Title: Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*

Authors: Kevin A. Bird[1,2], Chad Niederhuth[3], Shujun Ou[4], Malia Gehan[5], J. Chris Pires[6], Zhiyong Xiong[7,*], Robert VanBuren[1,8,*], Patrick P. Edger[1,2,*]

1. Department of Horticulture, Michigan State University, East Lansing, MI, 48824, USA
2. Ecology, Evolutionary Biology and Behavior, Michigan State University, East Lansing, MI, 48824, USA
3. Department of Plant Biology, Michigan State University, East Lansing, MI, 48824, USA
4. Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, 50011, USA
5. Donald Danforth Plant Science Center, St. Louis, MO, 63123, USA
6. Division of Biological Sciences, University of Missouri, Columbia, MO, 65211, USA
7. Key Laboratory of Herbage and Endemic Crop Biotechnology, Inner Mongolia University, Hohhot, 010070, China
8. Plant Resilience Institute, Michigan State University, East Lansing, MI, USA

* Correspondence to P.P.E. (edgerpat@msu.edu), R.V. (vanbur31@msu.edu), or Z.X. (xiongzy2003@aliyun.com).

Abstract:
One of the parental diploid genomes (subgenomes) in an allopolyploid often exhibits higher gene expression levels compared to the other subgenome(s) in the nucleus. However, the genetic basis and deterministic fate of subgenome expression dominance remains poorly understood. We examined the establishment of subgenome expression dominance in six isogenic resynthesized *Brassica napus* (rapeseed) allopolyploid lines over the first ten generations, and uncovered consistent expression dominance patterns that were biased towards the *Brassica oleracea* 'C' subgenome across each of the independent lines and generations. The number and direction of gene dosage changes from homoeologous exchanges (HEs) was highly variable between lines and generations, however, we recovered HE hotspots overlapping with those in multiple natural *B. napus* cultivars. Additionally, we found a greater number of 'C' subgenome regions replacing 'A' subgenome regions among resynthesized lines with rapid reduction in pollen counts and viability. Furthermore, DNA methylation differences between subgenomes mirrored the observed gene expression bias towards the 'C' subgenome in all lines and generations. Gene-interaction network analysis indicated an enrichment for network interactions and several biological functions for 'C' subgenome biased pairs, but no enrichment was observed for 'A' subgenome biased pairs. These findings demonstrate that "replaying the evolutionary tape" in allopolyploids results in repeatable and predictable subgenome expression dominance patterns based on preexisting genetic differences among the parental species.

Significance:

Interspecific hybridization merges evolutionarily distinct parental genomes (subgenomes) into a single nucleus. A frequent observation is that one subgenome is "dominant" over the other subgenome(s) including higher gene expression levels. Which subgenome becomes dominantly expressed in allopolyploids remains poorly understood. Here we "replayed the evolutionary tape" with six isogenic resynthesized allopolyploid *Brassica napus* lines and investigated subgenome dominance patterns over the first ten generations. We found that the same parental subgenome was consistently more dominantly expressed in all lines and generations. This suggests that subgenome expression dominance is largely predetermined based on differences among the parental diploid genomes. These findings have major implications regarding the genotypic and phenotypic diversity observed following plant hybridization in both ecological and agricultural contexts.

Introduction:

Hybridization among closely related species is a widespread and recurrent evolutionary process (1–3). By merging the genomes of independently evolved species into a single nucleus, hybridization creates a unique opportunity for immense variability that natural selection can act upon in subsequent generations (4, 5). Furthermore, hybridization is known to produce transgressive phenotypes, including heterosis, and novel phenotypic variation not observed in the parents (6, 7). However, the hybridization of highly diverged genomes, particularly those with different base chromosome numbers, can also lead to chromosome pairing issues during meiosis which greatly reduces fertility. Proper bivalent pairing of homologous chromosomes in such interspecific hybrids can be restored through whole genome duplication (i.e. polyploidization) resulting in the formation of an allopolyploid species (8). This may in part explain the high prevalence of polyploidy across flowering plants (9, 10).

Genome-scale analyses of recent and ancient allopolyploids led to the discovery that one of the parental species genomes (also known as subgenomes) often exhibits greater gene retention (11), more tandem gene duplications (12), higher gene expression (13) and lower DNA methylation (14). Collectively this phenomenon is referred to as 'subgenome dominance'. A previous study demonstrated that subgenome dominance at the gene expression level occurs at the moment of interspecific hybridization and increased over subsequent generations in the allopolyploid (15). This finding agrees with theoretical work of transcription factor binding and regulatory mismatch that predicts increasing subgenome dominance over generations in newly established allopolyploids (16). Preexisting differences between parental genomes has been shown to influence observed subgenome dynamics in an allopolyploid (17, 18). For example, analyses of diverse allopolyploids have revealed that gene expression differences among subgenomes mirrors differences in transposable element (TE) densities in flanking regions surrounding genes (12, 19, 20). These findings collectively suggest that subgenome dominance

may be largely predetermined based on subgenome differences in certain genomic features including TE densities.

Given that gene expression level dominance occurs instantly following the initial hybridization event (15), resynthesized allopolyploids are the ideal system to investigate the establishment and escalation of subgenome dominance. Few studies have used multiple independently derived resynthesized allopolyploids to investigate the variability of subgenome dominance (21–26). Thus it remains unclear the extent to which the emergence of subgenome dominance is a result of pre-existing characteristics of the diploid progenitors or due to independent and non-recurrent events during polyploid formation. In other words, will multiple independently established allopolyploids consistently exhibit the same patterns of subgenome dominance (e.g. towards the same subgenome)?

Here we analyzed subgenome dominance in six independent resynthesized allopolyploid *Brassica napus* (2n=4x=38) lines formed by hybridizing two doubled haploid parents from the progenitor species *Brassica rapa* (AA; 2n=2x=20) and *Brassica oleracea* (CC; 2n=2x=18) (27). The crop *B. napus* was formed between 7,500 and 12,500 years ago and is widely grown present-day as an oilseed crop (rapeseed), vegetable fodder crop (rutabaga) and vegetable crop (siberian kale) (28, 29). The strengths of the *B. napus* polyploid system include not only having high-quality reference genomes for both diploid progenitors and *B. napus*, but also being closely related to the model plant *Arabidopsis thaliana,* allowing for the integration of diverse genomic and bioinformatic resources (28, 30–32). Furthermore, a previous analysis of the *B. napus* reference genome identified a greater number of retained genes in the *B. oleracea* (C) subgenome compared to *B. rapa* (A) subgenome (28). This is consistent with patterns observed in older allopolyploids that exhibit subgenome dominance - dominant subgenome retaining a greater number of genes (33). Lastly, because the resynthesized *B. napus* lines were made with doubled haploids, each of the independent lines started out genetically identical (34). This permitted us to examine and compare the establishment of subgenome dominance across independently derived polyploid lines without the added influence of allelic variation segregating between different lines. We examined each of the six resynthesized polyploid lines with DNA-seq, RNA-seq, and Bisulfite-seq data to characterize gene dosage, expression, and methylation differences between high confidence homoeologs over the first ten generations. This permitted us to assess the variability of subgenome dominance during the earliest stages following allopolyploid formation.

Results:

Homoeolog Dosage Bias

This study utilized a previous generated population of 50 independently resynthesized B. napus lines, produced by hybridizing *B. oleracea* acc. TO1000DH and *B. rapa* acc. IMB-218DH. An individual from six of the fifty resynthesized lines was sequenced at the first (S1), fifth (S5), and tenth selfing generation (S10) and we analyzed the genomes of these 18 individuals to examine if changes in gene (homoeolog) dosage due to homoeologous exchanges are biased towards a particular subgenome. The six lines were selected to capture the highest and lowest amounts of pollen viability (three lines each; Table S1). The genome sequencing data was aligned to an *in silica* allopolyploid genome made by combining the reference genomes of the *B. oleracea* double-haploid parent line TO1000 (30) and a parental-SNP corrected *B. rapa* R500 reference genome. We identified 26,114 syntenic orthologs (homoeologs) between the A (BnA) and C (BnC) subgenome, and shifts in read depth coverage between these gene pairs allowed us to pinpoint changes in gene dosage across each of the six lines and over the ten generations. Changes in gene dosage occur through homoeologous exchanges, where non-reciprocal homoeologous recombination between syntenic regions of the parental subgenomes replace one homoeolog with another (35). Previous studies of this resynthesized *B. napus* population using a handful of DNA or cytogenetic markers identified extensive homoeologous exchanges that resulted in immense phenotypic variation in both plant height and pollen count (36, 37). However, the previous set of markers had limited resolution and small-scale exchanges were not identifiable. We used a whole genome resequencing approach to identify at higher resolution homoeologous exchanges that altered the relative dosage of homoeologs among individuals across this population.

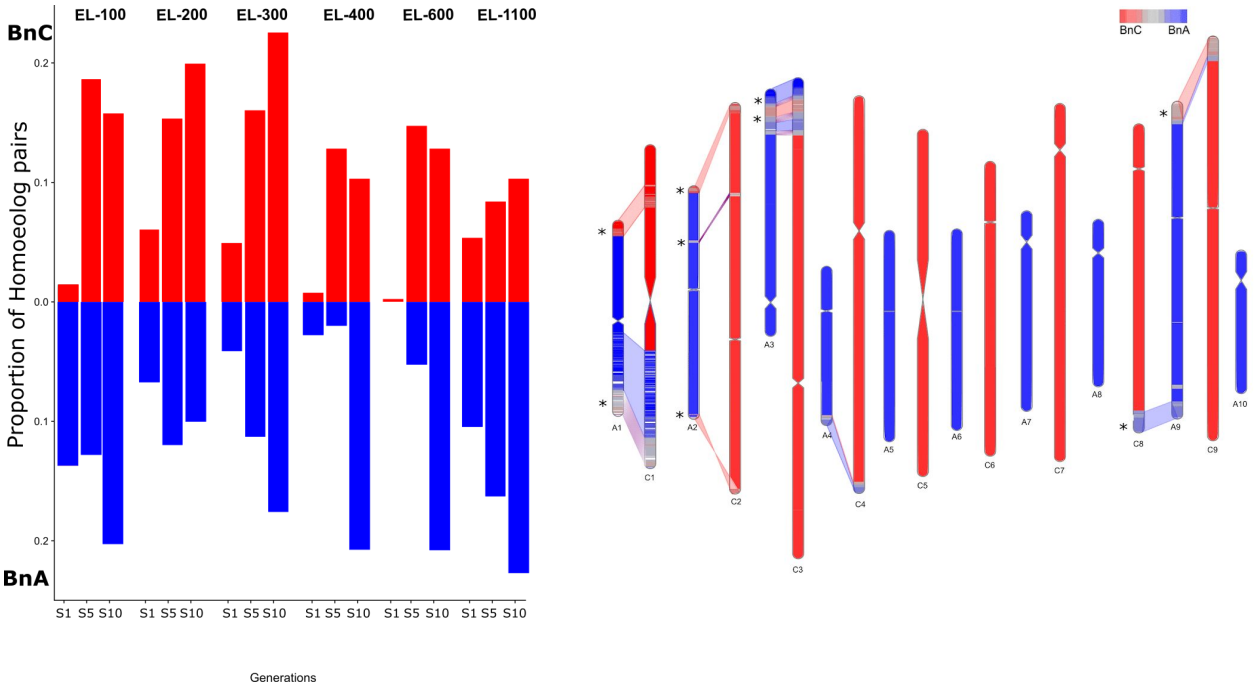


Figure 1: Homoeolog dosage changes across resyntensized lines. A) Proportion of homoeolog pairs showing ACCC or CCCC dosage (red) and AAAC or AAAA dosage (blue) for each resynthesized polyploid line (EL-100, EL-200, EL-300, EL-400, EL-600 and EL-1100) in the S1, S5 and S10 generations. **B)** Regions of dosage changes shared across all six lines. The color intensity indicates how many lines showed dosage changes in the same direction. Blue regions on C subgenome chromosomes show regions where in the majority of lines had A subgenome regions that replaced C subgenome regions, and red regions on A chromosomes represent regions where C subgenome regions replaced A subgenome regions. Grey regions were exchanged in both directions in an equal number of lines. Asterisks mark regions in common with HEs found in the natural *B. napus* lines 'Darmor' and 'Yudal' reported in Lloyd et al. (38).

The direction of homoeologous exchanges and proportion of homoeologous regions exchanged varied greatly between lines and generations with no consistent pattern significantly favoring the A or C subgenome **(**Fig 1A; Table S2; Fig S1-S18). However, in line with previous work linking fertility loss in individuals with more BnA to BnC exchange events (where BnC regions replaced BnA segments) (24), we found that the lines with rapid declines in pollen viability (EL 100, El 200, and EL 300) on average showed more BnC regions replacing BnA regions than lines with stable pollen count (EL 400, EL 600, and EL 1100) (Fig 1A; Table S1; Table S2). We saw an average of 1123 homoeolog pairs with C regions replacing A regions in unstable lines vs 585 homoeolog pairs in stable lines in the first generation, 4862 homoeolog pairs vs 2938 homoeolog pairs in the fifth generation, and 4208 homoeolog pairs vs 2991 homoeolog pairs in the tenth generation (Table S2). Individual lines ranged in the number of

exchanged homoeolog pairs from 114 to 10,231 (Table S2). Many regions are shared as 'non-exchanged' among all 6 lines in a given generation (S1=17,091; S5=8,715; and S10=5,438; Fig S19a-c), while fewer genomic rearrangements causing dosage shifts (either AAAA, AAAC, ACCC, or CCCC) are shared within a given generation (S1=62, S5=57, and S10=741; Fig S20a-c). Because of the high turnover of HEs resulted in few shared events within a single generation, we further analyzed shared gene dosage changes in any direction across all generations. We identified 3151 homoeolog pairs in these shared HE regions (Fig S20d). Parental species chromosomes with the highest collinearity (A1/C1, A2/C2, A3/C3, A9/C8:C9) exhibited the most gene dosage changes (Fig 1b), while A5, A6, A7, A8, A10, and C5, C6, C7 exhibited the fewest gene dosage changes. These shared gene dosage regions were also detected in past studies of fixed HEs identified in natural *B. napus* cultivars Darmor and Yudal (Fig 1B**;** (38)). This suggests that there are HE hotspots and that "replaying the evolutionary tape" has repeatedly yielded highly similar HE patterns at specific regions in *B. napus.*

Homoeolog Expression Bias

This population of resynthesized polyploids also provided a unique opportunity to examine if the same subgenome would repeatedly exhibit subgenome expression dominance. Gene expression was surveyed using RNAseq in sixteen of the eighteen resequenced individuals (six lines and three generations). Two samples failed during library construction, thus these were not able to be included in this analysis. However, all six lines were still able to be investigated in these set of analyses. Comparable tissues were also collected for RNAseq from the two diploid parental lines. Samples were again aligned to the aforementioned *in silica* polyploid reference genome. We restricted gene expression analyses to genomic regions with balanced gene dosage (2:2; AA:CC) for each individual to reduce the confounding factor of dosage changes in regions that have undergone HEs (Fig 1). Expression patterns of the six lines were also compared to the parental *B. rapa* and *B. oleracea* genotypes to test if expression differences may exist among the diploid progenitors.

The mean expression bias (Log2 FoldChange BnC expression /BnA expression) for homoeologs in balanced (2:2) regions ranged from 0.12 to 1.16 (median -0.50 to 0.96), with 15 of 16 individuals having mean expression bias significantly greater than 0 (One-way Wilcoxian-Mann-Whitney test, p<2.2e^-16). These results suggest a transcriptome-wide bias in favor of the C subgenome, however the magnitude of the expression bias was small. The homoeolog expression bias between the parents was also significantly greater than 0 in these balanced regions. Comparing bias difference between the parental lines and the synthetic polyploids revealed that only 5/16 were significantly different from the parents (two way Wilcoxian-Mann-Whitney test, p < 0.001).

We examined homoeolog pairs that were biased toward either subgenome (Log2 Fold change > 3.5 or < -3.5) and found there were significantly more BnC biased homoeolog pairs than expected in all individuals (χ^2 test , χ^2 > 170, df=1, p < 2.2e^16; Figure 2; Figure S21-26; Table S3). This indicates that there is a bias in homoeolog expression on a pair-by-pair basis towards the C subgenome. We also found that homoeolog bias in the synthetic polyploids was not independent of homoeolog bias in the parents (χ^2 test, χ^2=>9600, df =4, p < 2.2e^16;Figure

2; Figure S21-26; Table S4). Furthermore, a bimodal distribution was observed when comparing A and C subgenome expression; the rightmost distribution being largely due to homoeolog pairs with a lack of BnA expression.

Next we investigated whether individual gene pairs were biased in the same direction across the six lines. Due to the stochastic nature of HEs, dosage of a gene pair may differ between lines. To adjust for this, we first looked only at genes found in 2:2 dosage for all lines in a generation, resulting in 6917, 3574, and 2252 homoeologous pairs, for generations S1, S5, and S10 respectively. A majority of C biased Gene pairs were biased towards the C subgenome in all synthetic lines for each generation (S1=1806 [75%], S5=772 [71%], and S10=602 [70%]; Fig S27a-c) and in the parental comparison. In the first generation, roughly 36 gene pairs (1.5%) were uniquely dominant in only the parents and 32 (1.3%) were dominant in all six synthetic lines but not the parent (Fig S27a). In the fifth and tenth generation, there was a similar number of BnC dominant homoeologs that were dominant in only the parents (S5=17 [1.5%], S10=14 [1.6%]) and dominant in all six lines but not the parent (S5=17 [1.5%], S10=13 [1.5%] (Fig S27c). Similar patterns were observed for BnA biased homoeologs with most genes showing similar bias in all lines across generations and the parents (S1 = 698 [61%], S5 = 401 [58%], S10 = 221 [51%]; respectively; Fig S27d-g), and a consistently low number of genes biased in all six lines but not the parents in each generation (S1 = 28 [2.5%], S5 = 15 [2.2%], S10 = 10 [2.3%]) and only those biased in the parents (S1 = 41 [3.6%], S5 = 22 [3.2%], S10 = 18 [4.1%]; Fig S27d-g).

Lastly, we performed Gene Ontology (GO)(39) enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG)(40) pathway analysis, and protein-protein interaction (PPI) network analysis using STRING (Search Tool for the Retrieval of Interacting Genes)(41) to determine if BnC and BnA biased gene pairs were enriched with certain biological functions. In all generations, BnC biased gene pairs showed enrichment for interactions in the PPI network in generations S1, S5, and S10 ($p<e^{-16}$) with high average node degree (S1=20.6, S5=12.1 ,S10=7.99) and clustering coefficient (S1=0.386, S5=0.385, S10=0.385). GO term enrichment for BnC biased gene pairs related to core metabolism and photosynthesis. Cellular component GO terms related to the cytoplasm, plastid, and mitochondria, and ribosomes. BnC biased gene pairs were also significantly overrepresented in KEGG pathways annotated for amino acid biosynthesis, MAPK signalling and photosynthesis (Table S5-7). The enrichment for organellar functions for BnC biased pairs may be expected given that the C subgenome is the maternal progenitor of the resynthesized lines. This suggests that subgenome dominance may, in part, be related to maintenance of balanced nuclear-organellar interactions. On the contrary, BnA biased gene pairs showed no enrichment for any GO terms or KEGG pathways. However, there is still observed enrichment for interactions in the PPI network for BnA biased gene pairs in generations S1 ($p=9.75e^{-07}$) and S5 ($p=0.00681$), but not generation S10 ($p=0.16$). Lower network statistics were also observed for average node degree (S1=1.13, S5=0.912, S10=0.338) and average clustering coefficient (S1=0.270, S5=0.235, S10=0.209) for BnA biased gene pairs. In summary, subgenome expression dominance in *Brassica napus* are biased towards a set of highly interconnected BnC genes that are enriched for a wide variety of biological processes.

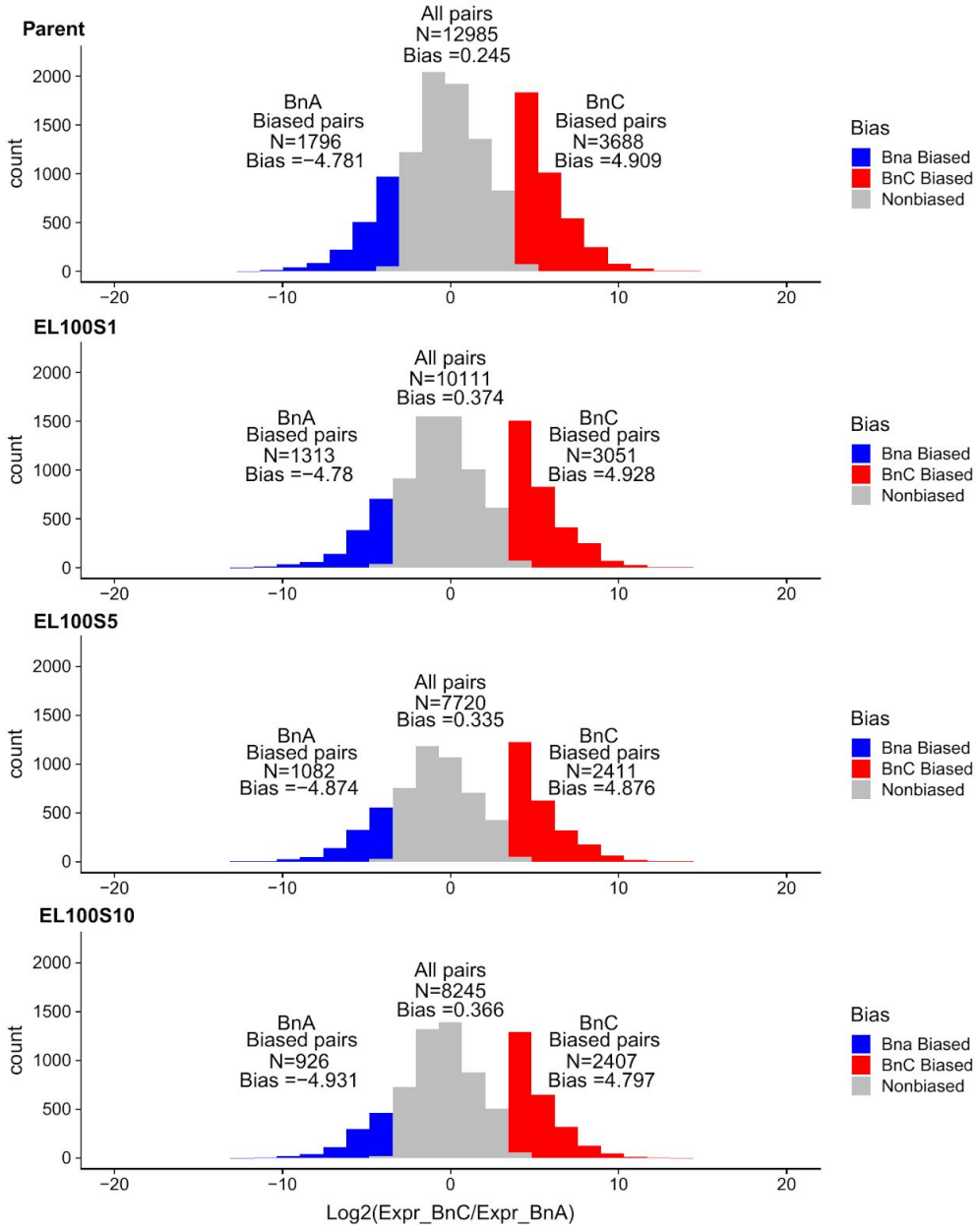


Figure 2: Homoeolog Expression Bias. Distribution of homoeolog bias in the parent and three generations of line EL100, red regions indicate BnC biased homoeologs with log2 expression fold change greater than 3.5 and blue regions indicate BnA biased homoeologs with log2 expression fold change less than -3.5.

DNA methylation

Cytosine methylations is involved in defining regions of chromatin, silencing transposons, maintaining genome integrity, and can affect gene expression. DNA methylation itself is shaped by factors such as gene expression and the underlying sequence (42) To understand how DNA methylation evolves following polyploidy and contributes to subgenome dominance, DNA methylation was assessed using whole-genome bisulfite sequencing (WGBS) in this resynthesized *B. napus* population(43, 44). In plants, three sequence contexts are generally recognized for DNA methylation, depending on second and third bases downstream of the methylated cytosine: CG (or CpG), CHG, and CHH (where H= A, T, or C). Methylation of these contexts are established and maintained by different molecular pathways and their associations with gene expression can differ based on the pattern of DNA methylation, hence they are typically analyzed separately (42).

All samples were mapped to the *in silica* polyploid reference genome to investigate homoeolog dosage and expression bias. To account for any biases resulting from mismapping to the wrong subgenome, the two parental datasets were combined to mimic the allopolyploid *in silico* (referred to as the "mock" sample) (see Methods). In all methylation contexts the C-subgenome progenitor showed higher methylation levels than the A-subgenome progenitor. When the entire genome is analyzed as a whole, CG methylation in the resynthesized lines typically fell between the two parents and showed little change in genes and Long-Terminal Repeat (LTR) retrotransposons, regardless of generation (Fig S28a, 29a). More variance in total methylation was observed for the non-CG contexts. In the first generation, CHG methylation was comparable to either parent. In generations five and ten, CHG methylation became increasingly variable (Fig S28b, Fig 29b). CHH methylation levels were even more striking. A slight increase in CHH methylation was observed in generation one and this trend became more pronounced in generations five and ten, with methylation levels surpassing the highest parent (Fig S28c, 29c).

Analyzed by subgenomes, methylation in the resynthesized lines was lower on the A subgenome (Fig S30-33) than the C subgenome (Fig 34-37). Additionally, a number of samples show increased CG methylation in genes and flanking regions for BnC genes, but little difference is observed for BnA genes (Fig S30a, S34a). In genic regions on the BnC subgenome, CG methylation increase occurred primarily for non-syntenic genes (FigS36a), however in flanking regions, CG methylation increases occurred for both syntenic and non-syntenic genes (Fig S36a, S37a). For CHG sites, methylation levels in the first generation samples again showed lower levels both in the bodies and in flanking regions of genes and LTR retrotransposons, and increased in subsequent generations of some lines in both the BnA and BnC subgenome (Fig S30b-S37b).

at the subgenome level, CHH methylation again showed the most striking differences. First generation allopolyploids showed little difference from the parents, when all genes were examined (Fig S28c). However, when examined by subgenome, BnC genes show increased CHH methylation in flanking regions (Fig S34c), while BnA genes did not (Fig S30c). In

subsequent generations, CHH methylation increased in both LTR retrotransposons and gene flanking regions (Fig S28c, S29c). For BnC genes, CHH methylation of gene flanking regions and LTR retrotransposons increased beyond the initial increase in the first generation (Fig S33c, S35c) Increased CHH methylation was not found in flanking regions of BnA genes (Fig S30c) but was observed to a lesser extent in LTR retrotransposons on the BnA subgenome (Fig S31c). These changes in flanking CHH methylation occurred for both syntenic and non-syntenic genes (Fig S32c, S33c, S36c, S37c). Additionally, we saw higher methylation of LTR retrotransposons in the dominant subgenome (BnC) than the non-dominant subgenome (BnA) (Fig S31c, S36c).

To investigate the potential roles and patterns of methylation related to subgenome dominance, we again focused on homoeologs identified in 2:2 dosage with biased expression in the previous analyses. For all lines there was higher methylation in all contexts in homoeologs on the BnC subgenome than the BnA subgenome. Additionally all lines showed higher methylation levels 2kb up- and downstream of the gene body in BnC biased homoeolog pairs in the CG, CHG, and CHH methylation context (Fig 3). Furthermore, there was a marked difference between BnA and BnC biased homoeologs at the transcription start site (TSS'), with BnC biased homoeologs showing lower CG methylation levels (Fig 3a,b). Similarly BnC biased homoeologs showed lower CHH and CHG gene body methylation levels.

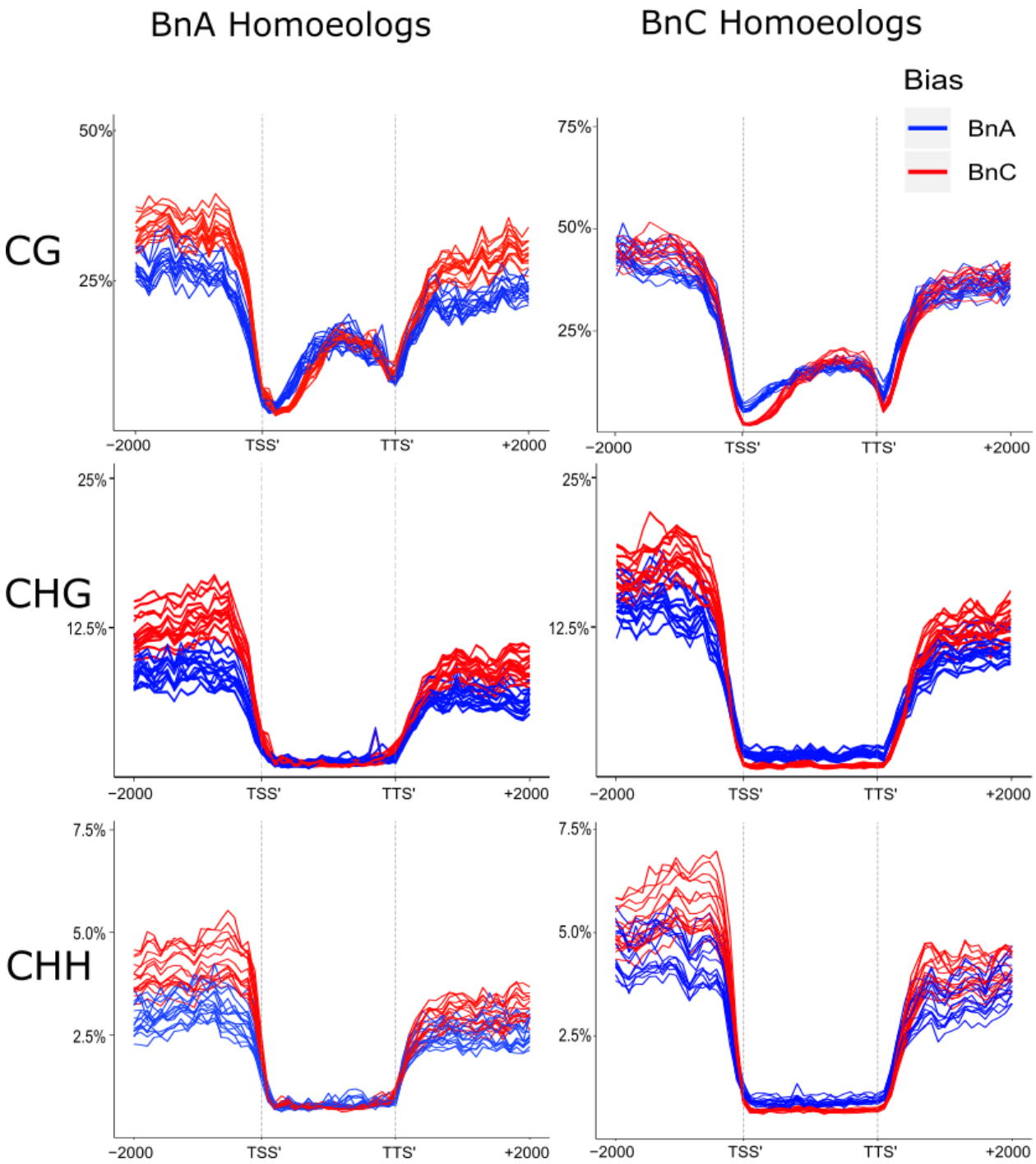


Fig 3: DNA Methylation Bias. Average weighted methylation for homoeolog pairs in 2:2 dosage for all 16 lines assessed using Bisulfite-seq. BnC biased expressed homoeologs in red and BnA biased expressed homoeologs in blue. Results are shown for methylation in CG, CHG, and CHH contexts. Methylation levels are shown for the transcription start site (TSS), gene body, transcription termination site (TTS) and 2kb up- and downstream of the TSS and TTS.

Discussion:

In summary, our findings suggest that “replaying the evolutionary tape” in *Brassica napus* yields nearly identical subgenome expression dominance patterns in the six resynthesized lines. We observed a significant expression bias toward the *Brassica oleracea* (BnC) subgenome in all six resynthesized lines, as well as comparisons of gene expression levels between the diploid parents. Thus, these results suggest that subgenome expression dominance is likely determined based on pre-existing genetic differences between the parental genomes. The gene ontology and pathway analyses revealed that BnC biased homoeologs are highly enriched with primary metabolic and organellar functions, while *Brassica rapa* (BnA) biased homoeologs were not enriched with any known biological functions.The enrichment of organellar functions towards the BnC subgenome, which is the maternally contributed subgenome, may be due to imprinting, but it is not definitive until similar analyses are repeated with reciprocal crosses. We also identified consistent DNA methylation pattern differences between subgenomes. BnC biased homeologs showed markedly lower CG methylation at the transcription start site and lower CHG and CHH methylation in the gene body compared to BnA biased homoeologs, patterns all associated with higher expression. (45). CHH methylation of LTR retrotransposons was also higher in the dominant subgenome. These patterns of biased subgenome methylation and expression were also seen in previous work in mimulus (15). It remains unclear whether these methylation patterns reflect the causal mechanism or are the result of expression differences of biased homoeologs rather than their cause, but this may be a promising new avenue to explore for improving our understanding of subgenome expression dominance.

In contrast, HE patterns across subgenomes were found to be a largely random between individuals. However, individuals with more BnA to BnC exchange events (where BnC regions replaced BnA segments) had lower pollen counts, potentially related to the meiotic instability they have been shown to introduce (24). This decline pollen viability may explain why natural and resynthesized *B. napus* lines grown in the field show more exchange events where BnA regions replace BnC regions (46). Furthermore, HE hotspots were identified in this population that are also shared with natural *B. napus* cultivars (38). Based on previous studies that HEs in nascent polyploids lead to phenotypic variability (36), variation and plasticity in resynthesized *B. napus* would appear to be the result of largely random HEs shifting the dosage of homoeologous regions with largely fixed expression bias toward a dominant subgenome. Thus, the interplay of deterministic and stochastic genomic changes give rise to immense phenotypic variation among individuals in this population that began as genetically identical.

Materials & Methods:

The resynthesized B. napus allopolyploid lines (CCAA) were obtained from a previous study (37) and grown at 23C during day and 20C at night with 16hr days in a growth chamber. True

leaf three was collected from all plants within one hour starting at 10am (4hrs into day) and immediately flash-frozen in liquid nitrogen. Leaves were split in half for RNA and DNA isolation. Total RNA and DNA was isolated using the respective KingFisher Pure Plant kits (Thermo Fisher Scientific, MA) and quantified using Qubit 3 Fluorometer (Thermo Fisher Scientific, MA). DNA and RNA libraries were prepared using the KAPA HyperPrep and mRNA HyperPrep kit protocol, respectively (KAPA Biosystems, Roche, USA). Bisulfite conversion was performed using the EZ DNA Methylation Kit (Zymo, CA) All libraries were submitted to a genomics facility (Beijing Nuohe Zhiyuan Technology Co., Beijing, China) and sequenced with paired-end 150bp reads on an Illumina HiSeq 4000 system. Pollen viability for each individual was assessed using MTT method of Rodriguez-Riano and Dafni (47).

For the “C” subgeome we used the *B. oleracea* reference genome TO1000 (30.) For the “A” subgenome the *Brassica rapa* R500 reference genome was SNP corrected by mapping paired-end 150bp genomic reads of *B. rapa* acc IMB-218 using Bowtie2 and generated a new reference with GATK. Syntenic orthologs between the two progenitor genomes were identified with MCScan (48). LTR retrotransposons were identified using LTR_Retriever v2.5 (49) as previously described (12). The IMB-218 reference genome was concatenated to the TO1000 reference genome (30) to create an *in silica* reference genome for *B.napus*. Paired-end 150bp genomic reads from the resynthesized lines were mapped to the *in silica* reference with Bowtie2 (50) and read depth of syntenic orthologs was quantified with the BedCov script from BEDtools (51). Homoeologous exchanges were identified based on skewed read depths of syntelog pairs averaged along a sliding window.

Paired-end 150bp RNAseq reads for the resynthesized lines and parents were aligned to the *in silica* reference using STAR(52). Only uniquely aligned reads were included downstream. Transcript abundance was quantified only for syntenic orthologs using StringTie (53) in units of transcripts per million (TPM). Regions of identified homoeologous exchanges and low expression were filtered out and biased homoeolog pairs were identified based on the log transformed ratio of homoeolog pair expression. Arabidopsis orthologs of biased homoeolog pairs shared across the 6 lines were tested for interaction and GO (39) and KEGG (40) enrichment using STRING (41) .

Paired-end 150bp whole-genome bisulfite sequencing was mapped to the *in silica* reference using MethylPy (54) and Bowtie2 (50). The chloroplast genome was used as in internal control for calculating the non-conversion rate of bisulfite treatment. To account for potential confounding of cross mapping we also generated an *in silico* allopolyploid by combining parental fastq files (see detailed methods in SI). Genome-wide levels of DNA methylation and DNA methylation metaplots were analyzed as previously described (42) using python v3.7.3. Pybedtools v (55) and Bedtools v2.25.0 (51). All plots and statistics were done in R v3.6.0 (56), plots made using ggplot2(57). All code and original analyzed data and plots are available on Github (https://github.com/niederhuth/Replaying-the-evolutionary-tape-to-investigate-subgenome-dominance).

References:

1. P. S. Soltis, D. E. Soltis, The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology* **60**, 561–588 (2009).

2. M. L. Arnold, A. Meyer, Natural hybridization in primates: one evolutionary mechanism. *Zoology* **109**, 261–276 (2006).

3. J. Mallet, Hybrid speciation. *Nature* **446**, 279–283 (2007).

4. E. Anderson, G. L. Stebbins, HYBRIDIZATION AS AN EVOLUTIONARY STIMULUS. *Evolution* **8**, 378–388 (1954).

5. L. H. Rieseberg, *et al.*, Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211–1216 (2003).

6. J. C. Pires, *et al.*, Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae). *Biol. J. Linn. Soc. Lond.* **82**, 675–688 (2004).

7. D. R. Dittrich-Reed, B. M. Fitzpatrick, Transgressive Hybrids as Hopeful Monsters. *Evol. Biol.* **40**, 310–315 (2013).

8. G. Charron, S. Marsit, M. Hénault, H. Martin, C. R. Landry, Spontaneous whole-genome duplication restores fertility in interspecific hybrids. *Nat. Commun.* **10**, 4126 (2019).

9. A. R. Leitch, I. J. Leitch, Genomic plasticity and the diversity of polyploid plants. *Science* **320**, 481–483 (2008).

10. Y. Van de Peer, S. Maere, A. Meyer, The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732 (2009).

11. B. C. Thomas, Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Research* **16**, 934–946 (2006).

12. P. P. Edger, *et al.*, Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* **51**, 541–547 (2019).

13. J. C. Schnable, N. M. Springer, M. Freeling, Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences* **108**, 4069–4074 (2011).

14. M. R. Woodhouse, *et al.*, Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5283–5288 (2014).

15. P. P. Edger, *et al.*, Subgenome Dominance in an Interspecific Hybrid, Synthetic

Allopolyploid, and a 140-Year-Old Naturally Established Neo-Allopolyploid Monkeyflower. *Plant Cell* **29**, 2150–2167 (2017).

16. S. Bottani, N. R. Zabet, J. F. Wendel, R. A. Veitia, Gene Expression Dominance in Allopolyploids: Hypotheses and Models. *Trends in Plant Science* **23**, 393–402 (2018).

17. R. J. A. Buggs, *et al.*, The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369** (2014).

18. D. Kryvokhyzha, *et al.*, Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid Capsella bursa-pastoris (Brassicaceae). *PLoS Genet.* **15**, e1007949 (2019).

19. M. Freeling, *et al.*, Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr. Opin. Plant Biol.* **15**, 131–139 (2012).

20. F. Cheng, *et al.*, Epigenetic regulation of subgenome dominance following whole genome triplication in Brassica rapa. *New Phytol.* **211**, 288–299 (2016).

21. M. Hao, *et al.*, The abundance of homoeologue transcripts is disrupted by hybridization and is partially restored by genome doubling in synthetic hexaploid wheat. *BMC Genomics* **18**, 149 (2017).

22. M.-C. Combes, *et al.*, Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biol. Evol.* **7**, 1110–1121 (2015).

23. N. Li, *et al.*, DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. *New Phytol.* **223**, 979–992 (2019).

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

25. J. Wu, *et al.*, Homoeolog expression bias and expression level dominance in resynthesized allopolyploid Brassica napus. *BMC Genomics* **19**, 586 (2018).

26. V. Chagué, *et al.*, Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. *New Phytol.* **187**, 1181–1194 (2010).

27. K. Song, P. Lu, K. Tang, T. C. Osborn, Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7719–7723 (1995).

28. B. Chalhoub, *et al.*, Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science* **345**, 950–953 (2014).

29. H. An, *et al.*, Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid Brassica napus. *Nat. Commun.* **10**, 2878 (2019).

30. I. A. P. Parkin, *et al.*, Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biol.* **15**, R77 (2014).

31. F. Cheng, *et al.*, Biased Gene Fractionation and Dominant Gene Expression among the Subgenomes of Brassica rapa. *PLoS ONE* **7**, e36442 (2012).

32. D. Koenig, D. Weigel, Beyond the thale: comparative genomics and genetics of Arabidopsis relatives. *Nat. Rev. Genet.* **16**, 285–298 (2015).

33. K. A. Bird, R. VanBuren, J. R. Puzey, P. P. Edger, The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* (2018) https:/doi.org/10.1111/nph.15256.

34. J. Ren, *et al.*, Novel technologies in doubled haploid line development. *Plant Biotechnol. J.* **15**, 1361–1370 (2017).

35. E. E. Higgins, W. E. Clarke, E. C. Howell, S. J. Armstrong, I. A. P. Parkin, Detecting de Novo Homoeologous Recombination Events in Cultivated Brassica napus Using a Genome-Wide SNP Array. *G3: Genes|Genomes|Genetics* **8**, 2673–2683 (2018).

36. R. T. Gaeta, J. C. Pires, F. Iniguez-Luy, E. Leon, T. C. Osborn, Genomic Changes in Resynthesized Brassica napus and Their Effect on Gene Expression and Phenotype. *THE PLANT CELL ONLINE* **19**, 3403–3417 (2007).

37. Z. Xiong, R. T. Gaeta, J. C. Pires, Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid Brassica napus. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7908–7913 (2011).

38. A. Lloyd, *et al.*, Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *New Phytol.* **217**, 367–377 (2018).

39. M. Ashburner, *et al.*, Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

40. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

41. D. Szklarczyk, *et al.*, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).

42. C. E. Niederhuth, R. J. Schmitz, Putting DNA methylation in context: from genomes to gene expression in plants. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1860**, 149–156 (2017).

43. R. Lister, *et al.*, Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523–536 (2008).

44. S. J. Cokus, *et al.*, Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).

45. C. E. Niederhuth, *et al.*, Widespread natural variation of DNA methylation within

angiosperms. *Genome Biol.* **17**, 194 (2016).

46. B. Samans, B. Chalhoub, R. J. Snowdon, Surviving a Genome Collision: Genomic Signatures of Allopolyploidization in the Recent Crop Species. *Plant Genome* **10** (2017).

47. T. Rodriguez-Riano, A. Dafni, A new procedure to asses pollen viability. *Sex. Plant Reprod.* **12**, 241–244 (2000).

48. H. Tang, *et al.*, Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research* **18**, 1944–1954 (2008).

49. S. Ou, N. Jiang, LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).

50. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).

51. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

52. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

53. M. Pertea, *et al.*, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

54. M. D. Schultz, *et al.*, Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).

55. R. K. Dale, B. S. Pedersen, A. R. Quinlan, Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).

56. Team, R.C. *R: A Language and Environment for Statistical Computing*. Available from: http://www.r-project.org/.

57. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer Science & Business Media, 2009).

58. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

59. D. W. Barnett, E. K. Garrison, A. R. Quinlan, M. P. Strömberg, G. T. Marth, BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691–1692 (2011).

60. Z. Hao, *et al.*, RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms https:/doi.org/10.7287/peerj.preprints.27928.

61. P. J. Kersey, *et al.*, Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Research* **46**, D802–D808 (2018).

62. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

63. M. D. Schultz, R. J. Schmitz, J. R. Ecker, “Leveling” the playing field for analyses of single-base resolution DNA methylomes. *Trends in Genetics* **28**, 583–585 (2012).

Supplemental Information:

Materials & Methods

Plant growth, tissue collection, library prep

The resynthesized B. napus allopolyploid lines (CCAA) were obtained from a previous study (37). Plants were grown at 23C during day and 20C at night with 16hr days in a growth chamber. True leaf three was collected from all plants within one hour starting at 10am (4hrs into day) and immediately flash-frozen in liquid nitrogen. Leaves were split in half for RNA and DNA isolation. Total RNA and DNA was isolated using the respective KingFisher Pure Plant kits (Thermo Fisher Scientific, MA) and quantified using Qubit 3 Fluorometer (Thermo Fisher Scientific, MA). DNA and RNA libraries were prepared using the KAPA HyperPrep and mRNA HyperPrep kit protocol, respectively (KAPA Biosystems, Roche, USA). Bisulfite conversion was performed using the EZ DNA Methylation Kit (Zymo, CA) All libraries were submitted to a genomics facility (Beijing Nuohe Zhiyuan Technology Co., Beijing, China) and sequenced with paired-end 150bp reads on an Illumina HiSeq 4000 system. Pollen viability for each individual was assessed using MTT method of Rodriguez-Riano and Dafni (47).

In silica reference genome construction

Paired end 150bp genomic illumina reads for the doubled haploid *Brassica rapa* accession IMB-218, were aligned to the *Brassica rapa* R500 reference genome using bowtie2 v. 2.3.4.1 (50) on default settings with the flag “--very-sensitive-local”. The resulting alignment files were sorted and had read groups added with PicardTools v. 2.8.1 and SNPs were called between the R500 reference and the IMB-218 alignment using GATK v 3.5.0 Unified Genotyper, filtered to only include homozygous SNPs, and a new fasta reference was made using GATK v 3.5.0 FastaAlternativeReferenceMaker. This IMB-218 reference genome was concatenated to the TO1000 reference genome to create an *in silica* reference genome for *B. napus*.

Homoeologous exchange analysis

Paired end 150bp genomic illumina reads were filtered with Trimmomatic v 0.33 (58) to remove illumina TruSeq3 adapters. Trimmed reads were aligned to the *in silica B.napus* reference genome with Bowtie2 v.2.3.4.1(50) on default settings with the flag "--very-sensitive-local". Bam files sorted with bamtools (59) for use in downstream analyses.

MCScan toolkit (48) was used to identify syntenic, homologous gene pairs (syntelogs) between *Brassica rapa* (reference genome R500) and *Brassica oleracea* (reference genome TO1000;30). In the synthetic polyploid these can be thought of as syntenic homoeologs. Bed files based on chromosome and start/stop position information for each subgenome were generated. For all 18 samples (6 individuals x 3 generations) read depth for the A subgenome (BnA) syntenic homoeologs was determined in Bedtools (51) with BedCov using the R500 syntelog bed file and for the C subgenome (BnC) using the TO1000 syntelog bed file. In R v 3.4.1, read depths for each syntenic homoeolog was normalized to reads per million for subgenome of origin and the ratio of reads mapping to a syntenic homoeolog compared to the overall read mapping for a syntenic homoeolog pair was averaged over a window of 50 genes with a step of one gene.

Homoeologous exchanged regions were identified by calculating average read depth for the BnC subgenome along a sliding window of 170 (85 up- and down stream) genes and step size of one. If 10 or more consecutive genes had a read depth within a pre-selected range it was called a homoelogous exchange. Regions 0 ≤ read depth < 0.2 were predicted to be in a 0BnC-to-4BnA ratio, 1BnC-to-3BnA was predicted for 0.2 ≤ read depth < 0.4, 2BnC-to-2BnA was predicted for 0.4 ≤ read depth < 0.6, 3BnC-to-1BnA for read depth between 0.6 ≤ read depth <0.8 and 4BnC-to-0BnA for read depth between 0.8 ≤ read depth < 1. HEs were plotted with the R package Rideogram (60)

RNASeq analysis

raw RNA-seq reads were filtered with Trimmomatic v 0.33 (58)to remove illumina TruSeq3 adapters and mapped to the *in silica* reference using STAR v 2.6.0 (52) on default settings. Transcripts were quantified in transcripts per million (TPM) from RNAseq alignments using StringTie v 1.3.5 (53). Because the syntelogs in the progenitor genomes are in the subgenomes of the synthetic polyploids, they can be thought of as syntenic homoeologs. To avoid dosage imbalance only syntenic homoeologs determined to be at a 2:2 dosage balance were analyzed for homeolog expression bias. Additionally, to remove lowly expressed genes that might be noise, syntenic homoeologs were only kept if the total TPM of the pair was greater than 10. Syntenic homoeolog pairs with Log2 Foldchange greater than 3.5 were called BnC biased, and less than 3.5 were called BnA biased. Because lack of subgenome dominance would follow a normal distribution where deviations from 0 FC is equal in either direction, a Chi-squared goodness of fit test was carried out to test for normality. The R package Upsetr was used to identify and plot syntenic homoeologs shared by all lines for a given generation. These gene

lists were investigated for GO and KEGG pathway enrichment (39, 40) in the STRING PPI network (41), clustering coefficients, and enrichment for network interactions.

DNA methylation analysis

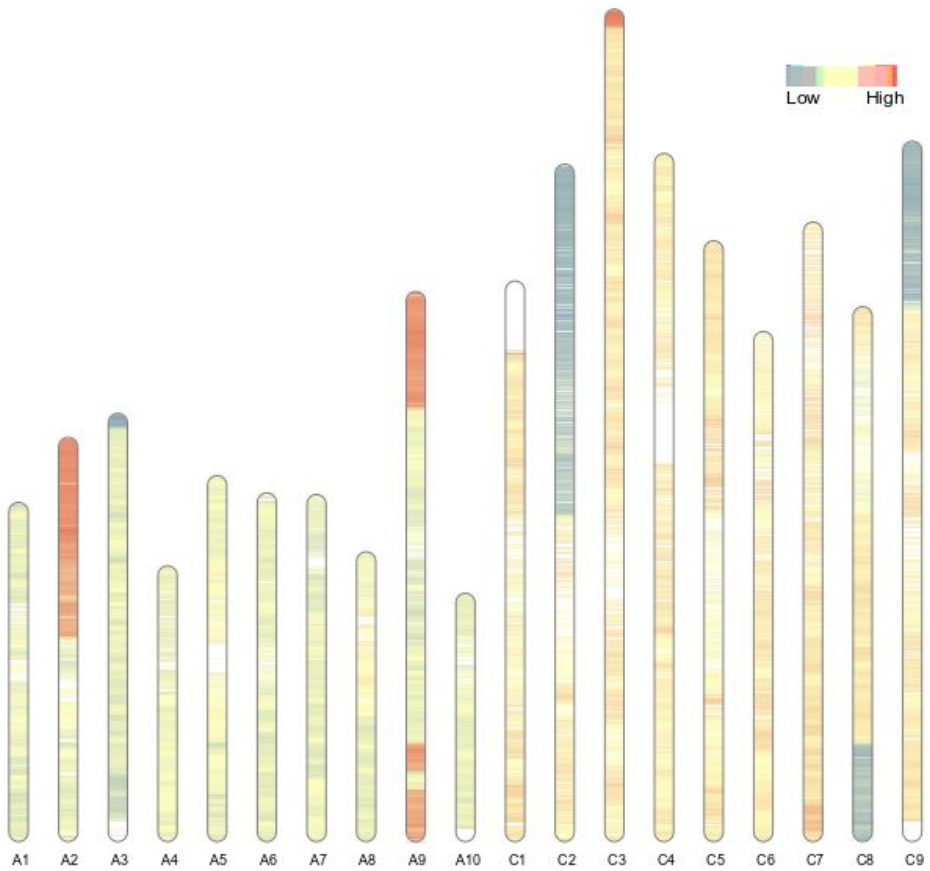
The genomes of both *B. oleracea* TO1000 (EnsemblPlants 43(30, 61) and *B. rapa* R500 were combined into a single fasta file to create an *in silico* allopolyploid genome. Whole genome bisulfite sequencing (WGBS) was mapped to the combined genome using methylpy v1.3.8 (54) (see Supplementary Table X); using cutadapt v2.3 (62) for adaptor trimming, Bowtie2 v2.3.5 (50) for alignment, and Picard tools v2.20.2 for marking duplicates. The chloroplast genome is unmethylated in plants and can be used as in internal control for calculating the non-conversion rate of bisulfite treatment (percentage of unmethylated sites that fail to be converted to uracil)(43). In mapping the parental TO1000 and R500 genomes, a small fraction of reads (~1.3% TO1000 and ~6.1% IMB218) mapped to the wrong genome, but seems to have had a limited impact on overall results.It is likely that there is a small percentage of mismapping in the resynthesized allopolyploids as well. As an additional control for this, we randomly down-sampled TO1000 to an equal number of read pairs as IMB218. These were combined with the IMB218 reads and mapped to the combined genome to mimic an *in silico* allopolyploid. By including this *in silico* allopolyploid alongside the individually mapped parents, we should be able to observe the influence of mismapping. DNA methylation levels in this combined dataset were either approximately half-way between the two parents for the whole genome or at a sub-genome level, approximately equal to that of the respective parent. This indicates that mismapping has little effect on genome-wide analyses, although it may still be a factor at specific regions.

Genome-wide levels of DNA methylation and DNA methylation metaplots were analyzed as previously described (45) using python v3.7.3. Pybedtools v (55) and Bedtools v2.25.0 (51). Briefly, for genome-wide DNA methylation levels, the weighted methylation level(63), which accounts for sequencing coverage, was calculated for each sequence context (CG, CHG, and CHH). This was done for the combined genome, and each individual subgenome. For gene metaplots, cytosines from 2 kbps upstream, 2 kbs downstream and within the gene body were extracted. Cytosines within gene bodies were restricted to those found in coding sequences, as the presence of TEs in introns and problems of proper UTR annotation can obscure start/stop sites and introduce misleadingly high levels of DNA methylation (45). Each of these three regions were then divided into 20 windows and the weighted methylation level for each window calculated and average for all genes. For LTR metaplots, the same analysis was performed, except the all cytosines within the LTR body were included. Metaplots were created for all genes in the combined genome, all genes within each subgenome, syntenic genes in each subgenome, and non-syntenic genes in each subgenome. All plots and statistics were done in R v3.6.0 (56), plots made using ggplot2 (57). All code and original analyzed data and plots are available on Github (https://github.com/niederhuth/Replaying-the-evolutionary-tape-to-investigate-subgenome-dominance).

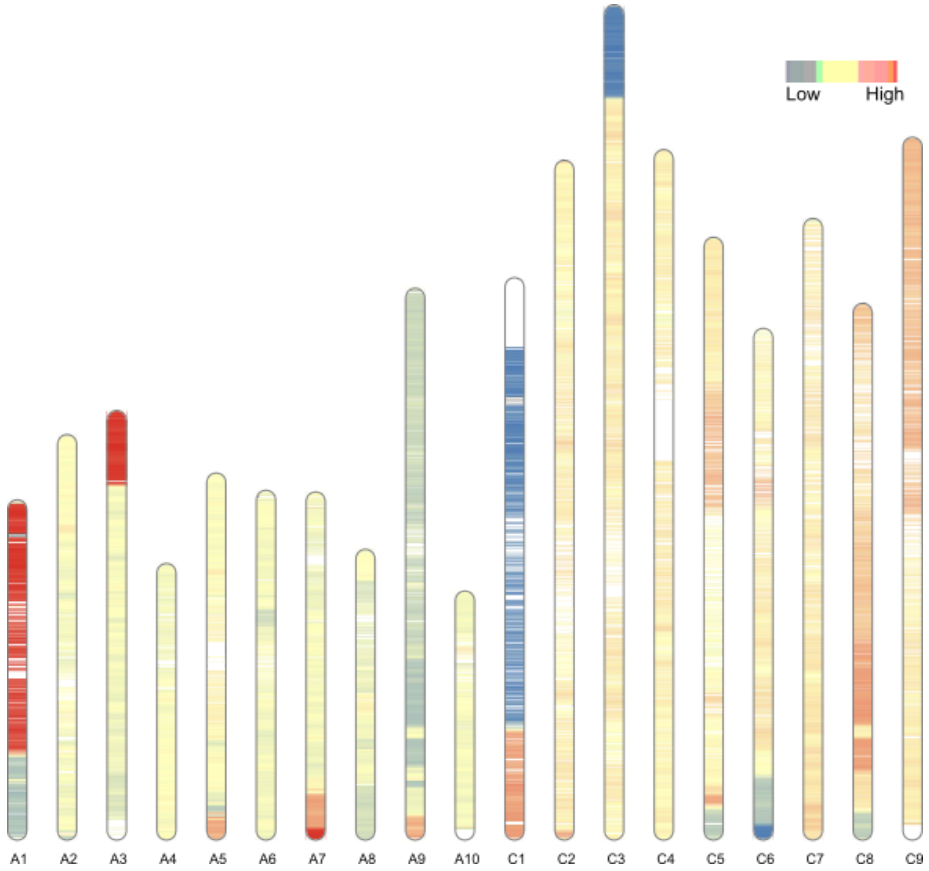
Data availability: Raw data from this project is available on the NCBI Sequence Read Archive (SRA) Project XXX

Figures 1-18 **Homoeologous exchanges**: Homoeologous exchange regions shown with karyotype plots of the normalized proportion of reads for a homoeolog pair deriving from the A-subgenome for chromosomes A1-A10 and the C-subgenome for chromosomes C1-C9. Red regions indicate high proportion of reads, blue indicates low, and yellow indicates approximately equal proportion. White regions are those where syntenic orthologs do not exist. Expected proportions should be ~50/50 for all pairs in the absence of non-reciprocal homoeologous exchange.

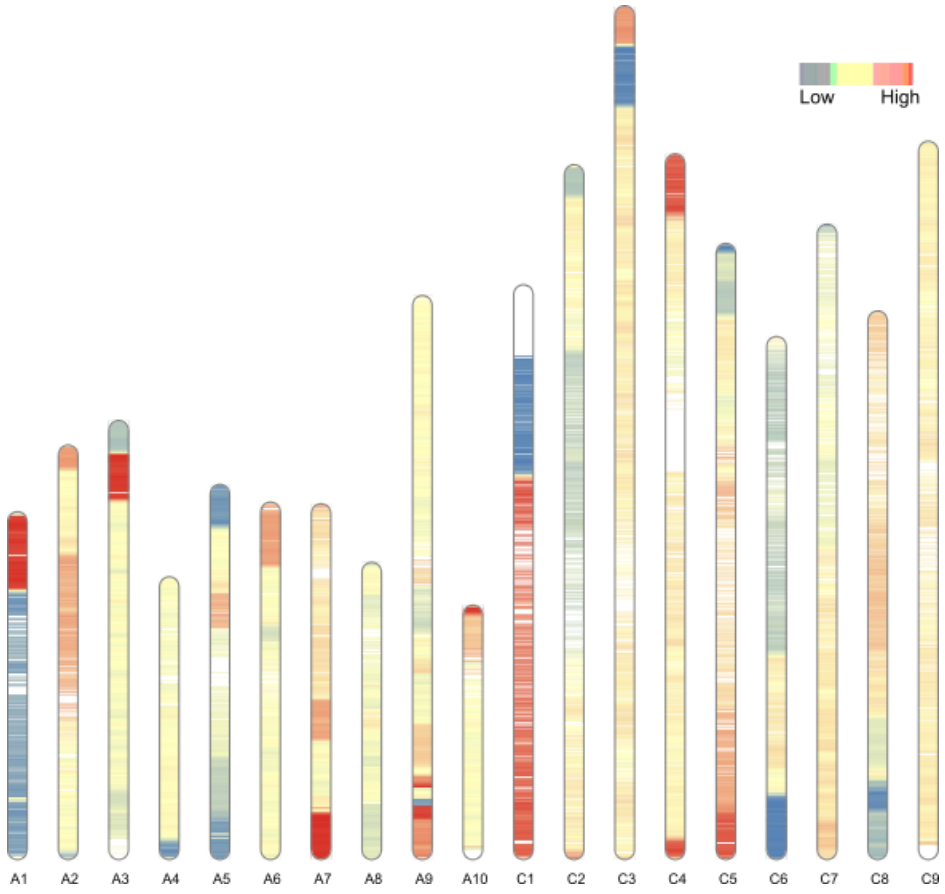
EL 100 S1



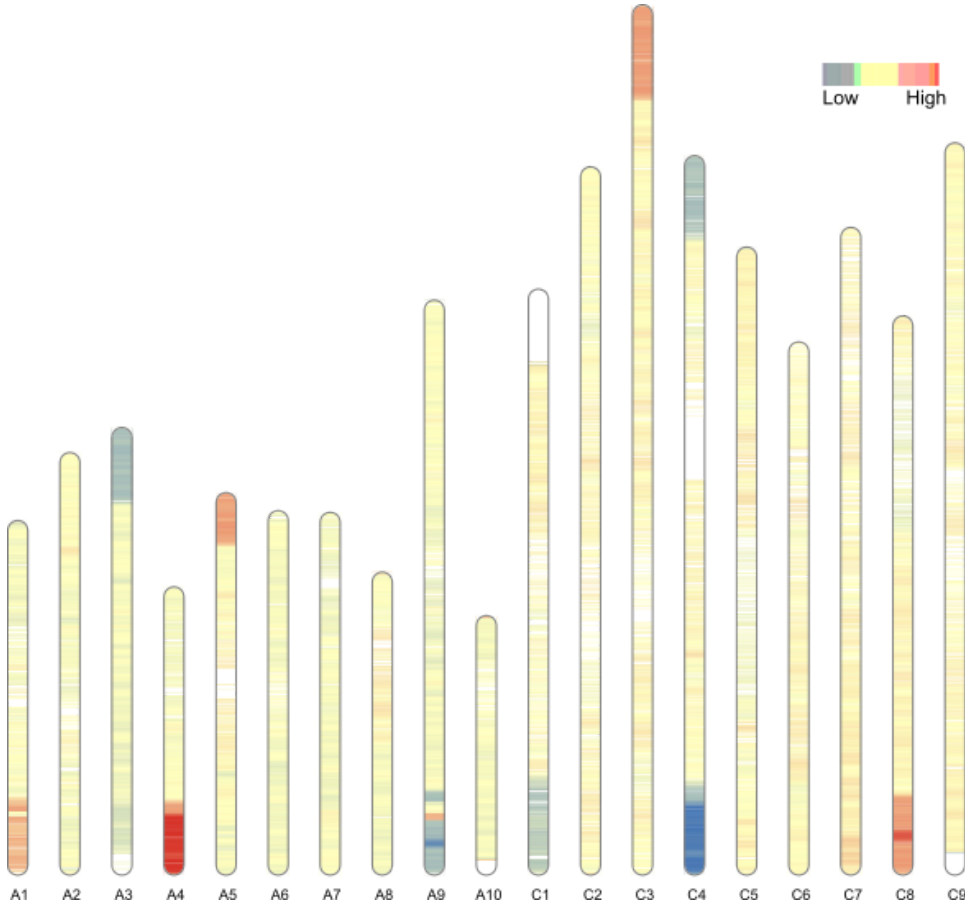
EL 100 S5



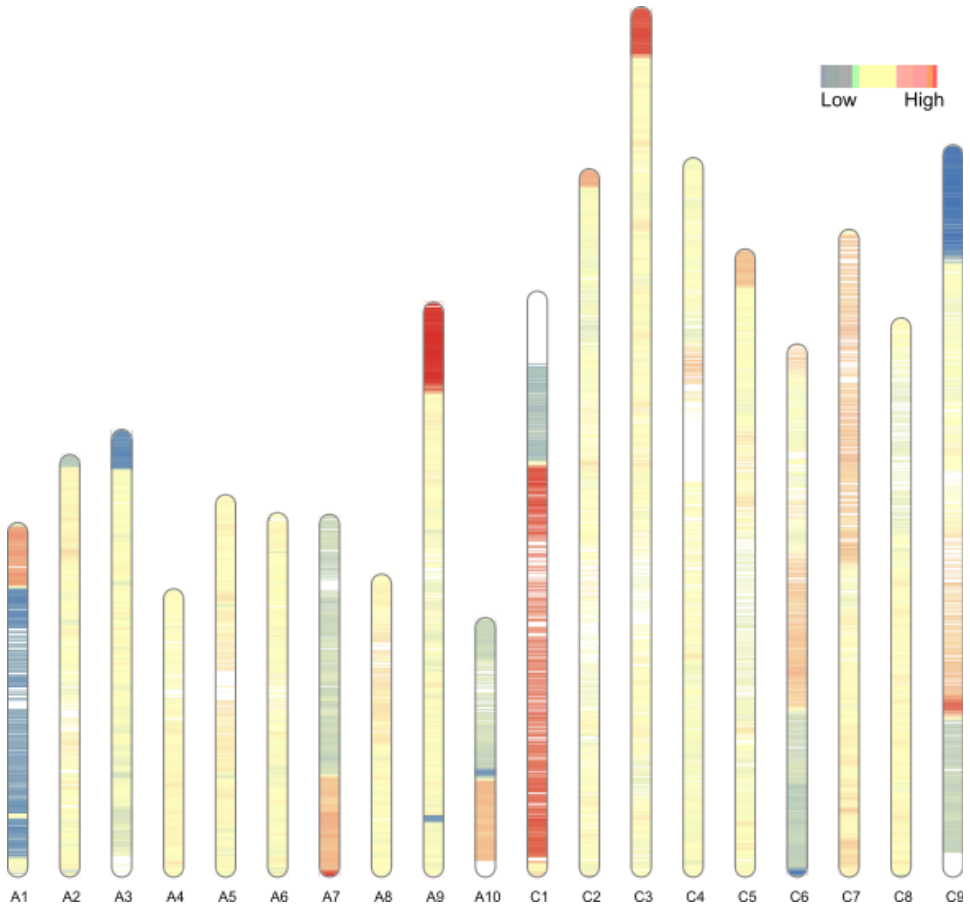
EL 100 S10



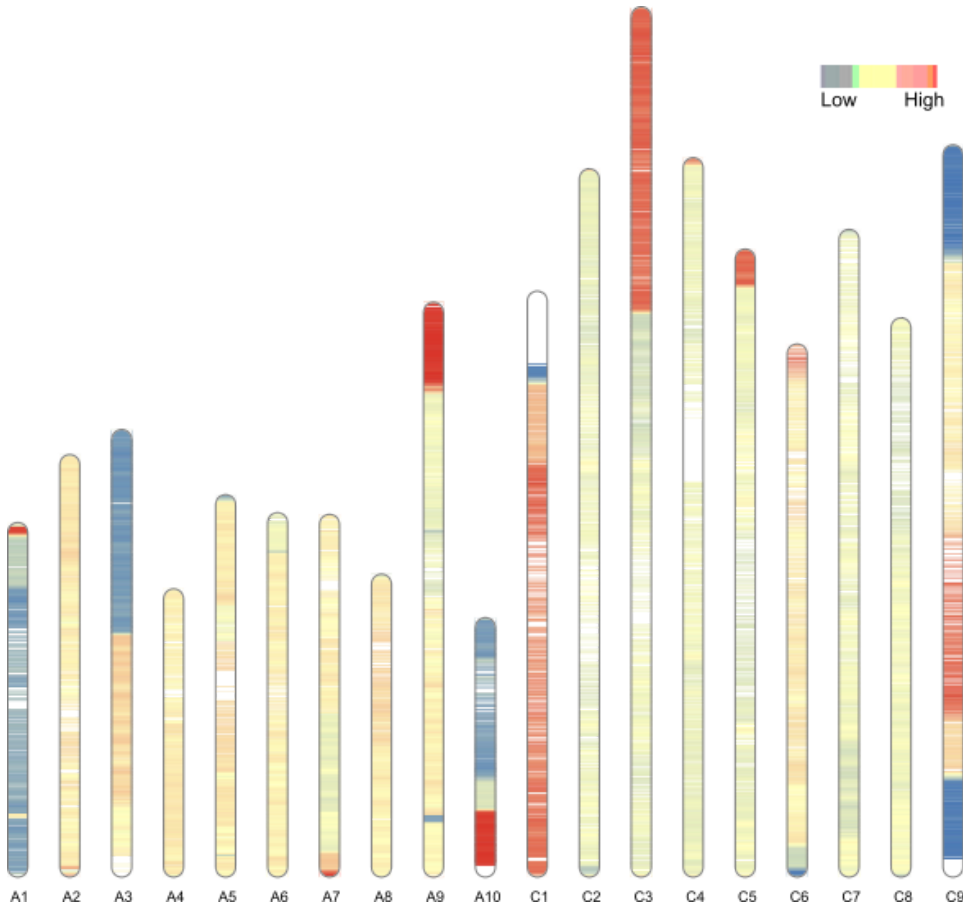
EL 200 S1



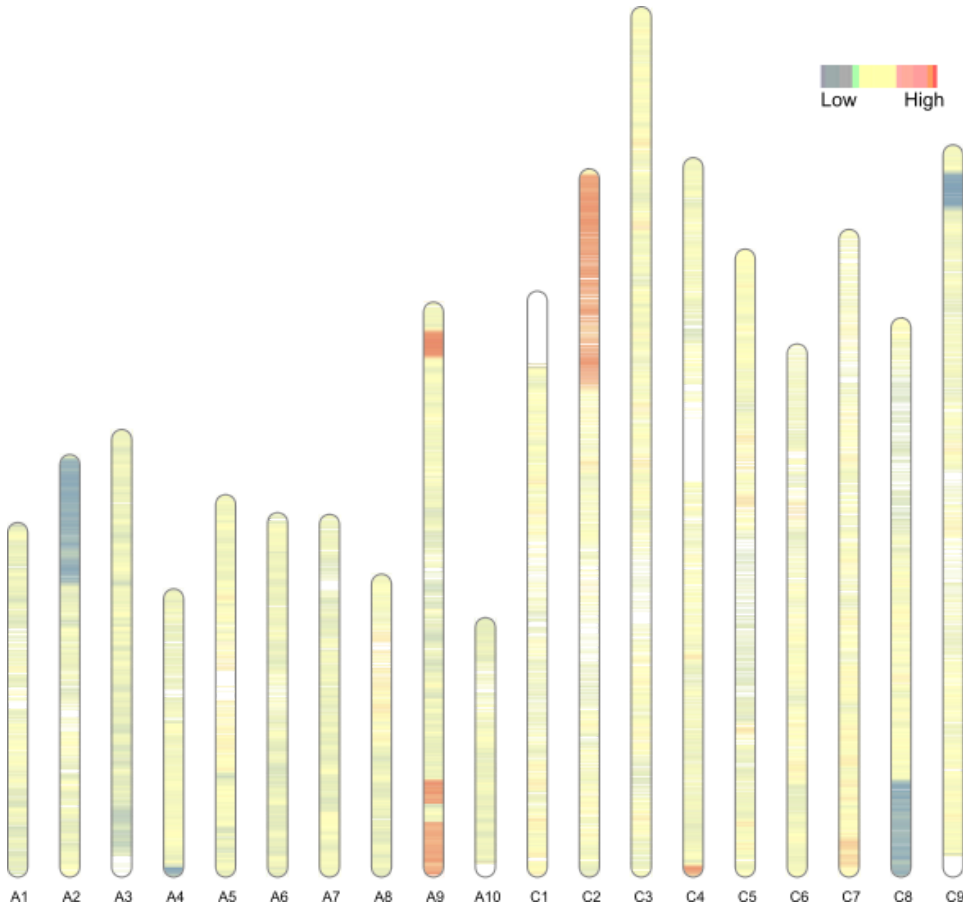
EL 200 S5



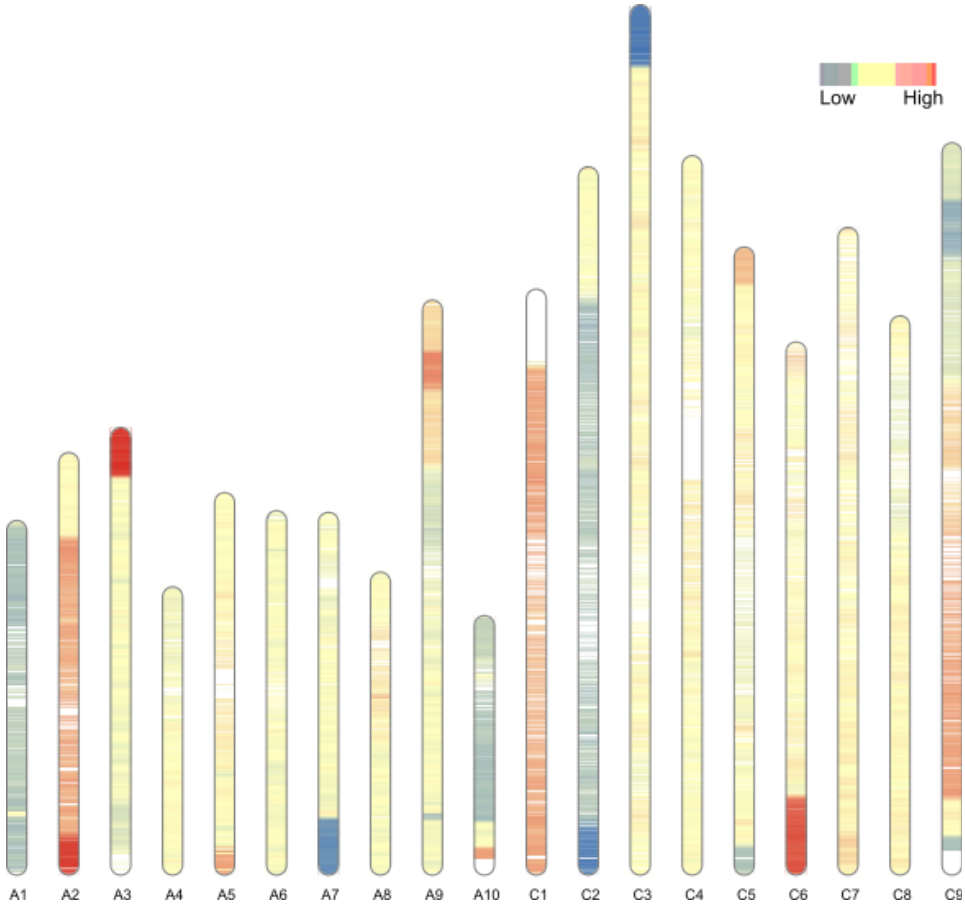
EL 200 S10



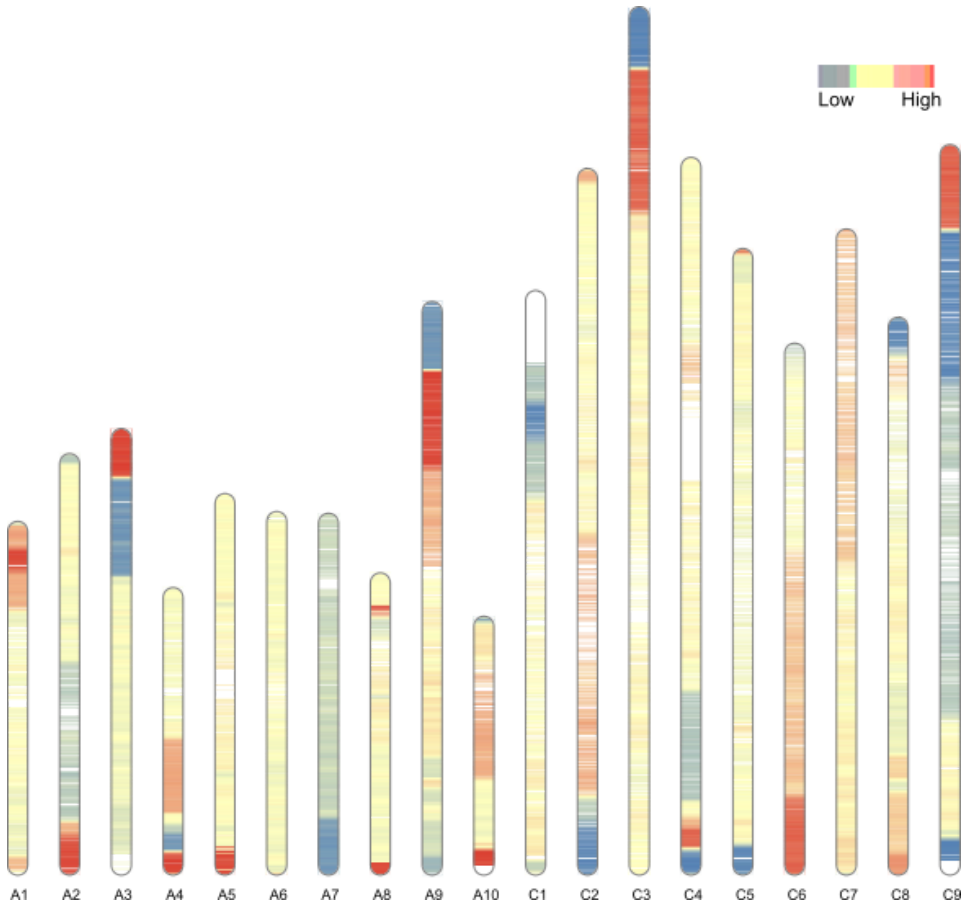
EL 300 S1



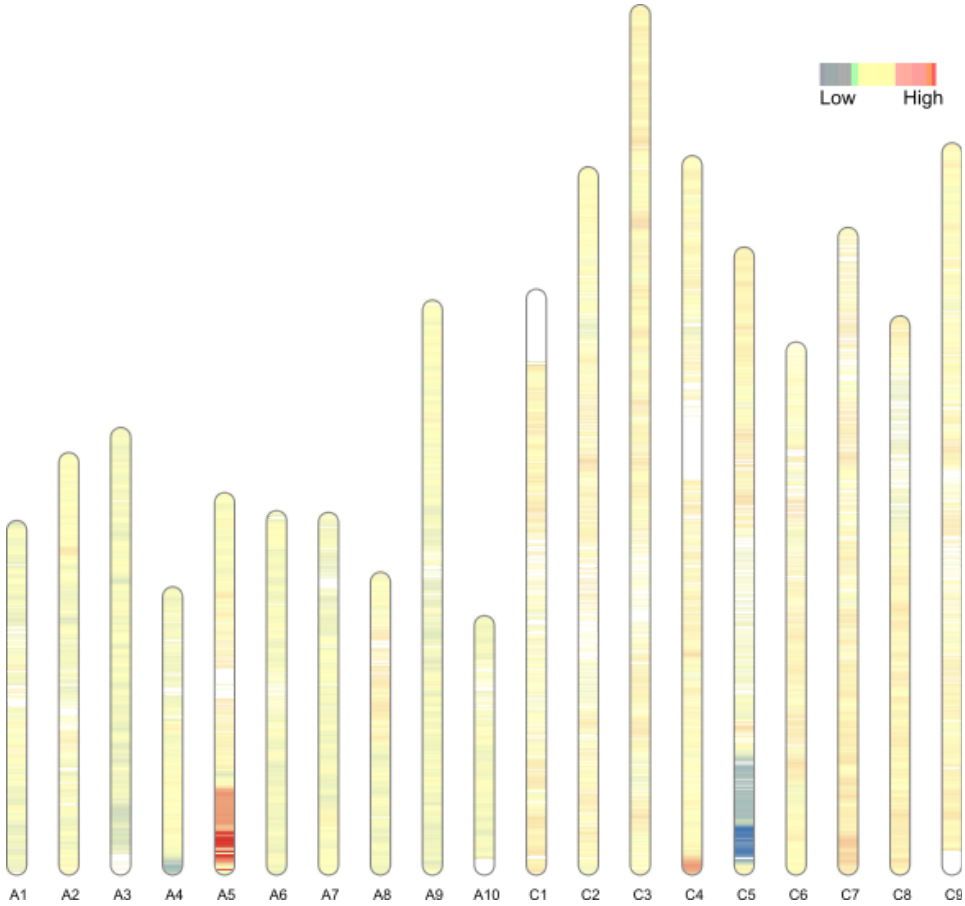
EL 300 S5



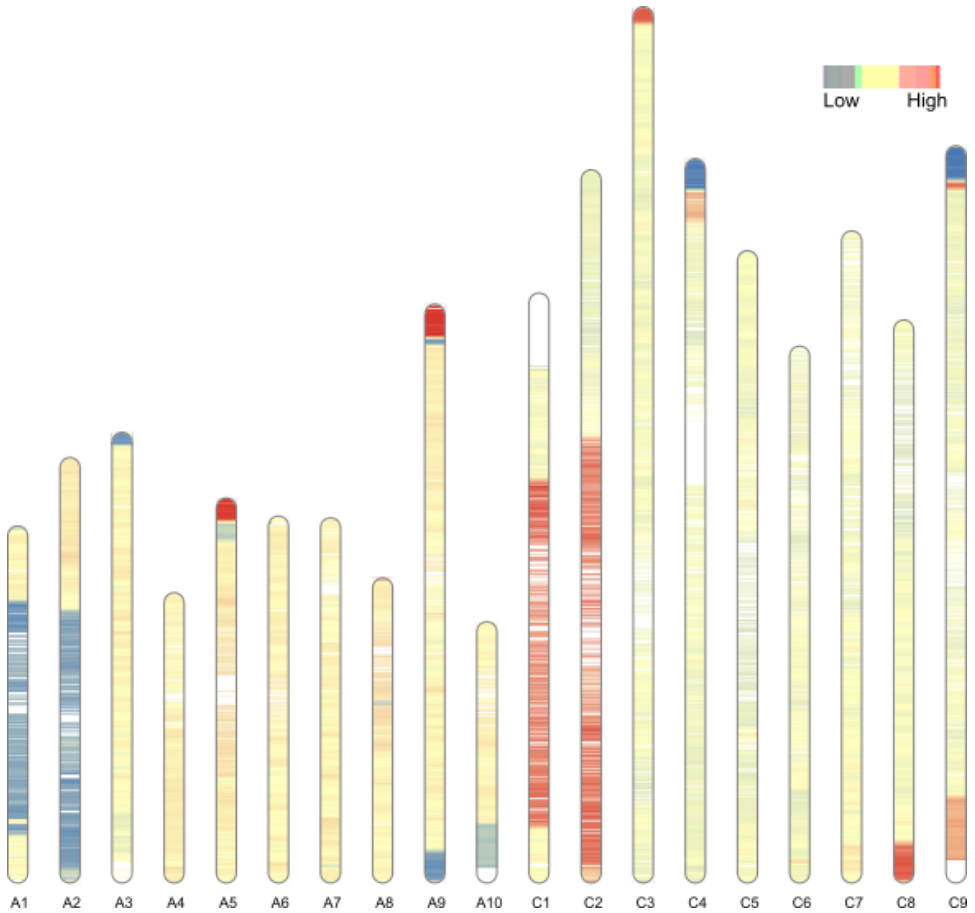
EL 300 S10



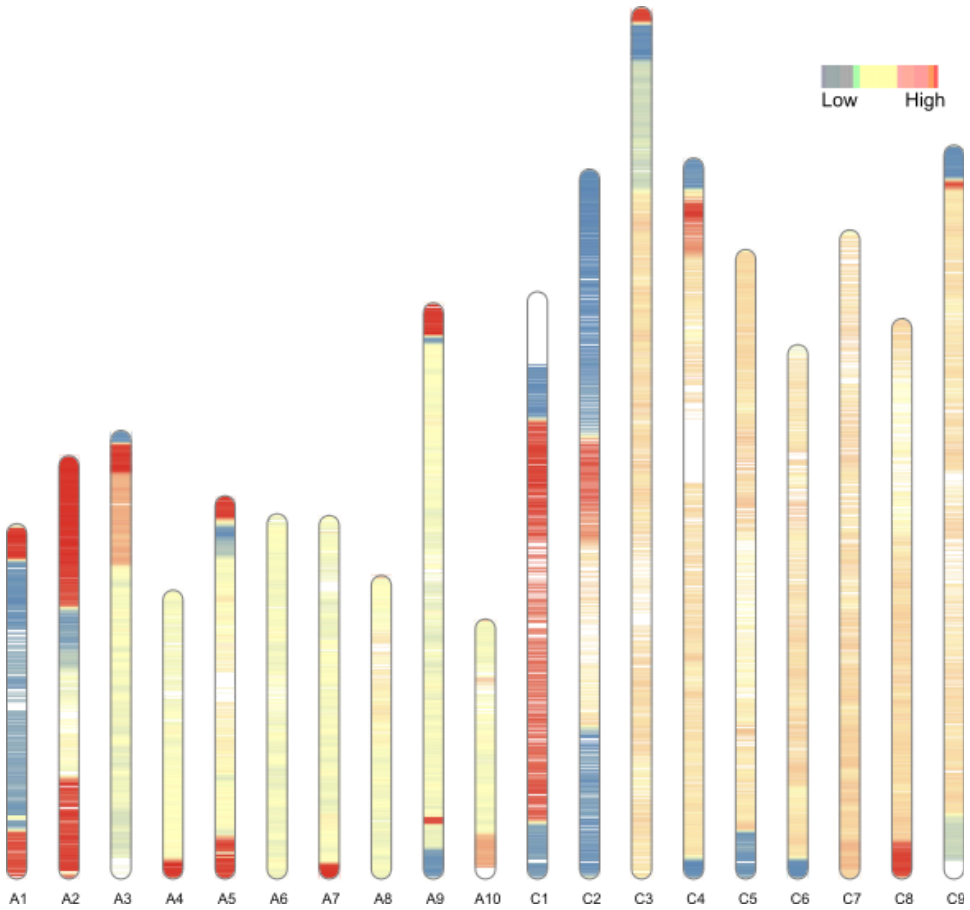
EL 400 S1



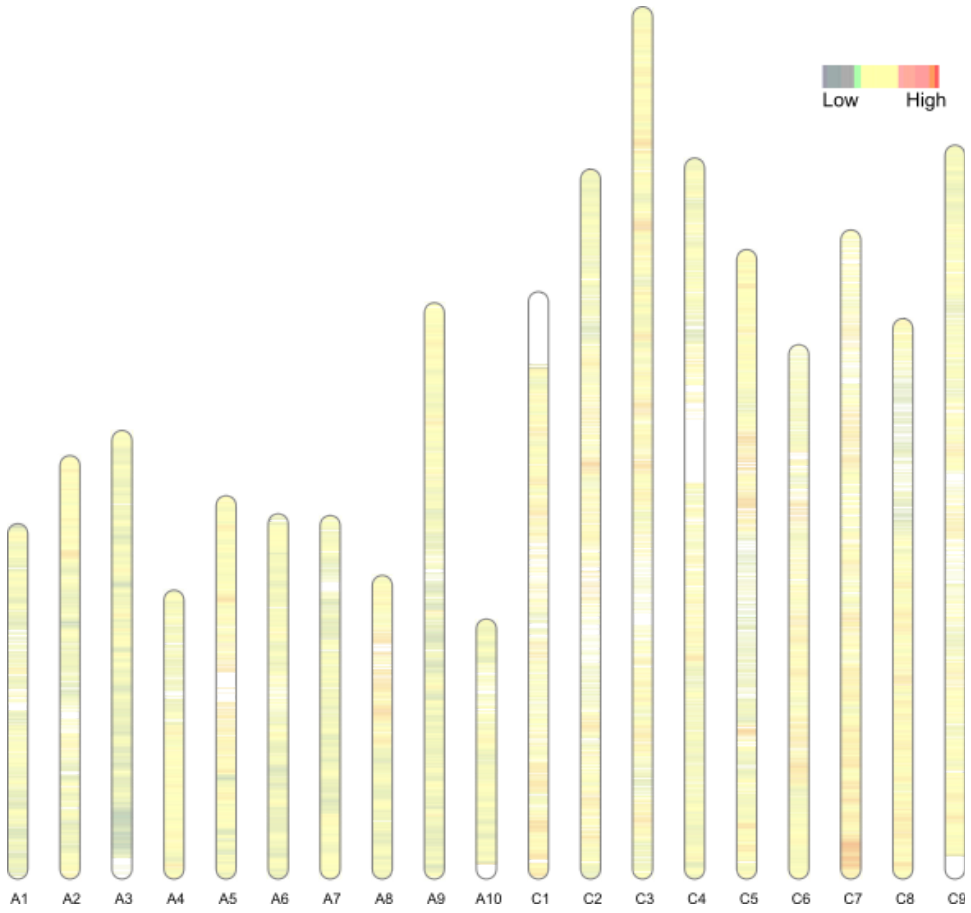
EL 400 S5



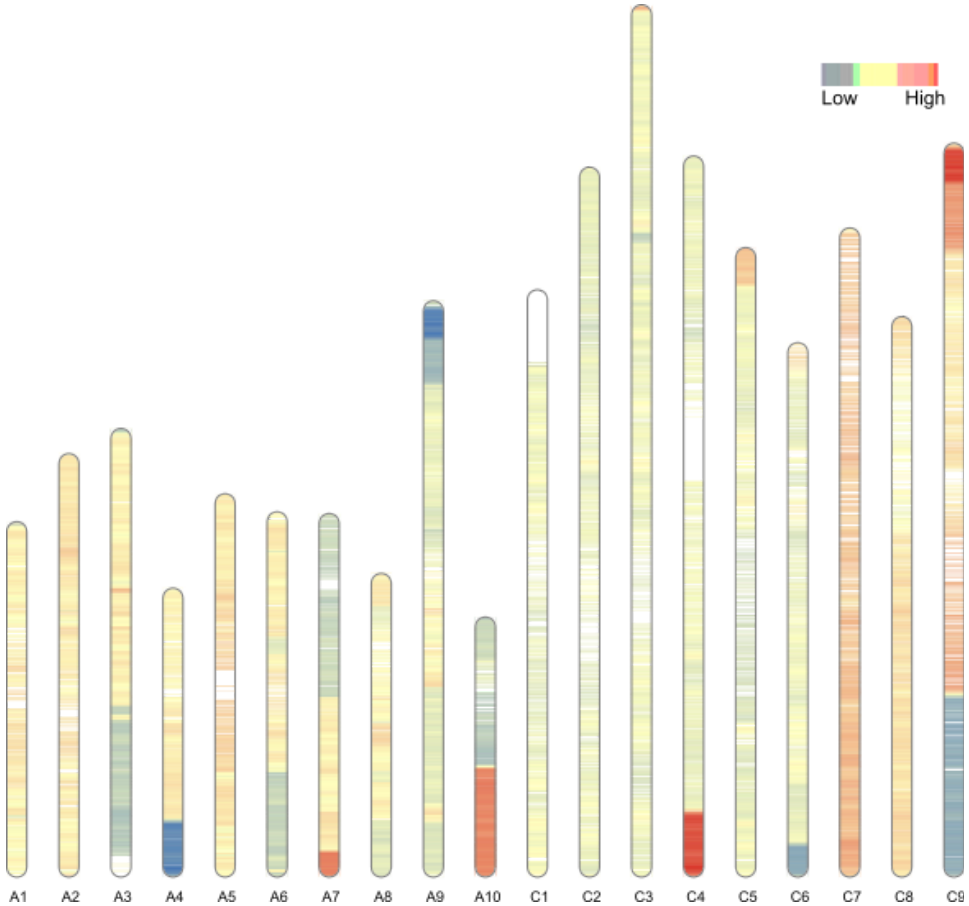
EL 400 S10



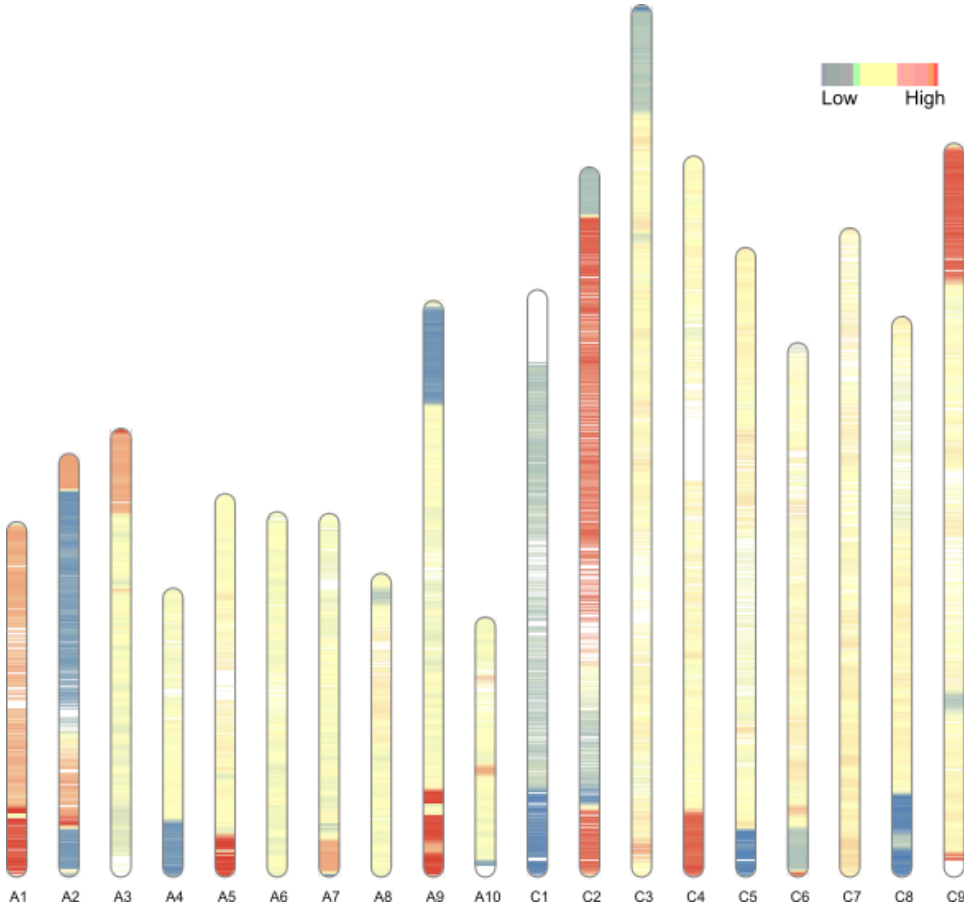
EL 600 S1



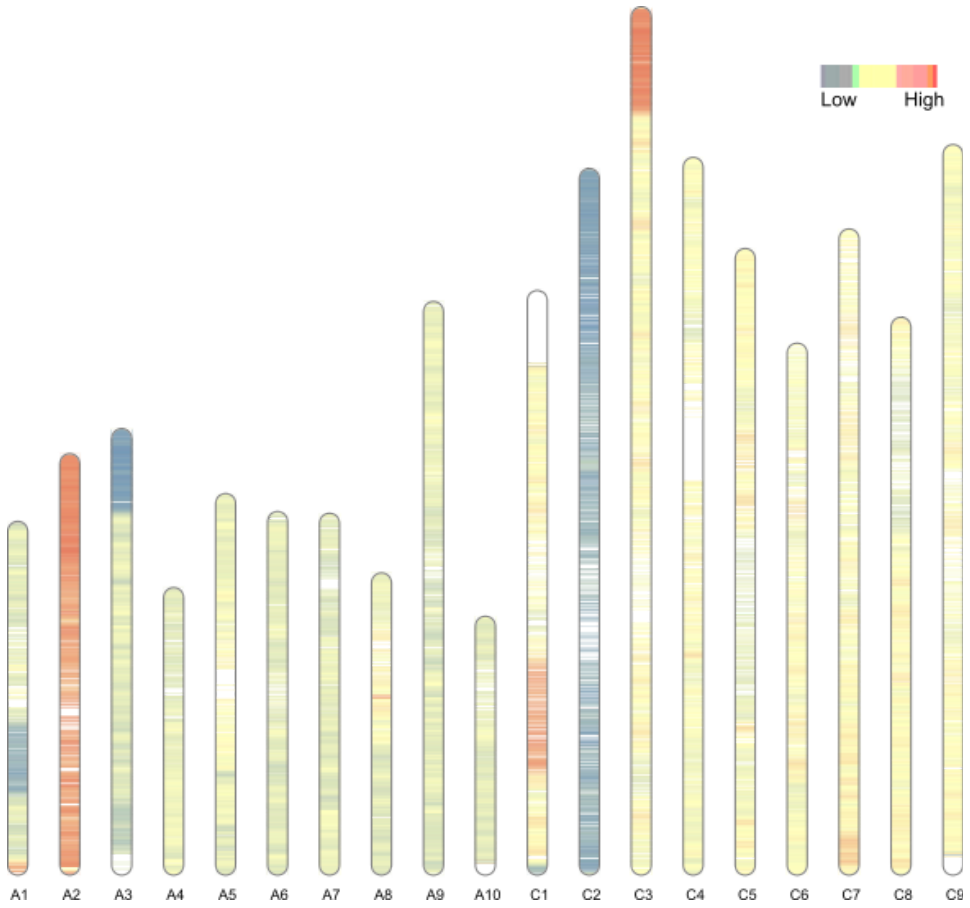
EL 600 S5



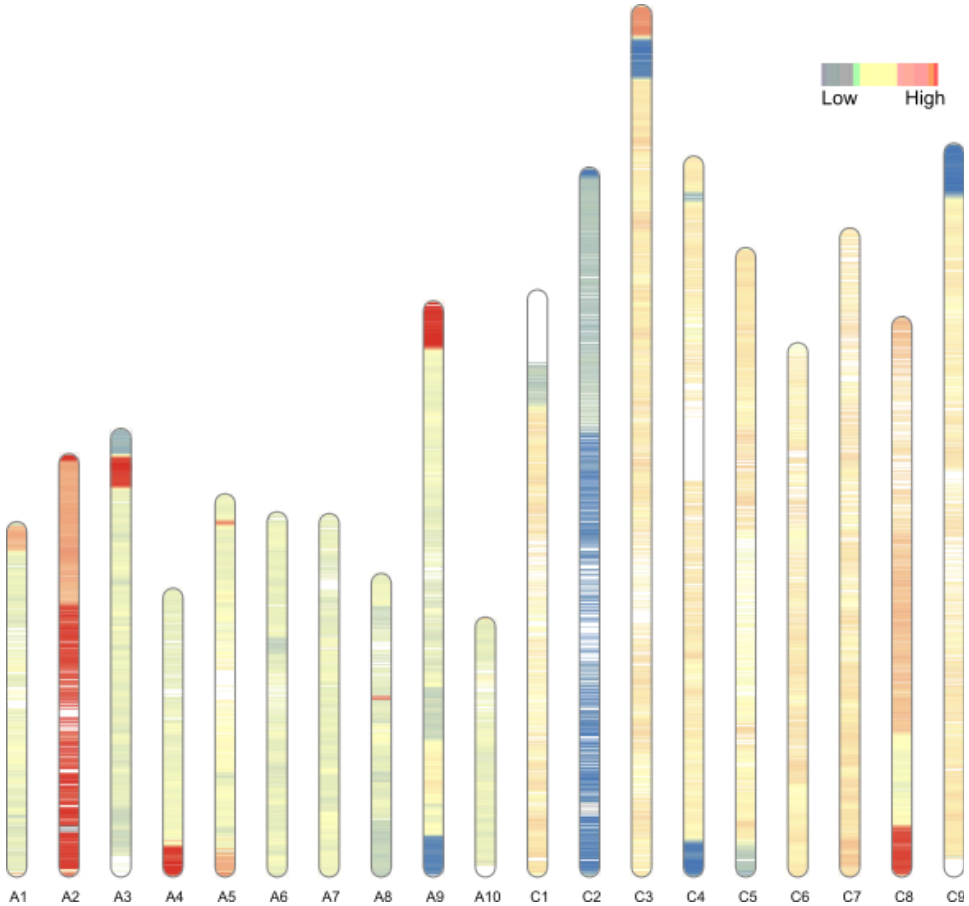
EL 600 S10



EL 1100 S1



EL 1100 S5



EL 1100 S10

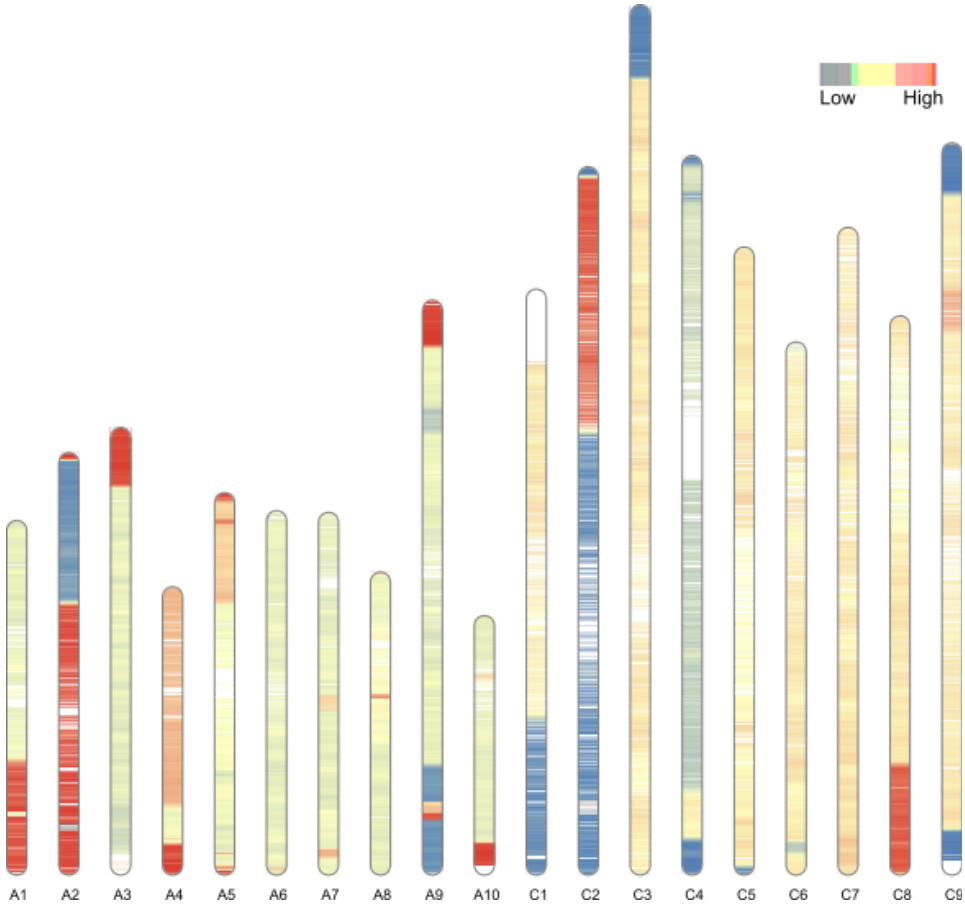
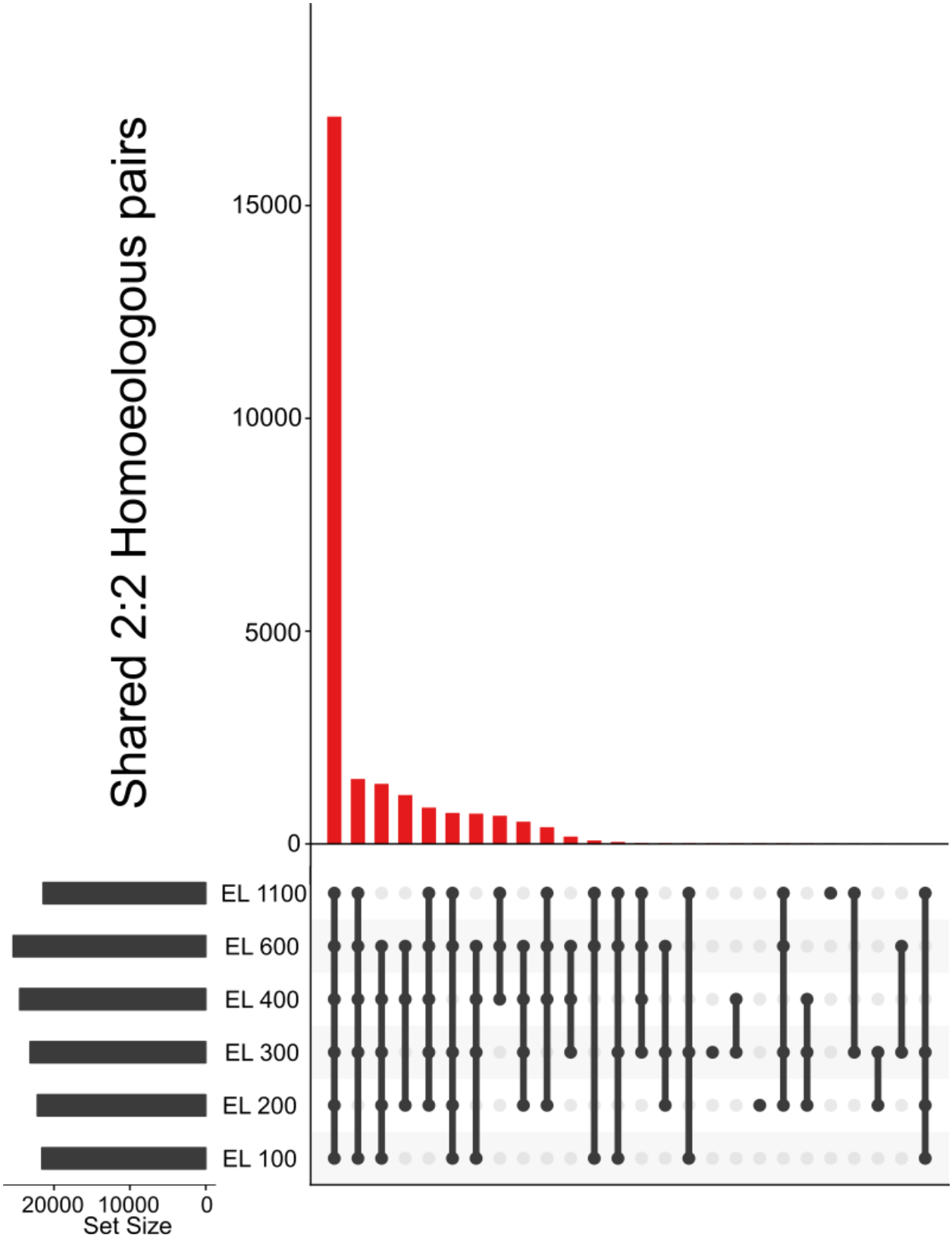
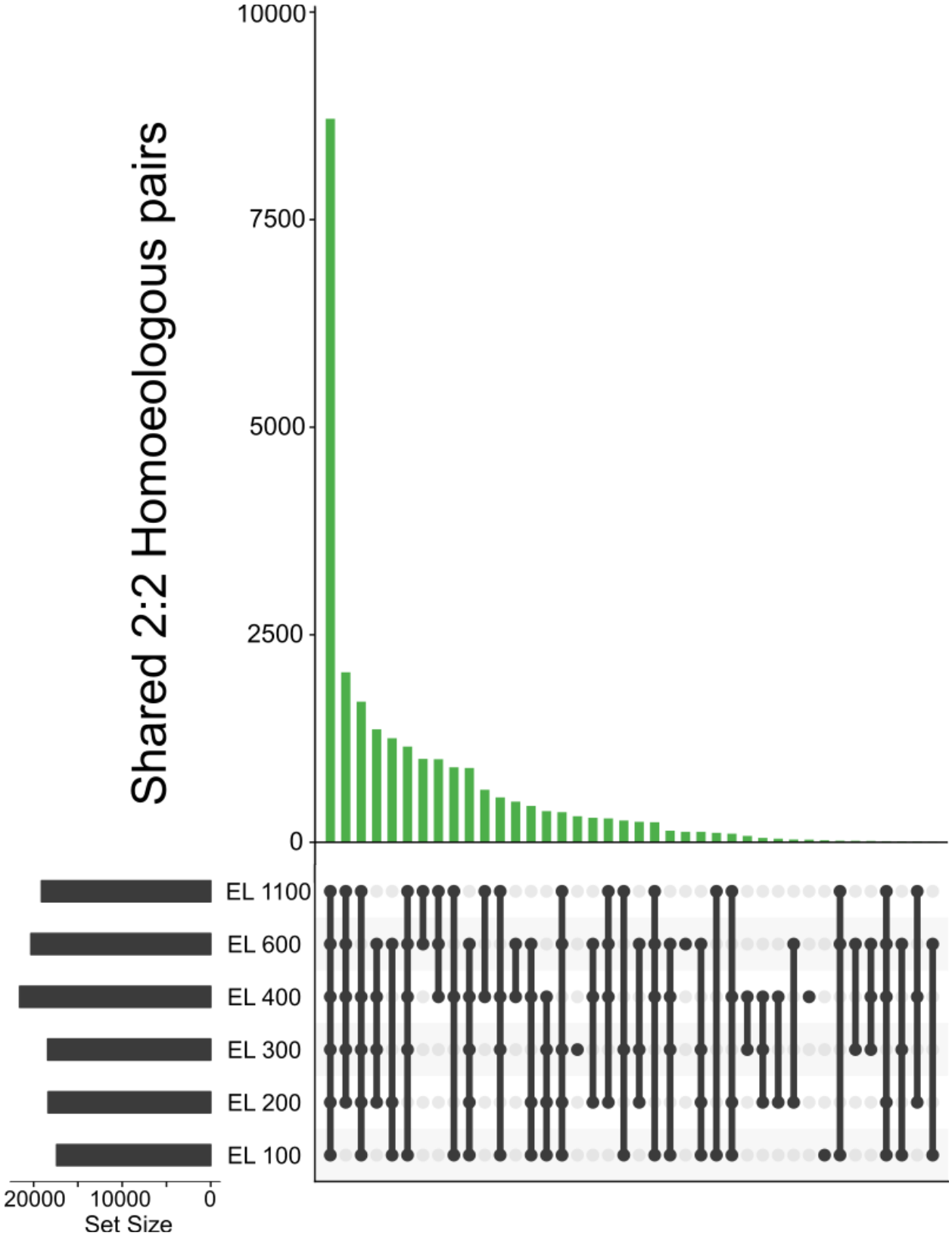


Figure 19a-c **Shared balanced homoelogous regions:** Upset plot of shared homoeologs in 2:2 gene dosage across the 6 lines in the three sampled generations.




Shared 2:2 Homoeologous pairs
10000
7500
5000
2500
0
EL 1100
EL 600
EL 400
EL 300
EL 200
EL 100
20000
10000
0
Set Size

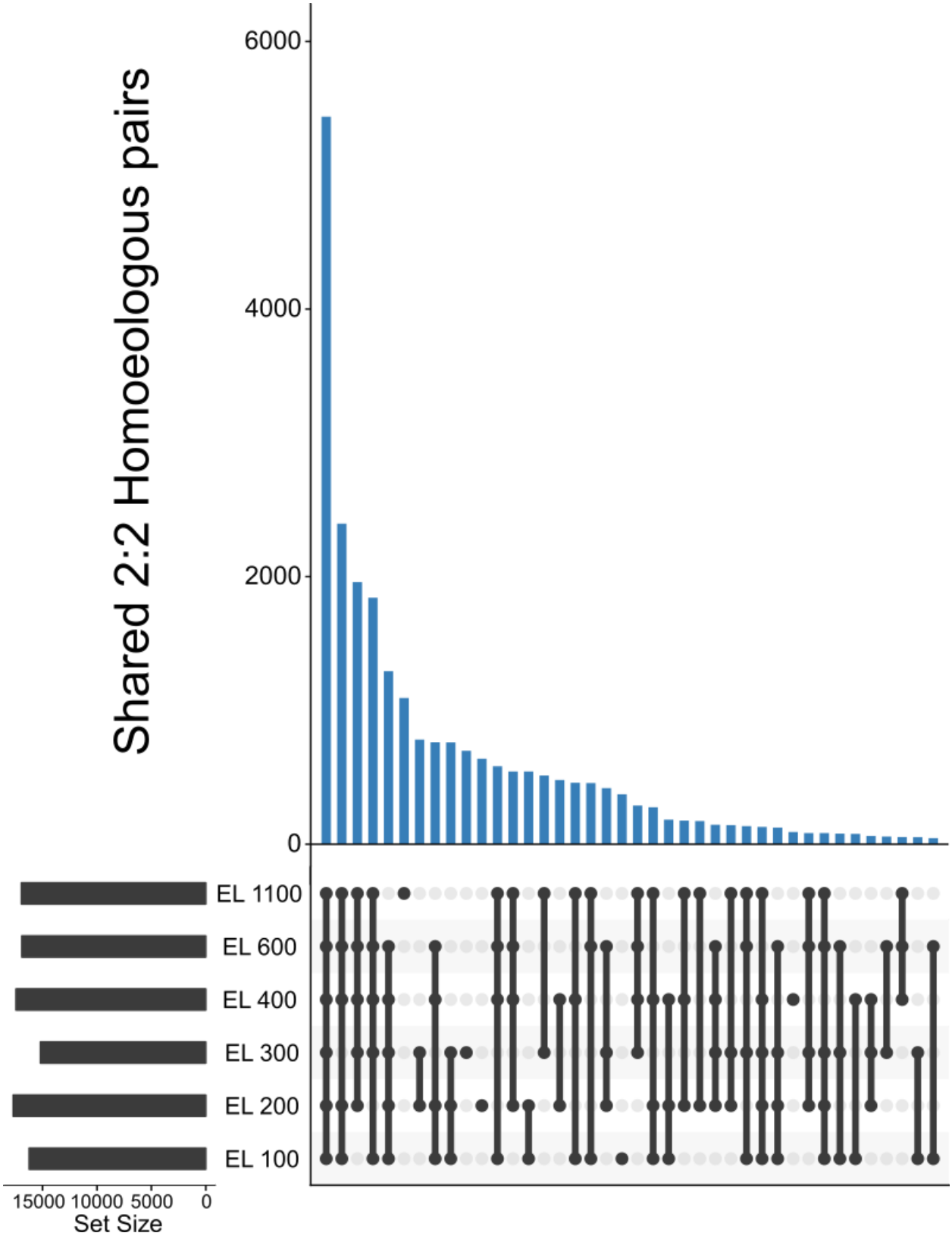
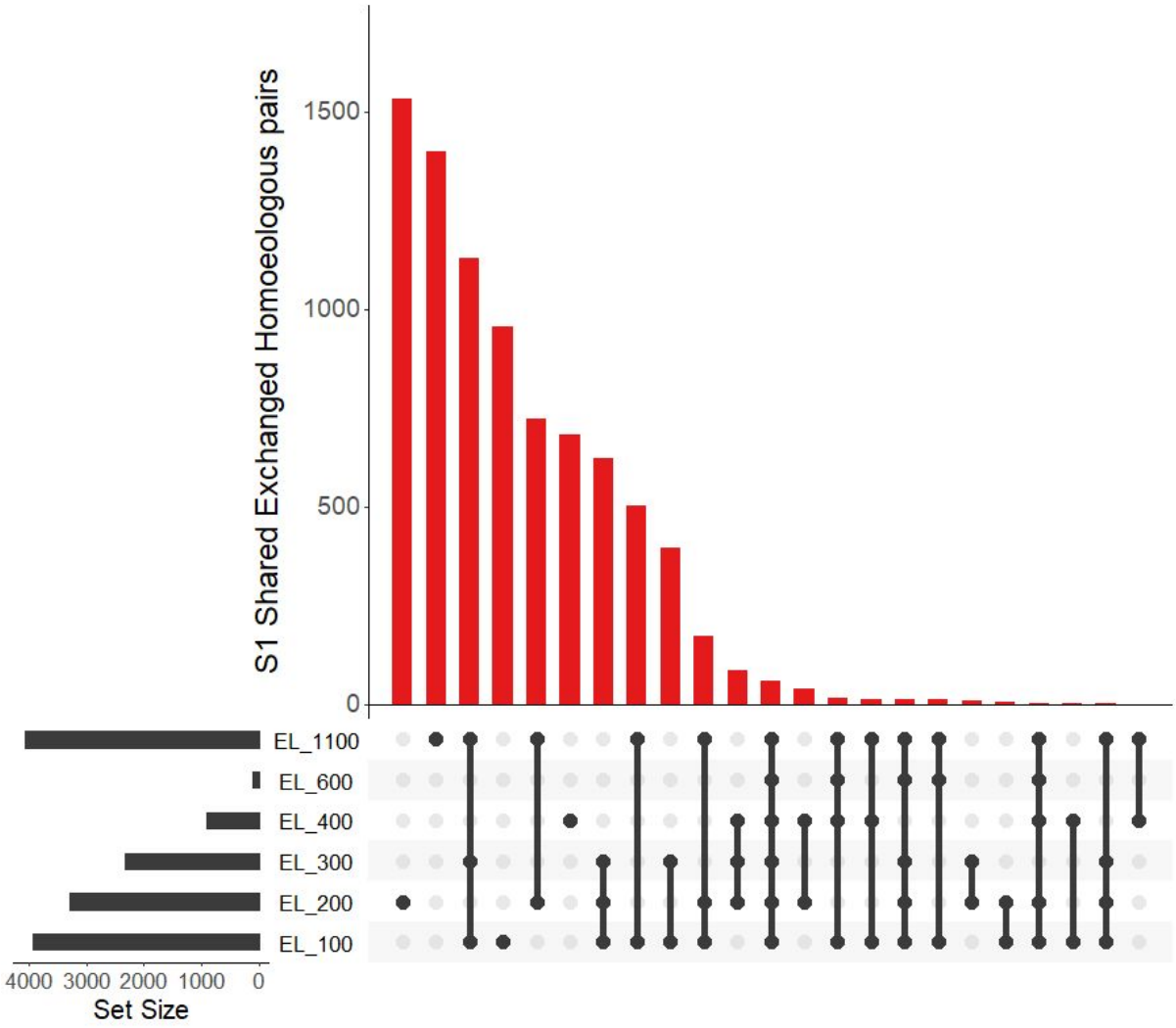
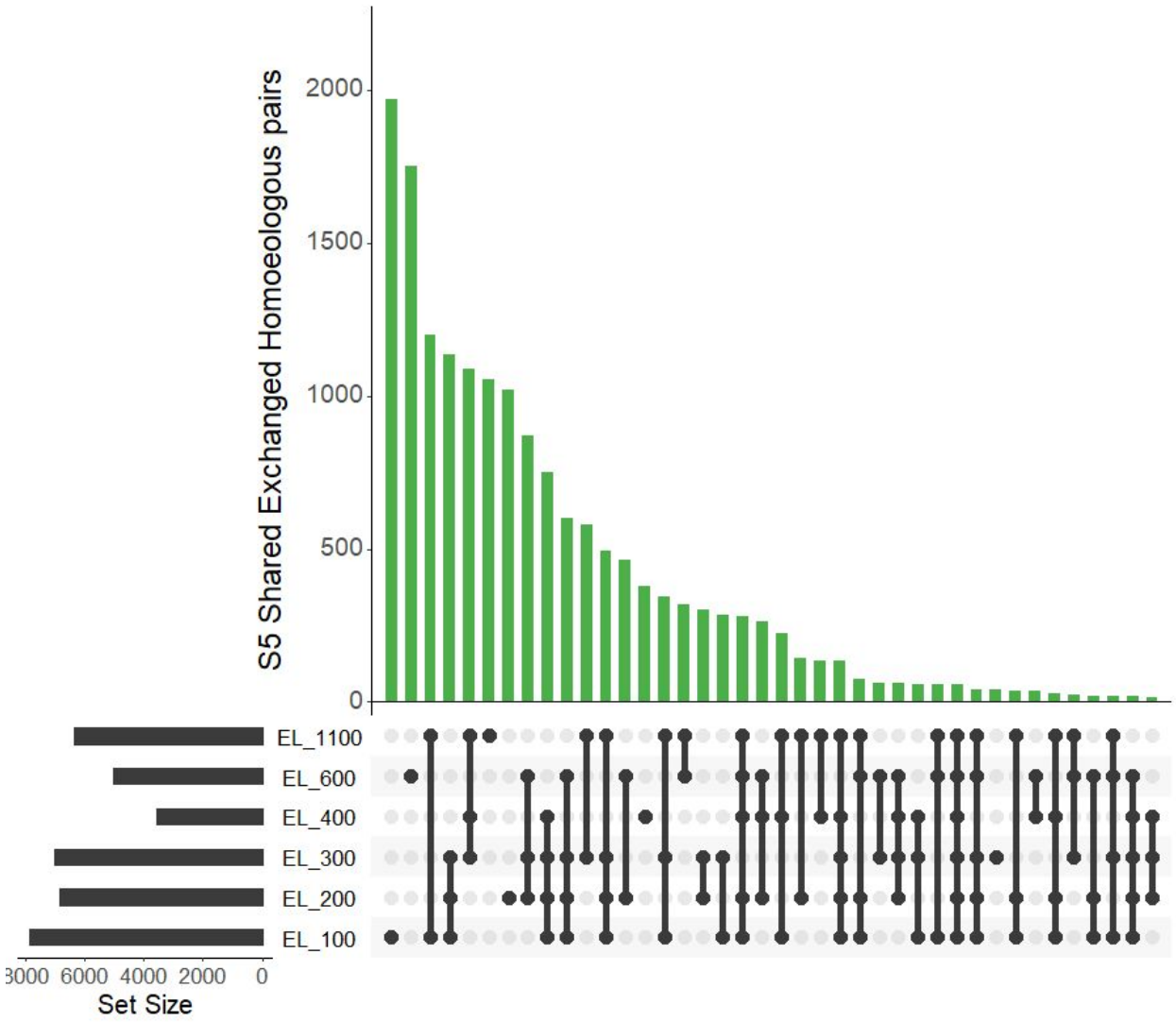
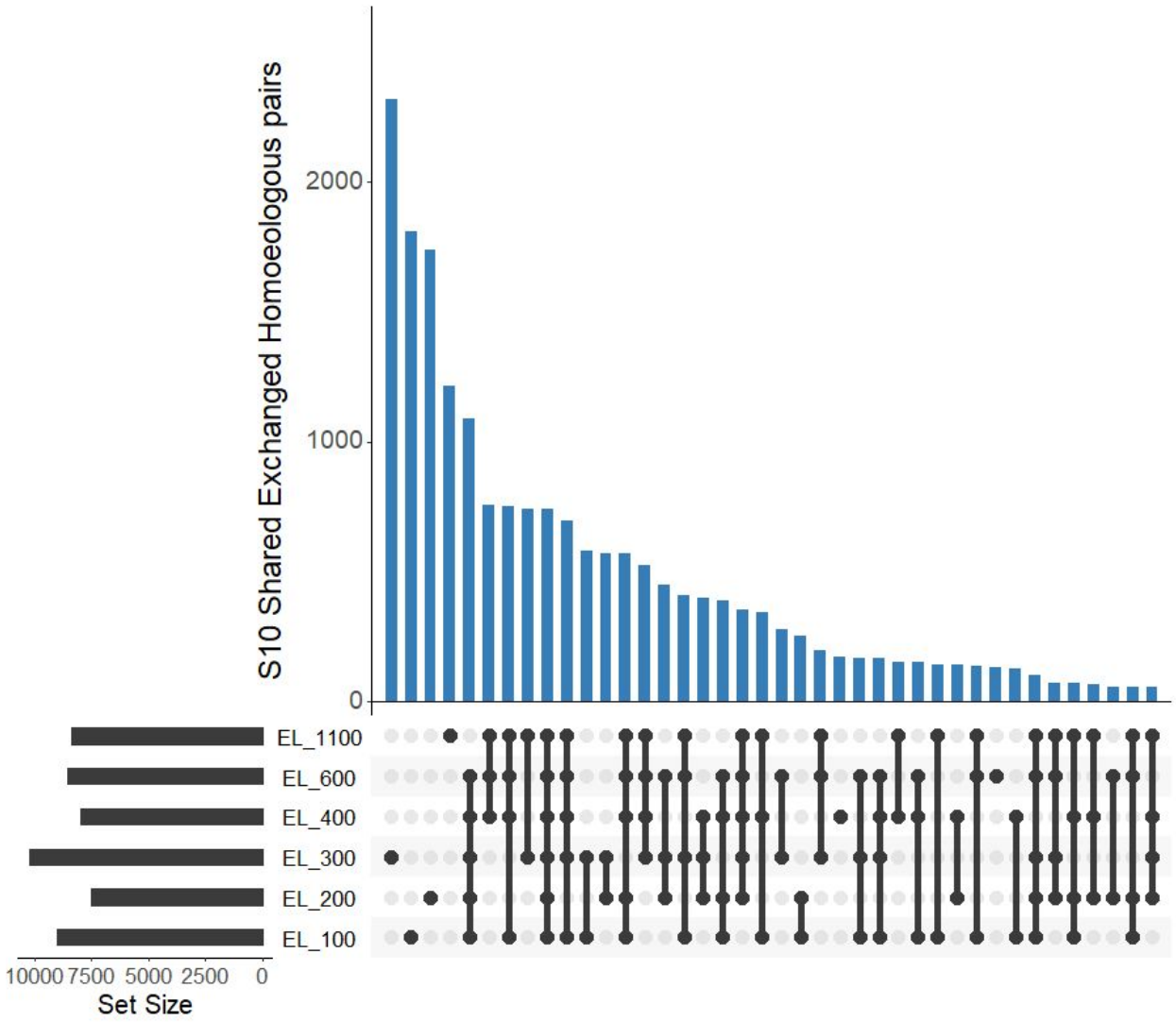

Shared 2:2 Homoeologous pairs
6000
4000
2000
0
EL 1100
EL 600
EL 400
EL 300
EL 200
EL 100
15000 10000 5000 0
Set Size

Figure 20 a-d **Shared Homoeologous exchanges** Homoeologous exchanged gene pairs shared across all six lines for a) generation , b) generation 5, c) generation 10, and d) shared across any of the 3 generations.




S5 Shared Exchanged Homoeologous pairs
2000
1500
1000
500
0
EL_1100
EL_600
EL_400
EL_300
EL_200
EL_100
8000 6000 4000 2000 0
Set Size


S10 Shared Exchanged Homoeologous pairs
2000
1000
0
EL_1100
EL_600
EL_400
EL_300
EL_200
EL_100
10000 7500 5000 2500 0
Set Size

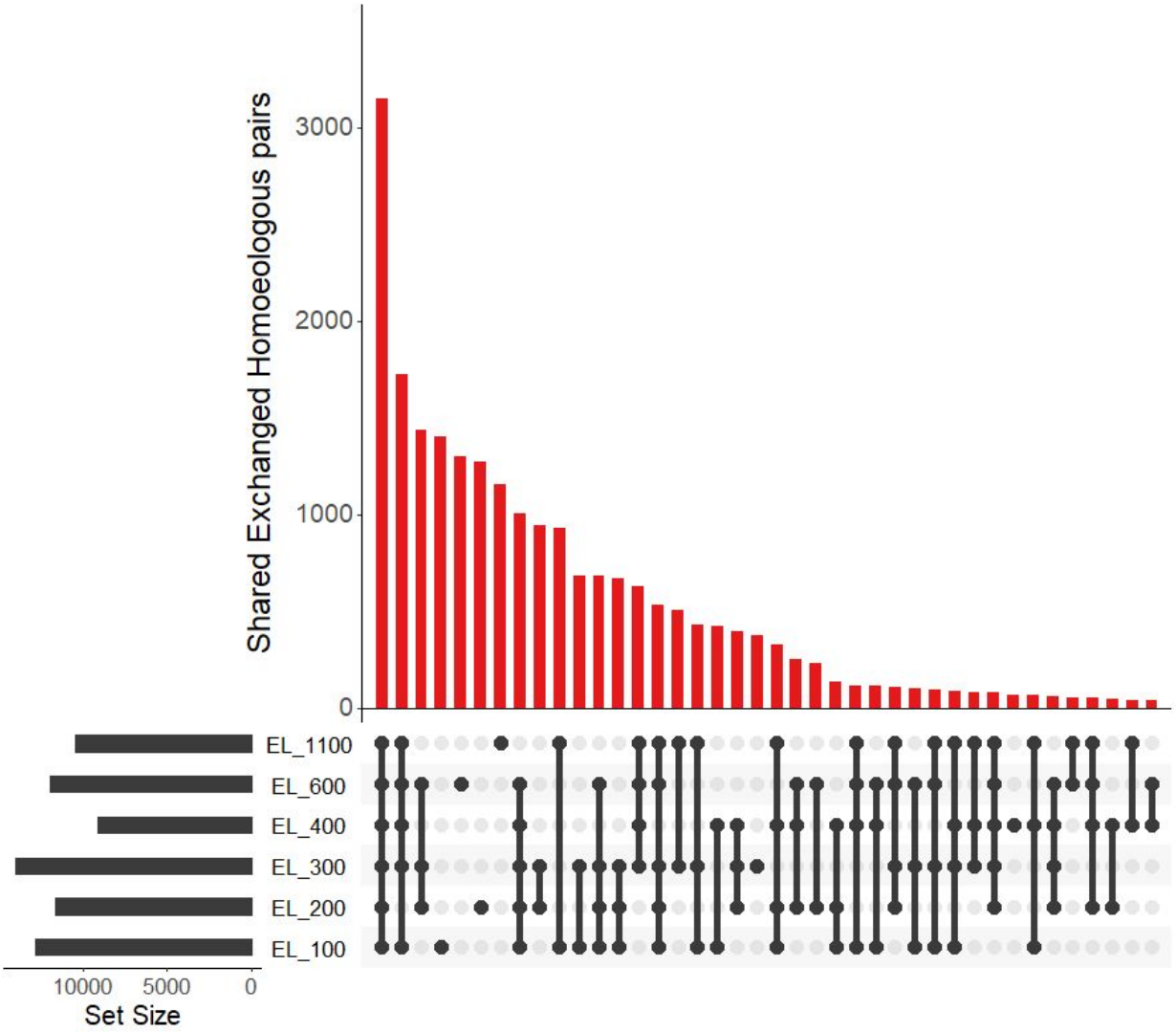
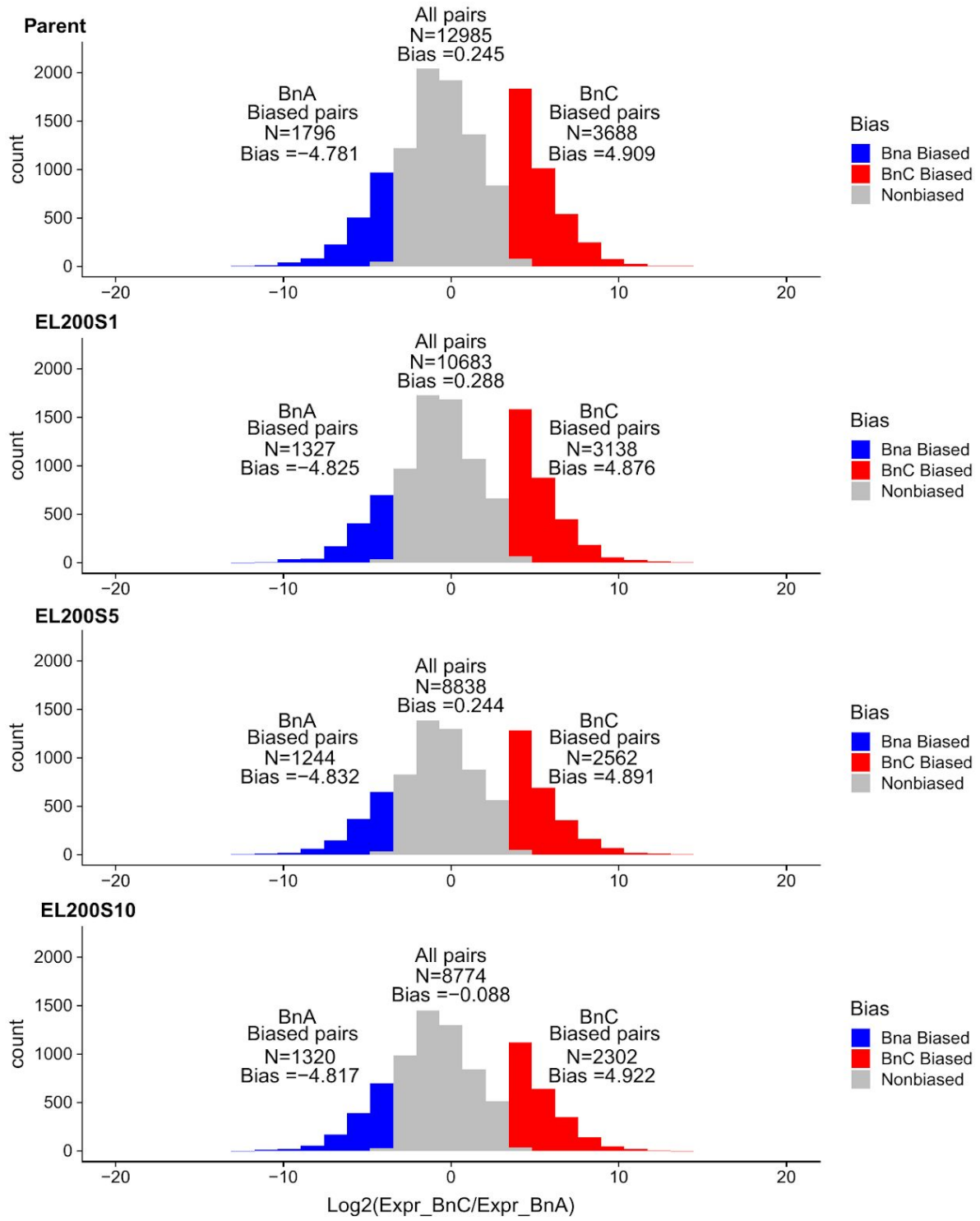
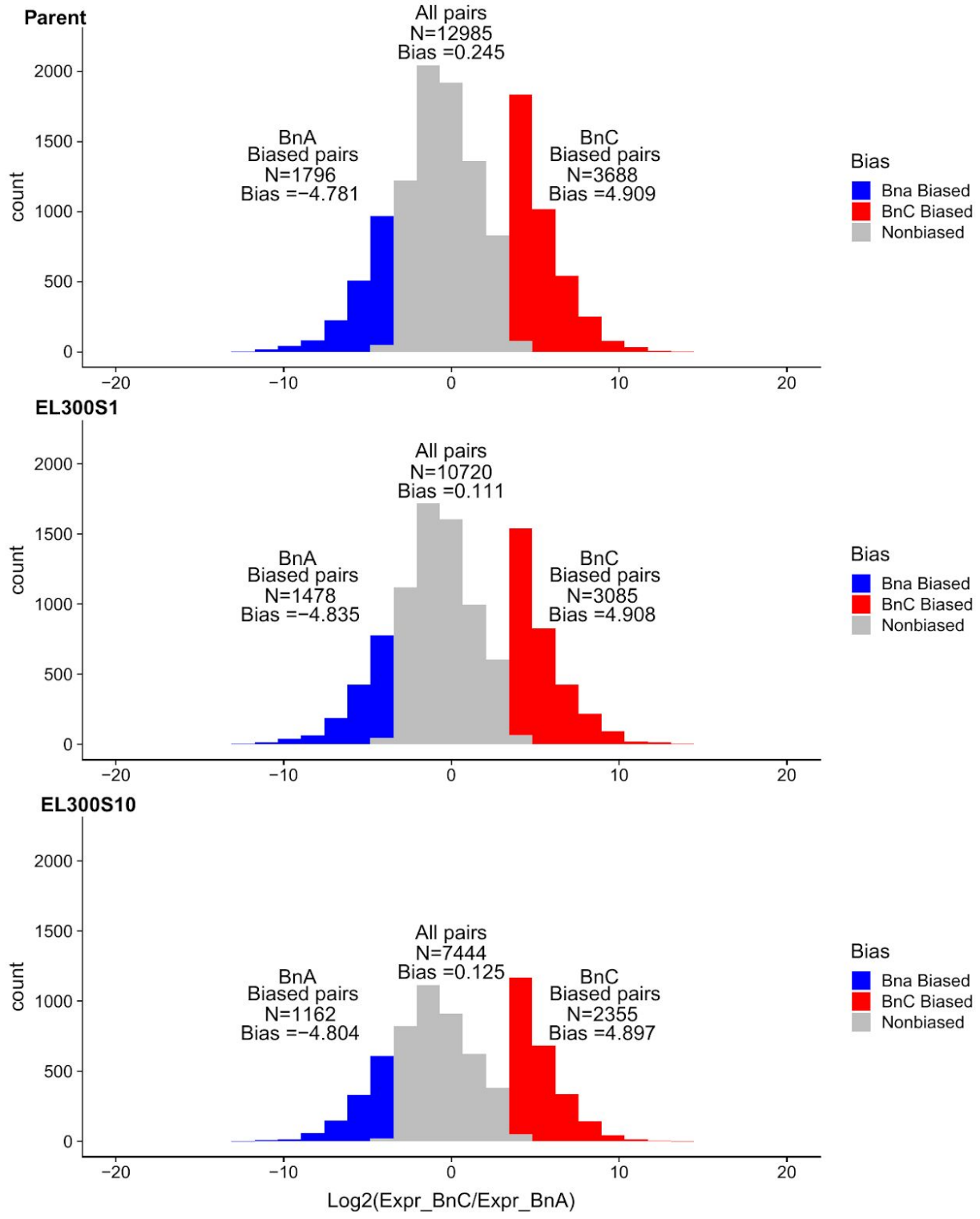

Shared Exchanged Homoeologous pairs
3000
2000
1000
0
EL_1100
EL_600
EL_400
EL_300
EL_200
EL_100
10000
5000
0
Set Size

Figure 21-26 **Homoeolog Expression Bias:** Distribution of homoeolog bias in the parent and three generations of line EL 200, EL 300, EL 400, EL 600, and EL 1100, red regions indicate BnC biased homeologos with log2 expression foldchange greater than 3.5 and blue regions indicate BnA biased homoeologs with log2 expression foldchange less than -3.5.

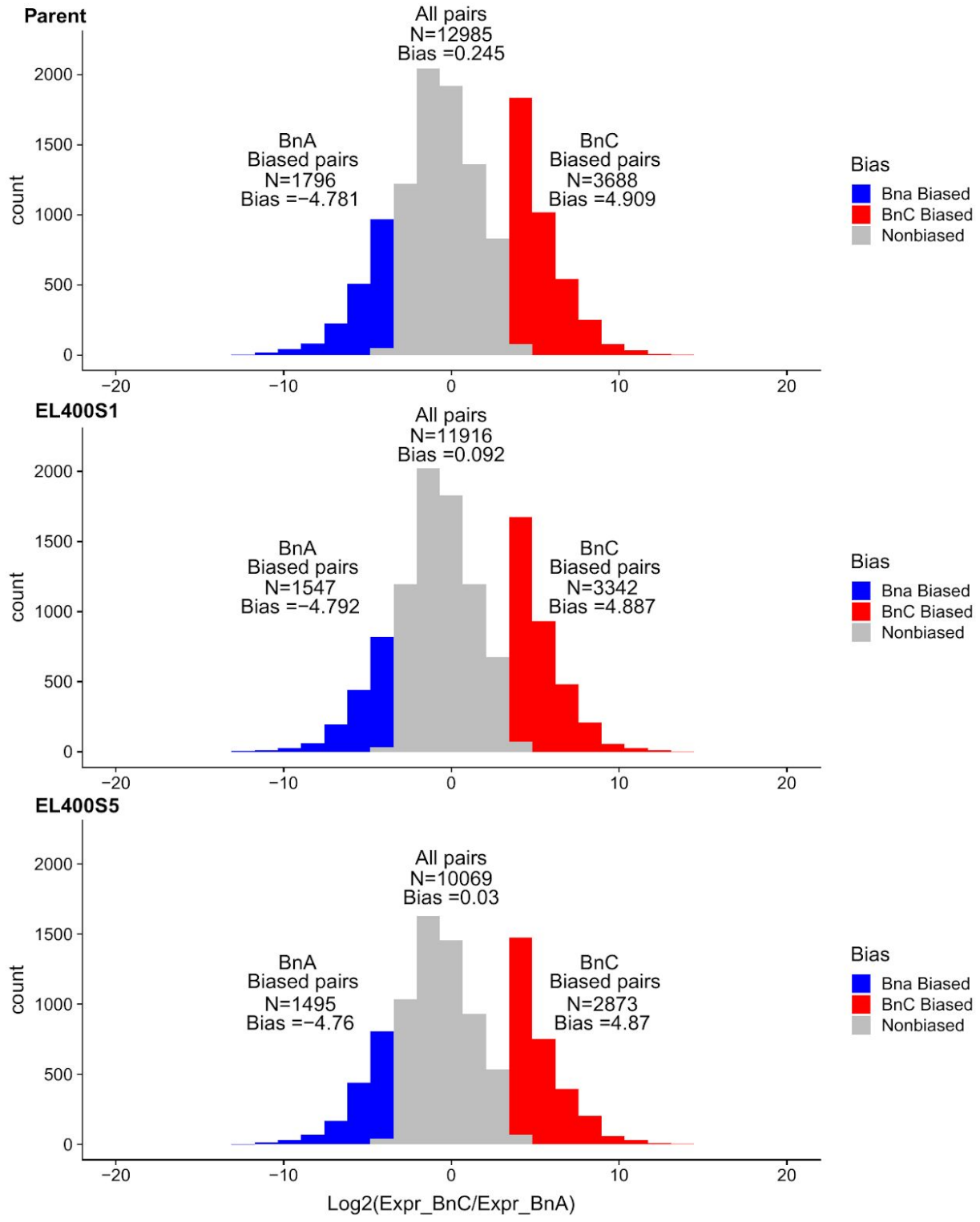
EL 200



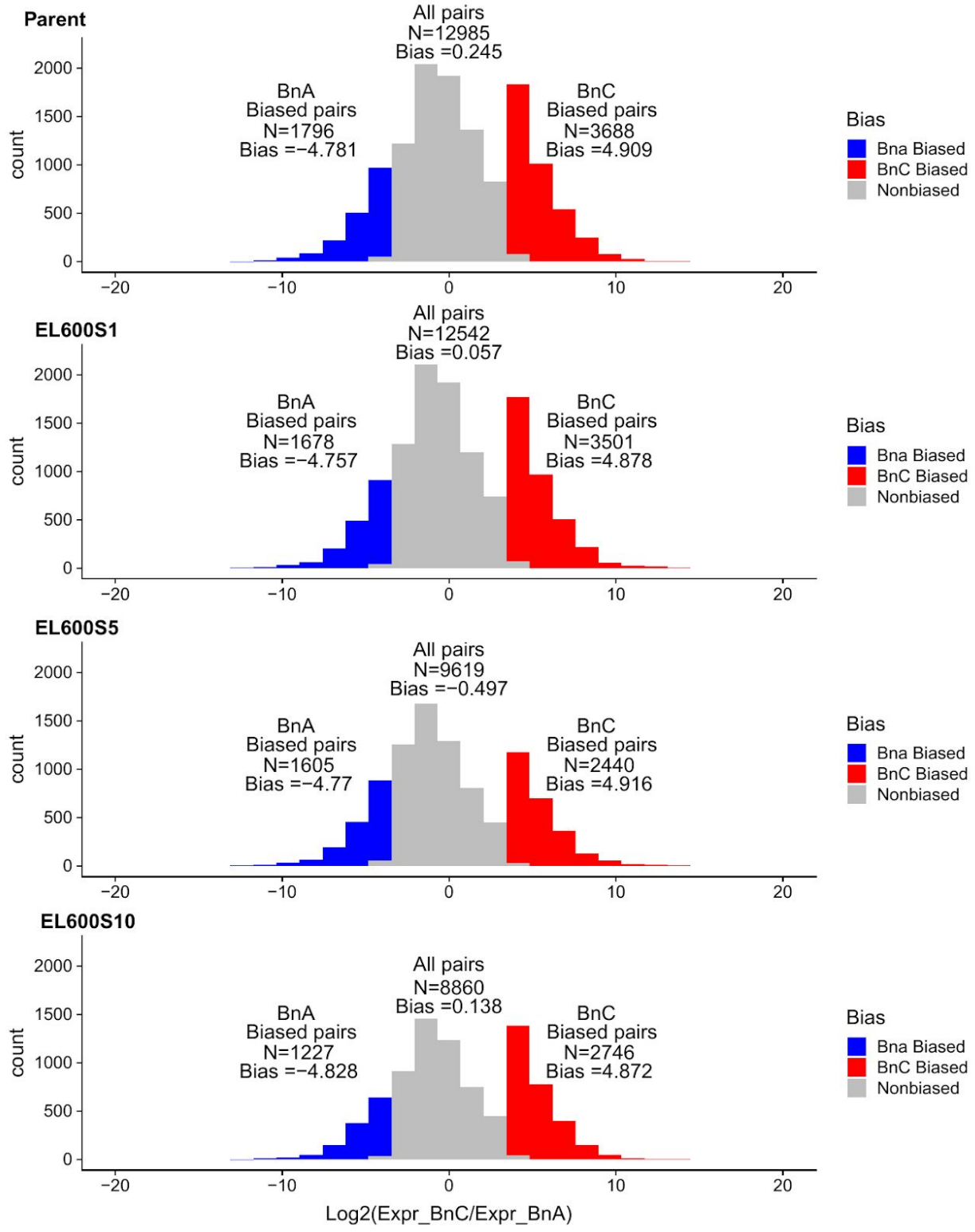
EL 300


Parent
All pairs
N=12985
Bias =0.245
BnA
Biased pairs
N=1796
Bias =−4.781
BnC
Biased pairs
N=3688
Bias =4.909
EL300S1
All pairs
N=10720
Bias =0.111
BnA
Biased pairs
N=1478
Bias =−4.835
BnC
Biased pairs
N=3085
Bias =4.908
EL300S10
All pairs
N=7444
Bias =0.125
BnA
Biased pairs
N=1162
Bias =−4.804
BnC
Biased pairs
N=2355
Bias =4.897
count
Log2(Expr_BnC/Expr_BnA)
Bias
Bna Biased
BnC Biased
Nonbiased

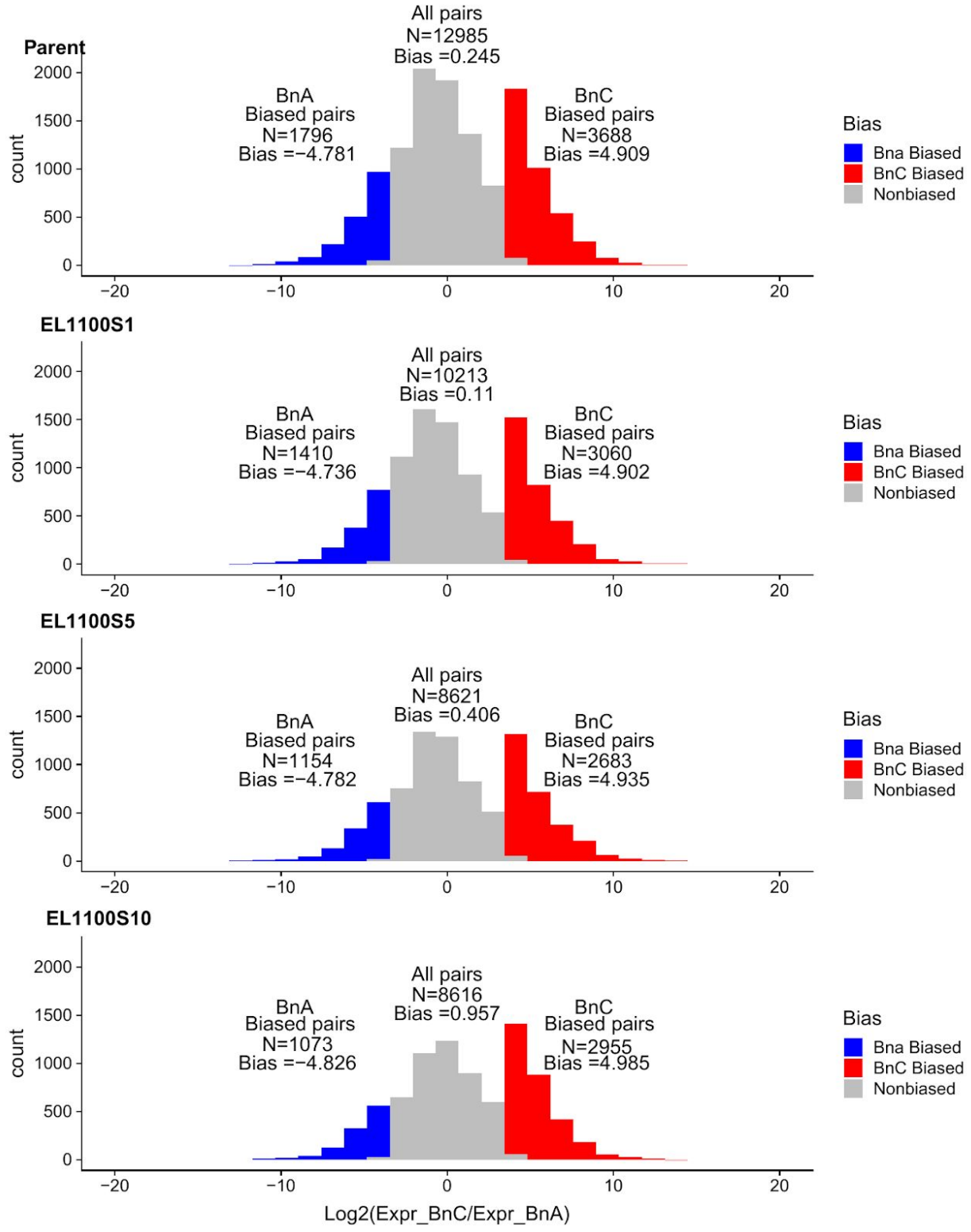
EL 400


Parent
All pairs
N=12985
Bias =0.245
BnA
Biased pairs
N=1796
Bias =−4.781
BnC
Biased pairs
N=3688
Bias =4.909
EL400S1
All pairs
N=11916
Bias =0.092
BnA
Biased pairs
N=1547
Bias =−4.792
BnC
Biased pairs
N=3342
Bias =4.887
EL400S5
All pairs
N=10069
Bias =0.03
BnA
Biased pairs
N=1495
Bias =−4.76
BnC
Biased pairs
N=2873
Bias =4.87
count
Log2(Expr_BnC/Expr_BnA)
Bias
Bna Biased
BnC Biased
Nonbiased

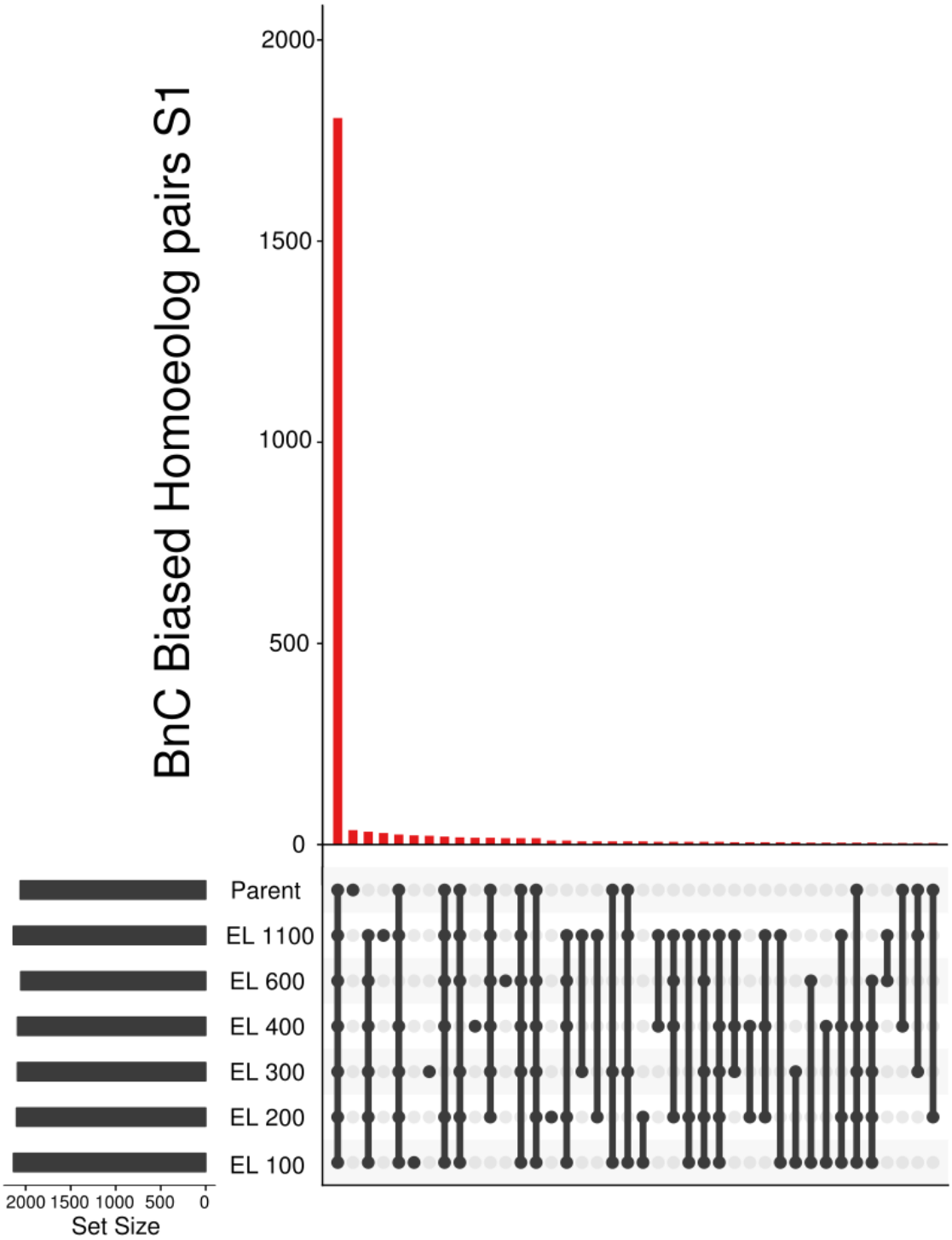
EL 600

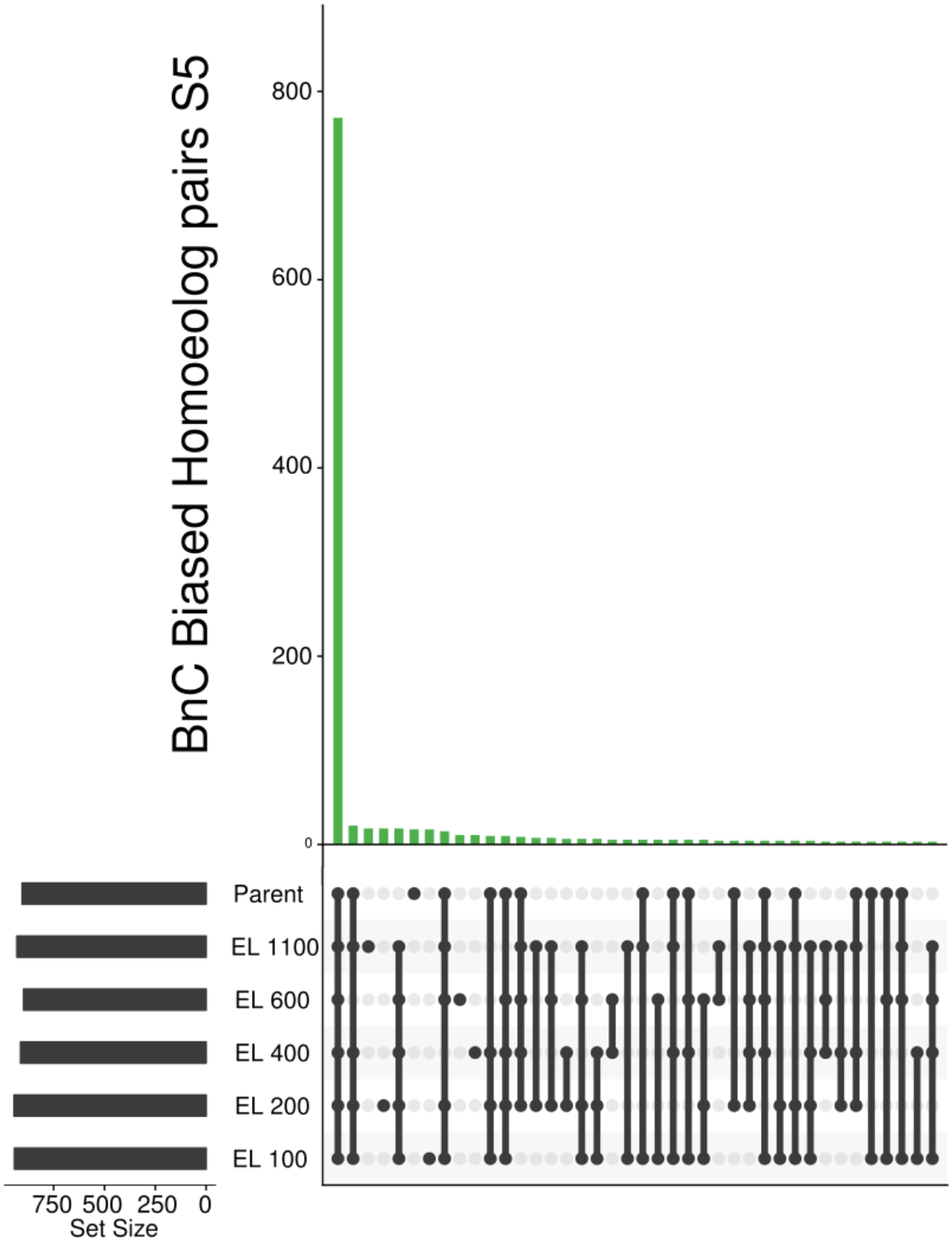

Parent
All pairs
N=12985
Bias =0.245
BnA
Biased pairs
N=1796
Bias =−4.781
BnC
Biased pairs
N=3688
Bias =4.909
EL600S1
All pairs
N=12542
Bias =0.057
BnA
Biased pairs
N=1678
Bias =−4.757
BnC
Biased pairs
N=3501
Bias =4.878
EL600S5
All pairs
N=9619
Bias =−0.497
BnA
Biased pairs
N=1605
Bias =−4.77
BnC
Biased pairs
N=2440
Bias =4.916
EL600S10
All pairs
N=8860
Bias =0.138
BnA
Biased pairs
N=1227
Bias =−4.828
BnC
Biased pairs
N=2746
Bias =4.872
count
Log2(Expr_BnC/Expr_BnA)
Bias
Bna Biased
BnC Biased
Nonbiased

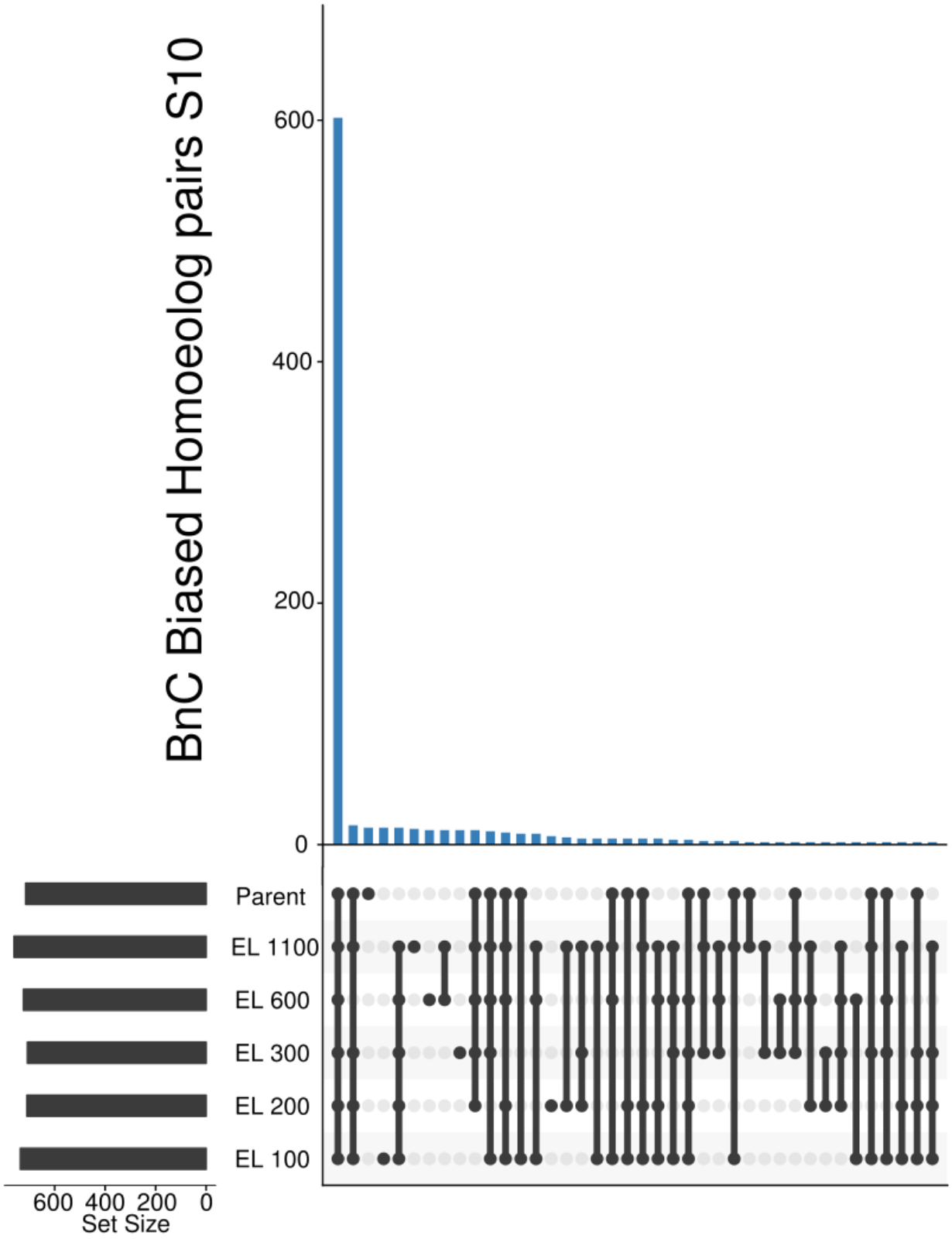
EL 1100

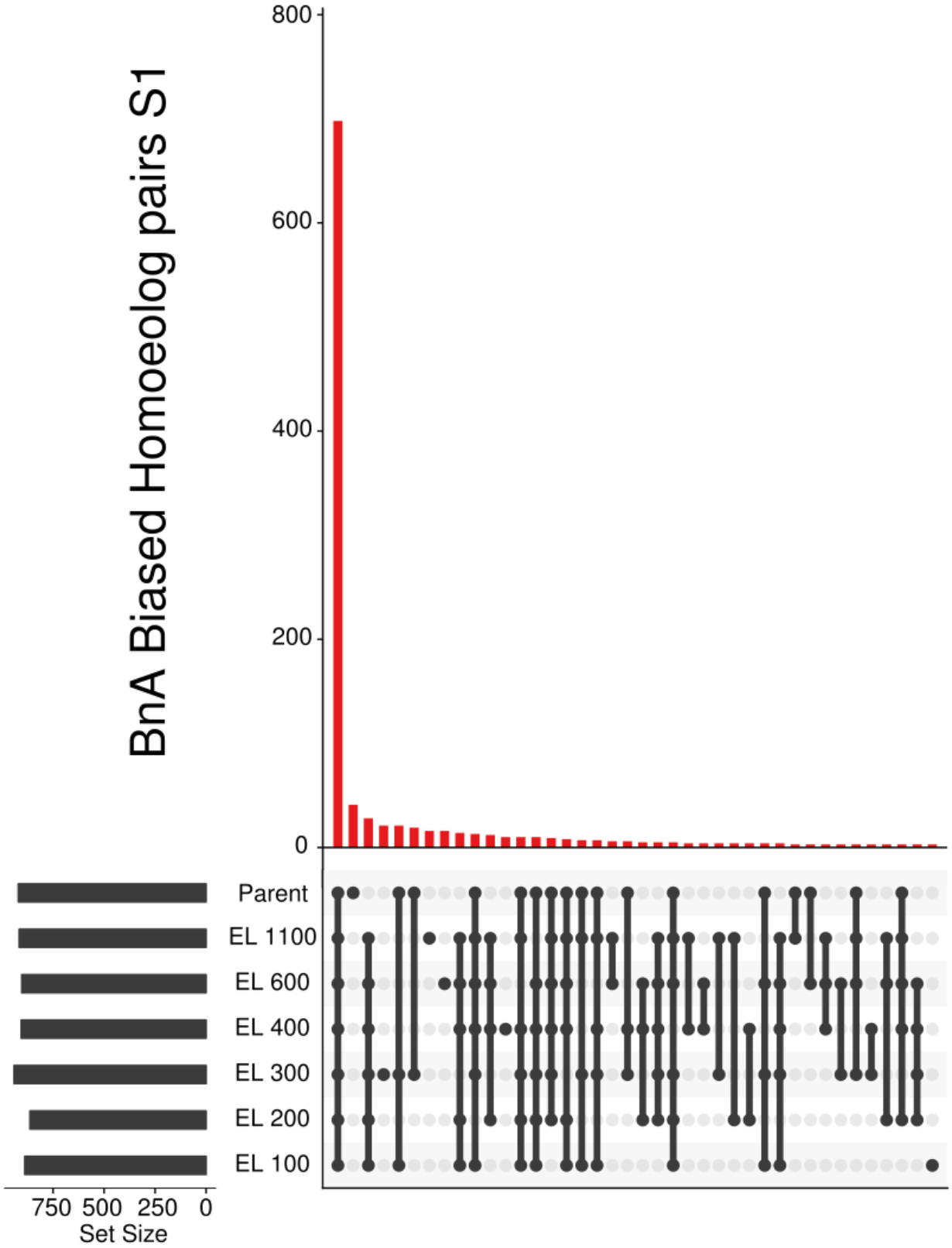

Parent
All pairs
N=12985
Bias =0.245
BnA
Biased pairs
N=1796
Bias =−4.781
BnC
Biased pairs
N=3688
Bias =4.909
EL1100S1
All pairs
N=10213
Bias =0.11
BnA
Biased pairs
N=1410
Bias =−4.736
BnC
Biased pairs
N=3060
Bias =4.902
EL1100S5
All pairs
N=8621
Bias =0.406
BnA
Biased pairs
N=1154
Bias =−4.782
BnC
Biased pairs
N=2683
Bias =4.935
EL1100S10
All pairs
N=8616
Bias =0.957
BnA
Biased pairs
N=1073
Bias =−4.826
BnC
Biased pairs
N=2955
Bias =4.985
count
Log2(Expr_BnC/Expr_BnA)
Bias
Bna Biased
BnC Biased
Nonbiased

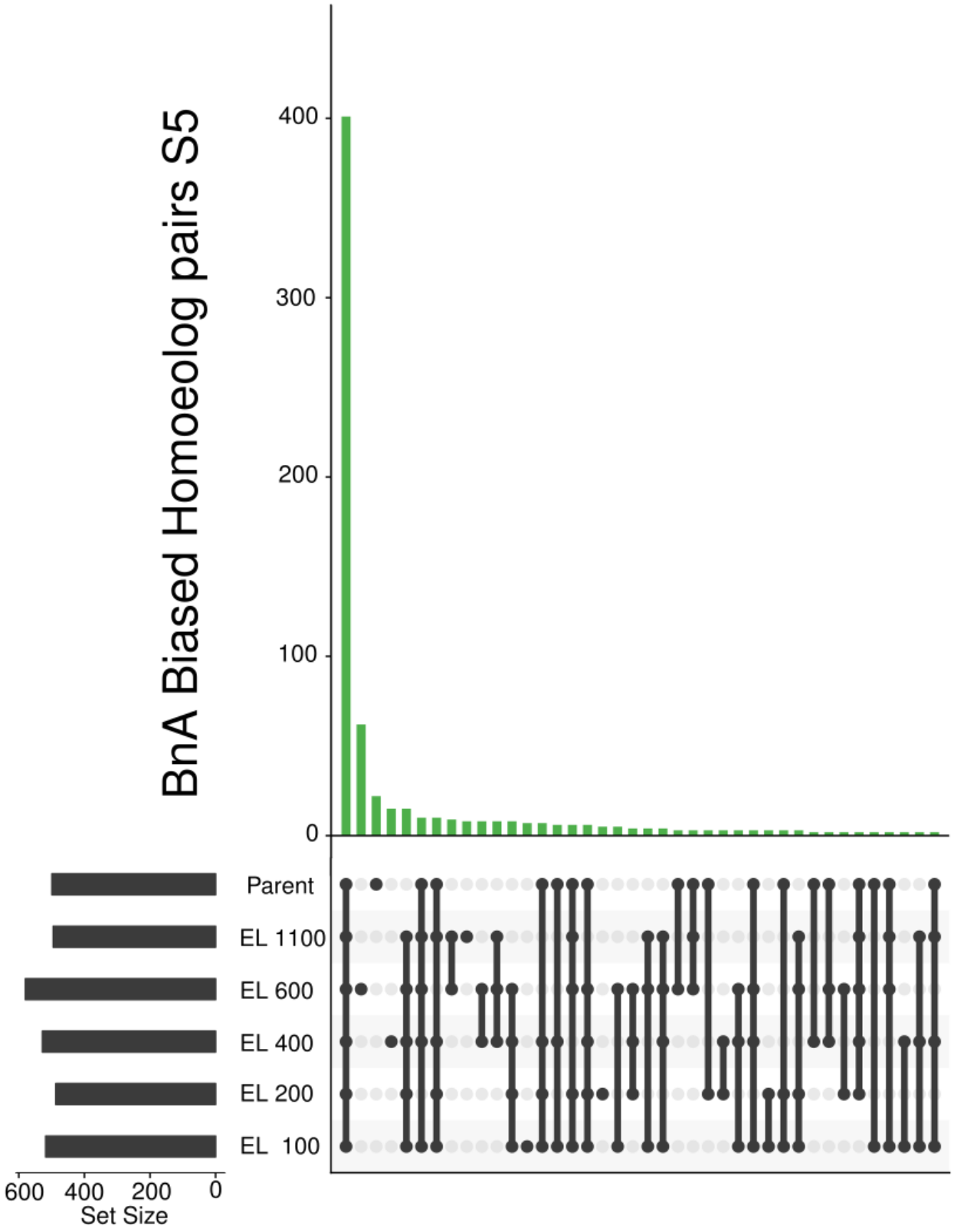
Figures 27 a-g **Common shared biased Homoeolog pairs:** Upset plot of how biased homoeologs pairs for BnC biased (a-c) and BnA biased (e-g) are shared among all six lines for the three sampled generations. These analysis was restricted only to homoeolog pairs in 2:2 balance in all 6 lines.




BnC Biased Homoeolog pairs S5
800
600
400
200
0
Parent
EL 1100
EL 600
EL 400
EL 200
EL 100
750 500 250 0
Set Size


BnC Biased Homoeolog pairs S10
600
400
200
0
Parent
EL 1100
EL 600
EL 300
EL 200
EL 100
600 400 200 0
Set Size


BnA Biased Homoeolog pairs S1
800
600
400
200
0
Parent
EL 1100
EL 600
EL 400
EL 300
EL 200
EL 100
750 500 250 0
Set Size


BnA Biased Homoeolog pairs S5
400
300
200
100
0
Parent
EL 1100
EL 600
EL 400
EL 200
EL 100
600
400
200
0
Set Size

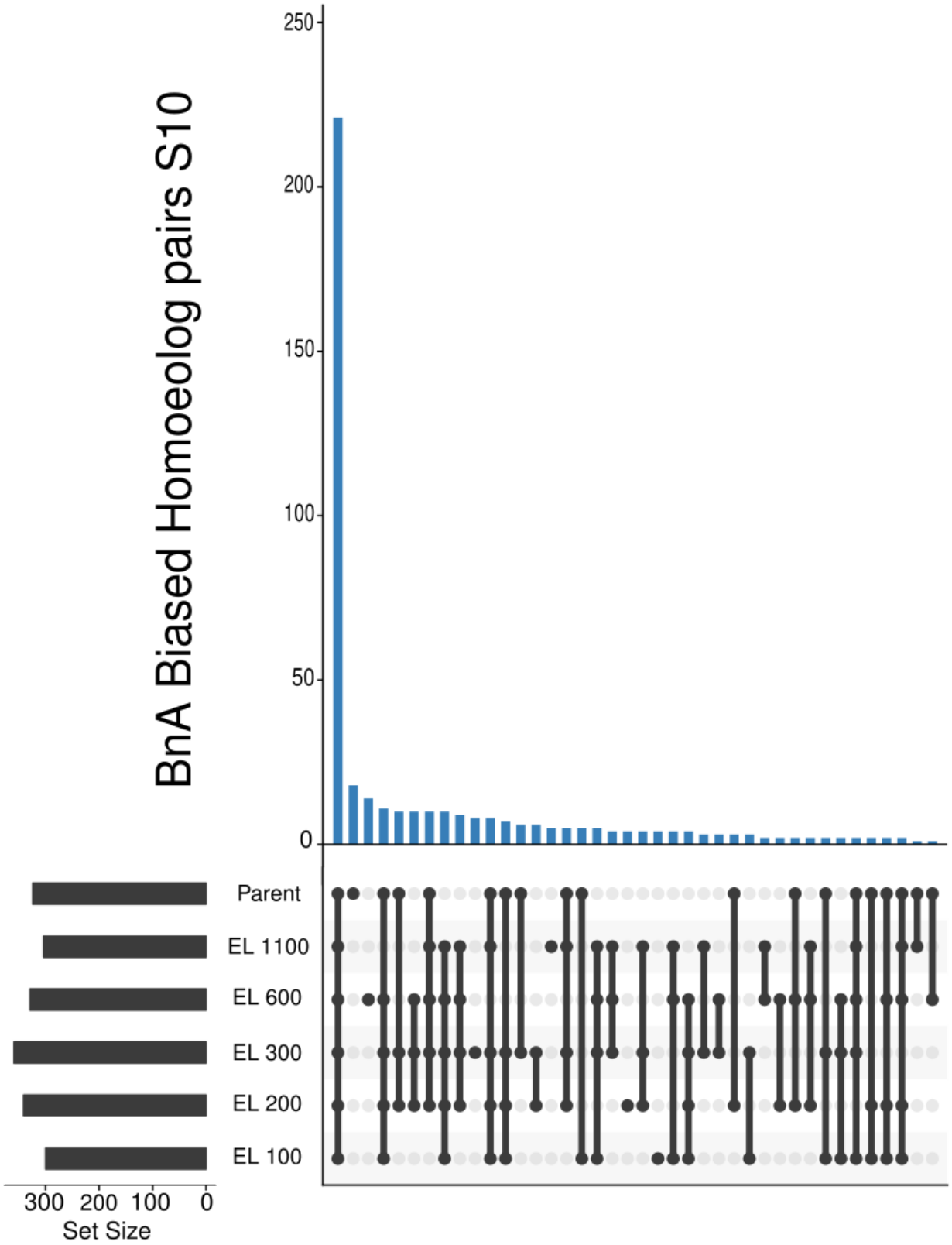
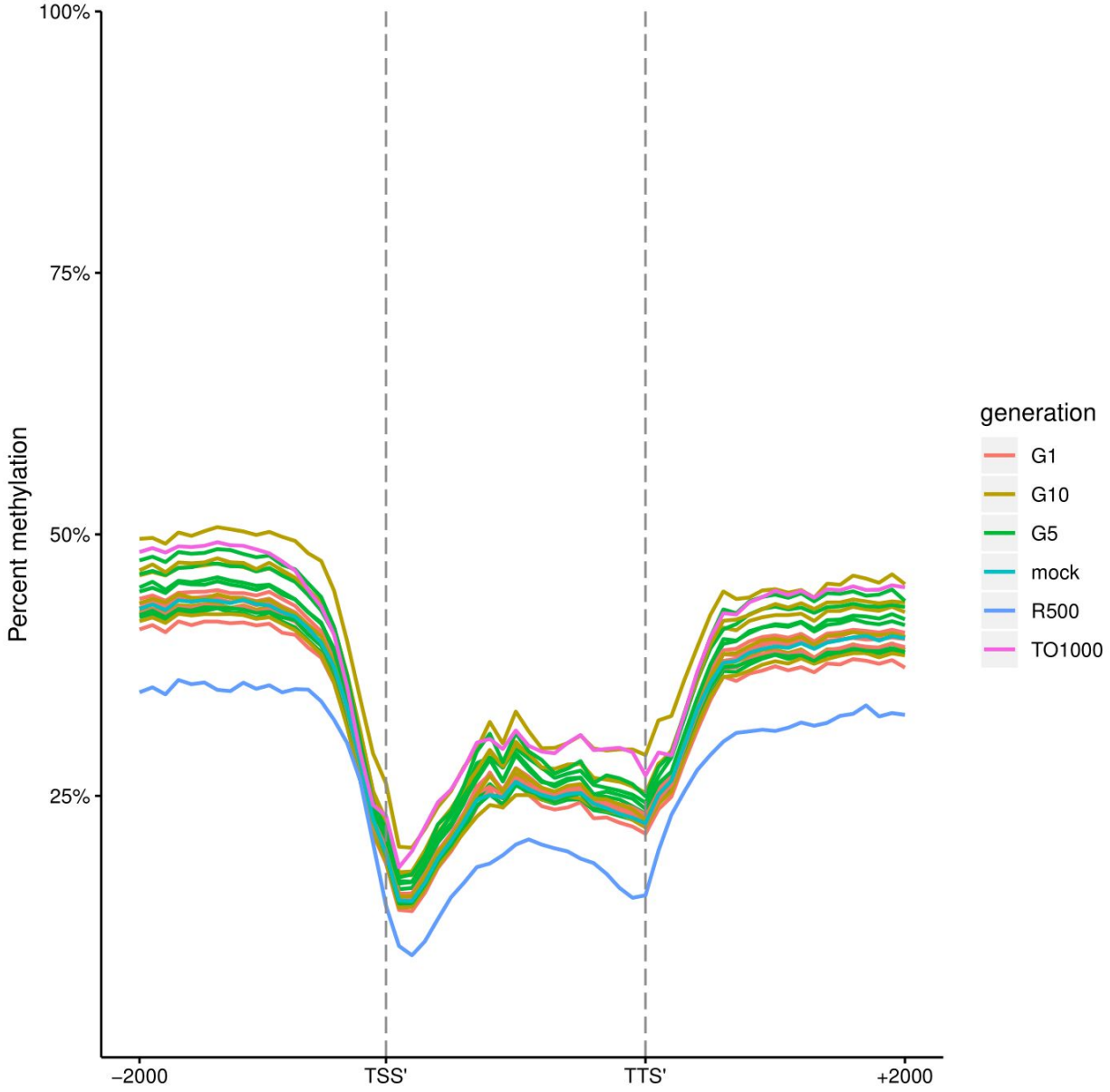
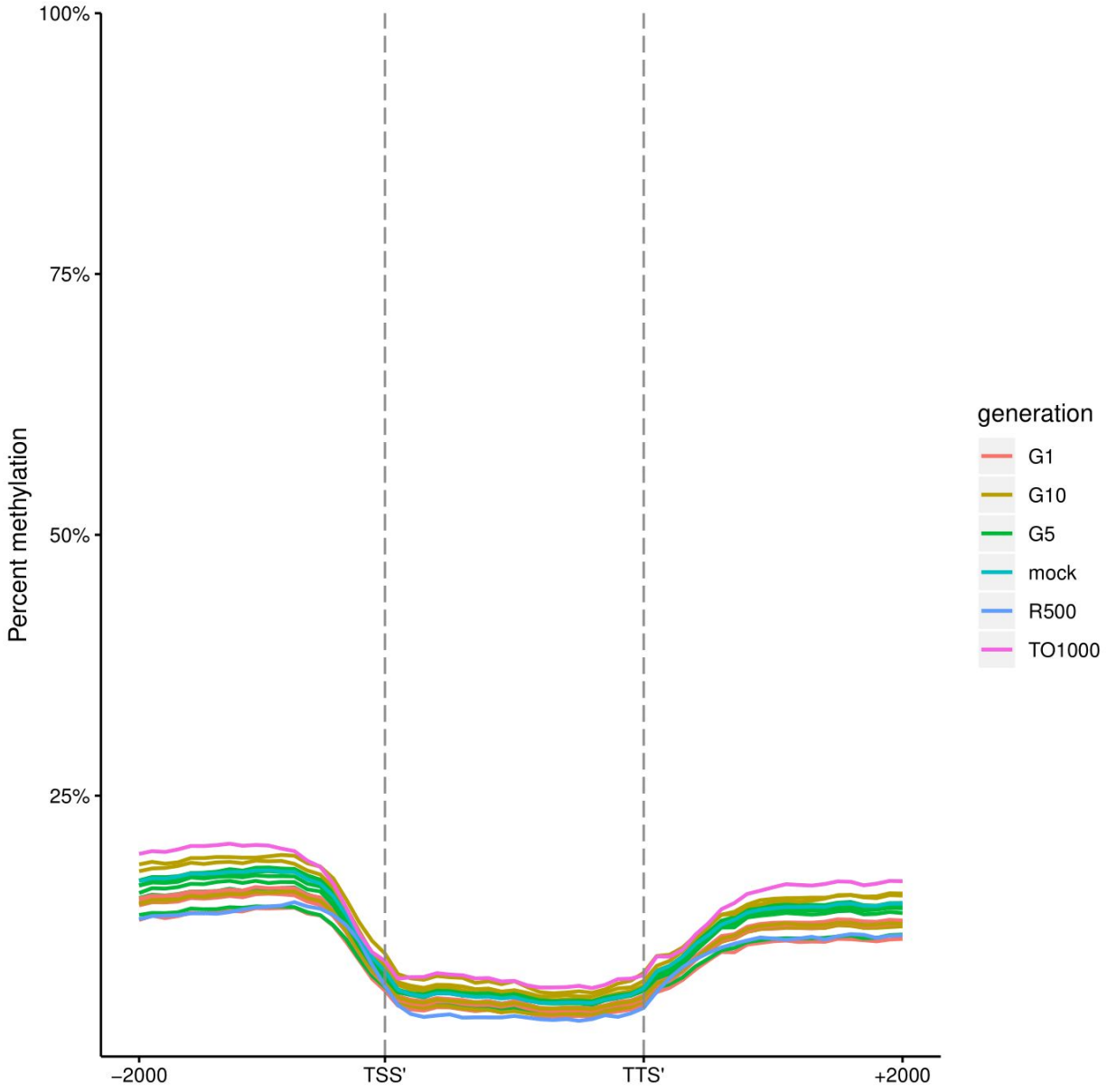

BnA Biased Homoeolog pairs S10
250
200
150
100
50
0
Parent
EL 1100
EL 600
EL 300
EL 200
EL 100
300
200
100
0
Set Size

Fig 28 a-c Metaplots of CG, CHG, and CHH mean weighted methylation of all annotated gene models 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

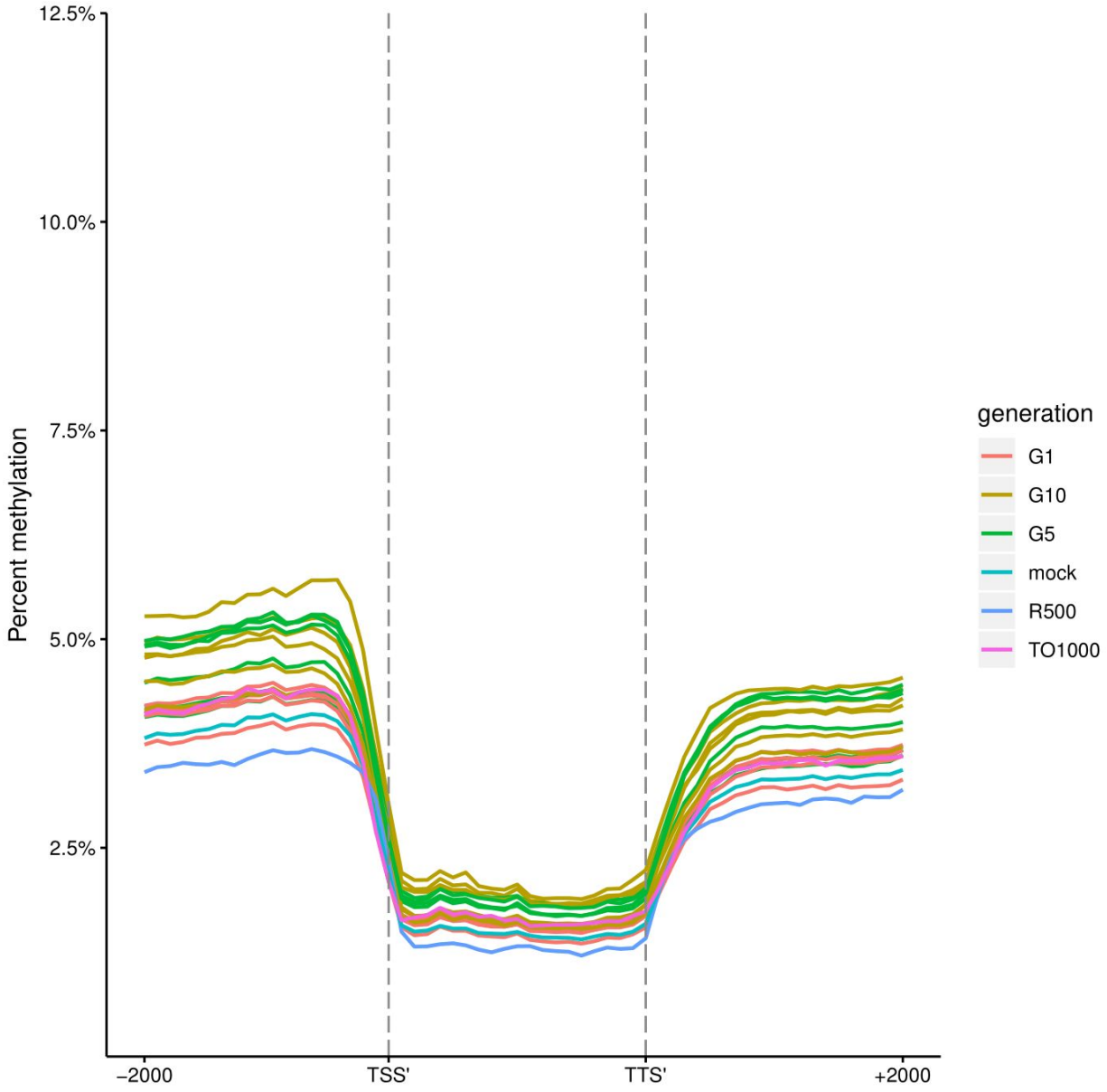
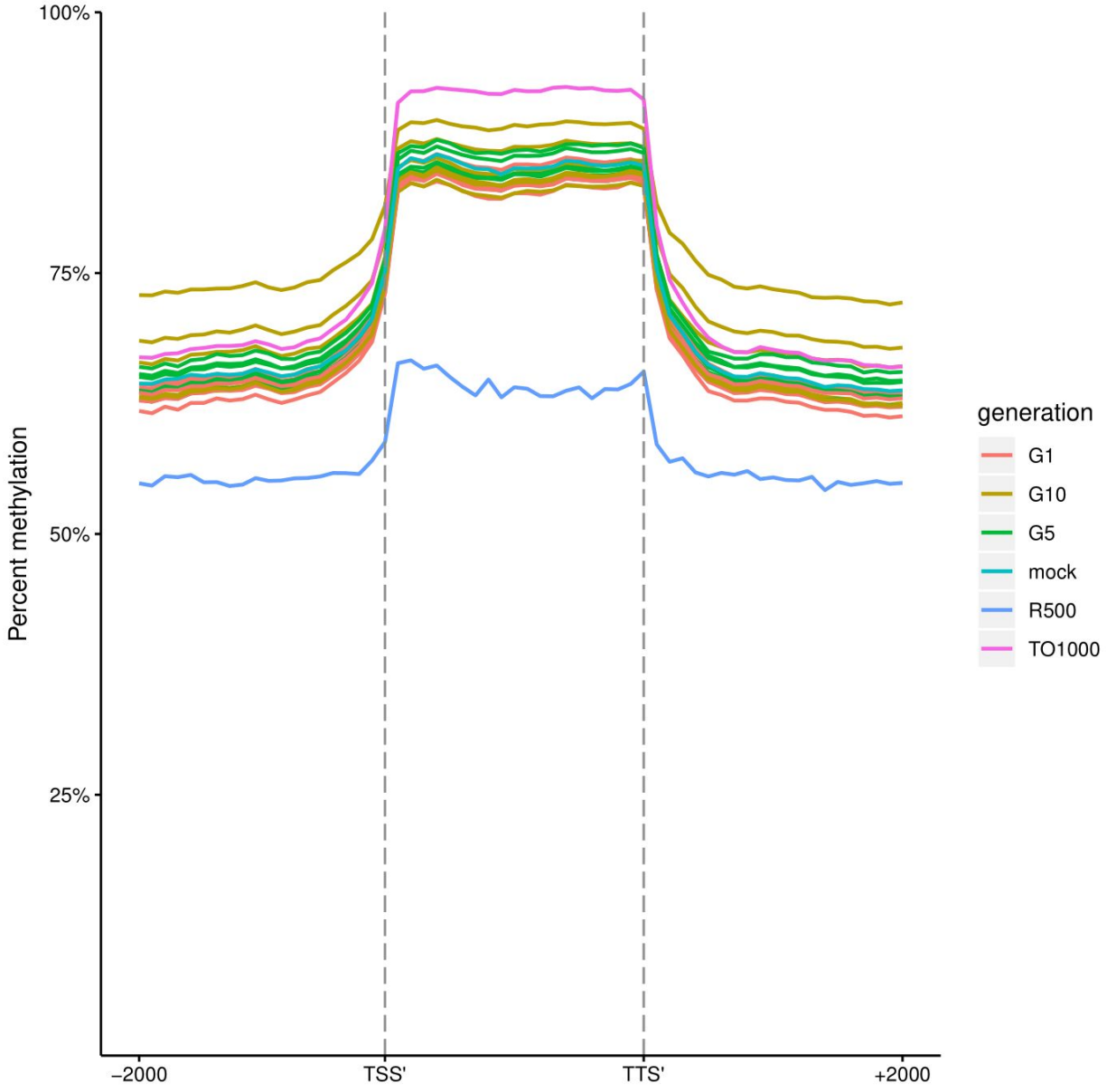
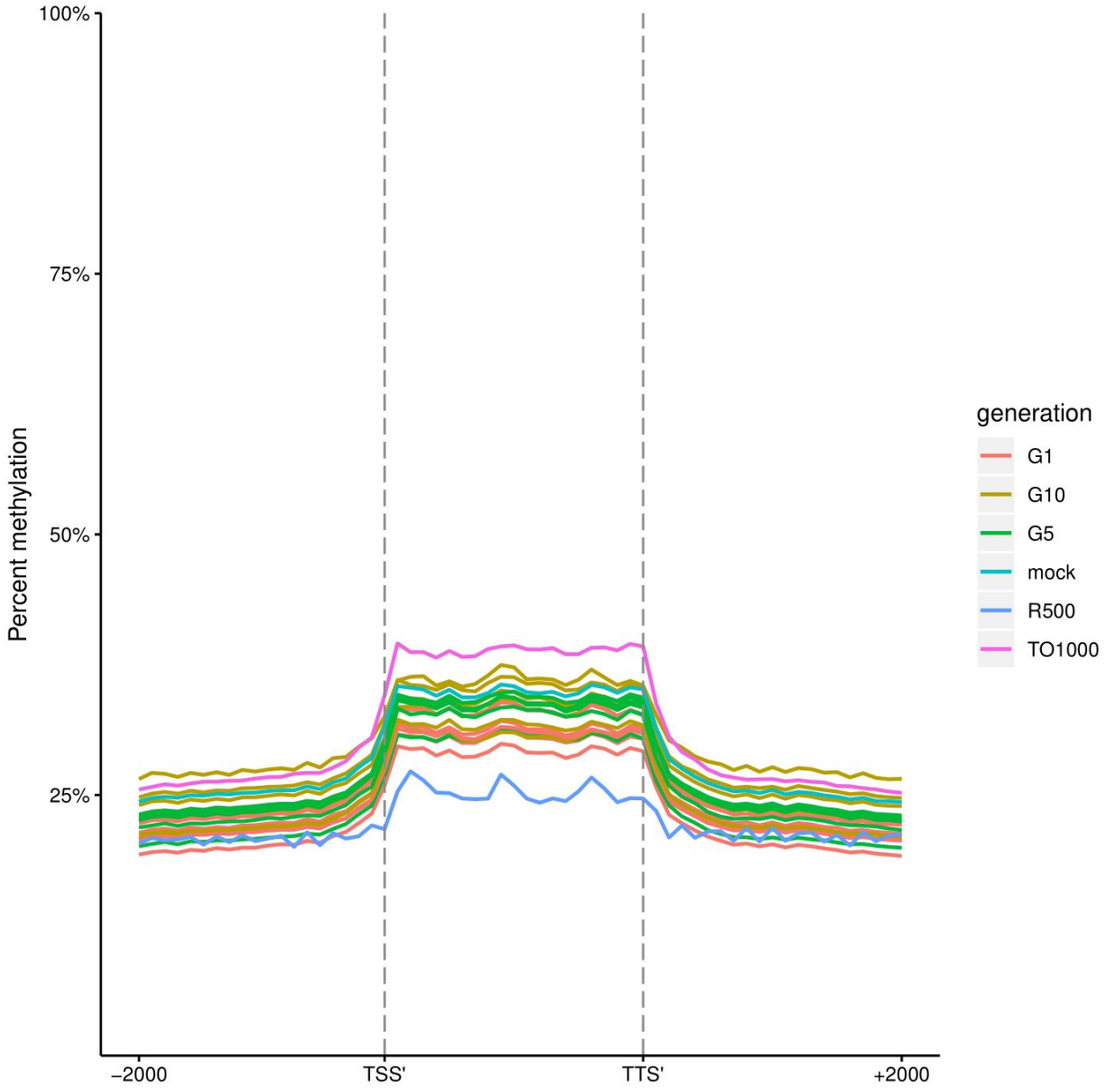

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Fig 29 a-c Metaplots of CG, CHG, and CHH mean weighted methylation of all annotated Long tandem repeat transposable elements 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

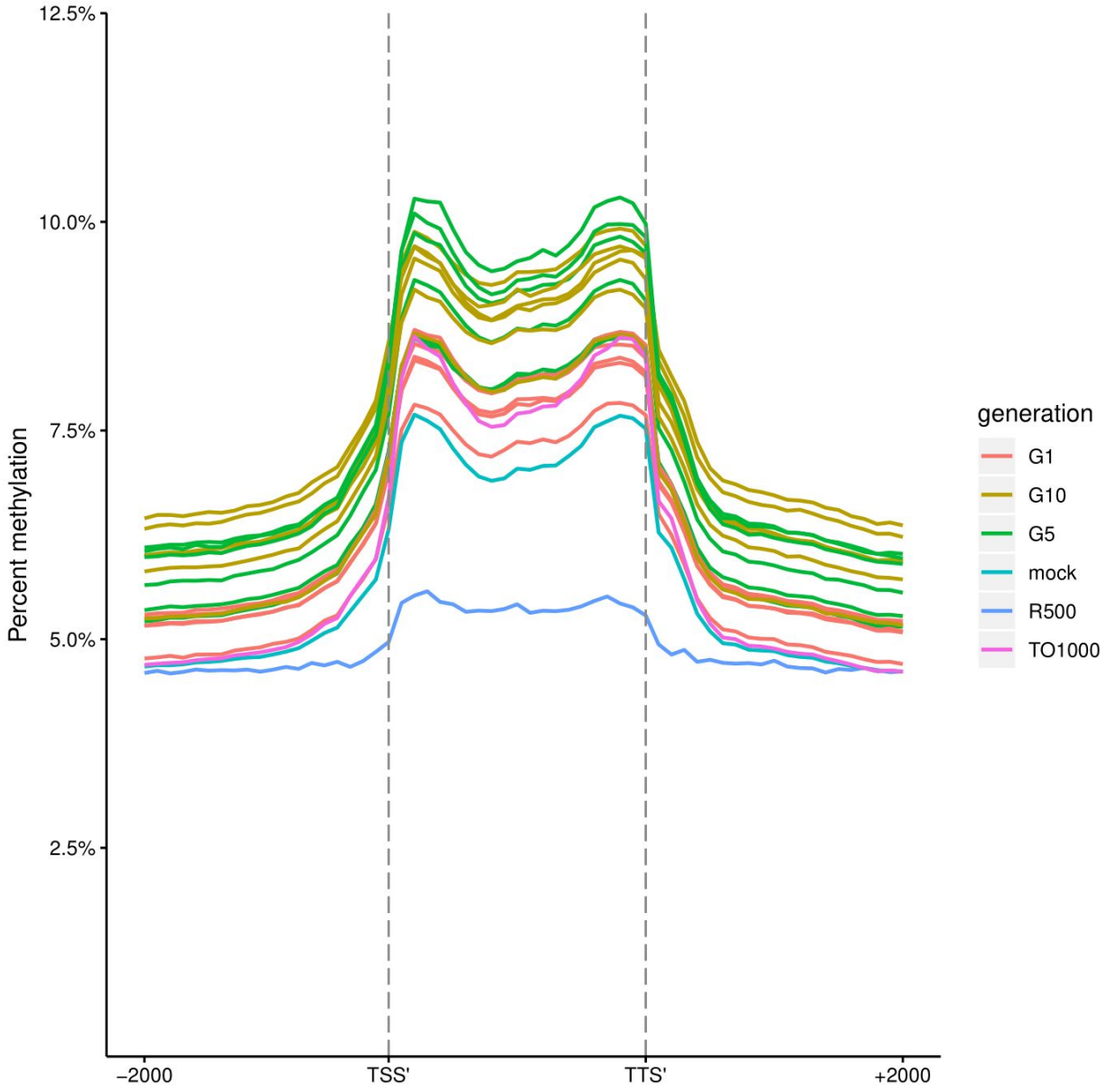
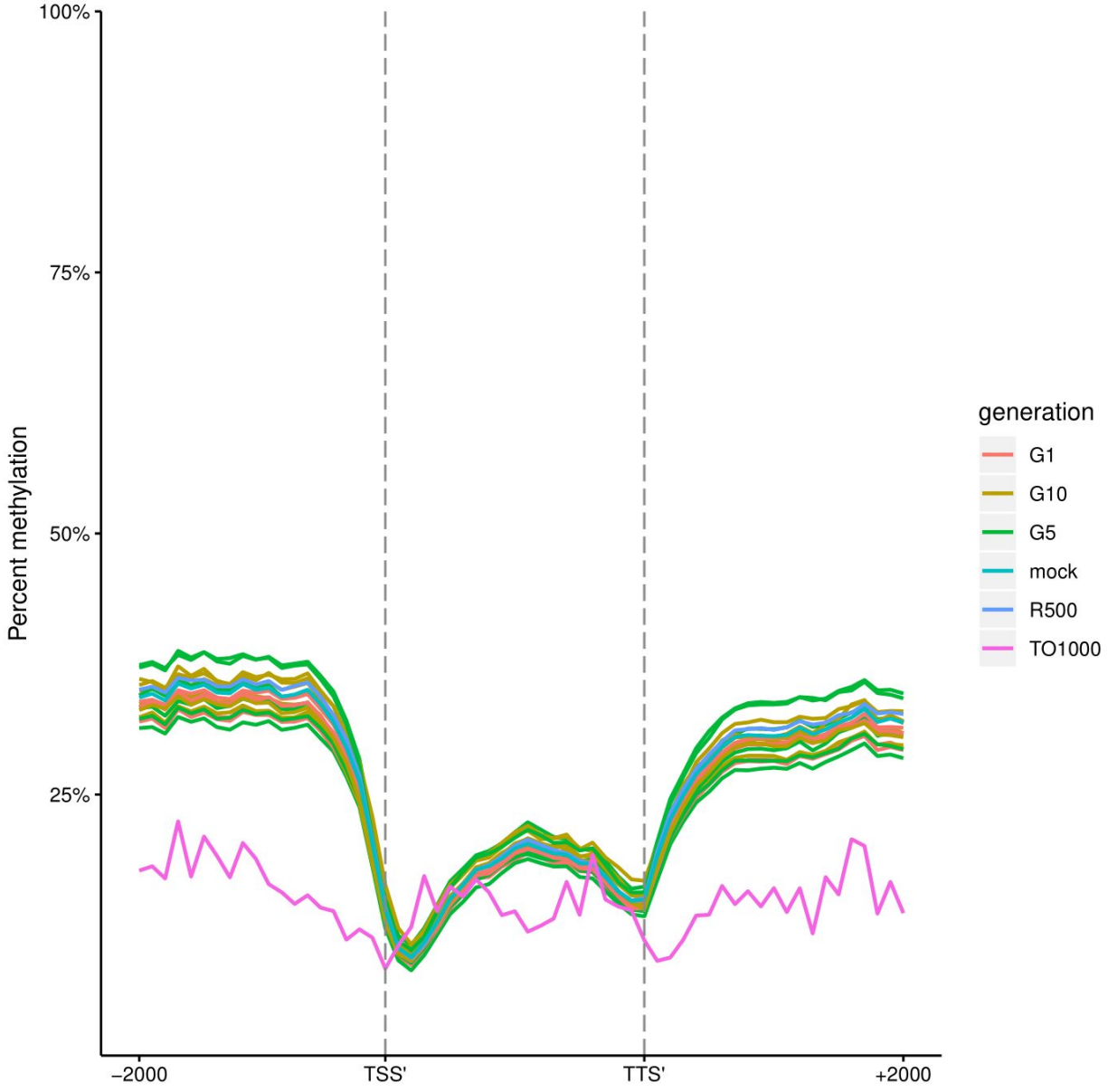
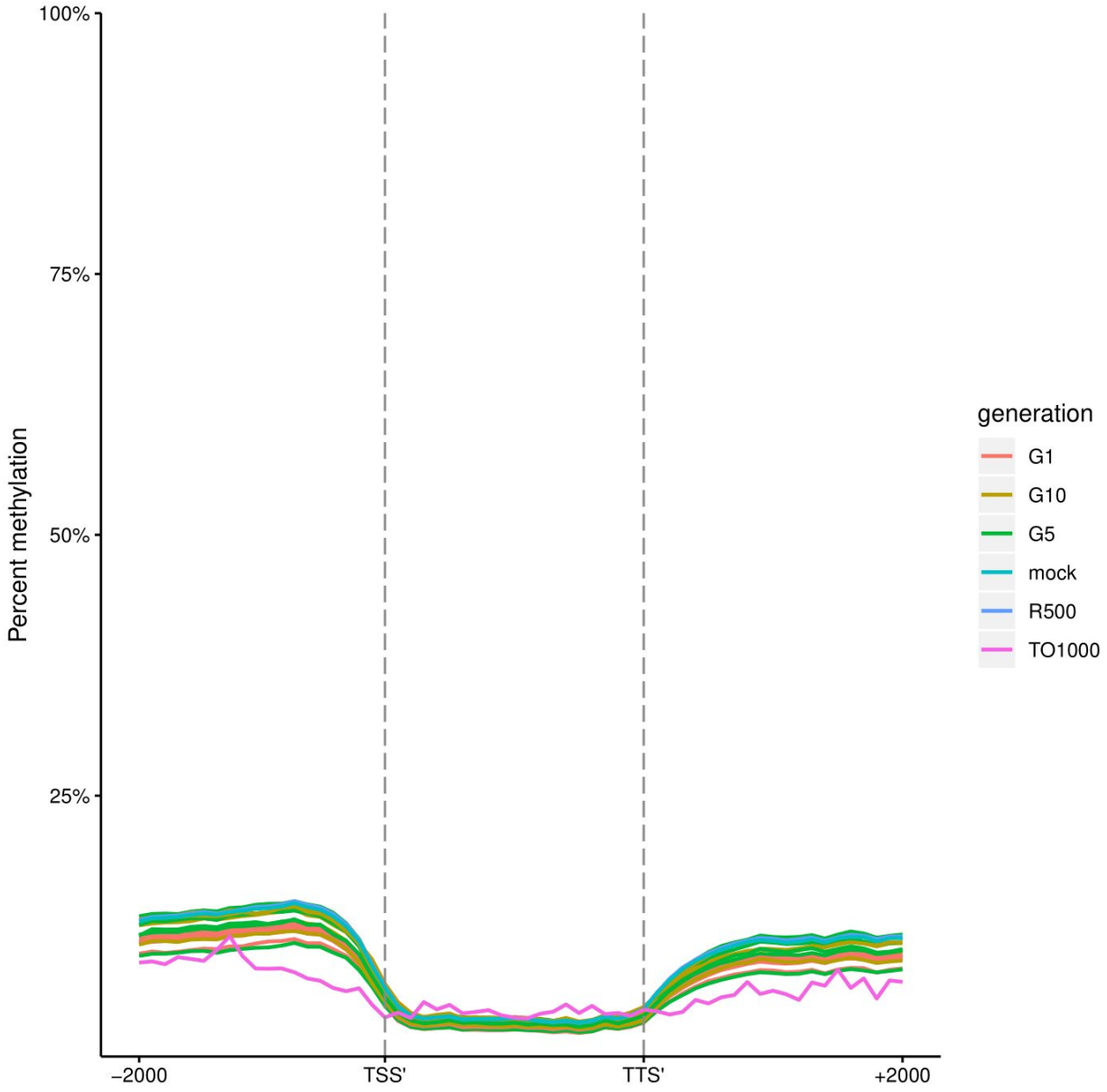

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Fig 30 a-c Metaplots of CG, CHG, and CHH mean weighted methylation of all annotated A subgenome gene models 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

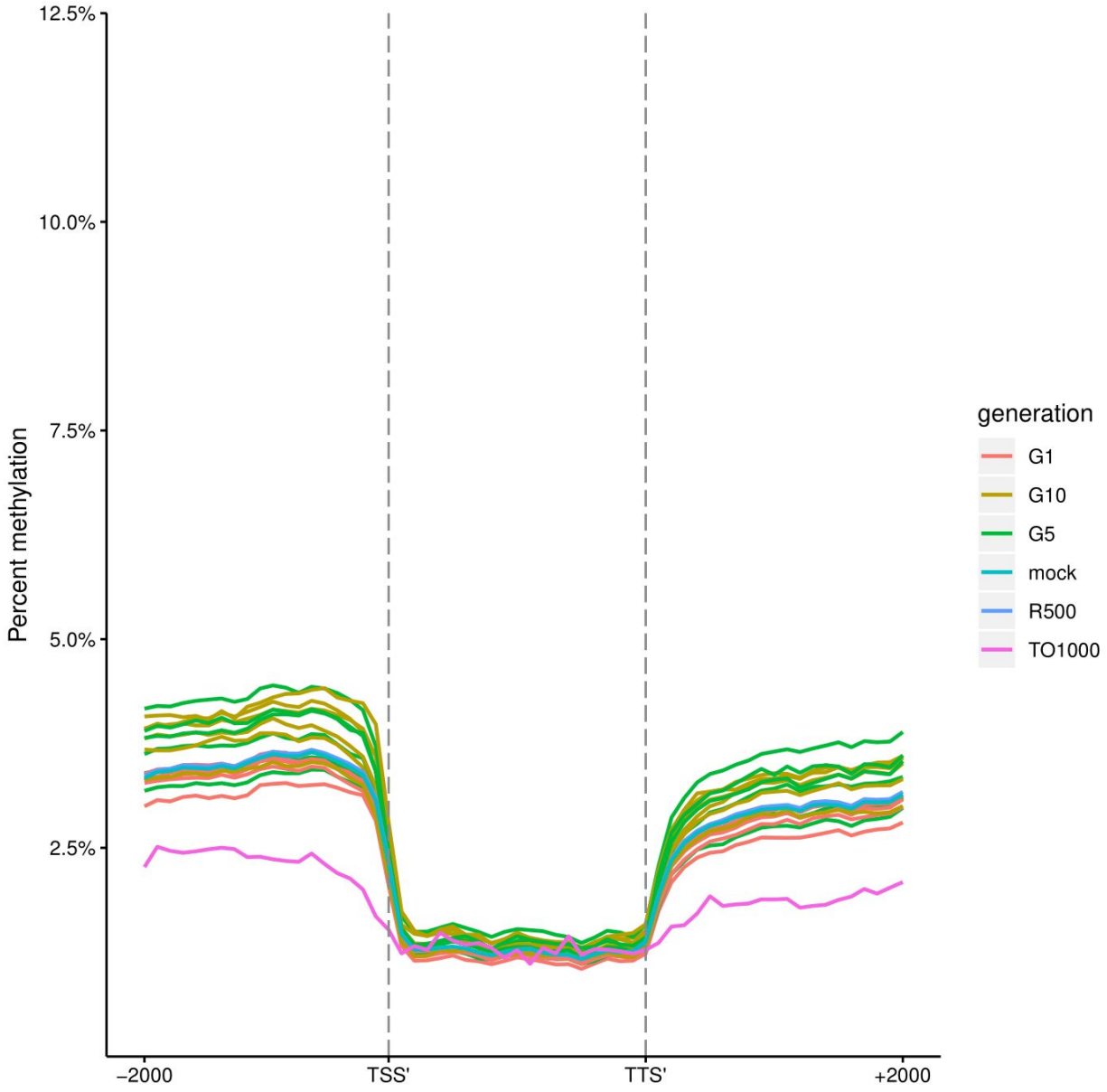
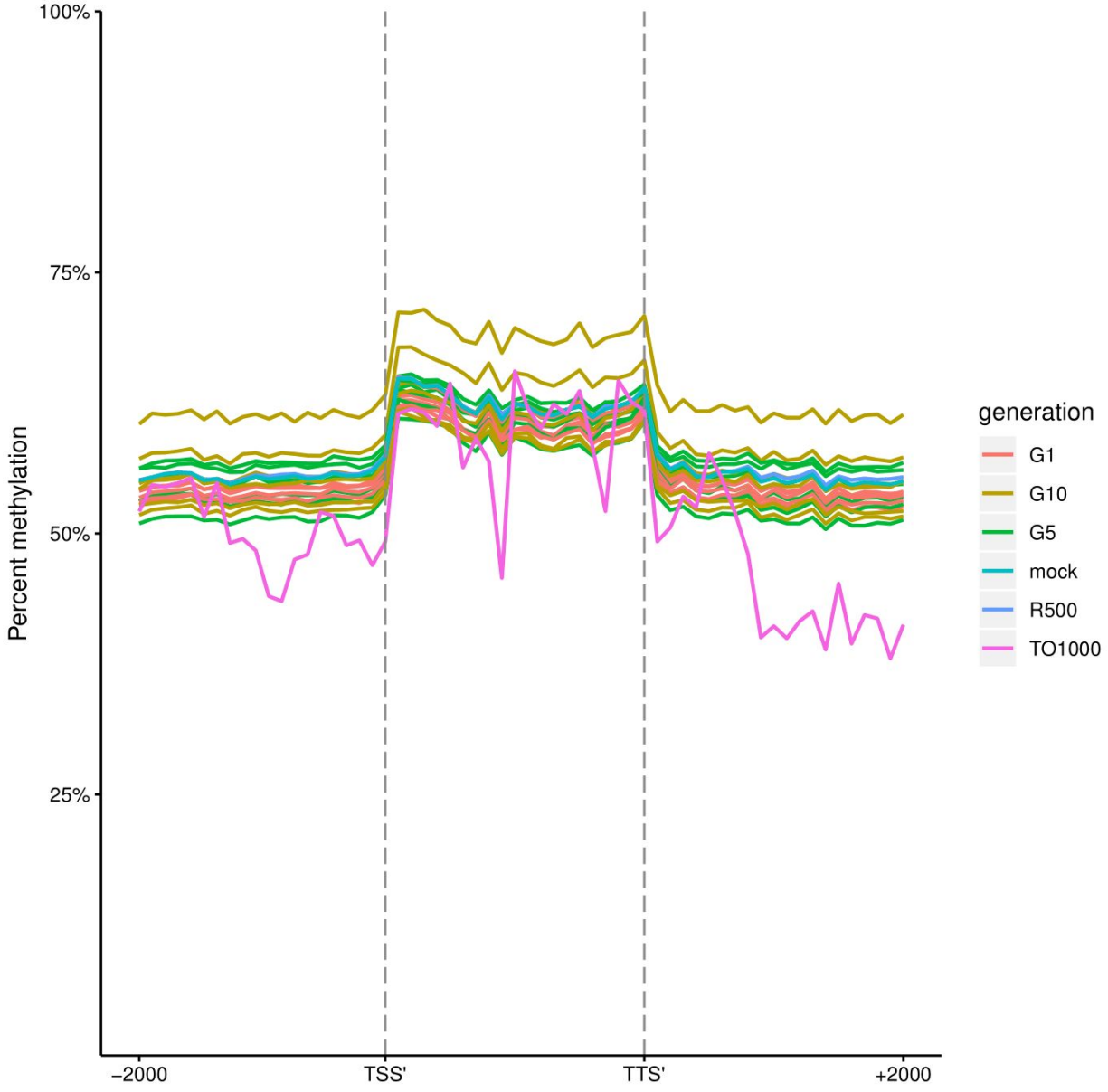
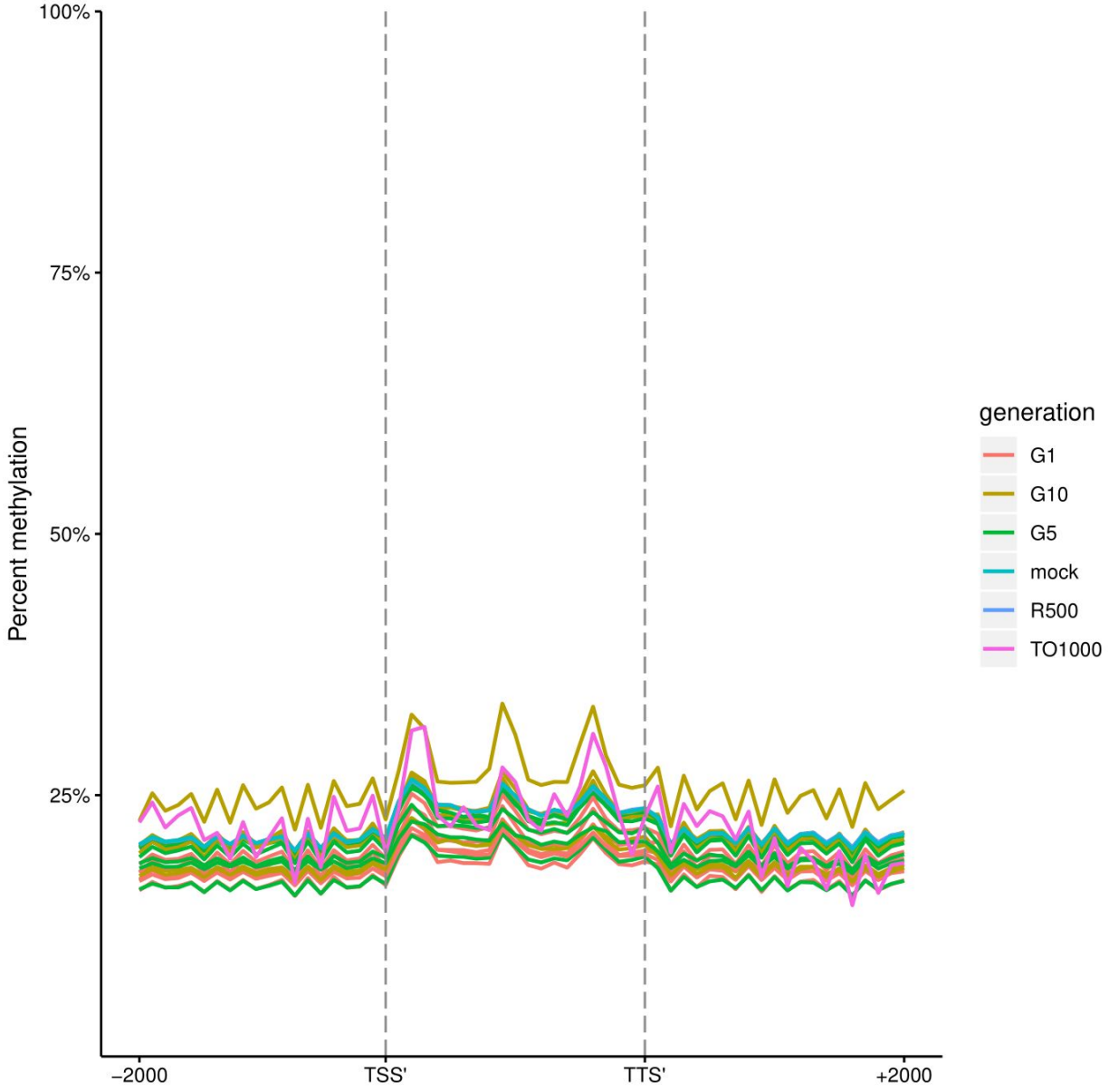

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Fig 31a-cMetaplots of CG, CHG, and CHH mean weighted methylation of A subgenome annotated Long tandem repeat transposable elements 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

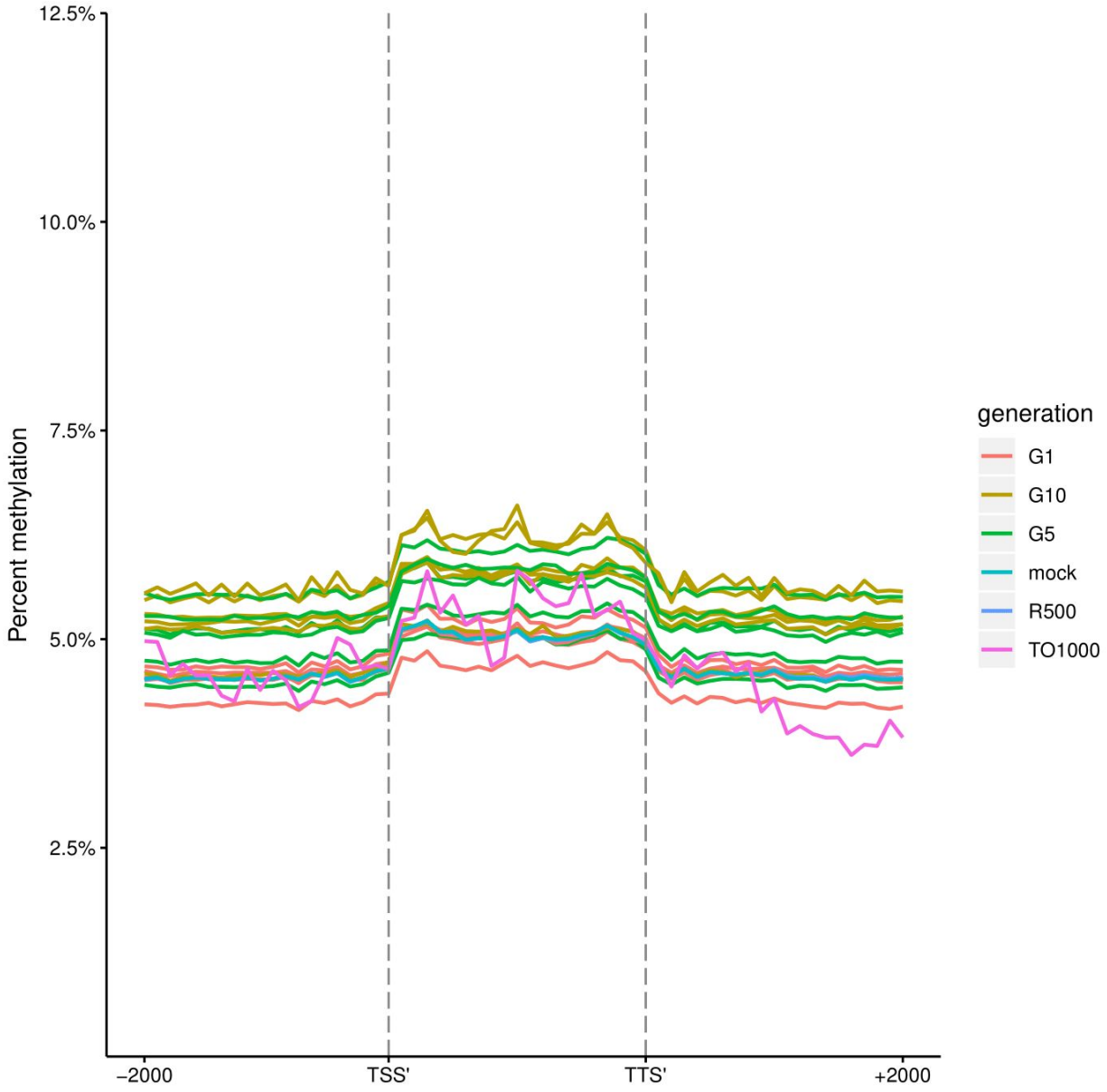
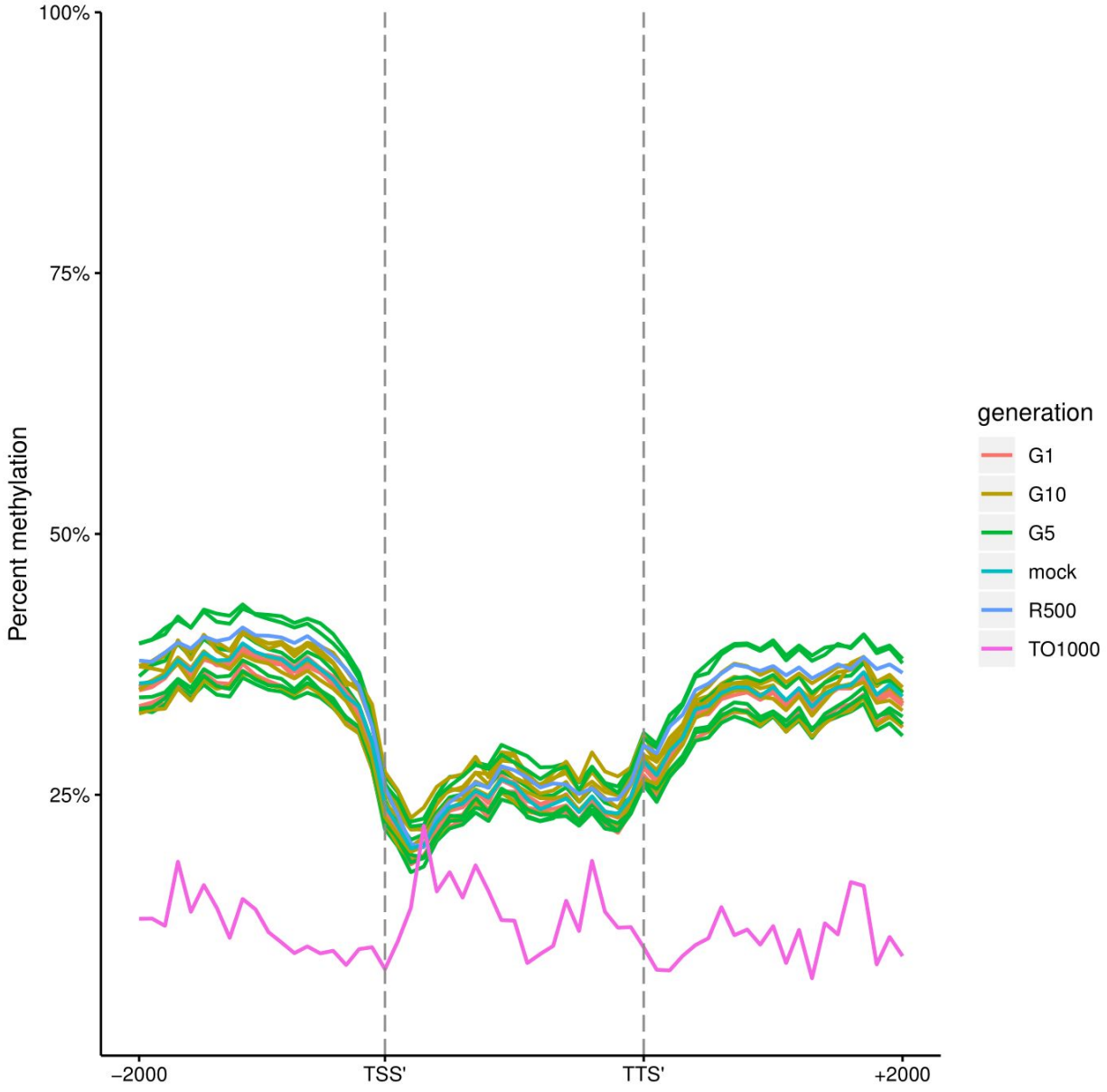
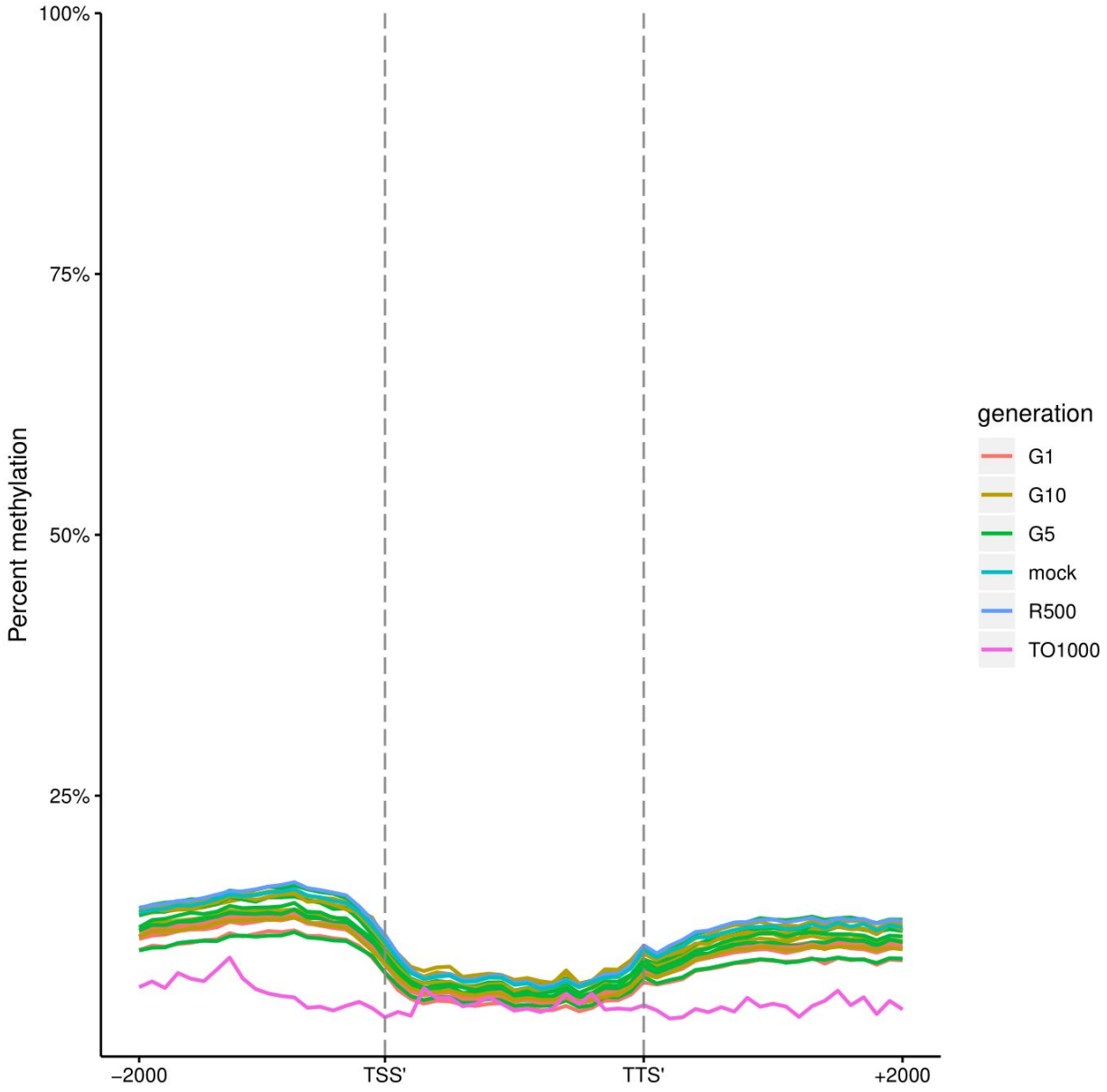

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Fig 32 a-c Metaplots of CG, CHG, and CHH mean weighted methylation of non-syntenic A subgenome gene models 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

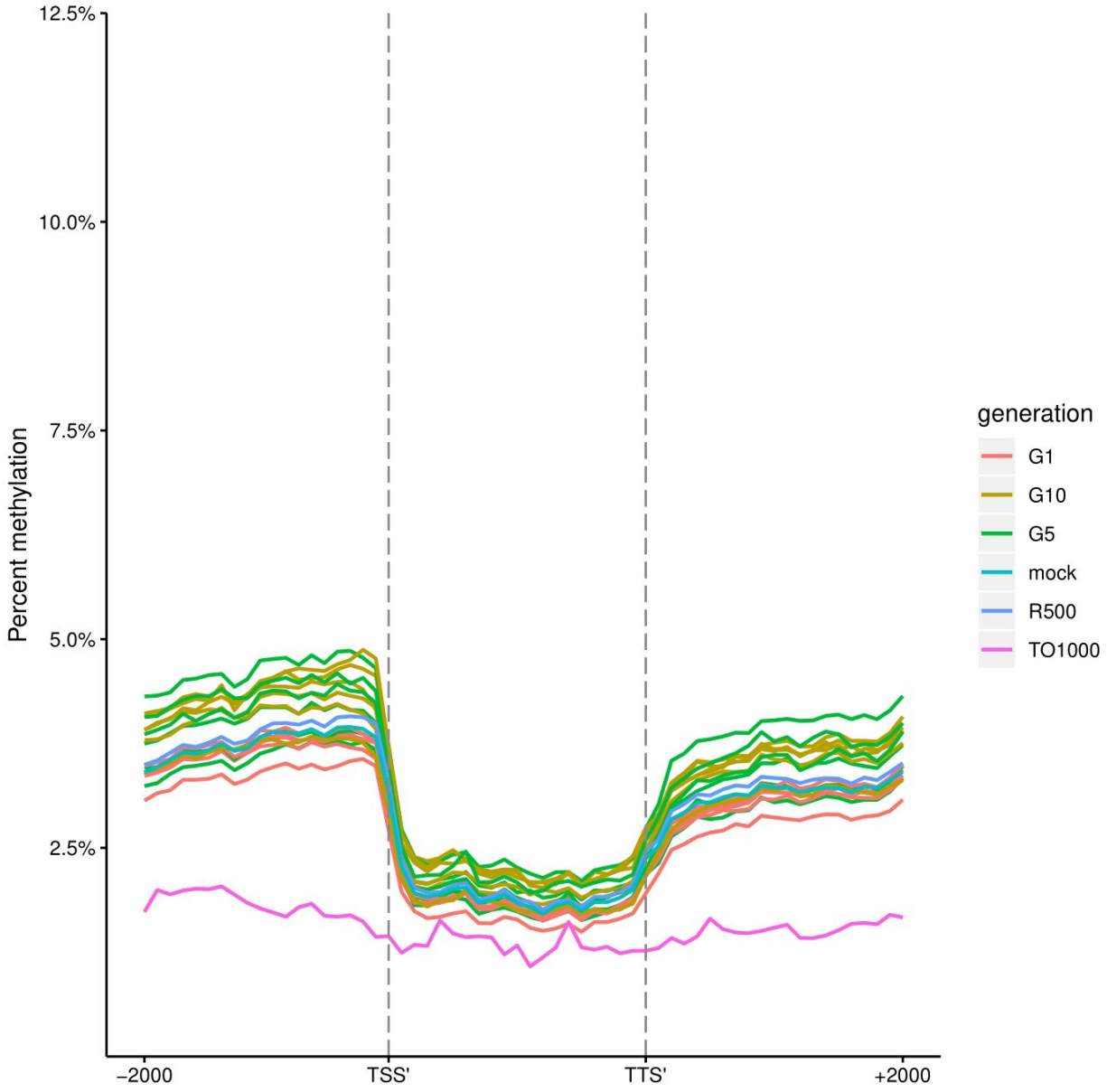
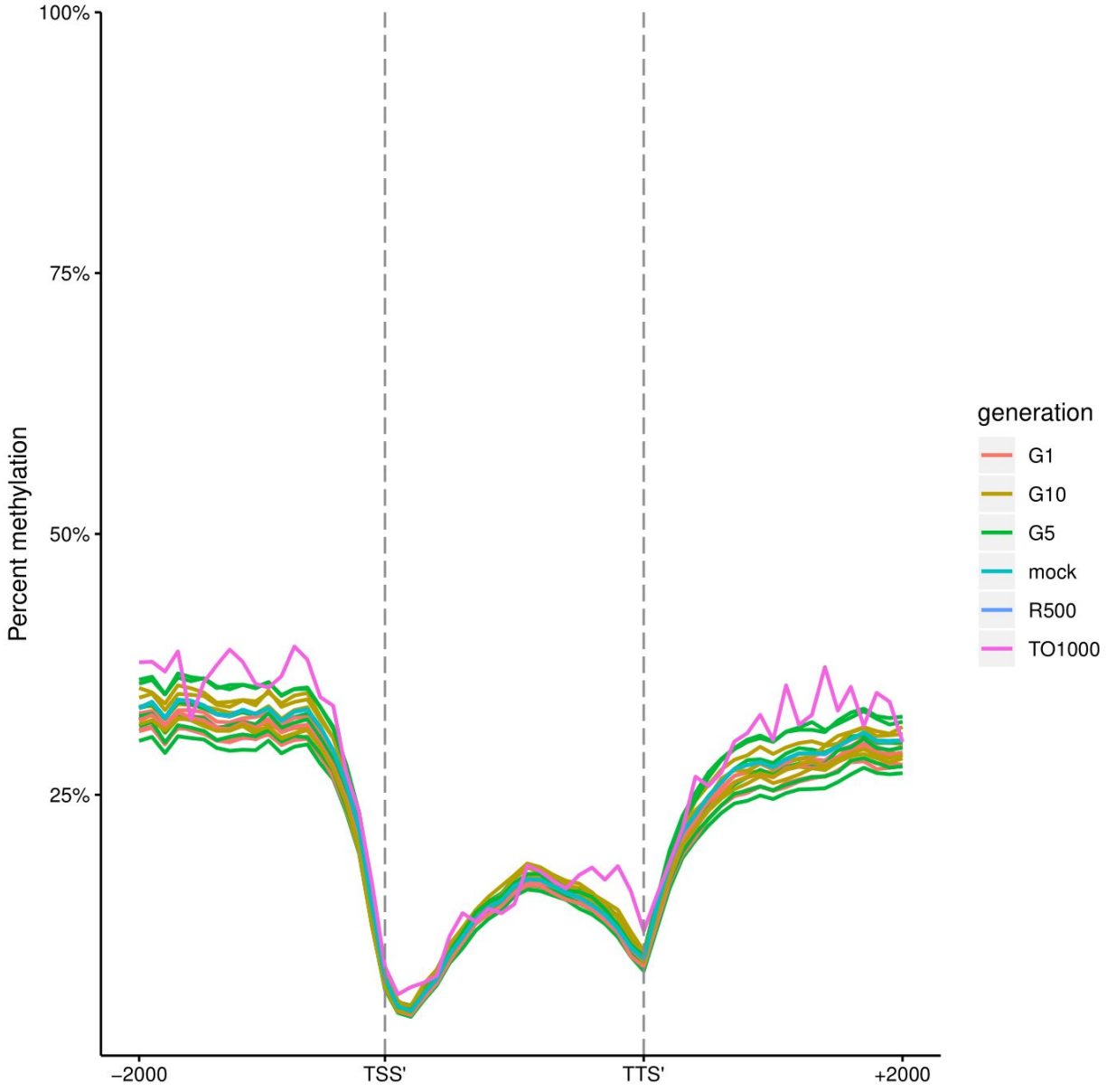
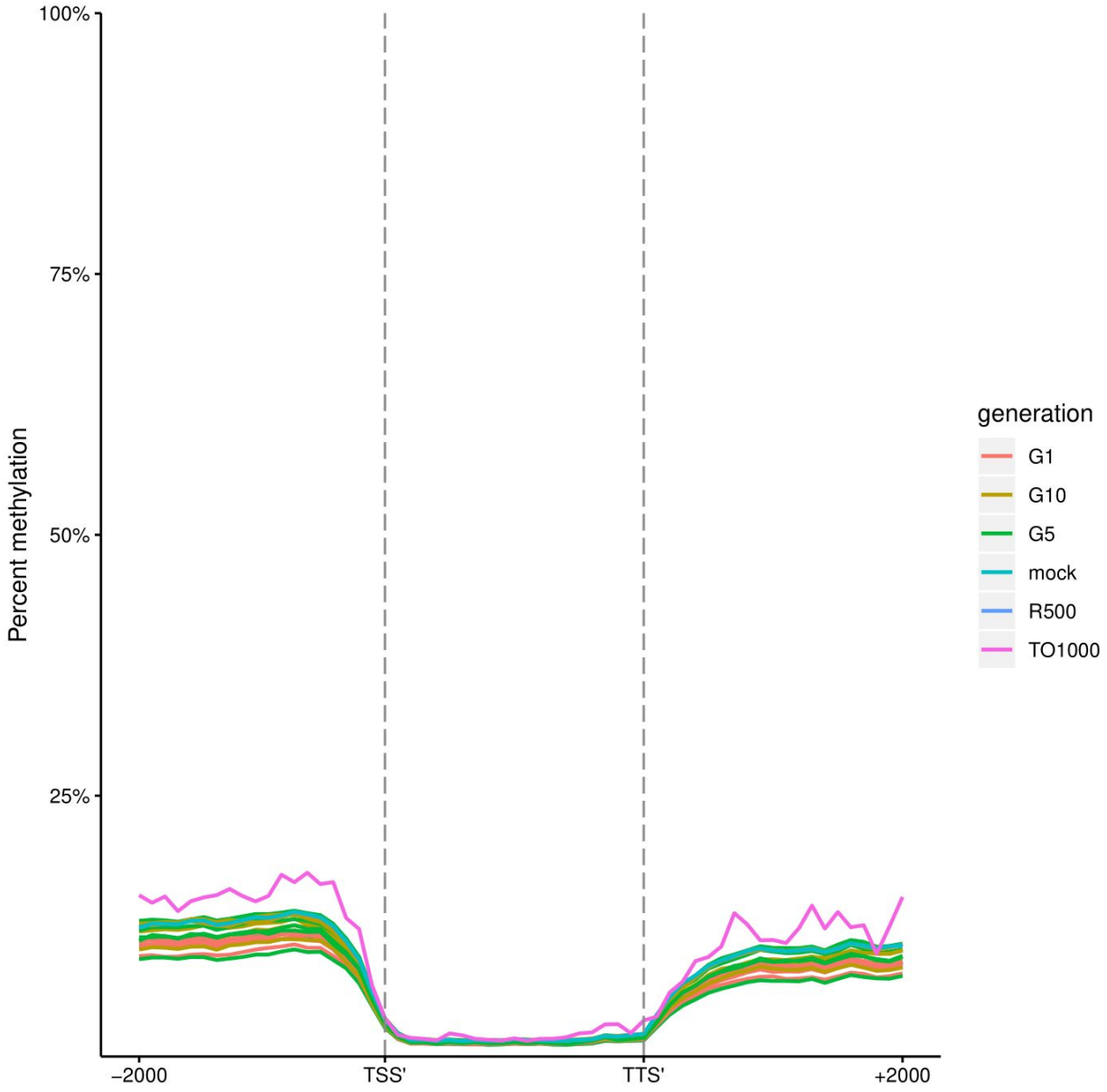

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Fig 33a-c Metaplots of CG, CHG, and CHH mean weighted methylation of syntenic A subgenome gene models 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

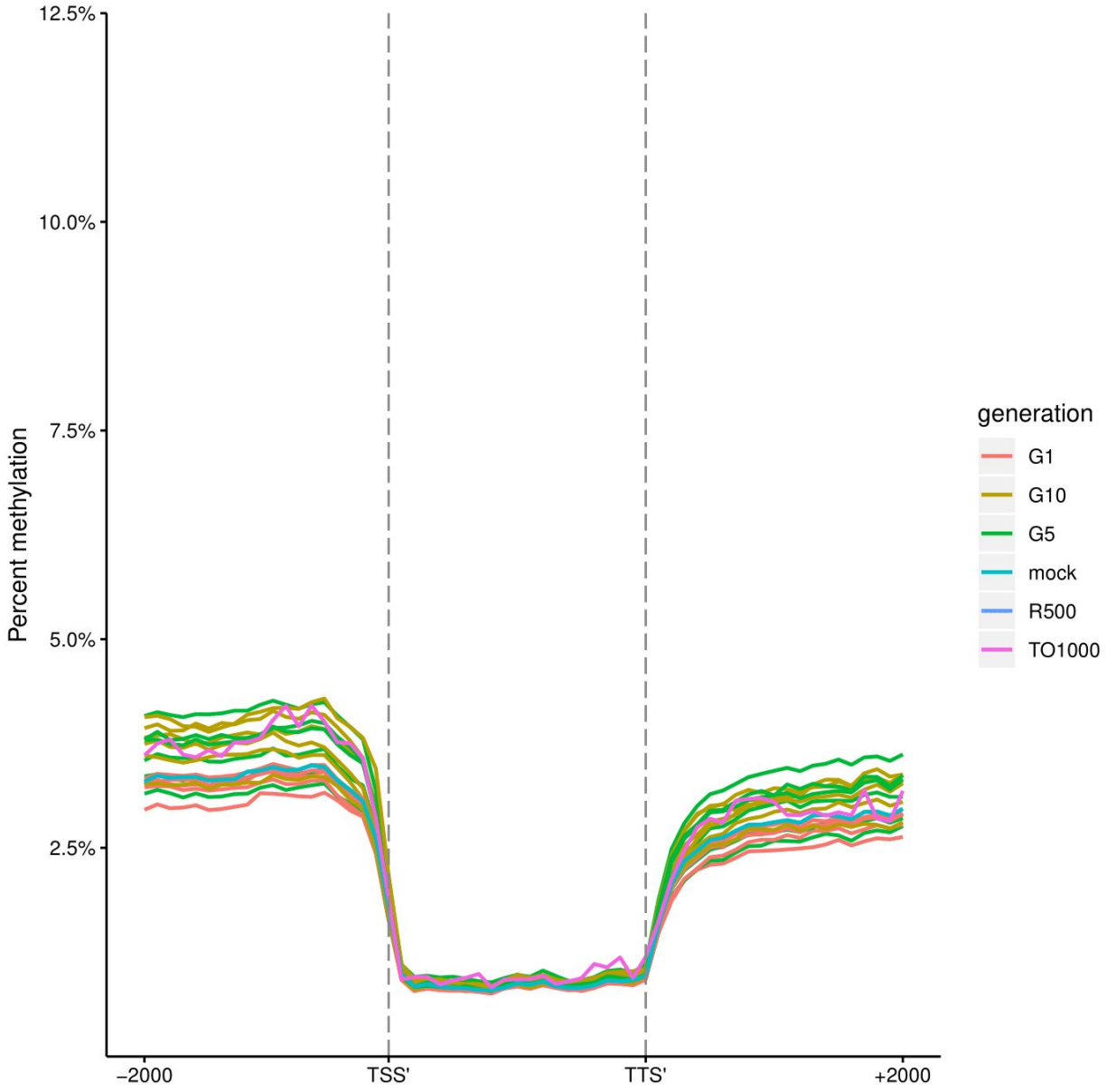
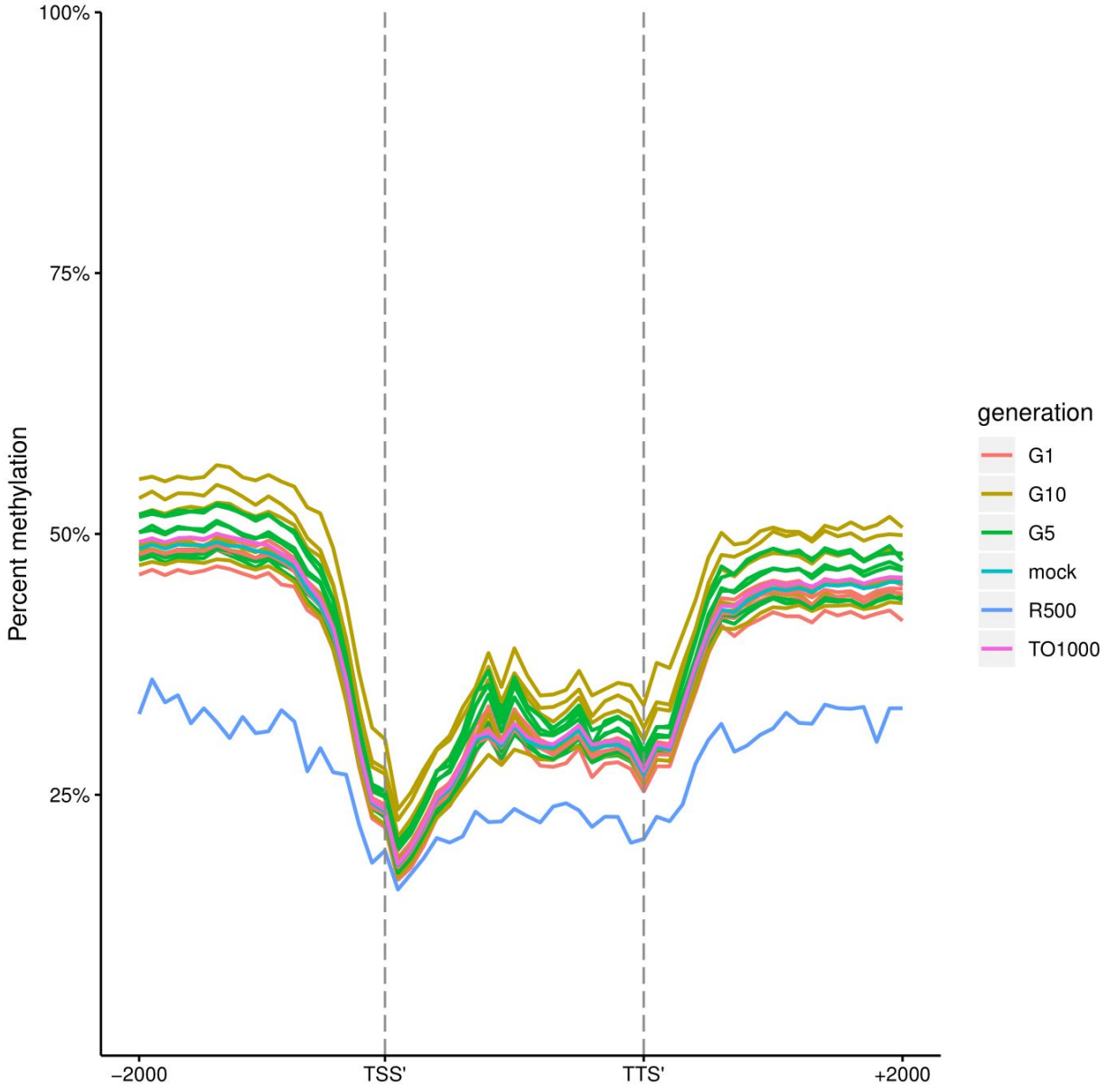
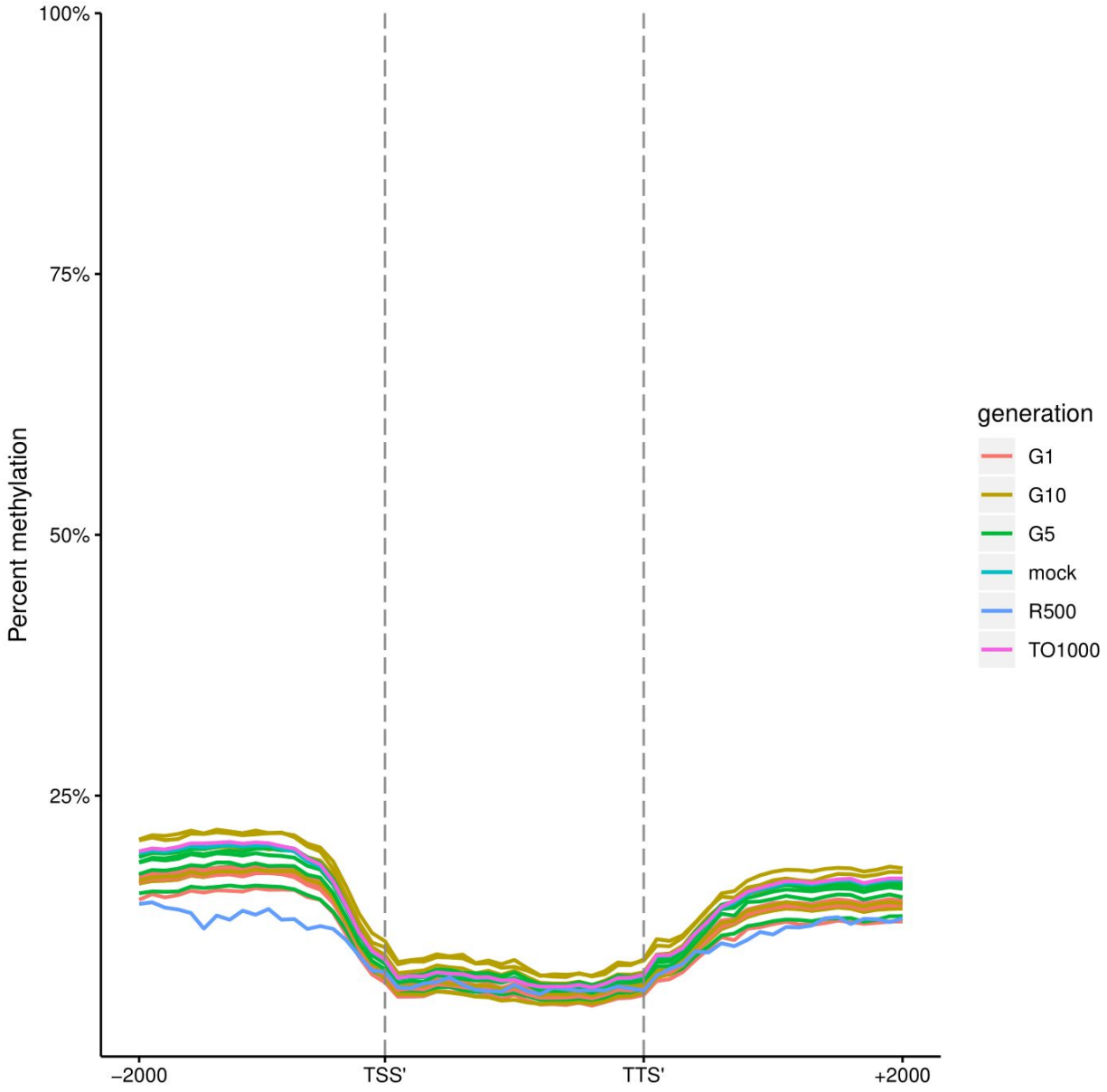

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Fig 34a-c Metaplots of CG, CHG, and CHH mean weighted methylation of all annotated C subgenome gene models 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

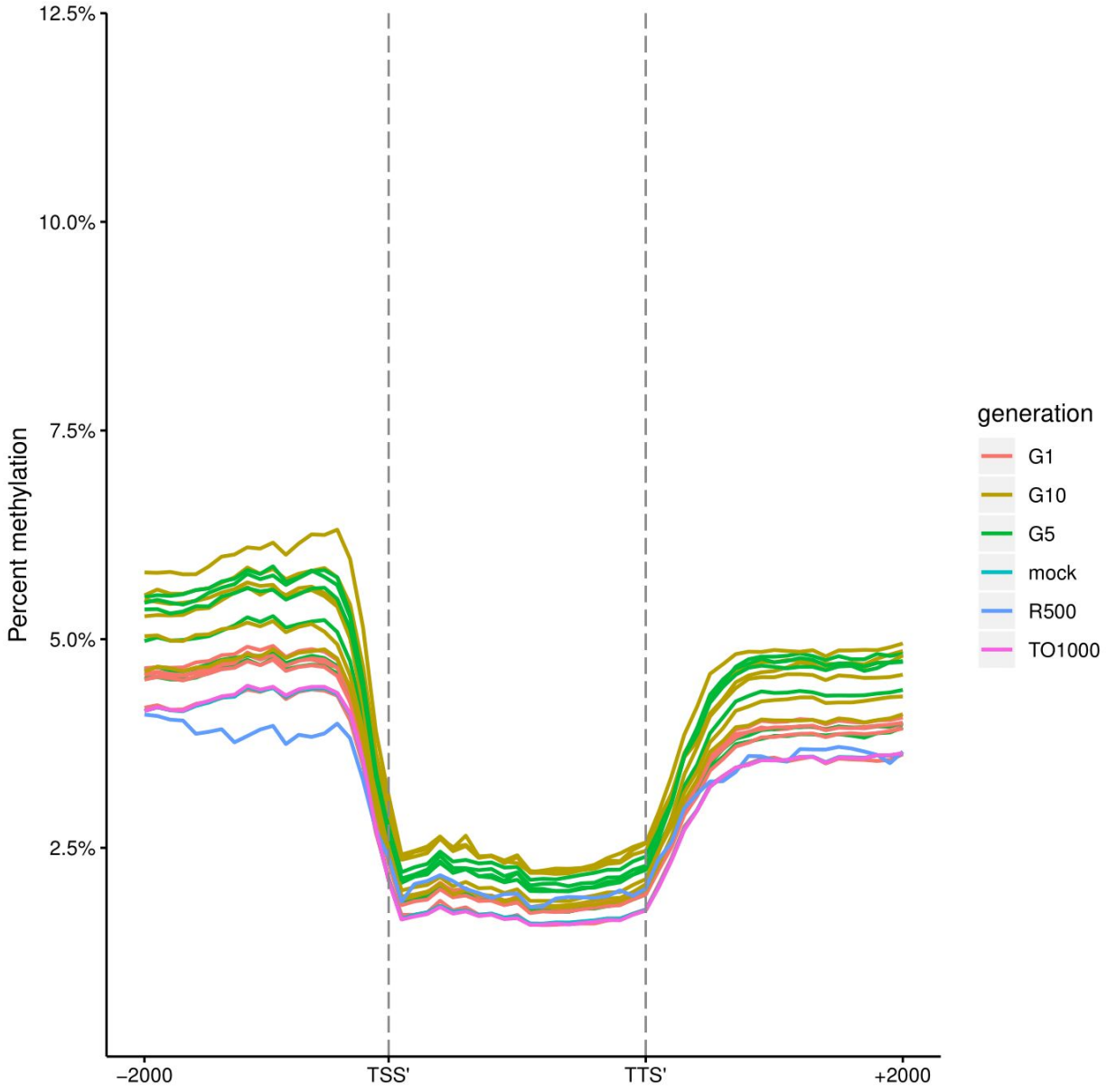
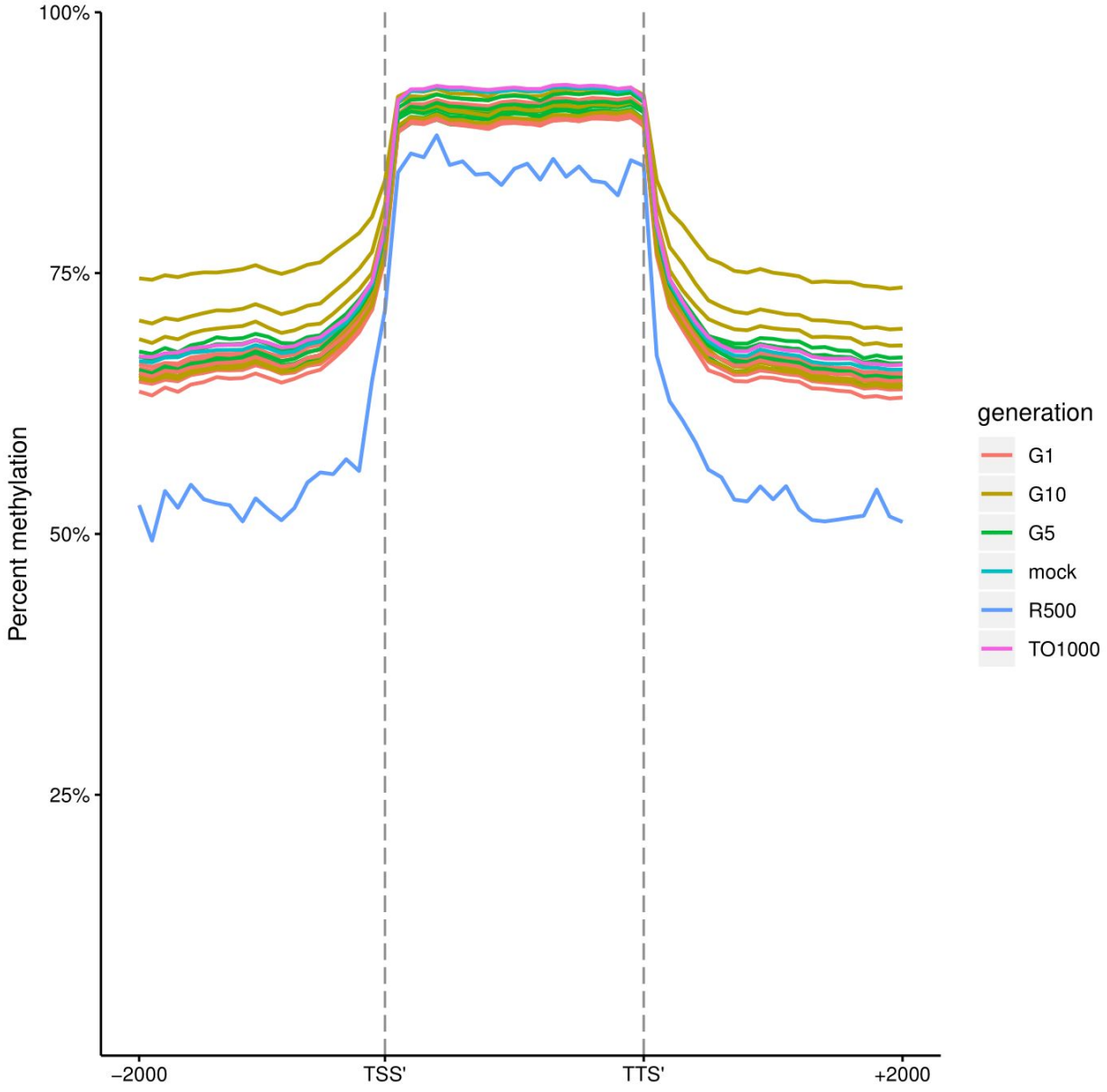
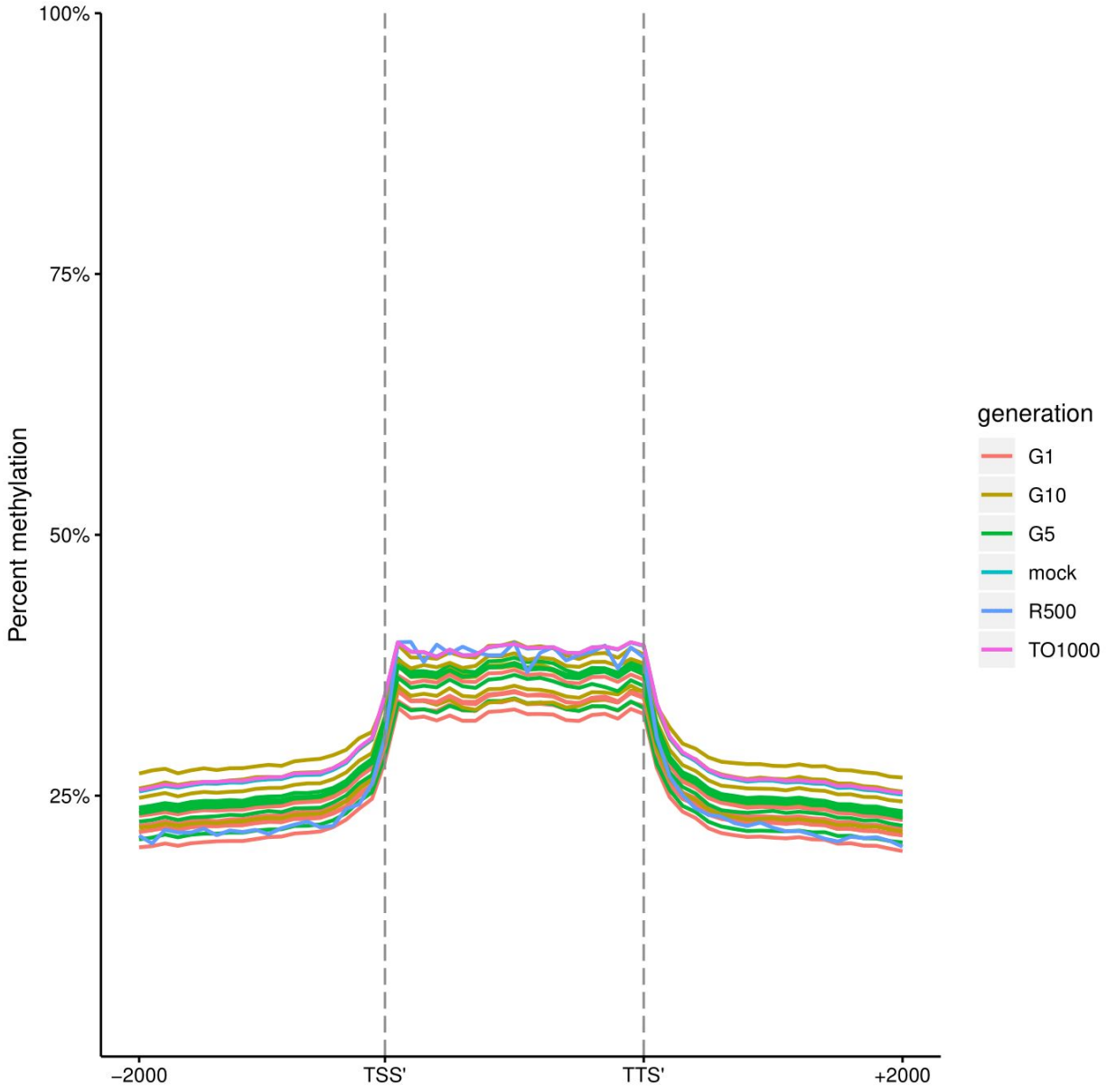

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Fig 35 a-c Metaplots of CG, CHG, and CHH mean weighted methylation of all annotated C Subgenome Long tandem repeat transposable elements 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

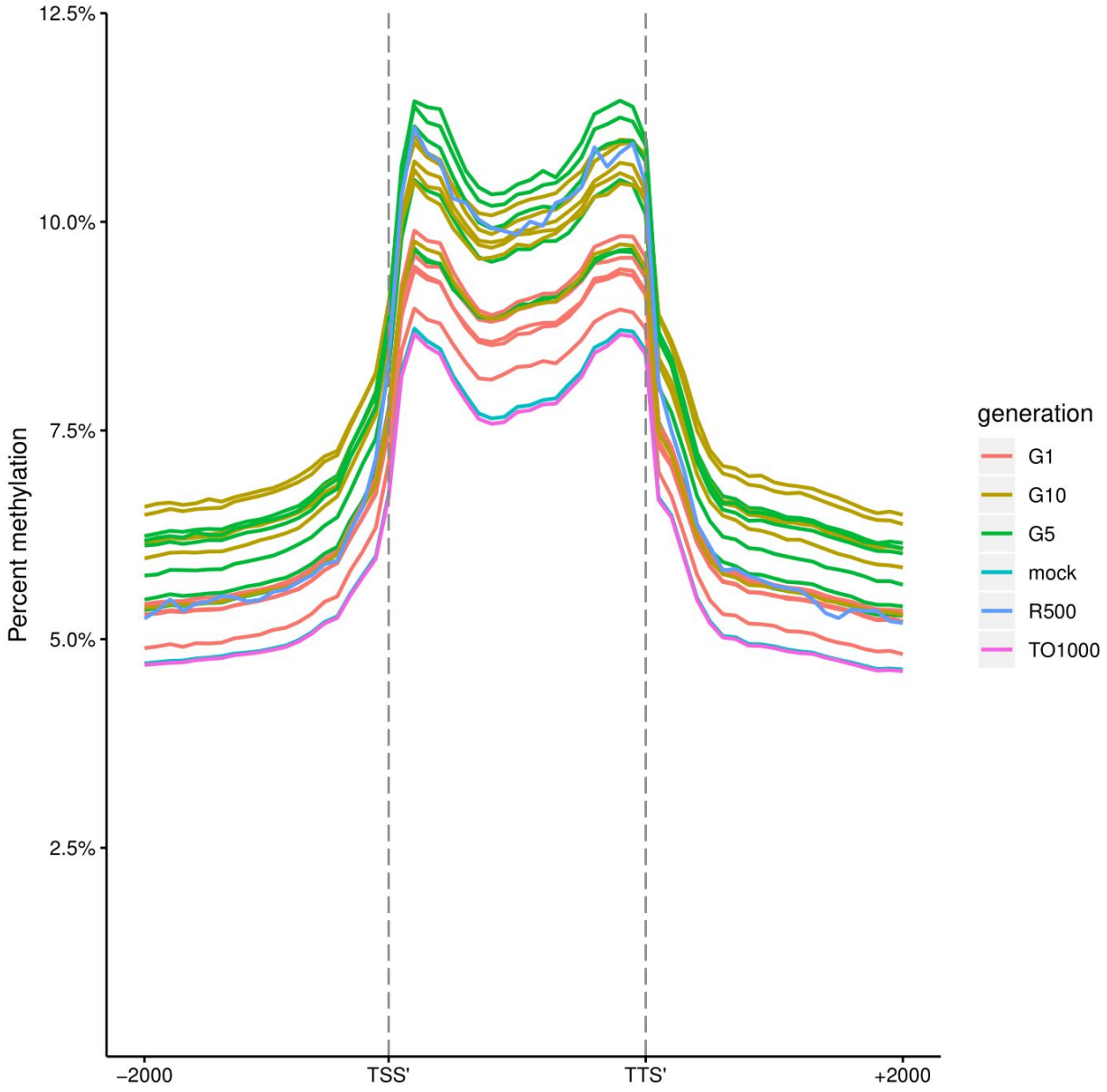
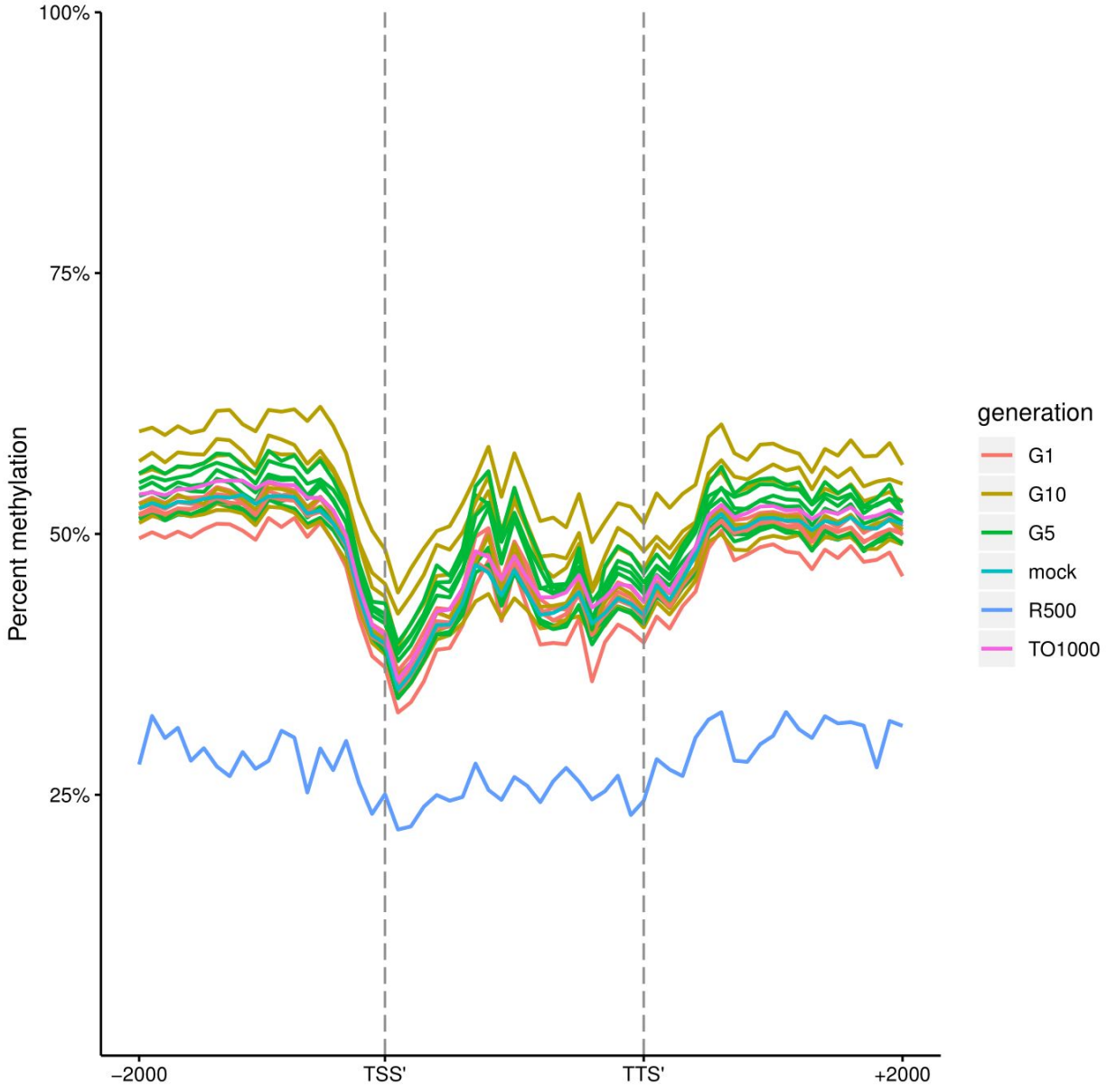
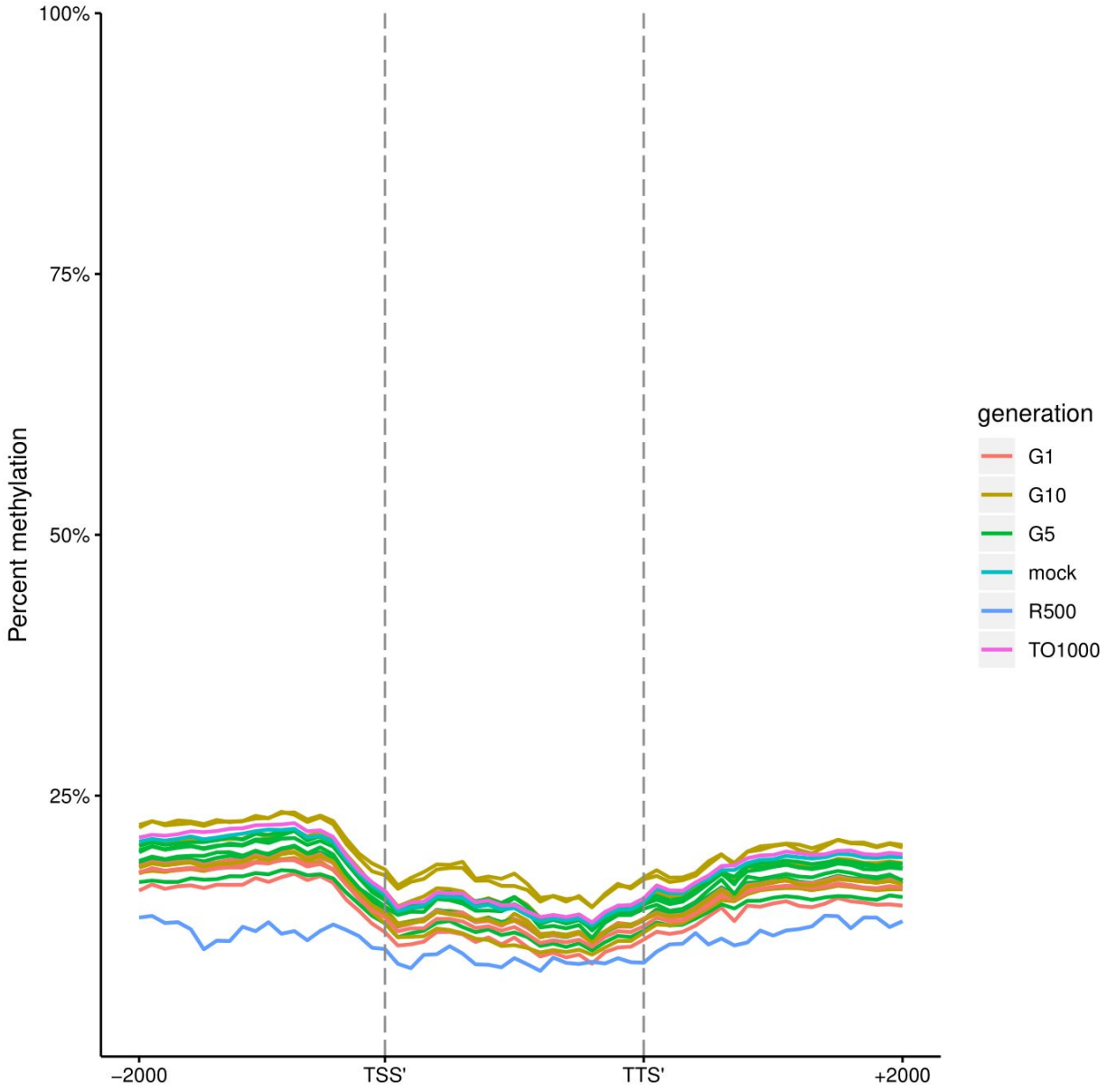

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Fig 36a-c Metaplots of CG, CHG, and CHH mean weighted methylation of non-sytenic C subgenome gene models 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

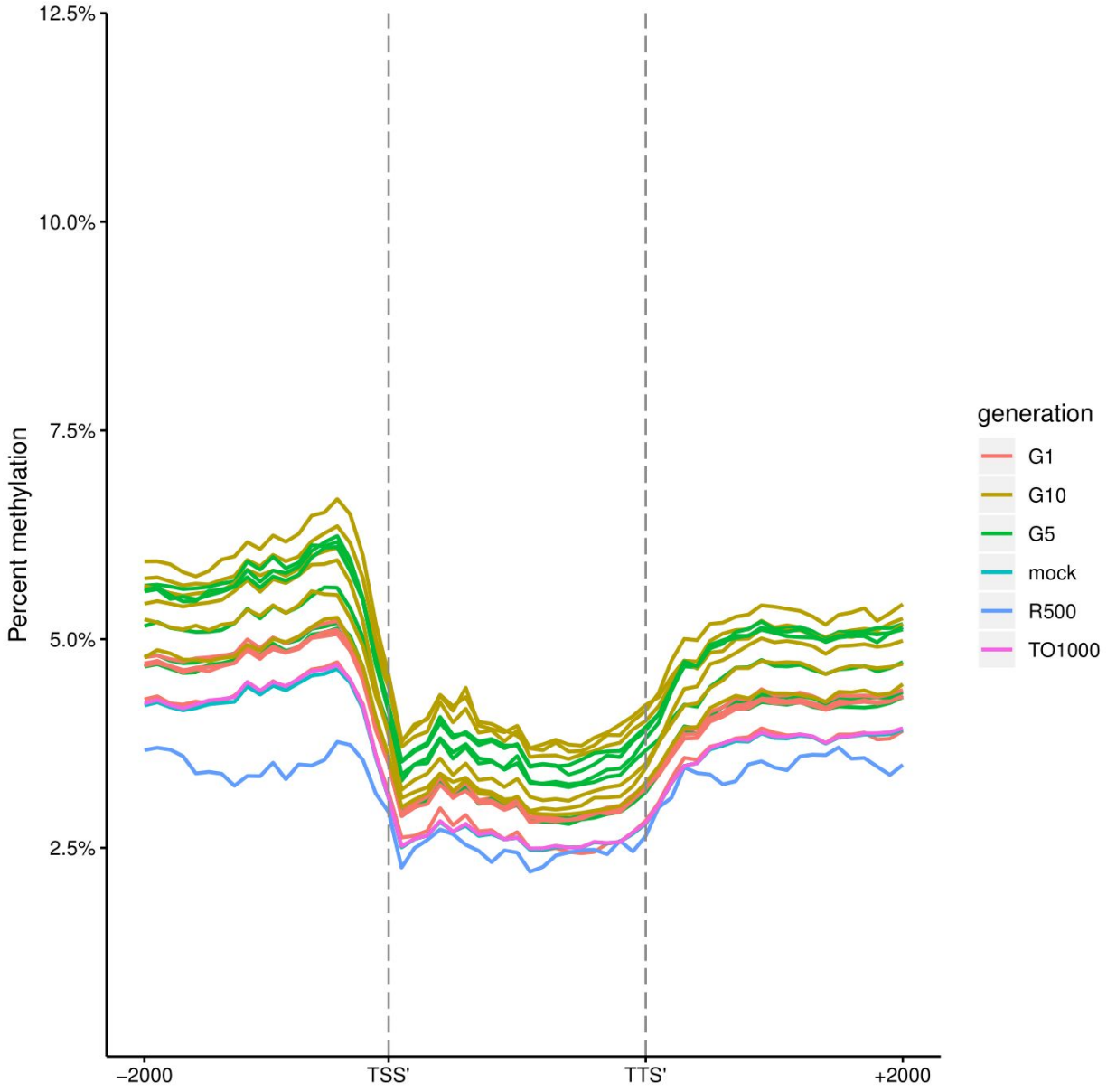
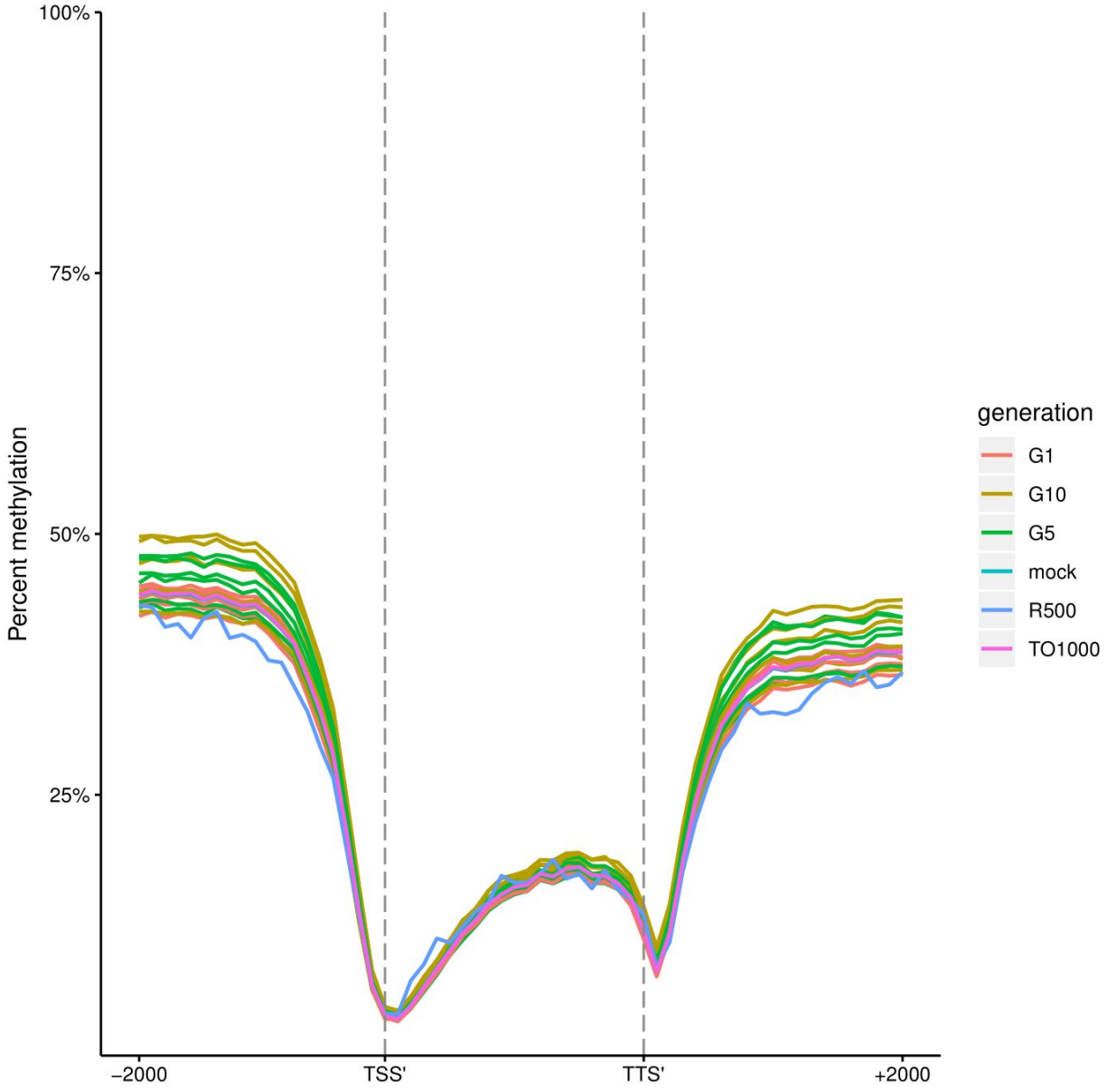
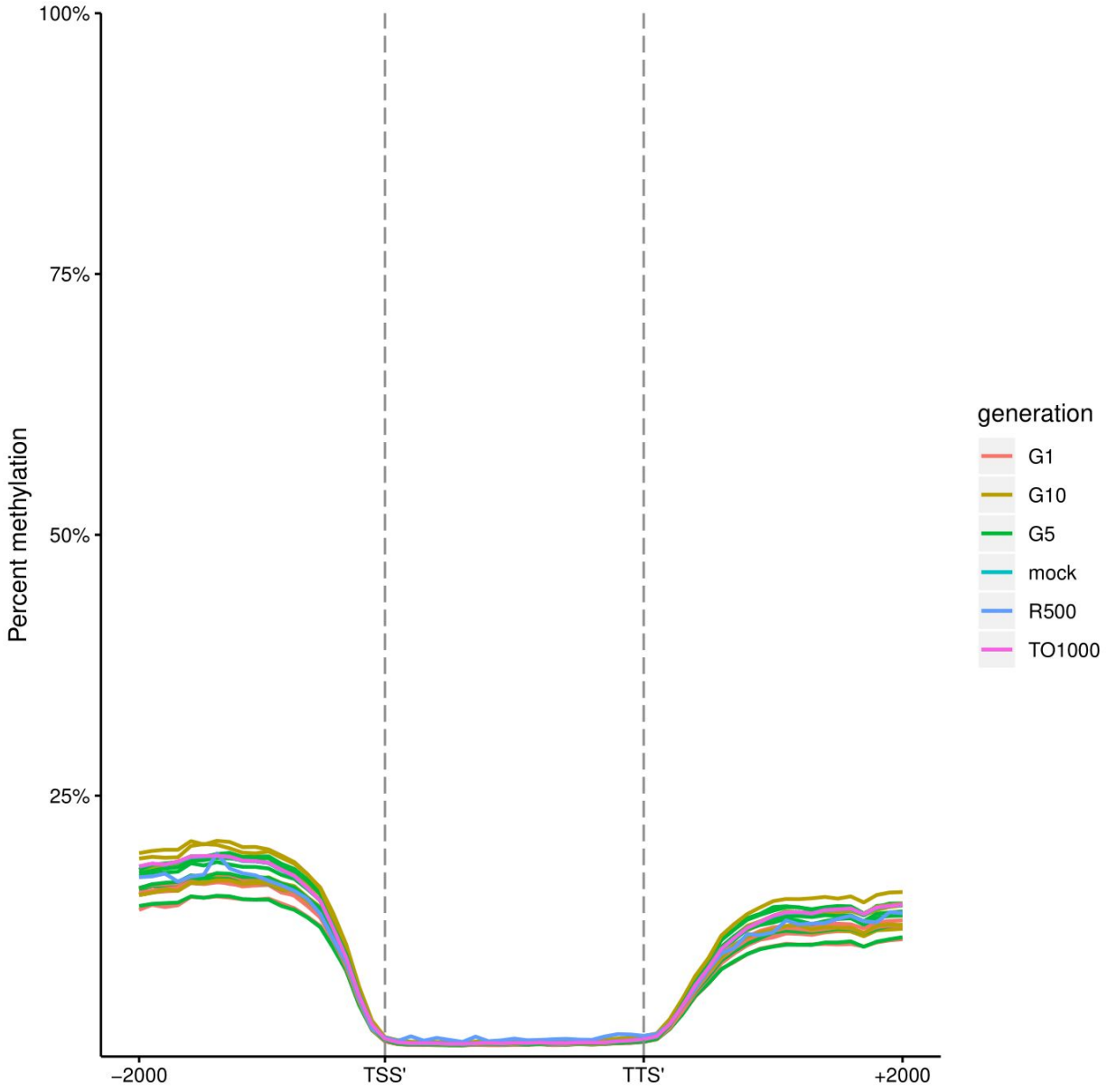

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Fig 37a-c Metaplots of CG, CHG, and CHH mean weighted methylation of syntenic C subgenome gene models 2kb upstream of the transcription start site, gene body, the transcription termination site, and 2kb downstream of the transcription termination site for both parents, an *in silica* “mock” polyploid and the three generations of the resynthesized polyploids for all six lines




100%
75%
50%
25%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

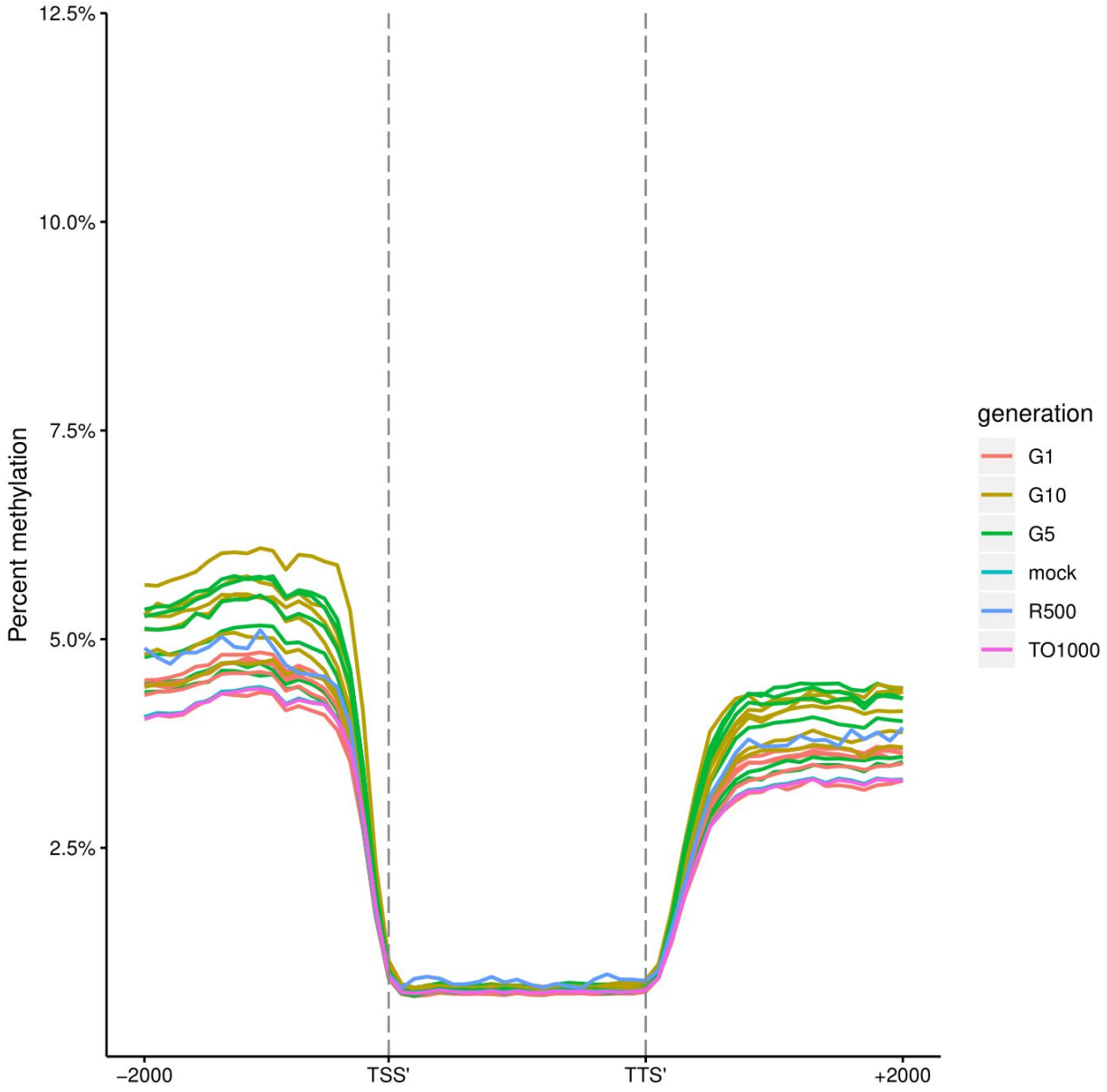

12.5%
10.0%
7.5%
5.0%
2.5%
Percent methylation
−2000
TSS'
TTS'
+2000
generation
G1
G10
G5
mock
R500
TO1000

Table S1:

Phenotype of Resynthesized Lines

Line	flower time-A	flower time-B	Pollen Activity	Pollen Activity-B	Plant Height-A	Plant Height-B
EL100 S1	5-Sep	30-Aug	90.5	93.6	77	87
EL100 S5	3-Sep	29-_Aug	86.5	76.2	79	53
EL100 S10	9-Oct	No	67.2	No	38	no
EL200 S1	10-Sep	31-Aug	93.5	93.4	82	78
EL200 S5	18-Sep	25-Aug	85.5	88	140	110
EL200 S10	25-Sep	7-Sep	76.1	85.6	68	99
EL300 S1	10-Sep	28-Aug	94.2	86.7	82	74
EL300 S5	9-Sep	13-Sep	78.2		90	104
EL300 S10	22-Sep	10-Sep	61.6		121	108
EL400 S1	8-Sep	2-Sep	93.7	91.9	82.5	78
EL400 S5	30-Aug	28-Aug	95.8	92.4	86	111
EL400 S10	28-Aug	16-Sep	89.9		79	103
EL600 S1	12-Sep	7-Sep	96	96.3	100	99
EL600 S5	16-Sep	31-Aug	87.9	86.7	80	89
EL600 S10	29-Aug	25-Aug	86.9	85.6	77	72
EL1100 S1	10-Sep	29-Aug	93.1		78	65
EL1100 S5	15-Sep	3-Sep	91.9	97.5	110	75
EL1100 S10	2-Sep	29-Aug	88.7	92.2	62	76

Table S2

Homoeologous Exchange Chi Squared table

Line	BnA. Observed	BnA. Expected	BnC. Observed	BnC. Expected	Chi.Squared	P.value
100S1	3455	1967	479	1967	2251.29	0
100S5	3266	3939.5	4613	3939.5	230.28	5.17e-52
100S10	5091	4507.5	3924	4507.5	151.07	1.01e-34
200S1	1696	1644	1592	1644	3.29	0.07
200S5	2975	3425	3875	3425	118.25	1.53e-27
200S10	2606	3762	4918	3762	710.44	1.61e-156
300S1	1033	1166	1299	1166	30.34	3.62e-08
300S5	2881	3508.5	4136	3508.5	224.46	9.64e-51
300S10	4487	5115.5	5744	5115.5	154.44	1.86e-35
400S1	684	458.5	233	458.5	221.81	3.64e-50
400S5	495	1788	3081	1788	1870.08	0
400S10	5179	3980	2781	3980	722.41	4.00e-159
600S1	0	57	114	57	114	1.30e-26
600S5	1395	2524	3653	2524	1010.02	1.19e-221
600S10	5320	4288.5	3257	4288.5	496.21	6.36e-110
1100S1	2662	2034.5	1407	2034.5	387.08	3.58e-86
1100S5	4132	3185.5	2239	3185.5	562.46	2.45e-124
1100S10	5636	4207.0	2778	4207.0	970.78	4.03e-213

Table S3:

Homeolog Expression Bias Chi Squared table

Line	BnC. Observed	BnC. Expected	BnA. Observed	BnA. Expected	Chi.Squared	P.value
100S1	3051	2182	1313	2182	692.17	1.51e-152
100S5	2411	1746.5	1082	1746.5	505.65	5.60e-112
100S10	2407	1666.5	926	1666.5	658.07	3.92e-145
200S1	3138	2232.5	1327	2232.5	734.54	9.23e-162
200S5	2562	1903	1244	1903	456.42	2.90e-101
200S10	2302	1811	1320	1811	266.24	7.49e-60
300S1	3085	2281.5	1478	2281.5	565.95	4.26e-125
300S10	2355	1758.5	1162	1758.5	404.68	5.28e-90
400S1	3342	2444.5	1547	2444.5	659.04	2.42e-145
400S5	2873	2184	1495	2184	434.73	1.52e-96
600S1	3501	2589.5	1678	2589.5	641.69	1.43e-141
600S5	2440	2022.5	1605	2022.5	172.37	2.25e-39
600S10	2746	1986.5	1227	1986.5	580.76	2.56e-128
1100S1	3060	2235	1410	2235	609.06	1.79e-134
1100S5	2683	1918.5	1154	1918.5	609.29	1.60e-134
1100S10	2955	2014	1073	2014	879.33	3.06e-193
Parent	3688	2742	1796	2742	652.75	5.65e-144

Table S4

Homeolog Expression Bias vs Parent Chi Squared table

Line	BnC Shared Observed	BnC Shared Expected	BnA Shared Observed	BnA Shared Expected	Chi.Squared	P.value
100S1	2665	853.4	1139	171.1	14789.48	0
100S5	2137	701.4	911	140.5	11593.29	0
100S10	1896	584.4	731	106.9	10259.22	0
200S1	2644	826.6	1079	160.9	14500.84	0
200S5	2194	665	1062	170.5	12914.21	0
200S10	1920	553.7	1011	171.1	11597.63	0
300S1	2777	849.7	1271	193.5	16379.92	0
300S10	1959	667.4	846	147.1	9617.91	0
400S1	2856	866	1256	188.9	16500.45	0
400S5	2502	763.6	1220	200	14507.02	0
600S1	2924	877.9	1300	203.3	16679.8	0
600S5	2000	553.5	1106	208	11856.82	0

600S10	2073	697.3	805	135.2	9719.97	0
1100S1	2579	821.1	1123	178.3	13949.4	0
1100S5	2307	760.6	948	145.4	12151.01	0
1100S10	2162	762.7	777	127.0	9648.66	0