guide to Hi-C Analysis: Practical guidelines. Methods. 72:65–75. doi:10.1016/j.ymeth.2014.10.031.

Langham RJ, Walsh J, Dunn M, Ko C, Goff SA, Freeling M. 2004. Genomic duplication, fractionation and the origin of regulatory novelty. Genetics. 166(2):935–945. doi:10.1534/genetics.166.2.935.

Lashermes P, Combes M-C, Hueber Y, Severac D, Dereeper A. 2014. Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee. Plant J. 78(4):674–685. doi:10.1111/tpj.12505.

Lazar NH, Nevonen KA, O'Connell B, McCann C, O'Neill RJ, Green RE, Meyer TJ, Okhovat M, Carbone L. 2018 Jun 18. Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. Genome Res. doi:10.1101/gr.233874.117. [accessed 2022 Mar 18]. https://genome.cshlp.org/content/early/2018/06/16/gr.233874.117.

Le Comber SC, Ainouche ML, Kovarik A, Leitch AR. 2010. Making a functional diploid: from polysomic to disomic inheritance. New Phytologist. 186(1):113–122. doi:10.1111/j.1469-8137.2009.03117.x.

Lee T-H, Kim J, Robertson JS, Paterson AH. 2017. Plant Genome Duplication Database. Methods Mol Biol. 1533:267–277. doi:10.1007/978-1-4939-6658-5_16.

Leitch IJ, Bennett MD. 2004. Genome downsizing in polyploid plants. Biological Journal of the Linnean Society. 82(4):651–663. doi:10.1111/j.1095-8312.2004.00349.x.

Leitch IJ, Bennett MD. 2007. Genome Size and its Uses: The Impact of Flow Cytometry. In: Flow Cytometry with Plant Cells. John Wiley & Sons, Ltd. p. 153–176. [accessed 2022 Jan 6]. http://onlinelibrary.wiley.com/doi/abs/10.1002/9783527610921.ch7.

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. 2008. The Ups and Downs of Genome Size Evolution in Polyploid Species of Nicotiana (Solanaceae). Ann Bot. 101(6):805–814. doi:10.1093/aob/mcm326.

Lewis WH, editor. 1980. Polyploidy - Biological Relevance. New York: Plenum Press.

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

Li Q, Qiao X, Yin H, Zhou Y, Dong H, Qi K, Li L, Zhang S. 2019. Unbiased subgenome evolution following a recent whole-genome duplication in pear (Pyrus bretschneideri Rehd.). Hortic Res. 6(1):1–12. doi:10.1038/s41438-018-0110-6.

Li Z, McKibben MTW, Finch GS, Blischak PD, Sutherland BL, Barker MS. 2021. Patterns and Processes of Diploidization in Land Plants. Annual Review of Plant Biology. 72(1):387–410. doi:10.1146/annurev-arplant-050718-100344.

Lim KY, Souckova-Skalicka K, Sarasan V, Clarkson JJ, Chase MW, Kovarik A, Leitch AR. 2006. A genetic appraisal of a new synthetic Nicotiana tabacum (Solanaceae) and the Kostoff synthetic tobacco. Am J Bot. 93(6):875–883. doi:10.3732/ajb.93.6.875.

Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF. 2001. Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome. 44(3):321–330. doi:10.1139/g01-011.

Liu B, Vega JM, Feldman M. 1998. Rapid genomic changes in newly synthesized amphiploids of Triticum and Aegilops. II. Changes in low-copy coding DNA sequences. Genome. 41(4):535–542. doi:10.1139/g98-052.

Liu S-L, Adams KL. 2010. Dramatic change in function and expression pattern of a gene duplicated by polyploidy created a paternal effect gene in the Brassicaceae. Mol Biol Evol. 27(12):2817–2828. doi:10.1093/molbev/msq169.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 15(12):550. doi:10.1186/s13059-014-0550-8.

Lovell JT, Jenkins J, Lowry DB, Mamidi S, Sreedasyam A, Weng X, Barry K, Bonnette J, Campitelli B, Daum C, et al. 2018. The genomic landscape of molecular responses to natural drought stress in Panicum hallii. Nat Commun. 9(1):5213. doi:10.1038/s41467-018-07669-x.

Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, Sreedasyam A, Healey A, Session A, Shu S, et al. 2021. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. Nature. 590(7846):438–444. doi:10.1038/s41586-020-03127-1.

Lovell JT, Sreedasyam A, Schranz ME, Wilson MA, Carlson JW, Harkess A, Emms D, Goodstein D, Schmutz J. 2022. GENESPACE: syntenic pan-genome annotations for eukaryotes. doi:10.1101/2022.03.09.483468. http://biorxiv.org/lookup/doi/10.1101/2022.03.09.483468.

Lusinska J, Majka J, Betekhtin A, Susek K, Wolny E, Hasterok R. 2018a. Chromosome identification and reconstruction of evolutionary rearrangements in Brachypodium distachyon, B. stacei and B. hybridum. Annals of Botany. 122(3):445–459. doi:10.1093/aob/mcy086.

Lusinska J, Majka J, Betekhtin A, Susek K, Wolny E, Hasterok R. 2018b. Chromosome identification and reconstruction of evolutionary rearrangements in Brachypodium distachyon, B. stacei and B. hybridum. Ann Bot. 122(3):445–459. doi:10.1093/aob/mcy086.

Ma J, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. PNAS. 101(34):12404–12410. doi:10.1073/pnas.0403715101.

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. 2002. Remodeling of DNA Methylation and Phenotypic and Transcriptional Changes in Synthetic Arabidopsis Allotetraploids. Plant Physiology. 129(2):733–746. doi:10.1104/pp.003095.

Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L. 2005. Genomic changes in synthetic Arabidopsis polyploids. The Plant Journal. 41(2):221–230. doi:10.1111/j.1365-313X.2004.02297.x.

Madlung A, Wendel JF. 2013. Genetic and Epigenetic Aspects of Polyploid Evolution in Plants. CGR. 140(2–4):270–285. doi:10.1159/000351430.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends in Ecology & Evolution. 20(5):229–237. doi:10.1016/j.tree.2005.02.010.

Manton I. 1950. Problems of Cytology and Evolution in the Pteridophyta. Cambridge: Cambridge University Press.

Marimuthu MPA, Maruthachalam R, Bondada R, Kuppu S, Tan EH, Britt A, Chan SWL, Comai L. 2021. Epigenetically mismatched parental centromeres trigger genome elimination in hybrids. Science Advances. 7(47):eabk1151. doi:10.1126/sciadv.abk1151.

Martienssen RA. 2010. Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of Arabidopsis. New Phytol. 186(1):46–53. doi:10.1111/j.1469-8137.2010.03193.x.

Martinez-Perez E, Shaw PJ, Moore G. 2000. Polyploidy Induces Centromere Association. J Cell Biol. 148(2):233–238.

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, et al. 2017. A chromosome conformation capture ordered sequence of the barley genome. Nature. 544(7651):427–433. doi:10.1038/nature22043.

Mason AS, Wendel JF. 2020. Homoeologous Exchanges, Segmental Allopolyploidy, and Polyploid Genome Evolution. Frontiers in Genetics. 11:1014. doi:10.3389/fgene.2020.01014.

McClintock B. 1984. The Significance of Responses of the Genome to Challenge. Science. 226(4676):792–801.

McCord RP, Kaplan N, Giorgetti L. 2020. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. Mol Cell. 77(4):688–708. doi:10.1016/j.molcel.2019.12.021.

Mestiri I, Chagué V, Tanguy A-M, Huneau C, Huteau V, Belcram H, Coriton O, Chalhoub B, Jahier J. 2010. Newly synthesized wheat allohexaploids display progenitor-dependent meiotic stability and aneuploidy but structural genomic additivity. New Phytol. 186(1):86–101. doi:10.1111/j.1469-8137.2010.03186.x.

Mirzaghaderi G, Mason AS. 2017. Revisiting Pivotal-Differential Genome Evolution in Wheat. Trends in Plant Science. 22(8):674–684. doi:10.1016/j.tplants.2017.06.003.

Misteli T. 2010. Higher-order Genome Organization in Human Disease. Cold Spring Harb Perspect Biol. 2(8):a000794. doi:10.1101/cshperspect.a000794.

Misteli T, Soutoglou E. 2009. The emerging role of nuclear architecture in DNA repair and genome maintenance. Nat Rev Mol Cell Biol. 10(4):243–254. doi:10.1038/nrm2651.

Mitros T, Session AM, James BT, Wu GA, Belaffif MB, Clark LV, Shu S, Dong H, Barling A, Holmes JR, et al. 2020. Genome biology of the paleotetraploid perennial biomass crop Miscanthus. Nat Commun. 11(1):5442. doi:10.1038/s41467-020-18923-6.

Murat F, Zhang R, Guizard S, Flores R, Armero A, Pont C, Steinbach D, Quesneville H, Cooke R, Salse J. 2014. Shared Subgenome Dominance Following Polyploidization Explains Grass Genome Evolutionary Plasticity from a Seven Protochromosome Ancestor with 16K Protogenes. Genome Biology and Evolution. 6(1):12–33. doi:10.1093/gbe/evt200.

Nelson MG, Linheiro RS, Bergman CM. 2017. McClintock: An Integrated Pipeline for Detecting Transposable Element Insertions in Whole-Genome Shotgun Sequencing Data. G3: Genes, Genomes, Genetics. 7(8):2763–2778. doi:10.1534/g3.117.043893.

Nicolas SD, Mignon GL, Eber F, Coriton O, Monod H, Clouet V, Huteau V, Lostanlen A, Delourme R, Chalhoub B, et al. 2007. Homeologous Recombination Plays a Major Role in Chromosome Rearrangements That Occur During Meiosis of Brassica napus Haploids. Genetics. 175(2):487–503. doi:10.1534/genetics.106.062968.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 485(7398):381–385. doi:10.1038/nature11049.

Ou S, Jiang N. 2018. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant Physiology. 176(2):1410–1422. doi:10.1104/pp.17.01310.

Ozkan H, Levy AA, Feldman M. 2001. Allopolyploidy-Induced Rapid Genome Evolution in the Wheat (Aegilops–Triticum) Group. The Plant Cell. 13(8):1735–1747. doi:10.1105/TPC.010082.

Ozkan H, Tuna M, Galbraith DW. 2006. No DNA loss in autotetraploids of Arabidopsis thaliana. Plant Breeding. 125(3):288–291. doi:10.1111/j.1439-0523.2006.01211.x.

Paape T, Briskine RV, Halstead-Nussloch G, Lischer HEL, Shimizu-Inatsugi R, Hatakeyama M, Tanaka K, Nishiyama T, Sabirov R, Sese J, et al. 2018. Patterns of polymorphism and selection in the subgenomes of the allopolyploid Arabidopsis kamchatica. Nat Commun. 9(1):3909. doi:10.1038/s41467-018-06108-1.

Papp B, Pál C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. Nature. 424(6945):194–197. doi:10.1038/nature01771.

Parada L, Misteli T. 2002. Chromosome positioning in the interphase nucleus. Trends Cell Biol. 12(9):425–432. doi:10.1016/s0962-8924(02)02351-6.

Parisod C, Mhiri C, Lim KY, Clarkson JJ, Chase MW, Leitch AR, Grandbastien M-A. 2012. Differential Dynamics of Transposable Elements during Long-Term Diploidization of Nicotiana Section Repandae (Solanaceae) Allopolyploid Genomes. PLOS ONE. 7(11):e50352. doi:10.1371/journal.pone.0050352.

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien M-A, Ainouche M. 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in Spartina. New Phytologist. 184(4):1003–1015. doi:10.1111/j.1469-8137.2009.03029.x.

Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, et al. 2014. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome Biology. 15(6):R77. doi:10.1186/gb-2014-15-6-r77.

Pelé A, Rousseau-Gueutin M, Chèvre A-M. 2018. Speciation Success of Polyploid Plants Closely Relates to the Regulation of Meiotic Recombination. Front Plant Sci. 9:907. doi:10.3389/fpls.2018.00907.

Pellicer J, Fay MF, Leitch IJ. 2010. The largest eukaryotic genome of them all?: THE LARGEST EUKARYOTIC GENOME? Botanical Journal of the Linnean Society. 164(1):10–15. doi:10.1111/j.1095-8339.2010.01072.x.

Prieto P, Santos AP, Moore G, Shaw P. 2004. Chromosomes associate premeiotically and in xylem vessel cells via their telomeres and centromeres in diploid rice (Oryza sativa). Chromosoma. 112(6):300–307. doi:10.1007/s00412-004-0274-8.

Pumphrey M, Bai J, Laudencia-Chingcuanco D, Anderson O, Gill BS. 2009. Nonadditive Expression of Homoeologous Genes Is Established Upon Polyploidization in Hexaploid Wheat. Genetics. 181(3):1147–1157. doi:10.1534/genetics.108.096941.

Quinlan AR. 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics. 47:11.12.1-11.12.34. doi:10.1002/0471250953.bi1112s47.

Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, et al. 2018. The transcriptional landscape of polyploid wheat. Science. 361(6403):eaar6089. doi:10.1126/science.aar6089.

Ramsey J. 2011. Polyploidy and ecological adaptation in wild yarrow. PNAS. 108(17):7096–7101. doi:10.1073/pnas.1016631108.

Ramsey J, Schemske DW. 1998. PATHWAYS, MECHANISMS, AND RATES OF POLYPLOID FORMATION IN FLOWERING PLANTS. Annu Rev Ecol Syst. 29(1):467–501. doi:10.1146/annurev.ecolsys.29.1.467.

Ramsey J, Schemske DW. 2002. Neopolyploidy in Flowering Plants. Annu Rev Ecol Syst. 33(1):589–639. doi:10.1146/annurev.ecolsys.33.010802.150437.

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. Molecular Biology and Evolution. 35(10):2582–2584. doi:10.1093/molbev/msy159.

Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. PLOS ONE. 6(9):e22594. doi:10.1371/journal.pone.0022594.

Renny-Byfield S, Gong L, Gallagher JP, Wendel JF. 2015. Persistence of Subgenomes in Paleopolyploid Cotton after 60 My of Evolution. Molecular Biology and Evolution. 32(4):1063–1071. doi:10.1093/molbev/msv001.

Renny-Byfield S, Kovařík A, Chester M, Nichols RA, Macas J, Novák P, Leitch AR. 2012. Independent, Rapid and Targeted Loss of Highly Repetitive DNA in Natural and Synthetic Allopolyploids of Nicotiana tabacum. PLOS ONE. 7(5):e36963. doi:10.1371/journal.pone.0036963.

Rigal M, Becker C, Pélissier T, Pogorelcnik R, Devos J, Ikeda Y, Weigel D, Mathieu O. 2016. Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in Arabidopsis thaliana F1 epihybrids. PNAS. 113(14):E2083–E2092. doi:10.1073/pnas.1600672113.

Salamov AA, Solovyev VV. 2000. Ab initio gene finding in Drosophila genomic DNA. Genome Res. 10(4):516–522. doi:10.1101/gr.10.4.516.

Salman-Minkov A, Sabath N, Mayrose I. 2016. Whole-genome duplication as a key factor in crop domestication. Nature Plants. 2(8):1–4. doi:10.1038/nplants.2016.115.

Salmon A, Ainouche ML, Wendel JF. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). Molecular Ecology. 14(4):1163–1175. doi:10.1111/j.1365-294X.2005.02488.x.

Samans B, Chalhoub B, Snowdon RJ. 2017. Surviving a Genome Collision: Genomic Signatures of Allopolyploidization in the Recent Crop Species Brassica napus. The Plant Genome. 10(3):plantgenome2017.02.0013. doi:10.3835/plantgenome2017.02.0013.

Sanei M, Pickering R, Kumke K, Nasuda S, Houben A. 2011. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proc Natl Acad Sci U S A. 108(33):E498-505. doi:10.1073/pnas.1103190108.

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. 1998. The paleontology of intergene retrotransposons of maize. Nat Genet. 20(1):43–45. doi:10.1038/1695.

Santos AP, Shaw P. 2004. Interphase chromosomes and the Rabl configuration: does genome size matter? Journal of Microscopy. 214(2):201–206. doi:10.1111/j.0022-2720.2004.01324.x.

Sarilar V, Palacios PM, Rousselet A, Ridel C, Falque M, Eber F, Chèvre A-M, Joets J, Brabant P, Alix K. 2013. Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized Brassica napus allotetraploids. New Phytologist. 198(2):593–604. doi:10.1111/nph.12156.

Satyaki PRV, Gehring M. 2019. Paternally Acting Canonical RNA-Directed DNA Methylation Pathway Genes Sensitize Arabidopsis Endosperm to Paternal Genome Dosage. The Plant Cell. 31(7):1563–1578. doi:10.1105/tpc.19.00047.

Schnable JC, Lyons E. 2012. Comparative genomics with maize and other grasses: from genes to genomes! Maydica. 56(2). [accessed 2021 Nov 10]. https://journals-crea.4science.it/index.php/maydica/article/view/690.

Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. PNAS. 108(10):4069–4074. doi:10.1073/pnas.1101368108.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science. 326(5956):1112–1115. doi:10.1126/science.1178534.

Senerchia N, Felber F, Parisod C. 2015. Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. Proceedings of the Royal Society B: Biological Sciences. 282(1804):20142874. doi:10.1098/rspb.2014.2874.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biology. 16(1):259. doi:10.1186/s13059-015-0831-x.

Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, Fukui A, Hikosaka A, Suzuki A, Kondo M, et al. 2016. Genome evolution in the allotetraploid frog Xenopus laevis. Nature. 538(7625):336–343. doi:10.1038/nature19840.

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001. Sequence Elimination and Cytosine Methylation Are Rapid and Reproducible Responses of the Genome to Wide Hybridization and Allopolyploidy in Wheat. The Plant Cell. 13(8):1749–1759. doi:10.1105/TPC.010083.

Sharma B, Kramer E. 2013. Sub- and neo-functionalization of APETALA3 paralogs have contributed to the evolution of novel floral organ identity in Aquilegia (columbine, Ranunculaceae). New Phytol. 197(3):949–957. doi:10.1111/nph.12078.

Shi X, Ng DW-K, Zhang C, Comai L, Ye W, Jeffrey Chen Z. 2012. Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. Nat Commun. 3(1):950. doi:10.1038/ncomms1954.

Shull GH. 1914. Duplicate genes for capsule-form in Bursa bursa-pastoris. Zeitschrift für induktive Abstammmungs und Vererbungslehre. 12(2):97–149.

Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 12(4):246–258. doi:10.1038/nrm3089.

Skalická K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A. 2005. Preferential elimination of repeated DNA sequences from the paternal, Nicotiana tomentosiformis genome donor of a synthetic, allotetraploid tobacco. New Phytologist. 166(1):291–303. doi:10.1111/j.1469-8137.2004.01297.x.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6(1):31. doi:10.1186/1471-2105-6-31.

Smit AF, Hubley R, Green P. 2013 2015. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Smith RD, Kinser TJ, Smith GDC, Puzey JR. 2019. A likelihood ratio test for changes in homeolog expression bias. BMC Bioinformatics. 20(1):149. doi:10.1186/s12859-019-2709-5.

Soltis DE, Buggs RJA, Barbazuk WB, Chamala S, Chester M, Gallagher JP, Schnable PS, Soltis PS. 2012. The Early Stages of Polyploidy: Rapid and Repeated Evolution in Tragopogon. In: Soltis PS, Soltis DE, editors. Polyploidy and Genome Evolution. Berlin, Heidelberg: Springer. p. 271–292. [accessed 2022 Jan 7]. https://doi.org/10.1007/978-3-642-31442-1_14.

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, Mei W, Cortez MB, Soltis PS, Gitzendanner MA. 2014. Are polyploids really evolutionary dead-ends

(again)? A critical reappraisal of Mayrose et al. (2011). New Phytologist. 202(4):1105–1117. doi:10.1111/nph.12756.

Soltis DE, Visger CJ, Marchant DB, Soltis PS. 2016. Polyploidy: Pitfalls and paths to a paradigm. American Journal of Botany. 103(7):1146–1166. doi:10.3732/ajb.1500501.

Song Q, Chen ZJ. 2015. Epigenetic and developmental regulation in plant polyploids. Current Opinion in Plant Biology. 24:101–109. doi:10.1016/j.pbi.2015.02.007.

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 7(1):62. doi:10.1186/1471-2105-7-62.

Stebbins GL. 1950. Variation and Evolution in Plants. New York Chichester, West Sussex: Columbia University Press.

Stebbins GL. 1971. Chromosomal Evolution in Higher Plants. London: Edward Arnold (Publishers) Ltd.

Steige KA, Slotte T. 2016. Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy. Curr Opin Plant Biol. 30:88–93. doi:10.1016/j.pbi.2016.02.006.

Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. 2018. Recent Activity in Expanding Populations and Purifying Selection Have Shaped Transposable Element Landscapes across Natural Accessions of the Mediterranean Grass Brachypodium distachyon. Genome Biology and Evolution. 10(1):304–318. doi:10.1093/gbe/evx276.

Stritt C, Wyler M, Gimmi EL, Pippel M, Roulin AC. 2020. Diversity, dynamics and effects of long terminal repeat retrotransposons in the model grass Brachypodium distachyon. New Phytologist. 227(6):1736–1748. doi:10.1111/nph.16308.

Sun H, Wu S, Zhang G, Jiao C, Guo S, Ren Y, Zhang J, Zhang H, Gong G, Jia Z, et al. 2017. Karyotype Stability and Unbiased Fractionation in the Paleo-Allotetraploid Cucurbita Genomes. Mol Plant. 10(10):1293–1306. doi:10.1016/j.molp.2017.09.003.

Sunder S, Wilson TE. 2019. Frequency of DNA end joining *in trans* is not determined by the predamage spatial proximity of double-strand breaks in yeast. Proc Natl Acad Sci USA. 116(19):9481–9490. doi:10.1073/pnas.1818595116.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLOS ONE. 6(7):e21800. doi:10.1371/journal.pone.0021800.

Szadkowski E, Eber F, Huteau V, Lodé M, Coriton O, Jenczewski E, Chèvre AM. 2011. Polyploid formation pathways have an impact on genetic rearrangements in resynthesized Brassica napus. New Phytologist. 191(3):884–894. doi:10.1111/j.1469-8137.2011.03729.x.

Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, et al. 2010. The first meiosis of resynthesized Brassica napus, a genome blender. New Phytol. 186(1):102–112. doi:10.1111/j.1469-8137.2010.03182.x.

Takaiwa F, Kikuchi S, Oono K. 1990. The complete nucleotide sequence of the intergenic spacer between 25S and 17S rDNAs in rice. Plant Mol Biol. 15(6):933–935. doi:10.1007/BF00039432.

Takaiwa F, Oono K, Sugiura M. 1984. The complete nucleotide sequence of a rice 17S rRNA gene. Nucleic Acids Res. 12(13):5441–5448.

Takaiwa F, Oono K, Sugiura M. 1985. Nucleotide sequence of the 17S-25S spacer region from rice rDNA. Plant Mol Biol. 4(6):355–364. doi:10.1007/BF02418257.

Tayalé A, Parisod C. 2013. Natural Pathways to Polyploidy in Plants and Consequences for Genome Reorganization. CGR. 140(2–4):79–96. doi:10.1159/000351318.

Thomas BC, 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 Res. 16(7):934–946. doi:10.1101/gr.4708406.

Thompson JD, Lumaret R. 1992. The evolutionary dynamics of polyploid plants: origins, establishment and persistence. Trends in Ecology & Evolution. 7(9):302–307. doi:10.1016/0169-5347(92)90228-4.

Tian E, Jiang Y, Chen L, Zou J, Liu F, Meng J. 2010. Synthesis of a Brassica trigenomic allohexaploid (B. carinata × B. rapa) de novo and its stability in subsequent generations. Theor Appl Genet. 121(8):1431–1440. doi:10.1007/s00122-010-1399-1.

Ungerer MC, Strakosh SC, Stimpson KM. 2009. Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. BMC Biology. 7(1):40. doi:10.1186/1741-7007-7-40.

Ungerer MC, Strakosh SC, Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol. 16(20):R872-873. doi:10.1016/j.cub.2006.09.020.

Usai G, Mascagni F, Vangelisti A, Giordani T, Ceccarelli M, Cavallini A, Natali L. 2020. Interspecific hybridisation and LTR-retrotransposon mobilisation-related structural variation in plants: A case study. Genomics. 112(2):1611–1621. doi:10.1016/j.ygeno.2019.09.010.

Van de Peer Y, Ashman T-L, Soltis PS, Soltis DE. 2021. Polyploidy: an evolutionary and ecological force in stressful times. The Plant Cell. 33(1):11–26. doi:10.1093/plcell/koaa015.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. Nat Rev Genet. 18(7):411–424. doi:10.1038/nrg.2017.26.

VanBuren R, Man Wai C, Wang X, Pardo J, Yocca AE, Wang H, Chaluvadi SR, Han G, Bryant D, Edger PP, et al. 2020. Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. Nat Commun. 11(1):884. doi:10.1038/s41467-020-14724-z.

Vicient CM, Casacuberta JM. 2017. Impact of transposable elements on polyploid plant genomes. Annals of Botany. 120(2):195–207. doi:10.1093/aob/mcx078.

Vimala Y, Lavania U. 2021. Genomic territories in inter-genomic hybrids: the winners and losers with hybrid fixation. The Nucleus. 64(1):1–6. doi:10.1007/s13237-021-00348-1.

Vogel JP. 2016. The Rise of Brachypodium as a Model System. In: Vogel JP, editor. Genetics and Genomics of Brachypodium. Cham: Springer International Publishing. (Plant Genetics and Genomics: Crops and Models). p. 1–7. [accessed 2022 Jan 20]. https://doi.org/10.1007/7397_2015_14.

Wang C, Liu C, Roqueiro D, Grimm D, Schwab R, Becker C, Lanz C, Weigel D. 2015. Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. Genome Res. 25(2):246–256. doi:10.1101/gr.170332.113.

Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, et al. 2006. Genomewide Nonadditive Gene Regulation in Arabidopsis Allotetraploids. Genetics. 172(1):507–517. doi:10.1534/genetics.105.047894.

Wang J, Tian L, Madlung A, Lee H-S, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ. 2004. Stochastic and Epigenetic Changes of Gene Expression in Arabidopsis Polyploids. Genetics. 167(4):1961–1973. doi:10.1534/genetics.104.027896.

Wang X, Morton JA, Pellicer J, Leitch IJ, Leitch AR. 2021. Genome downsizing after polyploidy: mechanisms, rates and selection pressures. The Plant Journal. 107(4):1003–1015. doi:10.1111/tpj.15363.

Wang X, Zhang H, Li Y, Zhang Z, Li L, Liu B. 2016. Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. New Phytologist. 209(3):1264–1277. doi:10.1111/nph.13678.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40(7):e49. doi:10.1093/nar/gkr1293.

Wei N, Cronn R, Liston A, Ashman T-L. 2019. Functional trait divergence and trait plasticity confer polyploid advantage in heterogeneous environments. New Phytologist. 221(4):2286–2297. doi:10.1111/nph.15508.

Wendel JF. 2015. The wondrous cycles of polyploidy in plants. American Journal of Botany. 102(11):1753–1756. doi:10.3732/ajb.1500320.

Wendel JF, Lisch D, Hu G, Mason AS. 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. doi:10.1016/j.gde.2018.01.004.

Wicker T, Gundlach H, Spannagl M, Uauy C, Borrill P, Ramírez-González RH, De Oliveira R, Mayer KFX, Paux E, Choulet F, et al. 2018. Impact of transposable elements on genome structure and evolution in bread wheat. Genome Biology. 19(1):103. doi:10.1186/s13059-018-1479-0.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8(12):973–982. doi:10.1038/nrg2165.

Wilson PB, Streich JC, Murray KD, Eichten SR, Cheng R, Aitken NC, Spokas K, Warthmann N, Gordon SP, Accession C, et al. 2019. Global Diversity of the Brachypodium Species Complex as a Resource for Genome-Wide Association Studies Demonstrated for Agronomic Traits in Response to Climate. Genetics. 211(1):317–331. doi:10.1534/genetics.118.301589.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. PNAS. 106(33):13875–13879. doi:10.1073/pnas.0811575106.

Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M, Wang X. 2014. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. PNAS. 111(14):5283–5288. doi:10.1073/pnas.1402475111.

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following Tetraploidy in Maize, a Short Deletion Mechanism Removed Genes Preferentially from One of the Two Homeologs. PLOS Biology. 8(6):e1000409. doi:10.1371/journal.pbio.1000409.

Wu H, Yu Q, Ran J-H, Wang X-Q. 2021. Unbiased Subgenome Evolution in Allotetraploid Species of Ephedra and Its Implications for the Evolution of Large Genomes in Gymnosperms. Genome Biology and Evolution. 13(2):evaa236. doi:10.1093/gbe/evaa236.

Wu J, Lin L, Xu M, Chen P, Liu D, Sun Q, Ran L, Wang Y. 2018. Homoeolog expression bias and expression level dominance in resynthesized allopolyploid Brassica napus. BMC Genomics. 19(1):586. doi:10.1186/s12864-018-4966-5.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 26(7):873–881. doi:10.1093/bioinformatics/btq057.

Xiao C-L, Chen Y, Xie S-Q, Chen K-N, Wang Y, Han Y, Luo F, Xie Z. 2017. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nat Methods. 14(11):1072–1074. doi:10.1038/nmeth.4432.

Xiong Z, Gaeta RT, Pires JC. 2011. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid Brassica napus. PNAS. 108(19):7908–7913. doi:10.1073/pnas.1014138108.

Yaakov B, Kashkush K. 2012. Mobilization of Stowaway-like MITEs in newly formed allohexaploid wheat species. Plant Mol Biol. 80(4–5):419–427. doi:10.1007/s11103-012-9957-3.

Yang S, Cheung F, Lee JJ, Ha M, Wei NE, Sze S-H, Stelly DM, Thaxton P, Triplett B, Town CD, et al. 2006. Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. Plant J. 47(5):761–775. doi:10.1111/j.1365-313X.2006.02829.x.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution. 24(8):1586–1591. doi:10.1093/molbev/msm088.

Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, Franklin FCH, Bomblies K. 2013. Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa. Current Biology. 23(21):2151–2156. doi:10.1016/j.cub.2013.08.059.

Yoo M-J, Szadkowski E, Wendel JF. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity. 110(2):171–180. doi:10.1038/hdy.2012.94.

Yu T, Huang X, Dou S, Tang X, Luo S, Theurkauf WE, Lu J, Weng Z. 2021. A benchmark and an algorithm for detecting germline transposon insertions and measuring de novo transposon insertion frequencies. Nucleic Acids Research. 49(8):e44. doi:10.1093/nar/gkab010.

Yu X, Wang P, Li J, Zhao Q, Ji C, Zhu Z, Zhai Y, Qin X, Zhou J, Yu H, et al. 2021. Whole-Genome Sequence of Synthesized Allopolyploids in Cucumis Reveals Insights into the Genome Evolution of Allopolyploidization. Advanced Science. 8(9):2004222. doi:10.1002/advs.202004222.

Yuan J, Jiao W, Liu Y, Ye W, Wang X, Liu B, Song Q, Chen ZJ. 2020. Dynamic and reversible DNA methylation changes induced by genome separation and merger of polyploid wheat. BMC Biology. 18(1):171. doi:10.1186/s12915-020-00909-x.

Zhang D, Pan Q, Cui C, Tan C, Ge X, Shao Y, Li Z. 2015. Genome-specific differential gene expressions in resynthesized Brassica allotetraploids from pair-wise crosses of three cultivated diploids revealed by RNA-seq. Frontiers in Plant Science. 6. [accessed 2022 Feb 4]. https://www.frontiersin.org/article/10.3389/fpls.2015.00957.

Zhang H, Bian Y, Gou X, Zhu B, Xu C, Qi B, Li N, Rustgi S, Zhou H, Han F, et al. 2013. Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. PNAS. 110(9):3447–3452. doi:10.1073/pnas.1300153110.

Zhang Y, Shen Q, Leng L, Zhang D, Chen Sha, Shi Y, Ning Z, Chen Shilin. 2021. Incipient diploidization of the medicinal plant Perilla within 10,000 years. Nat Commun. 12(1):5508. doi:10.1038/s41467-021-25681-6.

Zhang Z, Belcram H, Gornicki P, Charles M, Just J, Huneau C, Magdelenat G, Couloux A, Samain S, Gill BS, et al. 2011. Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. PNAS. 108(46):18737–18742. doi:10.1073/pnas.1110552108.

Zhao M, Zhang B, Lisch D, Ma J. 2017. Patterns and Consequences of Subgenome Differentiation Provide Insights into the Nature of Paleopolyploidy in Plants. The Plant Cell. 29(12):2974–2994. doi:10.1105/tpc.17.00595.

Zhao N, Zhu B, Li M, Wang L, Xu L, Zhang H, Zheng S, Qi B, Han F, Liu B. 2011. Extensive and Heritable Epigenetic Remodeling and Genetic Stability Accompany Allohexaploidization of Wheat. Genetics. 188(3):499–510. doi:10.1534/genetics.111.127688.