Discoveries

Biased retention of environment-responsive genes following genome fractionation

Marc Beringer[ab1], Rimjhim Roy Choudhury[ab1], Terezie Mandáková[c], Sandra Grünig[ab], Manuel Poretti[a], Ilia J. Leitch[d], Martin A. Lysak[c], Christian Parisod[ab*]

[a]Department of Biology, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland

[b]Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland

[c]Central European Institute of Technology, Masaryk University, Brno 625 00, Czech Republic

[d]Royal Botanic Gardens, Kew, Surrey, TW9 3AB, United Kingdom

[1]Joint first-authors having equally contributed.

*Corresponding author: Christian Parisod, Department of Biology, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland, Phone: +41 26 300 8852.

Email: christian.parisod@unifr.ch

This file includes:

Main Text

Legends of figures 1 to 4

Abstract

The molecular underpinnings and consequences of cycles of whole-genome duplication (WGD) and subsequent gene loss through subgenome fractionation remain largely elusive. Endogenous drivers, such as transposable elements, have been postulated to shape genome-wide dominance and biased fractionation leading to a conserved least-fractionated (LF) and a degenerated most-fractionated (MF) subgenome. In contrast, the role of exogenous factors, such as those induced by environmental stresses, has been overlooked. A chromosome-scale assembly of the alpine Buckler Mustard (*Biscutella laevigata*; Brassicaceae) that underwent a WGD event about 11 million years ago is here coupled with transcriptional responses to heat, cold, drought and herbivory to assess how gene expression is associated with differential gene retention across the MF and LF subgenomes. Counteracting the impact of transposable elements in reducing the expression and retention of nearby genes across the MF subgenome, dosage balance is highlighted as a main endogenous promoter of the retention of duplicated gene products under purifying selection. Consistent with the "turn a hobby into a job" model, about one third of environment-responsive duplicates exhibit novel expression patterns, with one copy typically remaining conditionally-expressed, whereas the other copy has evolved constitutive expression, highlighting exogenous factors as a major driver of gene retention. Showing uneven patterns of fractionation, with regions remaining unbiased while others show high bias and significant enrichment in environment-responsive genes, this mesopolyploid genome presents evolutionary signatures consistent with an interplay of endogenous and exogenous factors having driven gene content following WGD-fractionation cycles.

Keywords: conditionally-expressed genes; dosage balance; environmental stress; subgenome dominance; transposable elements; whole-genome duplication.

Introduction

Cycles of whole-genome duplication (WGD) followed by diploidization have been pervasive during the radiation of eukaryotes, especially in angiosperms (Leebens-Mack et al. 2019; Jiao et al. 2011; Schranz et al. 2012; van de Peer et al. 2017). Counteracting WGD events that increase the number of co-existing genomes in the nucleus and initially results in all loci being duplicated, genome fractionation (i.e. gene loss) and dysploidy (i.e. reduction of chromosome number) gradually lead to genome downsizing and a return to a diploid-like state (Lynch and Conery 2000; Mandáková and Lysak 2018; Tank et al. 2015). Despite their contribution to the architecture of genomes, neither the molecular underpinnings of such "wondrous cycles", nor the evolutionary mechanisms driving the fate of duplicated genes are fully understood (Freeling et al. 2012; Soltis et al. 2016; Wendel 2015).

Assuming an overarching connection between gene expression levels and the strength of selection acting on them, the differential expression of genes between subgenomes resulting from WGD has been postulated to drive genome fractionation by promoting the adaptive retention of specific duplicates against the accumulation of deleterious mutations and pseudogenization (Freeling 2009; Koonin and Wolf 2010). Following WGD, constraints due to the necessary dosage balance of interacting gene products are thus expected to promote the long-term retention of numerous duplicated genes with conserved functions (Birchler and Veitia 2012), whereas the partitioning of ancestral expression patterns between duplicates (i.e. sub-functionalization) supports their retention under purifying selection. In contrast, the evolution of novel functions or expression patterns is promoted by positive selection (i.e. neo-functionalization; Birchler and Yang 2022). Despite the null hypothesis that duplicated subgenomes undergo similar rates of sequence turnover, many studies have highlighted that one subgenome (coined as dominant) commonly retains more genes following WGD and is therefore "least fractionated" (LF) compared to the other subgenome(s) which appear more degenerated with fewer genes and described as the "most fractionated" (MF) (e.g. Chalhoub et al. 2014; Garsmeur et al. 2014). Although such biased fractionation is commonly regarded as non-random, underlying processes remain elusive and rely on partially overlapping hypotheses of genome-wide dominance against loci presenting lower expression (Alger and Edger 2020; Woodhouse et al. 2014). In particular, interspersed copies of

transposable elements (TEs) are expected to reduce the expression of nearby genes (Hollister et al. 2011) and have been predicted to influence subgenome-wide expression levels, determining dominance and patterns of biased fractionation between genomes with unbalanced TE loads (Freeling et al. 2015). Although an association between TE abundance and subgenome dominance has been documented in recently formed as well as ancient polyploid genomes (e.g. Edger et al. 2017; Garsmeur et al. 2014), several counterexamples indicate that other factors may also be at play (Douglas et al. 2015; Renny-Byfield et al. 2015; Zhao et al. 2017). Furthermore, the parental legacy of TEs associated with gene expression levels was recently shown to be insufficient to explain subgenome-wide dominance in experimental allotetraploids of *Brassica* (Zhang et al. 2023).

Beyond endogenous genomic features, exogenous factors such as different environmental conditions may also be involved in promoting the differential expression of loci between subgenomes (e.g. Shimizu-Inatsugi et al. 2017) and result in biased fractionation. The possible interactions between transcriptional plasticity in response to environmental changes and genome fractionation are indeed virtually unknown (Blischak et al. 2018).

The mustard family (Brassicaceae) that includes the model plant *Arabidopsis thaliana* contains numerous examples of taxa having undergone multiple rounds of WGD and thus offers pertinent model systems to investigate drivers and consequences of post-polyploid genome fractionation (Hendriks et al. 2023; Kagale et al. 2014; Mandáková et al. 2017). On top of the family-specific paleotetraploidy event (α-WGD) that occurred some 32 million years ago (mya; Hohmann et al. 2015) and left several duplicated genes in all extant diploid genomes of Brassicaceae, the genus *Biscutella* comprising more than 50 species of annual herbs or perennial shrublets has rapidly radiated across the Mediterranean basin following a presumably shared mesotetraploidy event (Geiser et al. 2016). Although this WGD event was associated with hybridization between two closely related, structurally similar genomes and resulted in a functionally redundant ancestral karyotype of *Biscutella* (Guo et al. 2020), comparative chromosome painting coupled with transcriptomics in Buckler Mustard (i.e. *B. laevigata*; x = 9) has shown that chromosomal segments conserved as duplicates are side-by-side with loci having undergone fractionation in this mesopolyploid (Geiser et al. 2016). Although many recently active TEs were identified in this

species (Bardil et al. 2015), the lack of appropriate genomic resources to anchor genes and TEs under scrutiny to specific loci and subgenomes precluded the characterization of genome fractionation. Using an annotated chromosome-level assembly of the alpine Buckler Mustard genome (*Biscutella laevigata* subsp. *austriaca*), this work thus addresses the role of endogenous TEs and environmental factors on gene expression and long-term retention. Specifically, we (i) characterized the mesopolyploid WGD event in the context of TE activity in the Buckler Mustard, (ii) assessed biased fractionation by comparing syntenic genes in the LF and MF subgenomes in relation to their expression, TE loads and patterns of selection, and (iii) quantified how transcriptional changes in response to exogenous factors support the retention of duplicated genes and shapes genomic regions with low versus high levels of biased fractionation.

Results

Assembly and annotation of the Buckler Mustard genome

The allogamous sample of Buckler Mustard (*Biscutella laevigata* subsp. *austriaca*) here sequenced with a combination of long and short reads (total coverage of 282X; Table S1) was estimated to have a haploid genome size of 904 Mb based on the flow cytometry analysis. K-mers estimated a total size of 832 Mb and a moderate heterozygosity rate of 1.88-1.92%, which is in between the Col-0 accession of *A. thaliana* (0.22%; Kang el al. 2023) and the T16 accession of *Brassica oleracea* (up to 5.78%; Li et al. 2024), and matches estimates for typically outcrossing diploid plants such as *Arabidopsis lyrata* (i.e. 1.4% to 2.1%; Ross-Ibarra et al. 2008) as well as the sampled population of *B. laevigata* (i.e. 3.3%; Grünig et al. 2024). Following long-range scaffolding (Fig. S1), collapsing of similar scaffolds to reduce heterozygosity while retaining duplicates arising from WGD events (Fig. S1c) and gap filling, a final assembly consisting of 6,350 scaffolds with a total length of 873.75 Mb (https://genomevolution.org/coge/GenomeInfo.pl?gid=67230; N50 = 71.38 Mb; Fig. S2) was produced, showing 98.2% of complete BUSCO genes (i.e. 78.1% single copy and 20.1% duplicated; Table S2). The Hi-C contact map was manually curated and scaffolded into chromosomes that showed a band of high contact density along their diagonal, reflecting the well-ordered underlying assembly submitted as a supplementary dataset (Figure S2.b, Dataset S1). Synteny comparisons with the ancestral genomic blocks of Brassicaceae (Lysak et al. 2016) and comparative chromosome painting confirmed the structure of the 13 largest scaffolds and supported their

arrangement into the nine nearly-complete chromosomes (total length: 764.46 Mb; Fig. 1a). The chromosome structure of Ba5 and Ba6, characterized by the combination of genomic blocks O+P+W+R and an inactive paleocentromere (Fig. S3), align with chromosome AK6/8 of the ancestral Proto-Calepineae Karyotype (n = 8), identified as the common ancestor of the tribe Biscutelleae (Guo et al. 2020). With its largely duplicated blocks, the assembly presented significant synteny with the closely related *B. laevigata* subsp. *varia* (Geiser et al. 2016), including two interstitial 5S rDNA loci identified at the pericentromeric heterochromatin of chromosomes Ba1 and Ba4, and two terminal 35S rDNA loci on chromosomes Ba2 and Ba3 (Fig. S3).

Annotated transposable elements (TEs) comprised 539.14 Mb, encompassing 68.63% of the nine chromosome-scale scaffolds (Table S3), with a majority (ca. 40%) of long terminal repeat (LTR) retrotransposons. Among them, a total of 2,993 full-length copies were identified (Table S4) among the main lineages of Copia (e.g. 864 Ale, 301 Ivana) and Gypsy (e.g. 293 Athila, 21 Tekay) which are predicted to have been active recently in Brassicaceae (Zhou et al. 2021). The distribution of TEs supported the structure of the assembly, with a higher abundance of LTR-retrotransposons located towards the centromeric regions (Fig. 1a). Two centromeric tandem repeats were identified (213- and 468-bp; Fig. S3). *Ab initio*-predicted genes supported by RNAseq data from seven tissues under mesic environments as well as from leaf tissue of clone plants subjected to cold, drought, heat and herbivory resulted in the high-quality annotation of 43,632 gene models, and 86.3% of the complete BUSCO genes. Despite the basal split of Biscutelleae from other Brassicaceae clades, here dated to be c. 20 mya based on the plastid phylogeny (Fig. 1b), at least 13,221 orthogroups (23,247 genes) presented a clear ortholog in all or all-but-one of the representative genomes of Brassicaceae (Fig. S4; Table S5). Together with 864 such orthogroups being absent from the *B. laevigata* assembly, these genes were classified as conserved orthologs across Brassicaceae.

Main molecular drivers of mesopolyploid genome evolution

The mesopolyploid nature of the Buckler Mustard genome seen in the structure of the assembly and the karyotype was confirmed by a peak in the distribution of synonymous substitutions (Ks)

among paralogs at 0.34 (± 0.15 standard deviation) in addition to the peak around 0.96 (± 0.42) that is indicative of the α-WGD event shared with *A. thaliana* (Fig. 1c + Fig. S5). Consistent with prior estimates, the younger meso-WGD event was dated to be younger than 11.5 mya. To assess whether it promoted the concomitant activation of TEs according to the genome-shock hypothesis or subsequently supported effective transposition due to relaxed selection (Parisod et al. 2010), we dated the amplification of TE copies by estimating their divergence from consensus sequences in the main TE lineages of the Buckler Mustard genome (Maumus and Quesneville 2014). Identified peaks indicative of transposition bursts were observed to range from 10 to 6% sequence divergence and indicated the ongoing transposition of several LTR retrotransposons between 10 and 5 mya (Fig. 1d). The lower divergence of TE copies than duplicated genes is consistent with a more recent onset of TE proliferation than the mesopolyploid WGD event and hence high TE dynamics during the early stages of genome fractionation in *B. laevigata*.

Synteny between the *B. laevigata* assembly and sequences of *A. thaliana* identified 14,923 genes retained in the high-confidence chromosomal segments of syntenic duplicates derived from the meso-WGD event among a total of 122 non-overlapping windows spanning 81.2% of the nine main scaffolds (i.e. chromosomes; Fig. S6). Downstream analyses were performed on these syntenic windows, which were shown to be mostly shorter than 2 Mb (median length 1.58 Mb; Fig. S6b), to ensure that our conclusions were largely unaffected by potential misassemblies that were shown to possibly span genome segments ≥ 7 Mb (Fig. S2c). These duplicated windows included 6,436 duplicates in 3,218 pairs (Fig. 2a, Table S6) that were assigned to the LF and the MF subgenomes based on corresponding gene trees with high node support (Fig. S7). Beyond these retained duplicates representing 43.1% of the genes among syntenic windows, the LF and MF subgenomes presented a total of 8,542 and 6,381 intact genes, respectively, supporting differential gene retention which is consistent with biased fractionation.

To further assess the underpinnings of genome evolution, we turned to expression data and confirmed the predicted association between expression and selection by showing that expressed genes had lower ratios of non-synonymous (Ka) per synonymous substitution rates (Ks) than genes considered to be unexpressed (Wilcoxon test, $p < 0.001$). The maximum level of gene expression was shown to be linearly associated with Ka/Ks values (slope = -0.10, $p < 0.001$, Fig. S8), matching

the expectation that highly expressed genes are conserved under stronger purifying selection (i.e. lower values of Ka/Ks). The majority of genes under scrutiny in the Buckler Mustard indeed showed conserved coding sequences under pervasive purifying selection (i.e. 14,745 genes with Ka/Ks < 0.79 as compared to *A. thaliana*). Despite similar signals of purifying selection across both subgenomes, pairs of duplicated genes presented significantly lower Ka/Ks ratios than genes that had returned to singleton state and this difference was particularly pronounced in the LF subgenome (Fig. 2b; $p < 0.001$).

Addressing how expression and selection shaped gene retention between subgenomes, we characterized the expressed genes in the LF (6,771) compared with the MF subgenome (4,920; z-test, $p < 0.01$) and showed significantly higher gene expression in the LF than in the MF subgenome (Wilcoxon test, $p < 0.001$; Fig. 2c and Fig. S9). As expected, expressed genes in the MF subgenome accordingly showed stronger signals of purifying selection than those in the LF subgenome (average Ka/Ks of 0.170 and 0.175, respectively; Wilcoxon test, $p < 0.01$), suggesting that high gene expression associated with strong purifying selection has been necessary to support gene retention in the MF subgenome.

In contrast to genes, the density of TEs was significantly lower in the LF subgenome than in the MF subgenome (Wilcoxon test p-value < 0.001; Fig. 2d). Given that the presence of TE copies up to 2,000 bp upstream or downstream of genes was significantly associated with their lower levels of expression (Fig. S10 and Table S7), such differential TE load likely played a role in shaping gene expression and hence selection across the LF and MF subgenomes. The presence of a TE copy within 201 to 2000 bp was seen to reduce the median expression of nearby genes by 28.14% (1.83 $\log_2$ fold change), as expected by their epigenetic silencing locally affecting flanking loci. Although such indirect effects of TEs may have contributed to the long-term biased fractionation under pervasive purifying selection, the observed association between TE density and gene retention does not exclude other drivers of fractionation or other triggers of gene expression across subgenomes as drivers of fractionation.

Environmental triggers of duplicate retention following WGD

To what extent conditional gene expression in response to environmental cues has been shaping genome fractionation was first assessed by inspecting the 977 single-copy and 760 duplicated genes from 966 pairs that were differentially expressed (DE) in response to experimental cold, heat, drought and herbivory treatments out of the 14,923 syntenic genes under scrutiny in the LF and MF subgenomes (Fig. 3a; Fig. S10). Noticeably, the Ka/Ks ratios of DE genes were significantly lower than for genes with other expression patterns (Wilcoxon tests, $p < 0.001$; see Fig. 3b), indicating that coding sequences conditionally-expressed in response to environmental cues were more likely to be retained under stronger purifying selection than constitutively-expressed ones. It is notable that the patterns of selection in DE duplicated genes were consistent with retention under even stronger purifying selection than DE single copy genes (i.e. Ka/Ks of 0.155 and 0.176, respectively; Wilcoxon test, $p < 0.001$), indicating that abiotic and biotic cues promoted the long-term adaptive retention of duplicates in this mesopolyploid genome. Matching the genome-wide pattern of fractionation, retained DE genes were significantly more numerous in the LF than the MF subgenome, as shown for conserved orthologs across Brassicaceae in Fig. 3c.

By cross-matching patterns of expression and signals of selection in the Buckler Mustard, we assessed evolutionary underpinnings of 2,196 expressed pairs of retained duplicates across syntenic windows, considering their orthologs in *A. thaliana* as "progenitor singletons" (Table S6). Consistent with the advanced fractionation of the mesopolyploid genome, only eight pairs of duplicated genes (0.36%) presented signals of neutral divergence ($0.79 < Ka/Ks \leq 1.21$) for either both or one member of the pair and were hence possibly retained without selection. Only four pairs of duplicates (0.18%) presented one member with a clear signal of positive selection ($Ka/Ks > 1.21$) pointing to possible neofunctionalization, while the other member was retained under purifying selection.

In contrast, the vast majority of expressed duplicates retained in the Buckler Mustard presented both copies under purifying selection (99.4%). Considering their expression in response to cold, heat, drought and/or herbivory treatments to highlight possible changes in the environmental trigger(s) compared with their progenitor singletons, almost two thirds of the retained duplicates in the Buckler Mustard showed either no environmental trigger (1,367 pairs) or were both

differentially expressed under similar conditions compared with their progenitor singletons (33 pairs) and were thus considered consistent with retention of conserved function under dosage balance constraints (63.8% of expressed duplicates). The remaining third of duplicates retained under purifying selection (796 pairs; 36.2%) showed a change in response to an environmental trigger compared to the progenitor singletons. A total of 296 duplicate pairs showed both members having lost their ability to respond to the environmental trigger(s), whereas 27 and 31 pairs presented one member with a conserved environmental trigger on the LF and MF subgenome respectively, with the other member being constitutively expressed. Nevertheless, the majority of the expressed duplicates (i.e. 391 pairs) presented gain(s) in environmental trigger(s) in the Buckler Mustard, affecting either one or both members and indicating considerable expression repatterning during fractionation.

Pointing to dosage balance as the chief constraint driving the retention of duplicated genes, our results also highlight the importance of expression changes among environment-responding duplicates following fractionation. Expression changes chiefly resulted in the constitutive expression of one or both member(s) and hence increased dosage of conserved coding sequences that were ancestrally stress-responding among retained duplicates in the Buckler Mustard. A similar evolutionary response of constitutive expression was apparent beyond the duplicated segments under scrutiny here, with most of the 2,098 progenitor singletons previously shown to be environmentally responsive in *A. thaliana* being identified as singleton (1,032, 49.2%) or retained as duplicate (353) presenting constitutive expression in the Buckler Mustard. Among the 222 progenitor singletons retained as environment-responsive duplicates in *B. laevigata*, only 57 had both members responding to at least one environmental treatment, while 70 presented constitutive expression of one member. Such canalization of ancestrally environment-responsive genes towards constitutive expression is consistent with the "turn a hobby into a job" model (Conant and Wolfe 2008) and likely promoted the increased tolerance of the Buckler Mustard to the stressful conditions that are typical of alpine environments where it currently thrives.

Biased retention of environment-responsive genes during genome fractionation

To further assess how the environment-responding genes evolved across subgenomes, we partitioned the 122 windows of syntenic duplicates into (i) 36 regions comprising those with "low-bias" (i.e. regions with non-significant differences in the proportion of retained genes between the LF and MF subgenome), and (ii) 86 "high-bias" regions (i.e. regions characterized by a significantly reduced proportion of retained genes in the MF subgenome compared with the LF subgenome; Chi-square test, $p < 0.05$; Fig. 4a). "Low-bias" regions were shown to contain duplicated genes characterized by significantly lower divergence than in "high-bias" regions (Fig. 4b) and they did not differ in TE density, unlike "high-bias" regions (Fig. 4c). These "low-bias" regions are therefore considered to have undergone limited TE-driven biased fractionation and possibly may have had prolonged exchanges between subgenomes that supported unbiased fractionation in the absence of differential TE load. Overall, "high-bias" regions of the LF subgenome were significantly enriched in constitutively-expressed duplicates (Fig. 4d).

The retention of genes responding to environmental triggers was otherwise consistent across "high-bias" regions, showing pervasive conservation of DE and constitutively-expressed genes on the LF subgenome. Biased fractionation hence supported the retention of specific environment-responding genes, as seen across the genomic block U (Fig. 4e) that is duplicated on chromosomes Ba1 and Ba2 and whose "high-bias" region on the LF subgenome appeared significantly enriched in DE genes responding to cold (Gene ratio: 58/724, qvalue < 0.01; Table S8) and herbivory (Gene ratio: 36/724, qvalue < 0.01). Genes related to the isopentenyl diphosphate biosynthetic process (GO:0019288, comprising 15 genes; Table S9) were notably abundant across that segment of the LF subgenome showing an enrichment of KEGG terms related to plant-pathogen interactions (comprising eight genes; Table S10) and suggesting specialization of the locus in terpenoid biosynthesis such as reported in a previous study of the Buckler Mustard (Knauer et al. 2018). Genome fractionation in such regions is hence biased towards the retention of genes essential for survival under harsh conditions that only polyploids can harness through an abundance of gene copies being adaptively sorted (van de Peer et al. 2021).

Among the 14,085 orthogroups conserved in all but one of the considered Brassicaceae species, 93.8% were present in *B. laevigata*, supporting the necessary presence of most genes in each

progenitor genome. Among the syntenic windows, the LF subgenome indeed presented significantly more of these conserved genes (i.e. 58.0%) than the MF subgenome (i.e. 42.0%; p-value < 0.001, Fisher's exact test) and a similar enrichment was also reflected among singletons (i.e. 64.1% in the LF subgenome compared to 51.7% in the MF subgenome; p-value < 0.001, Fisher's exact test). This pattern held true for the specific genomic block U, in which the LF subgenome harboured a significantly higher proportion of conserved genes (59.1%, p-value < 0.001, Fisher's exact test), among which were a higher percentage of singletons (60.3%) compared to the MF subgenome (45.9%, p-value < 0.001, Fisher's exact test). Such overall and locus-specific enrichments of conserved duplicates and singletons in the LF subgenome strongly support that progenitor genomes have contributed similar sets of genes and, despite possible subtle differences in their regulatory circuits before hybridization, had undergone post-WGD sorting that chiefly shaped the two subgenomes of *B. laevigata*.

Discussion

The mesotetraploid genome of the Buckler Mustard originated by WGD coupled with hybridization between two closely related progenitors which contributed similar gene sets before the polyploid genome started to undergo diploidization over about 11.5 million years in association with descending dysploidy towards 9 pairs of chromosomes (Geiser et al. 2016; Guo et al. 2020). Despite bioinformatic challenges arising from aiming to reduce heterozygosity while maintaining WGD-derived duplicates in a haploid assembly and those inherent to reconstructing the history of polyploids and their long-extinct progenitors (Kellogg 2016), the outcomes of long-term biased fractionation are still visible across the majority of 122 duplicated segments in the mesotetraploid genome today. Although our results appear consistent with predictions of TE-driven subgenome dominance (Alger and Edger 2020), 36 of these duplicated segments (27.0%, spanning 241 Mb) actually show unbiased fractionation suggestive of locus-specific rather than (sub)genome-wide drivers. Further, contrasting with the legacy of progenitor TEs determining subgenome dominance, our data show that the proliferation of several types of TEs took place during the early stages of genome fractionation in the Buckler Mustard and instead support the prediction that relaxed selection on the initially redundant loci cumulatively fostered the biased

genomic divergence towards a least and a most fractionated (LF vs MF) subgenome (Bird et al. 2018; Woodhouse et al. 2014). Although the exact TE composition of the long-extinct progenitors is unknown and their role in driving subgenome dominance immediately after WGD remains elusive, the partially biased fractionation of the mesopolyploid genome of *B. laevigata* appears consistent with runaway pseudogenization coupled with the loss of lowly expressed genes that could only have been antagonized by strong selection resulting in the retention of highly expressed genes, including duplicates mostly constrained by dosage balance (Blanc and Wolfe 2004).

Here we show that, in addition to the role of endogenous factors such as TEs and genes involved in dosage balance, exogenous factors, i.e. different environmental conditions driving the conditional expression of genes (as shown in some previous studies; e.g. Shimizu-Inatsugi et al. 2017; Lee and Adams 2020) have also substantially contributed to genome fractionation. Despite the many challenges inherent to distinguishing the partitioning of ancestral functions from gain(s) of novel environmental triggers (Birchler and Yang 2022; Innan and Kondrashov 2010), numerous ancestral environment-responsive genes with conserved coding sequences were identified as having promoted increased dosage through the evolution of constitutive expression and/or the retention of duplicates as a pervasive outcome of long-term fractionation. Although expression changes can be expected to evolve neutrally through time (Khaitovitch et al. 2004), transcriptional plasticity in response to environmental conditions was generally retained by only one member of the duplicate pair, with the other showing constitutive expression that likely supported general survival under stressful environmental conditions (Conant and Wolfe 2008; van de Peer et al. 2021). While such co-option of transcriptionally plastic genes that promoted constitutive adaptation to exogenous factors may have been instrumental in shaping the current mesopolyploid genome, it likely imposed costs and hence may contribute to explaining the slow growth of the perennial *B. laevigata* under alpine conditions. Connections between WGD *per se* and stressful conditions in the short-term remain elusive, although insights from our analysis of Buckler Mustard's mesopolyploid genome point to post-WGD fractionation and particularly the retention of environment-responsive duplicates coupled with expression changes as key to their possible radiation across harsh environments (Dodsworth et al. 2016). Although here we have

unravelled plausible mechanisms linking WGD and increased stress tolerance that have operated over millions of years of evolution, future work using experimentally resynthesized and recently established polyploids will be needed to address how genome fractionation unfolds through time and affects the fate of duplicated genes from the initial WGD event to the highly-fractionated mesopolyploid genomes entering new rounds of WGD (Bird et al. 2020; Soltis et al. 2016; Parisod 2024).

Materials and Methods

This section gives a summary of the methodology, which is detailed in the Supplementary Methods.

Plant material, sequencing, assembly and annotation

The same individual sample of *Biscutella laevigata* subsp. *austriaca* grown from a seed collected near Schneealpe (Steiermark, Austria: 47.6968°N, 15.6100°E; 1740 m asl) was used throughout, from *de novo* genome assembly to RNAseq data, using regenerated cuttings (i.e. clonal ramets).

The genome size was estimated by flow cytometry and high molecular weight DNA was sequenced with short Illumina 10X genomics linked reads (75X) which has been shown to produce reliable assembly in maize, a species that also went through WGD some 5–12 million years ago (Visendi 2022). Linked reads dataset was complemented with a combination of long Pacbio reads (12X) of an average length of 5.3 kb, and paired-end reads (75X; Table S1). The hybrid assembler Platanus-allee, which marks better performance in highly heterozygous genomes (Kajitani et al. 2019), produced a draft genome that was scaffolded by using ChicagoTM (52X coverage) and Hi-C (68X coverage) methods (Dovetail Genomics, Santa Cruz, CA). Hi-C maps may contain errors or inaccuracies that were carefully evaluated and, combined with evidence from cytogenetic maps, refined to ensure a more accurate genomic assembly. K-mers (k=21) were counted using Jellyfish, and the resulting histogram was analyzed with GenomeScope2 (Ranallo-Benavidez et al. 2020) to estimate genome size, heterozygosity, and repeat content. After removal of uncollapsed haplotigs and gap filling, the completeness of the final assembly was assessed with the BUSCO from embryophyte odb10. Merqury (Rhie et al. 2020) was then employed to compare the heterozygous

k-mer content before and after removal of uncollapsed haplotigs. Curation of the 13 largest scaffolds into the 9 main chromosome-level scaffolds was further validated through comparative chromosome painting as described in Geiser et al. (2016).

Repetitive elements across the assembly were first identified based on TE structural features using EDTA (Ou et al. 2019). The dynamics of TEs were estimated based on the percentage of divergence of each copy to the consensus according to Maumus and Quesneville (2014) and dated using 8.22×10^{-9} substitutions/synonymous site/year for Brassicaceae species (Kagale et al. 2014).

Genes were annotated using *ab initio* and mapped RNA-seq reads from seven tissues (i.e. roots, young leaves, senescent leaves, stems, apical meristem, floral buds, and open flowers; European Nucleotide Archive accession: PRJEB48599) and leaf tissues under different environmental conditions (see *Gene expression in response to environmental changes*) as well as Swissprot protein sequences from Viridiplantae used as homology-based support. Only annotations with an edit distance below 0.5 and coding for proteins longer than 20 amino acids were considered.

Gene expression in response to environmental changes

Replicated leaf transcriptomes in response to environmental treatments (European Nucleotide Archive accession: PRJEB48469) were generated from clones of the sequenced individual subjected to control (22°C, 16/8h light/dark cycle), cold (24h at 4°C, 16/8h light/dark), heat (3h gradual increase from 22-42°C and 6h at 42°C), drought (11.5 days without watering) and herbivory condition (30h of feeding by the moth *Plutella xylostella*). Those treatments were designed to mimic data available for *Arabidopsis thaliana* (Dubois et al. 2017; Klepikova et al. 2016; Nallu et al. 2018) as the only other plant species whose transcriptional responses to several environments has been investigated.

Gene expression was quantified using RSEM (Li and Dewey 2011), with only genes expressed at > 1 transcript per million considered as "expressed". Differentially expressed genes (DEGs) presenting a log2-fold change > 1 were identified using edgeR (Robinson et al. 2010).

Analysis of duplicated chromosome segments

The "SynMap" algorithm within CoGe (https://genomevolution.org/CoGe/GEvo.pl) identified duplicated genes from the mesopolyploid WGD event through collinearity within the Buckler Mustard genome and with ancestral genomic blocks of Brassicaeae extracted from *A. thaliana*. Following Woodhouse et al. (2011), windows of syntenic duplicates were seeded with ten collinear genes, comprised 100 *A. thaliana* genes and corresponding duplicates of the Buckler Mustard genome, with a maximum of 20 non-syntenic genes to be considered.

Windows of syntenic duplicates were assigned to the two sub-genomes according to RAxML phylogenetic trees of orthologous coding sequences from *A. thaliana*, *Megadenia pygmaea* and the sister genus *Heldreichia bupleurifolia* and accordingly classified as "Least Fractionated" (LF) and "Most Fractionated" (MF) following Guo et al. (2020).

SynMap further determined synonymous substitution rates (Ks) and non-synonymous substitution rates (Ka) as compared to *A. thaliana* orthologs. Approximate Gaussian distributions of Ks between duplicates marking WGD events were detected by mixture models using mixtools (https://github.com/dsy109/mixtools). The α-WGD event (mean Ks = 0.96) dated at 32.42 mya (Hohmann et al. 2015) was used as a calibration point to estimate the minimum age of the mesopolyploid WGD event.

A signal of selection was assessed using Ka/Ks values as compared to orthologous loci in *A. thaliana*, considering genes to be under purifying selection when Ka/Ks ≤ 1-SD (i.e. ≤ 0.787), neutral when 1-SD < Ka/Ks ≤ 1+SD, and under positive selection when Ka/Ks > 1+SD (i.e. > 1.213).

Duplicated windows from each subgenome were further partitioned into "low-bias" regions presenting quasi-unbiased fractionation (i.e. with similar number of retained syntenic genes in the MF and LF subgenomes) and "high-bias" regions undergoing heavily biased fractionation (i.e. significantly different number of retained syntenic genes between the MF and LF subgenomes) based on chi-squared tests (non-significant difference in proportion of retained genes in MF and LF (p-value > 0.05) classified as "low-bias", and significant difference classified as "high-bias").

Acknowledgments

We thank Noa Parisod for her help during field work, Christopher Ball and Jasmin Sekulovski for heir help with plant cultivation, and Viona Ernst as well as three anonymous reviewers for valuable comments on the manuscript. Sequence data were produced at the Next Generation Sequencing Platform of the University of Bern. Hi-C and Chicago data were produced by Dovetail. This research was funded by the Czech Science Foundation (project no. 21-07748L to MAL) and the Swiss National Science Foundation (Grants 31003A_178938 and 310030L_197839 to CP).

Author Contributions: RRC, MB, and CP designed the research; RRC collected and analyzed genomic data; MB collected and analyzed transcriptomic data; IL collected and analyzed flow cytometry data; TM and MAL collected and analyzed molecular cytogenetics data; SG collected and analyzed phylogenetic data; MP analyzed HiC data; MB, RRC and CP integrated and interpreted data and drafted the manuscript.

Competing Interest Statement: The authors declare no competing interests.

Data and Resource Availability: Raw sequence data are available on the European Nucleotide Archive repository (https://www.ebi.ac.uk/ena/browser/home), as follows:

Genomic DNA sequence data: PacBio long-reads (SRR26423064), Paired-end Illumina short reads (SRX8787129), 10X genomics linked reads (SRX8815186), Chicago Illumina short reads (SRR26396391), Hi-C Illumina short reads (SRR26404274).

Transcriptomic data: RNAseq among tissues (ERP132985), RNAseq among environmental treatments (ERP132838).

The genome assembly is available at: https://genomevolution.org/coge/GenomeInfo.pl?gid=67230

References

Alger EI, Edger PP. 2020. One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Current Opinion in Plant Biology*. 54:108–113.

Bardil A, Tayalé A, Parisod C. 2015. Evolutionary dynamics of retrotransposons following autopolyploidy in the Buckler Mustard species complex. *Plant Journal*. 82:621–631.

Birchler JA, Veitia RA. 2012. Gene balance hypothesis. Connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences of the United States of America*. 109:14746–14753.

Birchler JA, Yang H. 2022. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell*. 34:2466-2474.

Bird KA, Niederhuth CE, Ou S, Gehan M, Pires JC, Xiong Z, VanBuren R, Edger PP. 2020. Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *New Phytologist*. 230:354-371.

Bird KA, VanBuren R, Puzey JR, Edger PP. 2018. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist*. 220:87-93.

Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell*. 16:1679–1691.

Blischak PD, Mabry ME, Conant GC, Pires JC. 2018. Integrating networks, phylogenomics, and population genomics for the study of polyploidy. *Annual Reviews in Ecology Evolution and Systematics.* 49:253–278.

Chalhoub B, Denoeud F, Liu S, et al. (79 co-authors). 2014. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science*. 345:950–953.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nature reviews. Genetics*. 9:938–950.

Dodsworth S, Chase MW, Leitch AR. 2016. Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Botanical Journal of the Linnean Society*. 180: 1-5.

Douglas GM, Gos G, Steige KA, et al. (13 co-authors). 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proceedings of the National Academy of Sciences of the United States of America*. 112:2806–2811.

Dubois M, Claeys H, van den Broeck L, Inzé D. 2017. Time of day determines Arabidopsis transcriptome and growth dynamics under mild drought. *Plant, cell and environment*. 40:180–189.

Edger PP, Smith RD, Mckain MR, et al. (13 co-authors). 2017. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year- old naturally established neo-allopolyploid monkeyflower. *Plant Cell*. 29:2150-2167.

Freeling M. 2009. Bias in plant gene content following different sorts of duplication. Tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology*. 60:433–453.
Freeling M, 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*. 15:131–139.
Freeling M, Xu J, Woodhouse M, Lisch D. 2015. A solution to the C-value paradox and the function of junk DNA: The genome balance hypothesis. *Molecular Plant*. 8:899–910.
Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M. 2014. Two evolutionarily distinct classes of paleopolyploidy. *Molecular Biology and Evolution*. 31:448–454.
Geiser C, Mandakova T, Arrigo N, Lysak MA, Parisod C. 2016. Repeated whole-genome duplication, karyotype reshuffling, and biased retention of stress-responding genes in Buckler Mustard. *Plant Cell*. 28:17–27.
Grünig S, Patsiou T-S, Parisod C. 2024. Ice age-driven range shifts of diploids and expanding autotetraploids of Biscutella laevigata within a conserved niche. *New Phytologist*. in press.
Guo X, Mandáková T, Trachtová K, Özüdoğru B, Liu J, Lysak MA. 2020. Linked by ancestral bonds: multiple whole-genome duplications and reticulate evolution in a Brassicaceae tribe. *Molecular Biology and Evolution*. 38:1695-1714.
Hendriks KP, Kiefer C, Al-Shehbaz IA, et al. (43 co-authors). 2023. Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. *Current Biology*. 33(19):4052-4068.e6.
Hohmann N, Wolf EM, Lysak MA, Koch MA. 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell*. 27:2770–2784.
Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, Gaut BS. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences of the United States of America*. 108:2322–2327.
Innan H, Kondrashov F. 2010. The evolution of gene duplications. Classifying and distinguishing between models. *Nature Reviews Genetics*. 11:97–108.
Leebens-Mack JH., Barker MS, Carpenter EJ, et al. (191 co-authors). 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. 574:679–685.
Jiao Y, Wickett NJ, Ayyampalayam S, et al. (14 co-authors). 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature*. 473:97–100.
Kagale S, Robinson SJ, Nixon J, et al. (9 co-authors). 2014. Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell*. 26:2777–2791.
Kajitani R, Yoshimura D, Okuno M, et al. (7 co-authors). 2019. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nature Communications*. 10:1702.
Kang M, Wu H, Liu H, Liu W, Zhu M, Han Y, Liu W, Chen C, Song Y, Tan L, Yin K. 2023.The pan-genome and local adaptation of Arabidopsis thaliana. *Nature Communications*. 14:6259.

Kellogg EA. 2016. Has the connection between polyploidy and diversification actually been tested? *Current Opinion in Plant Biology*. 30: 25–32.
Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Pääbo S. 2004. A neutral model of transcriptome evolution. *PLoS Biology* 2: E132.
Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. 2016. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant Journal*. 88:1058–1070.
Knauer AC, Bakhtiari M, Schiestl FP. 2018. Crab spiders impact floral-signal evolution indirectly through removal of florivores. *Nature Communications*. 9:1367.
Koonin EV, Wolf YI. 2010. Constraints and plasticity in genome and molecular-phenome evolution. *Nature Review Genetics*. 11:487–498.
Lee JS, Adams KL. 2020. Global insights into duplicated gene expression and alternative splicing in polyploid Brassica napus under heat, cold, and drought stress. *The Plant Genome* 13:e20057.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12:323.
Li X, Wang Y, Cai C, Ji J, Han F, Zhang L, Chen S, Zhang L, Yang Y, Tang Q, Bucher J. 2024. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in Brassica oleracea. *Nature Genetics*. 56:517-29.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science*. 290:1151–1155.
Lysak MA, Mandakova T, Schranz ME. 2016. Comparative paleogenomics of crucifers. Ancestral genomic blocks revisited. *Current Opinion in Plant Biology*. 30:108–115.
Mandáková T, Li Z, Barker MS, Lysak MA. 2017. Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *Plant Journal*. 91:3–21.
Mandáková T, Lysak MA. 2018. Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology*. 42:55–65.
Maumus F, Quesneville H. 2014. Ancestral repeats have shaped epigenome and genome composition for millions of years in *Arabidopsis thaliana*. *Nature Communications*. 5:4104.
Nallu S, Hill JA, Don K, et al. (9 co-authors). 2018. The molecular genetic basis of herbivory between butterflies and their host plants. *Nature Ecology and Evolution*. 2:1418–1427.
Ou S, Su W, Liao Y, et al. (10 co-authors). 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*. 20:275.
Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA. 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytologist*. 186:37–45.

Parisod C. 2024. Duplicated gene networks promote 'hopeful' phenotypic variation. *Trends in Genetics*. 40:109–111.
Ranallo-Benavidez T. R., Jaron K. S., Schatz M. C. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature communications*. 11:1–10.
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*. 32:1063–1071.
Rhie A., Walenz B.P., Koren S., Philippy A.M. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*. 21:245.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26:139–140.
Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS. 2008. Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. PLoS One. 3:e2411.
Schranz ME, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification. the WGD Radiation Lag-Time Model. *Current Opinion in Plant Biology*. 15:147–153.
Shimizu-Inatsugi R, Terada A, Hirose K, Kudoh H, Sese J, Shimizu KK. 2017. Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Molecular Ecology*. 26:193–207.
Soltis DE, Visger CJ, Marchant DB, Soltis PS. 2016. Polyploidy. Pitfalls and paths to a paradigm. *American Journal of Botany*. 103:1146–1166.
Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchliff CE, Brown JW, Sessa EB, Harmon LJ. 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist*. 207:454–467.
van de Peer Y, Ashman T-L, Soltis PS, Soltis DE. 2021. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell*. 33:11–26.
van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics*. 18:411–424.
Visendi, P. (2022). De novo assembly of linked reads using supernova 2.0. In Plant bioinformatics: Methods and protocols (pp. 233-243). New York, NY: Springer US.
Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *American Journal of Botany*. 102:1753–1756.
Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M, Wang X. 2014. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proceedings of the National Academy of Sciences of the United States of America*. 111:5283–5288.
Woodhouse MR, Tang H, Freeling M. 2011. Different gene families in *Arabidopsis thaliana* transposed in different epochs and at different frequencies throughout the rosids. *Plant Cell*. 23:4241–4253.

Zhang K, Zhang L, Cui Y, et al. (13 co-authors). 2023. The lack of negative association between TE load and subgenome dominance in synthesized *Brassica* allotetraploids. *Proceedings of the National Academy of Sciences of the United States of America*. 120:e2305208120.
Zhao M, Zhang B, Lisch D, Ma J. 2017. Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. *Plant Cell*. 29:2974–2994.
Zhou S-S, Yan X-M, Zhang K-F, et al. (11 co-authors). 2021. A comprehensive annotation dataset of intact LTR retrotransposons of 300 plant genomes. *Scientific Data*. 8:174.

Legends of figures

Figure 1. Assembly and annotation of the Buckler Mustard mesopolyploid genome. A) Circos plot showing (a) the nine main scaffolds (chromosomes in Mb) of *Biscutella laevigata* subsp. *austriaca* with syntenic genes shown as lines coloured and labelled by capital letters according to the ancestral genomic blocks in Brassicaceae; (b) differentially expressed duplicated genes across the least fractionated (LF) subgenome, shown as dots coloured based on expression in response to environmental treatments (blue = cold, red = heat, yellow = drought and green = herbivory); (c) differentially expressed genes across the most fractionated (MF) subgenome as in (b); (d) gene density per Mb; (e) LTR-retrotransposon density per Mb. **B)** Phylogenetic placement of Buckler Mustard among Brassicaceae and within the tribe Biscutellae based on the analysis of whole plastid genome sequences. **C)** Synonymous substitutions (Ks) among paralogs, with significant Ks peaks corresponding to the α-WGD event shown in blue and to the meso-WGD event in red. The green line indicates the Ks-based divergence between *B. laevigata* and *A. thaliana* (see Fig. S5). **D)** Dynamics of main types of transposable elements (TEs) with dated peaks indicative of transposition bursts in relation to the mean Ks values of the α-WGD and the meso-WGD events as in (C). LTR = class I long-terminal repeat retrotransposons; DNA = class II DNA transposons; DTM = mutator; DTC = CACTA; DTH = PIF-Harbinger; DTA = hAT; DTT = Tc1-Mariner; LINE = class I long interspersed nuclear element.

Figure 2. Patterns of fractionation in the least fractionated (LF) and the most fractionated (MF) subgenomes of the mesopolyploid Buckler Mustard. A) Number of retained duplicates and genes that have returned to singleton state following WGD among the 122 hi-confidence duplicated

segments, showing biased fractionation with a higher number of intact genes in the LF than on the MF subgenome. **B)** Non-synonymous (Ka) per synonymous (Ks) substitutions showing stronger signals of purifying selection among retained duplicates than singletons in each subgenome. **C)** Maximum expression levels of genes in leaf transcriptomes under cold, heat, drought and herbivory treatments showing significantly higher levels of expression in the LF than in the MF subgenome. **D)** Density of transposable elements (TE) in basepairs per Mb showing a significantly lower TE load in the LF compared to the MF subgenome. Significance of Wilcoxon test represented by adjusted p-value < 0.001 (***) and non significant (ns).

Figure 3. Expression and selection of environmentally-responsive genes in the mesopolyploid Buckler Mustard. A) Distribution of differentially-expressed (DE) genes in response to cold, heat, drought and herbivory treatments. **B)** Patterns of selection based on Ka/Ks values between duplicated and single-copy genes according to their type of expression, showing stronger conservation of coding sequences among environmentally-responsive (DE) genes and constitutively-expressed genes compared to unresponsive genes (i.e. genes that were either unexpressed or were expressed but were not DE). Significance of Wilcoxon tests represented as adjusted p-value < 0.001 (***), < 0.01 (**) and non-significant (ns). **C)** Proportions of DE genes conserved across Brassicaceae showing biased distribution across the least-fractionated as compared to the most-fractionated subgenome. Significance of Fisher's exact tests represented as adjusted p-value < 0.001 (***) and < 0.01 (**).

Figure 4. Biased retention of environment-responsive genes across regions of the least-fractionated (LF) vs the most-fractionated (MF) subgenomes. A) Analysis of the 122 windows partitioned into 36 "low-bias" and 86 "high-bias" regions showing similar versus significantly different numbers of retained duplicate genes between the LF (open circles) and MF (filled grey circles) subgenomes, respectively. **B)** Divergence based on synonymous substitutions (Ks) between duplicated genes in the "low-bias" vs "high-bias" fractionation regions. **C)** Density of transposable elements (TE) in basepairs per Mb in the LF and MF subgenomes is not significantly

different in “low-bias” regions, whereas the “high-bias” regions of the MF subgenome have significantly higher TE density than the LF subgenome. **D)** Enrichment of gene set as compared to the whole genome showing constitutively-expressed genes are significantly over-represented in the “high-bias” LF regions. **E)** Genomic segment (genomic block U) showing a pattern of biased retention of genes responding to cold (depicted in blue) and herbivory (depicted in green) within the "high-bias" LF subgenome. Constitutively-expressed genes are represented in black, while lost or unexpressed genes are depicted in grey. The first panel illustrates differentially-expressed genes related to cold and herbivory in *Arabidopsis thaliana* (labelled A. tha ortho), whereas the second and third panels show the LF and MF subgenomes in *Biscutella laevigata*, respectively. The dashed red line delineates the segment into its "low-bias" and its "high-bias" region. Significance of Wilcoxon tests represented as adjusted p-value < 0.001 (***), < 0.01 (**) and non-significant (ns).

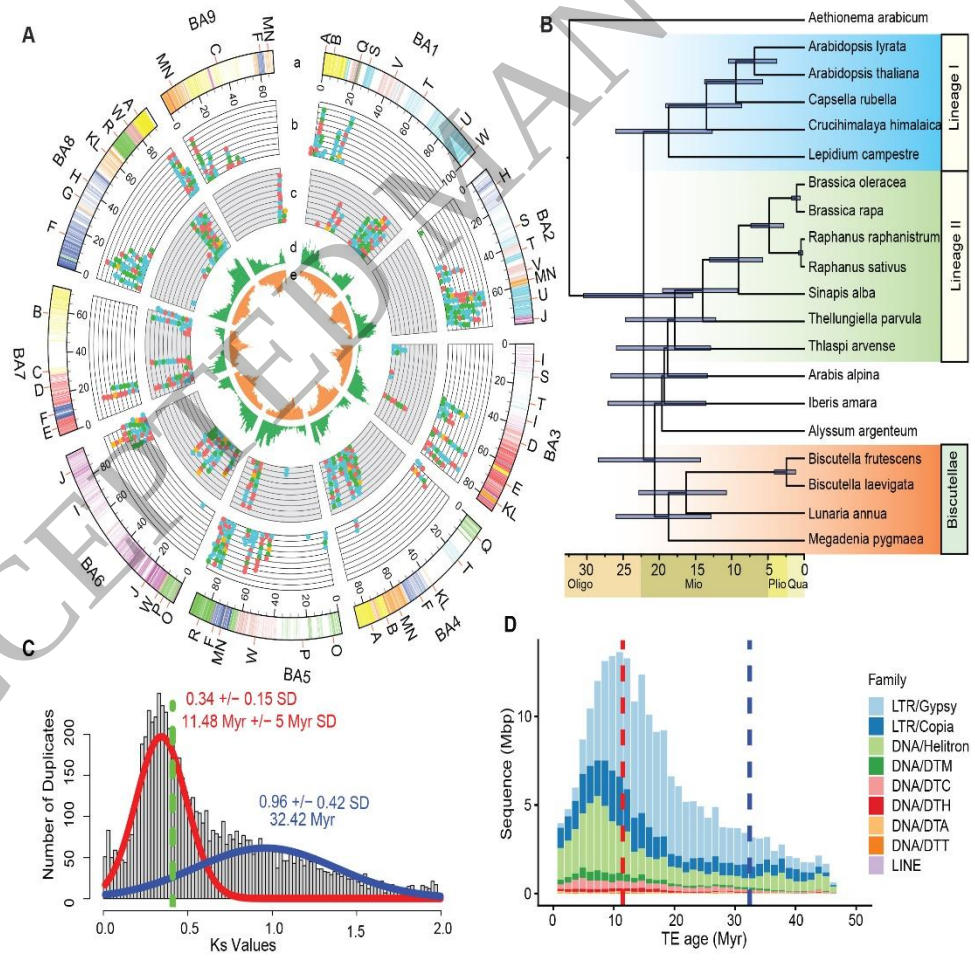

Figure 1
220x180 mm (x DPI)

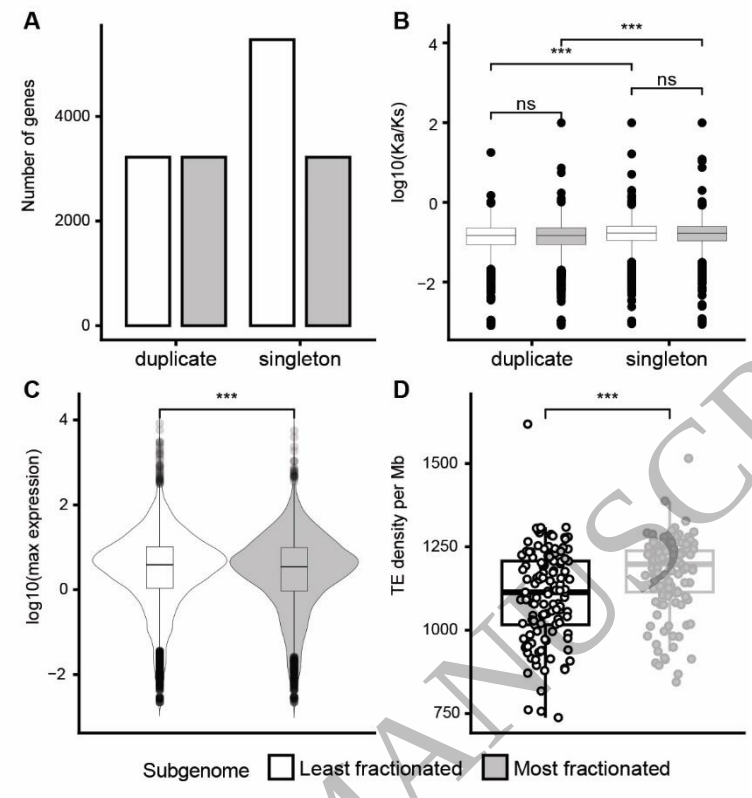

Figure 2
99x105 mm (x DPI)

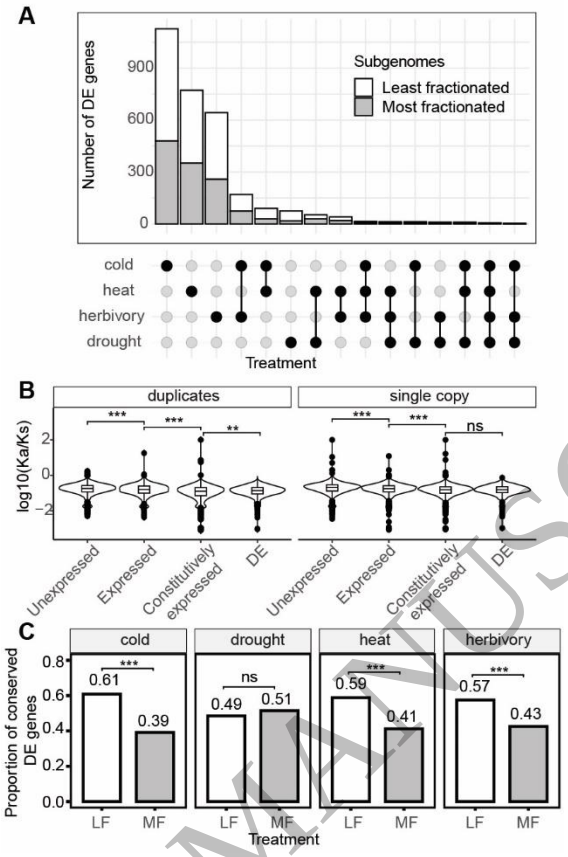

Figure 3
74x113 mm (x DPI)

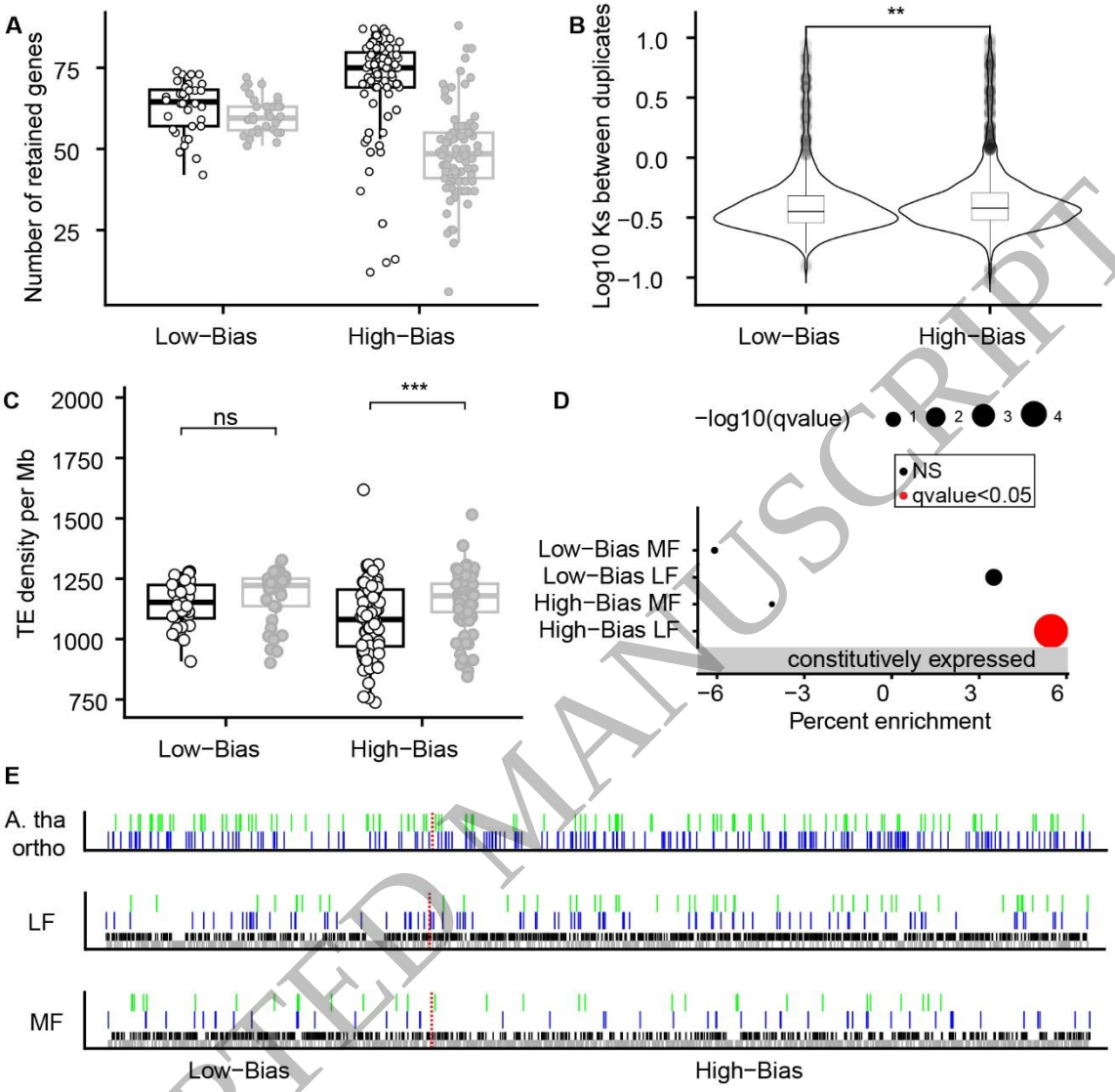

Figure 4
151x149 mm (x DPI)