



HAL
open science

DNA methylation around transcription start sites is not globally associated with transcription in the grain of natural and synthetic hexaploid wheat

Meriem Banouh, Mamadou Dia Sow, Caroline Pont, Michael A Seidel, Jérôme Salse, Peter Civaň

► To cite this version:

Meriem Banouh, Mamadou Dia Sow, Caroline Pont, Michael A Seidel, Jérôme Salse, et al.. DNA methylation around transcription start sites is not globally associated with transcription in the grain of natural and synthetic hexaploid wheat. *BMC Plant Biology*, 2026, 26, <10.1186/s12870-026-08607-6>. <hal-05613681>

HAL Id: hal-05613681

<https://hal.inrae.fr/hal-05613681v1>

Submitted on 6 May 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

RESEARCH

Open Access



DNA methylation around transcription start sites is not globally associated with transcription in the grain of natural and synthetic hexaploid wheat

Meriem Banouh¹, Mamadou Dia Sow¹, Caroline Pont¹, Michael A. Seidel², Jérôme Salse¹ and Peter Cíváň^{1*}

Abstract

Background Epigenetic mechanisms including DNA methylation are assumed to play crucial roles in the maintenance of genome integrity, regulation of gene expression and development, and their increasing exploitation in breeding applications is anticipated. However, the relationship between DNA methylation and gene expression remains ambiguous and difficult to generalize. Here we explored the hypothesized causality between the level of transcription and cytosine methylation at the 5' end of genes (around transcription start sites and start codons) in relation to whole-genome duplication in natural and synthetic allohexaploid wheat (*Triticum/Aegilops* complex).

Results Using transcriptomes and a sequence capture protocol coupled with bisulfite sequencing, we observed sometimes significant, but overall very weak associations between gene expression and 5' end methylation on a genome-wide scale. In synthetic wheat allohexaploids, global methylation differences between subgenomes are not triggered by the polyploidization, as the subgenome patterns are rather faithfully inherited from parents. A small number of genes differentially methylated between the parents and synthetics was consistently recovered in reciprocal synthetics and subsequent generations. Differences in transcription between homeologs are not clearly associated with 5' end methylation in either natural or synthetic wheat. Overall, allopolyploidization triggers only minor methylation changes around transcription start sites and start codons of nascent wheat allopolyploids, and these are not statistically associated with differential expression. Although there is a measurable methylation difference between silent and expressed genes in the developing grain, our results do not support the hypothesis that 5' end DNA methylation is engaged in the regulation of gene expression in natural and synthetic wheat.

Conclusions While a 'genome shock' hypothesis predicts extensive transcriptomic and epigenetic reorganization after polyploidization, DNA methylation patterns around transcription start sites are generally undisturbed in nascent wheat allohexaploids. Although this stability might indicate importance for gene regulation, a clear relationship between DNA methylation and transcription was not observed either on a genome-wide scale, or among triads of homeologous genes.

*Correspondence:
Peter Cíváň
peter.civan@inrae.fr

Full list of author information is available at the end of the article



Keywords *Triticum* sp., *Aegilops tauschii*, BSseq, Polyploidization, RNA-seq, Genome shock, DEG, DMR, Correlation

Background

DNA methylation commonly refers to the addition of a methyl group to the C5 position of cytosines in the DNA molecule. This covalent modification widespread in the plant genome and across species is crucial for maintaining genome integrity by silencing transposable elements (TEs) [1], and is thought to be the main component of the epigenetic regulation of gene expression [2, 3]. In plants, methylated cytosines (mCs) occur in any sequence context (i.e. on cytosines followed by any nucleotide), and have been shown to be involved in various biological processes, including germline and embryo development, parental imprinting, fruit ripening [3, 4] and stress responses [5, 6].

The regulatory role of DNA methylation is exerted on at least two different levels. Firstly, DNA methylation affects chromatin states and plays an instrumental role in heterochromatin formation [3], with general impacts on transcriptional levels across larger genomic regions. Together with other epigenetic modifications, namely the histone variants H3K9me2, H3K27me1 and H2A.W, mCs determine a repressive chromatin state to ensure stable silencing [7, 8]. Secondly, DNA methylation affects the binding of transcription factors (TF) and other proteins to DNA. A group of methyl-binding domain proteins binds to methylated CG-dense sequences and subsequently interacts with an array of additional proteins to regulate the silencing or activation of associated genes through modulation of the state of protein aggregation [3]. In *Arabidopsis*, most TFs are sensitive to the methylation state of TF recognition motifs *in vitro*, with 72% and 4.3% of tested TFs preferentially binding to unmethylated and methylated DNA motifs, respectively [9]. These findings lead to an expectation that heavy cytosine methylation is associated with transcriptional silencing, and that promoters of actively transcribed genes should be unmethylated in most cases. Cytosine methylation patterns of TEs (Transposable Elements), which should remain transcriptionally silent in most if not all tissues to maintain genome integrity, are in agreement with this expectation. In all examined plant genomes, CpG and CHG methylation (H representing any of A, C or T), and usually also CHH methylation, is distinctly elevated in TEs compared to surrounding regions [10]. However, the relationship between DNA methylation and transcription is more difficult to untangle in the case of non-TE transcripts. Constitutively unmethylated genes can have a very wide range of expression levels. On the other hand, genes characterized by TE-like methylation (i.e. high methylation levels in all sequence contexts) are often times silent, but some of them can be highly expressed

in specific tissues [11]. In most species, highly expressed genes are characterized with high CpG methylation and low CHG/CHH methylation in their bodies, forming a pattern called gene-body methylation (gbM). Nonetheless, gbM genes have the widest range of expression levels in most species [10], leading to doubts about the role of cytosine methylation in gene expression. Since the loss of gbM in a *met1* mutant of *Arabidopsis* (defective in DNA methylation maintenance) does not affect gbM gene expression patterns [12], it has been suggested that gbM has a minimal or no effect on gene expression, but possibly has a 'homeostatic' role in preventing aberrant transcripts [13]. Nonetheless, one general feature of DNA methylation in relation to transcription is recognized. Averaged across all genes, including gbM genes, CpG methylation shows a marked dip around transcription start sites (TSS), and to a lesser extent also around transcription termination sites [10]. For many species, including grasses *Zea mays* and *Brachypodium distachyon*, CpG methylation at TSS is associated with almost complete silencing [10]. It therefore appears that the vicinity of TSS needs to stay free of any CpG methylation in order to enable the initiation of transcription.

Bread wheat (*Triticum aestivum* L.) is a monocotyledonous crop with a large genome (14.6 Gb pseudomolecule-level assembly of Chinese Spring v2.1; [14]). The large genome size can be attributed to the combined effect of high TE content (85% of the assembly) and recent polyploidizations. Bread wheat is an allohexaploid ($2n=6x=42$ chromosomes; genomic constitution AABBDD) that emerged approximately ~10,000 years ago from hybridizations and subsequent polyploidization involving tetraploid *T. turgidum* ($2n=4x=28$; AABB) and diploid *Aegilops tauschii* ($2n=2x=14$; DD) [15, 16]. Due to its economic significance and the status of a worldwide staple crop, substantial efforts have been directed to characterize the bread wheat genome [17]. However, much less attention has been paid to the epigenome and the methylome (the complete set of DNA methylation modifications) in particular. A few large-scale studies [18–20] have reported that a genome-wide overview of DNA methylation in wheat matches the general patterns known from other species. However, more in-depth analyses are in demand, as the epigenetic regulation of gene expression is of special interest for epibreeding applications [21–23], but also in relation to fundamental questions on the evolution of polyploid species. While most genes of bread wheat are in three homeologous copies (i.e. syntenic paralogs located on the A, B and D subgenomes of the hexaploid), a significant fraction of them is silenced [19, 24–26], leading to frequent situations

where the total expression of a ‘triad’ (the set of three homeologous genes) is supplied by one or two copies. This phenomenon of homeolog expression imbalance is potentially significant for both evolutionary adaptation (homeolog neo- and sub-functionalization) and crop performance related to the ‘fixed’ heterosis of allopolyploids [27, 28]. It may also be related to the transcriptional reprogramming commonly observed in nascent polyploids [29, 30] and hypothesized to have an epigenetic basis [31]. DNA methylation changes have been observed to accompany polyploidization in multiple species [32–35], including *Triticeae* [36–39], but their genome-wide assessment in relation to gene expression in allohexaploid wheat is missing.

With these considerations, we developed a hybridization-based sequence capture design intended for methylome analysis *via* bisulfite sequencing (BSseq), targeting the 5′ end region (centered on TSS or start codon - SC) of most high-confidence (HC) genes of the wheat genome. We specifically hypothesize that cytosine methylation around TSSs is employed by plants as a transcriptional on-off switch, and we tested this hypothesis by comparing the BSseq and transcriptomic data from developing grain. We particularly focus on DNA methylation patterns in the polyploid context, exploring the relationship between DNA methylation and homeolog expression imbalance, monitoring DNA methylation changes in nascent allohexaploid wheat triggered by the whole-genome duplication, and examining possible links between differential gene expression and differential DNA methylation. We discuss our findings in the context of the underwhelming genomic response to polyploidization in wheat, and the persistently obscure relationship between DNA methylation and gene expression.

Methods

Plant material

All seeds used in this study were provided by the ‘Centre de Ressources Biologiques Céréales à Paille’ (Small Grain Cereals Biological Resources Centre), INRAE, Clermont-Ferrand, France. All data reported here were obtained from the developing grain sampled at 250 degree-days from both natural and synthetic hexaploid wheat lines (AABBDD) as well as from the parental genotypes *Ae. tauschii* (DD) and *T. turgidum* subsp. *durum* (AABB) (Supplementary Table 1). Natural hexaploids are represented by the bread wheat cultivar ‘Recital’. The synthetic wheats were prepared by crossing *Ae. tauschii* genotype ‘Tauschii-109’ with the durum cultivar ‘Langdon’ (producing synthetics ‘109xL-C2’, ‘109xL-C4’ and ‘Lx109-C2’; with identifiers indicating the direction of the cross and the sampled generation after colchicine-induced chromosome doubling). Additional spontaneous allohexaploids ‘Jx87-S5’ and ‘Jx109-S5’ produced by a different research

group (INRAE - Agrocampus Rennes) from the durum cultivar ‘Joyau’ and *Ae. tauschii* genotypes ‘Tauschii-109’ and ‘Tauschii-87’ were also analyzed, although direct descendants of the parental lines used for the crosses were not available for this study. Additional details about the allohexaploid synthesis, nomenclature and growth conditions are given in our previous paper [26] reporting on the transcriptomic analysis of these samples.

DNA extraction

The nucleic acid extracts that had been used previously in Banouh et al. [26] for RNA purification were processed here to obtain genomic DNA. Briefly, the grain tissue was ground into a fine powder using liquid nitrogen, a mortar and a pestle. Up to 1 g of the powder was dissolved in 4.5 ml of extraction buffer (0.1 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, 1% SDS). Nucleic acids were extracted twice with phenol:chloroform:isoamylalcohol (25:24:1) and subsequently precipitated using 3 M sodium acetate (1/10 volume) and two volumes of absolute ethanol. Following centrifugation, the resulting pellet was resuspended in nuclease-free water. An aliquot of this crude nucleic acid extract (50–75 µg) was used for the RNA-seq purification [26], while another one was treated with 30 µg RNase A and subsequently purified with silica-based spin columns (Qiagen) in order to obtain purified DNA. DNA was obtained from two biological replicates per sample.

Capture-BSseq probe design

Transcription start sites were identified on the basis of the Chinese Spring v1.1 gene annotation, treating the first position of 5′ untranslated regions (5′-UTR) as TSS. Out of 107,891 annotated HC genes, 68,731 (63.7%) have 5′-UTR annotations. In the remaining 36.3% of HC genes, TSS information is missing due to the absence of full-length mRNA data, and we aimed to capture the beginning of the coding sequence instead, i.e. the SC of those genes. While the SC position is not relevant for the initiation of transcription, the SC regions provide an insight into 5′-UTRs (upstream of SC) and the coding sequence (downstream of SC). Each gene was targeted by two probes within a 300 bp window around TSS or SC (one probe upstream and another one downstream). In order to increase the applicability of the probe design to varied *Triticeae* genomes (diploid *Ae. tauschii*, tetraploid *T. turgidum* sp., hexaploid *T. aestivum*), the probe design was adjusted according to 15 different chromosome-level assemblies - Chinese Spring v1.1 (IWGSC 2018); 10+ Wheat Genome Project [40]; *T. turgidum* subsp. *durum* cv. Svevo v2 [41]; *T. turgidum* subsp. *dicoccoides* [42]; and two assemblies of *Ae. tauschii* v4.0 [43] and AOCO02000000 [44]. Our design aimed for probes that correspond to consensus TSS/SC regions of all these

assemblies rather than just the Chinese Spring reference. To achieve this, all assemblies were shredded into 500 bp fragments with 250 bp sliding step using GenomeTools shredder, and all fragments were mapped onto the Chinese Spring v1.1 reference with *bwa mem* [45], creating a single bam file. Reads with mapping quality < 20 were filtered out (*samtools*). Subsequently, fragments (or rather a single consensus sequence per genome built from overlapping fragments) around each TSS/SC were extracted from the bam file in the form of multiple sequence alignments, using a custom python script (150 bp upstream and 150 bp downstream). All upstream fragments (for all genes and genomes) were then re-assembled into contigs with Geneious 6.1 [46] allowing no gaps and up to 3% mismatches within ≥ 120 bp overlaps. The same procedure was repeated for all downstream fragments. These contigs were considered as probe candidates, and further filtered. We removed probe candidates (i) redundant among the upstream and downstream sets; (ii) mapping to the wheat chloroplast genome with $\leq 10\%$ mismatches; (iii) having valid alignments (*bwa mem*) with < 8 of the used assemblies; and (iv) having more off-target than on-target blat hits (using *-fastMap* and *-minScore = 80*). This process resulted in 200,495 consensus probes covering ~ 100 k genes of wheat (typically with 2 probes per TSS/SC; Supplementary Table 2). Chinese Spring genes corresponding to all the removed probe candidates were not enriched in any GO (Gene Ontology) terms using TGT website [47, 48], indicating that our probe design is unbiased in respect to gene function. The capture probes were synthesized by Agilent (SureSelect custom design), covering both strands of the targets.

Preparation, sequencing and mapping of the capture-BSseq libraries

The sequence capture protocol with SureSelect Custom Tier6 probes (Agilent) was applied on genomic DNA prior to the bisulfite conversion with EZ DNA Methylation-Gold Kit (ZYMO Research). Illumina-compatible libraries were prepared with SureSelectXT Methyl Reagent kit (Agilent), with post-conversion multiplexing for parallel sequencing of all samples. The libraries were prepared by PGTB (France) and sequenced on NovaSeq6000 (2 \times 150) by Genotoul (France) to a theoretical depth of target coverage 40x. We obtained raw reads from 18 directional libraries (with both the top and bottom DNA strands). These reads underwent initial processing with Trim Galore to eliminate adapter sequences, reads with length < 20, and unpaired reads. Reads passing these quality filters were aligned to the Chinese spring reference genome v1.0 using Bismark [49] in the directional mode, employing the Bowtie2 mapper with predefined settings. Methylation data were extracted with *Bismark_methylation_extractor*, retaining only reads

with a unique best alignment score. Methylation data were then analyzed in the MethyKit R package [50].

Estimation of bisulfite conversion efficiency

We also mapped the filtered reads to chloroplast genomes (assumed to be completely unmethylated) in order to assess the bisulfite conversion efficiency. Reads from Recital, Langdon, Jx109, Jx87 and Lx109-C2 were aligned to the chloroplast genome of Chinese Spring (GeneBank ID NC_002762.1). Reads from 109 \times L-C2, 109 \times L-C4, Tauschii-109 and Tauschii-87 were mapped to the *Ae. tauschii* chloroplast genome (NC_022133.1). The conversion rate (Supplementary Table 3) was then calculated following the formula: conversion rate = (number of converted cytosines/total number of analyzed cytosines) $\times 100$.

Data analysis

We further filtered out presumed PCR artifacts by removing bases with coverage above the 99.99th percentile, using the 'filterByCoverage' function in methylKit. Cytosine methylation was summarized using two different coverage thresholds, depending on the type of the analysis. For the Principal Component Analysis (PCA) and the detection of differentially methylated cytosines, we only considered positions with a minimum of 10x coverage. PCA was calculated in R, using the function *prcomp* with data centering and scaling enabled. For global methylation averages across all genes summarized per subgenome and sequence context, mean methylation level [51] was calculated from all positions with ≥ 10 x coverage. For region-based methylation analyses (i.e. DNA methylation scans along chromosomes, all TSS/SC-based analyses in relation to gene expression and differential methylation), weighted methylation level [51] was calculated from all cytosines with coverage ≥ 3 . We consider such summarization optimal for the region-based analyses, since it maximizes the number of evaluated cytosines in an unbiased way (a low-coverage site contributes proportionally less to the weighted methylation level compared to a high-coverage site). Fine-scale methylation around TSS/SCs was summarized in 10 bp bins using the *ScoreMatrixBin* function from genomation package [52]. To explore the relationship between TSS methylation and transcription in leaves, complementing our data from grain, we downloaded public data for two-weeks old leaf of the bread wheat cultivar Chinese Spring (leaf_9 in SRA studies SRP133674 and SRP133837) published by Ramírez-González et al. [10]. These whole-genome BSseq and RNA-seq data were processed with the same pipelines used for our own data, restricting the methylation analysis to our target windows around TSS/SC.

Differentially methylated regions (DMRs) and differentially methylated cytosines (DMCs) were identified between hexaploid synthetics and parents (except for the synthetics Jx87-S5 and Jx109-S5, where BSseq data of the tetraploid parent are not available) using the 'calculateDiffMeth' function in methylKit that accounts for variation between replicates. It employs a logistic regression-based method with overdispersion correction and adjusts the p-values using the Benjamini-Hochberg method (FDR) [50]. Regions/cytosines with a methylation level difference greater than 25% and $FDR < 0.01$ were considered differentially methylated. For the DMR identification, we employed a model-based method using the methseg function in methylKit, which divides the target regions into variable-length segments of homogeneous methylation levels [50]. Segments identified in the parents (Tauschii-109 and Langdon) were used to screen the parents and their derived synthetics.

Results

Capture-BSseq efficiency

Assuming that DNA methylation of promoters and TSSs is important for the transcriptional level of genes, we aimed to capture the 5' ends of all HC genes annotated in the assembly of the Chinese Spring wheat genome. These intended targets are centered around TSS in most cases (150 bp upstream and downstream), but our capture probes targeted SC (150 bp upstream and downstream) in HC genes that lack 5'UTR annotation. We analyzed the methylation in the developing grain of natural hexaploid wheat cultivar (*cv.*) Recital, tetraploid durum wheat *cv.* Langdon, two diploid *Ae. tauschii* genotypes and five synthetic allohexaploids (Supplementary Table 1), targeting 108,196 TSS/SCs in 95,883 non-redundant genes of the Chinese Spring reference genome (Supplementary Table 2). The synthetics include polyploid progeny from reciprocal crosses and two successive generations (109xL-C2, 109xL-C4 and Lx109-C2), and synthetics derived from alternative parental genotypes (Jx87-S5 and Jx109-S5; BSseq data from the tetraploid parent missing). With two biological replicates per sample, we sequenced 18 libraries containing between 25 and 33 million read pairs (2×150 bp; Supplementary Table 4). The mapping efficiency (i.e. the percentage of uniquely mapped reads in respect to the raw reads count) ranged from 36% to 54%. Out of the uniquely mapped reads, 43% to 51% overlap with the intended targets comprising only 0.26% of the Chinese Spring assembly, which demonstrates strong target enrichment. Overall, 18%-26% of raw BSseq reads were used in the methylomic analysis (reads uniquely mapped to the target space). The bisulfite conversion rates estimated from the chloroplast genome-mapping reads averaged to 0.9787 (0.966–0.9844; standard deviation 0.0057; Supplementary Table 3). Within the target

regions, 69% to 88% of cytosine positions were covered with the minimum depth of 3x, while 39%-79% were covered with the minimum depth of 10x, per sample (Supplementary Fig. 1). The BSseq data reported here covers 75–88% of differentially expressed genes (DEGs) identified previously from the same samples [26], with depths 3x or higher (Supplementary Fig. 2).

Subgenome differences in mean methylation are inherited from lower-ploidy parents

Cytosine methylation occurs in the three sequence contexts (CpG, CHG and CHH) with different frequencies, as observed in other plant species. The highest mean TSS/SC methylation (considering only cytosines with $\geq 10x$ coverage) was observed at CpG sites (20%-27% across samples). Methylation of the CHG sites ranges from 12% to 15%, while the methylation at CHH sites is the lowest, at 2%-3% (Supplementary Fig. 3). Since the mean methylation of the CHH context is close to the non-conversion rate (i.e. the proportion of unmethylated cytosines where bisulfite conversion failed), we ascertained the proportion of CHH sites where the methylation is statistically significant according to the binomial test. Across samples, the methylation level of 84%-90% CHH cytosines is not statistically different from the failed conversion rate, and only 10%-16% of all CHH sites around TSS/SC show significant levels of methylation (Supplementary Table 5).

We calculated mean methylation levels per subgenome in all synthetic hexaploids, parental lines (Langdon and Tauschii-109), and the natural hexaploid wheat, using only cytosine positions consistently covered in all these samples (Fig. 1; Supplementary Fig. 4). In all three sequence contexts, the B subgenome showed the highest mean methylation level, followed by A and D (Fig. 1; Supplementary Fig. 4). This above-average B-methylation and below-average D-methylation is observed in both natural and synthetic hexaploids, as well as in the combined parental genomes. This suggests that the overall differences in subgenome methylation in the allopolyploids are not triggered by the whole-genome duplication, but rather are the legacy of the parental differences.

TSS/SC methylation varies along chromosomes, but is stable across polyploidization

Principal Component Analysis (PCA) calculated from all sequence contexts shows that the top two principal components of DNA methylation distinguish samples according to their genetic identity (Fig. 2a). The methylomes of synthetics do not diverge much from the parental methylomes, but are clearly separated from the methylome of natural bread wheat (Recital).

To examine DNA methylation landscapes along chromosomes, we calculated weighted mean methylation across TSS and SC in 2 Mb tiling windows and 1 Mb

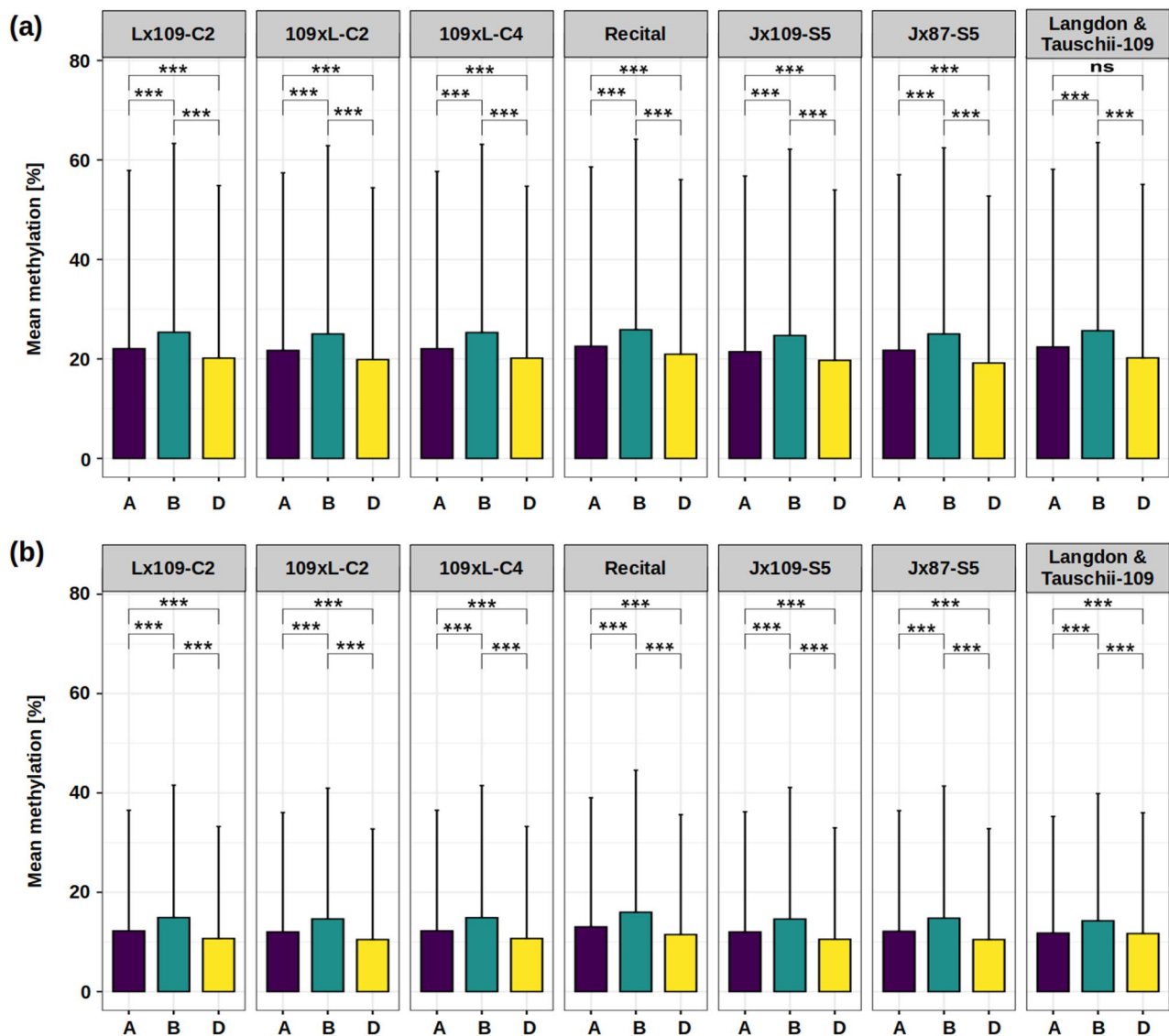


Fig. 1 Mean methylation levels per subgenome (A, B and D) across genotypes. **a** CpG context; **b** CHG contexts. Asterisks indicate significant differences (Wilcoxon test); ns, not significant at $\alpha=0.05$; *** $p < 0.001$. Error bars represent the standard deviation

sliding intervals, merging all sequence contexts into a single track. DNA methylation of the target regions varies considerably along chromosomes, with TSS/SC around centromeres showing elevated levels. Nonetheless, maximum values can sometimes be found away from centromeres (Fig. 2b).

Across samples, the chromosomes display highly similar methylation patterns, especially on A and B, with more apparent differences along the D chromosomes. Universally, differences between genotypes (Tauschii-87 vs. Tauschii-109; Recital vs. all other) are greater than differences between parents and their derived synthetics. The methylation landscapes of Langdon are highly correlated with the AB patterns of the derived synthetics (Pearson's correlation in the range 0.992–0.995; Fig. 2b).

While the D methylomes of Tauschii-109-derived synthetics (109xL-C2, 109xL-C4, Lx109-C2, Jx109-S5) differ only slightly among each other (r range 0.991–0.995), all of them diverged more clearly from their D-parent (r range 0.963–0.978) (Fig. 2c). This suggests that the D methylome encountered modifications that are consistent across multiple polyploidization events. Overall, the D methylome appears to be more diverse than the AB methylomes, and more prone to changes after polyploidization.

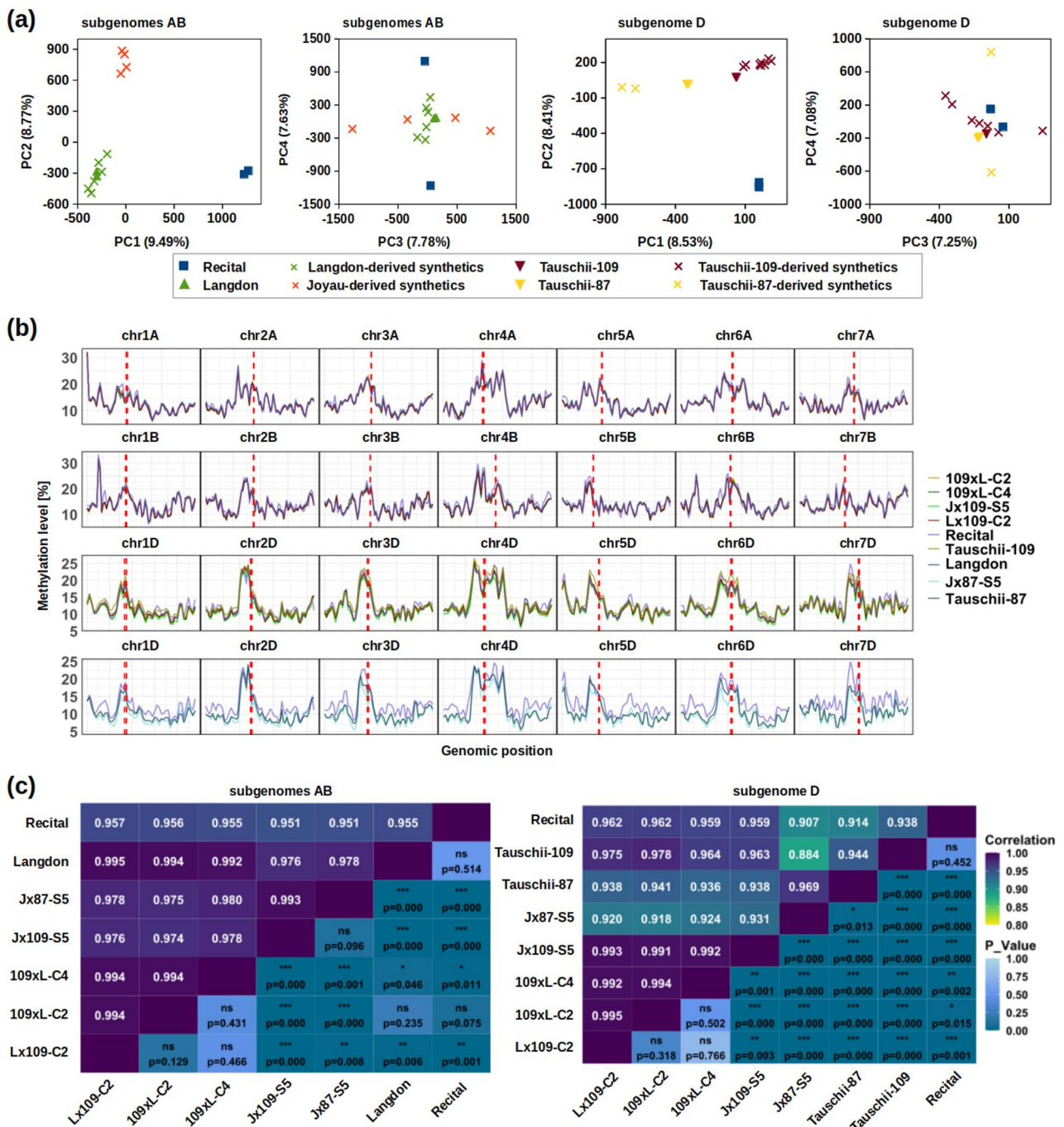


Fig. 2 DNA methylation profiles and correlations across chromosomes and samples. **a** The top four principal components of methylome variation across all samples, calculated from all genes and contexts for the AB subgenome (panels to the left of center) and the D subgenome (panels to the right). **b** An overview of total DNA methylation along chromosomes (2 Mb windows with 1 Mb step), with each row of panels containing synthetic lines, their corresponding parental lines and the natural hexaploid Recital. The first and second rows display the A and B chromosomes, respectively, and both include the same synthetic lines derived from the AB parent Langdon as well as the corresponding AB chromosomes from Recital. The third row shows chromosomes from Tauschii-109, the D subgenome of the derived synthetics, and the D chromosomes of Recital. The fourth row represents chromosomes from Tauschii-87, the D subgenome of the derived synthetics, and the D chromosomes of Recital. **c** Heatmaps showing chromosome-level methylome differences between genotypes, separately for the D and AB subgenomes, respectively. Pearson's correlation coefficients and Wilcoxon test p-values are given above and below the diagonal, respectively. Asterisks indicate significant differences: ns, not significant at $\alpha=0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

TSS/SC methylation does not clearly depend on centromere distance and is a poor predictor of transcriptional levels

In the following two sections, we investigated the TSS/SC methylation in relation to transcription and the distance from the centromere within trivial one-sample analyses (i.e. across all genes in a single sample, or jointly in all samples), not considering the polyploid context.

Since centromeric regions are rich in TEs and have relatively low gene densities, it can be expected that DNA methylation around centromeres is higher compared to more distal regions, as we have observed on the chromosome scans (Fig. 2b). Accordingly, we observed a consistently significant negative correlation between CHG methylation and centromere distance for the group of genes with SC-centered data (Fig. 3; Supplementary Fig. 5). However, this is not the case for CpG methylation around TSS, where we observed positive correlation (weak, but significant in all samples; Supplementary Fig. 5). In both cases, the proportion of shared variance is very low ($r^2 < 0.01$), which means that TSS/SC methylation barely depends on the distance from the centromere. This contrasts somewhat with the relationship between

gene expression [logarithmic form of transcripts per million (TPM)] and centromere distance, which is consistently negative and significant for both groups of genes (Supplementary Fig. 5). It may seem surprising that genes closer to the centromere have higher transcript levels, but this observation is consistent with findings of Zhao et al. [53], who explained this observation by the high accessibility of centromeric chromatin.

The relationship between TSS/SC methylation and gene expression is consistently inverse and statistically significant in the CpG and CHG contexts of all samples (Fig. 3; Supplementary Fig. 5). However, this relationship is very weak, with the proportion of shared variation (r^2) between 0.004 and 0.025. For each sample, we observed large numbers of silent genes with both methylated and unmethylated TSS, as well as hundreds of moderately/highly expressed genes (TPM > 10) with TSS methylation > 25%. A group of genes with high expression (TPM > 200) and medium-to-high methylation around SC (> 30% in CpG and simultaneously > 15% in CHG) can be recognized, with most of them (87) identified in at least two samples (Supplementary Table 6). Among these genes, 34 are associated with the gene ontology term

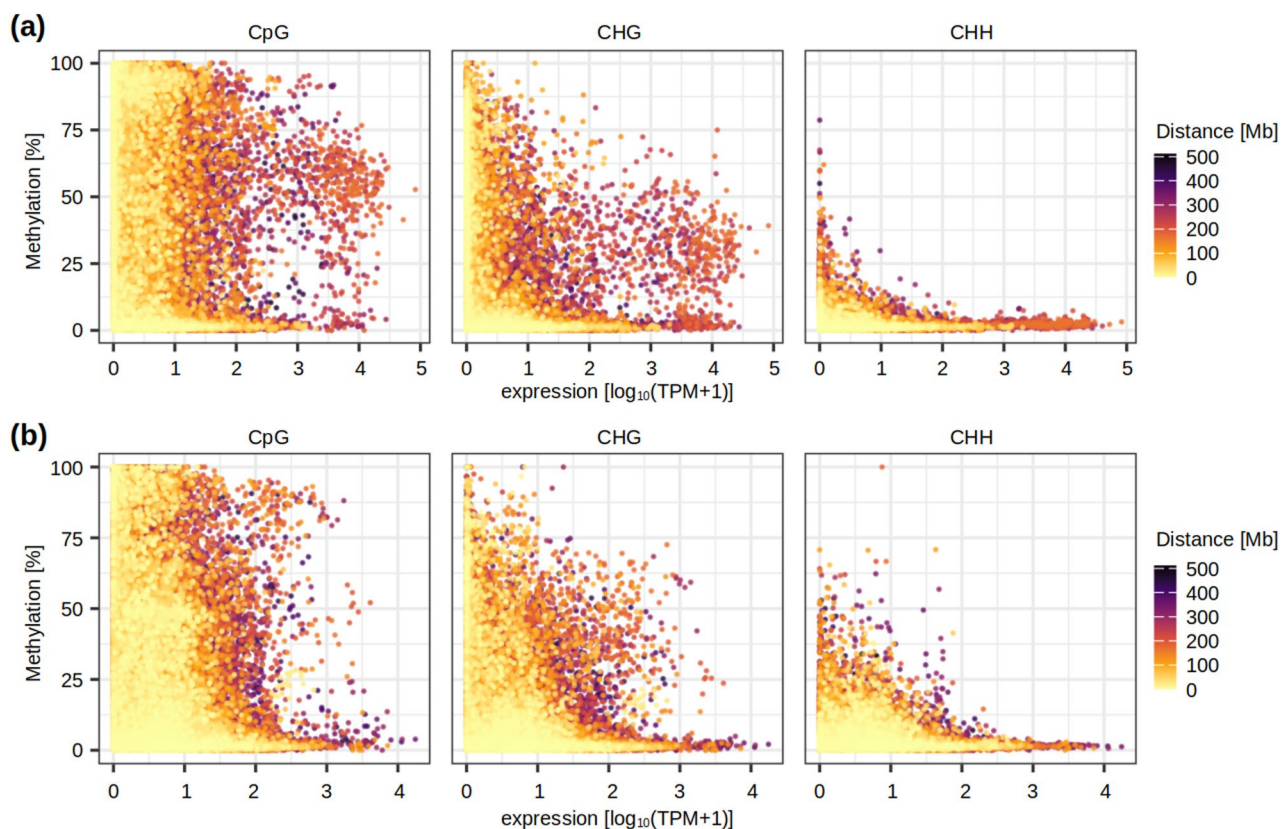


Fig. 3 The relationship between transcription, DNA methylation level and centromere distance. Transcription is shown on the x-axis, DNA methylation on the y-axis, and the centromere distance is indicated with a color code. All samples with all three types of data were overlaid on the same scatter plot, and each dot represents a single gene. Data for individual samples, together with correlation coefficients for methylation-distance, methylation-expression, and expression-distance, are shown on Supplementary Fig. 5. **a** SC-centered dataset; **b** TSS-centered dataset

'nutrient reservoir activity' (syn. 'storage protein'; enrichment FDR $1.39E-85$), which implies that genes producing seed storage proteins, which are highly up-regulated in the developing grain, have relatively high SC methylation. CHH methylation shows only a weak relationship to transcription in the SC-centered dataset, and no consistent relationship in other comparisons (Fig. 3; Supplementary Fig. 5).

The absence of a strong correlation between TSS methylation and transcript abundance on the whole genome level does not indicate a common role of TSS methylation in the regulation of gene expression, as initially hypothesized. We further explored a *post-hoc* hypothesis that TSS methylation may be involved in switching alternative transcription start sites. Common read-summarization methods (like featureCounts used with our data) cannot determine a specific start of transcription, whether it lies within or outside of annotated features. Therefore, a possible role of DNA methylation in switching among alternative TSSs would be invisible on a simple methylation-expression scatter plot. While most genes in the TSS set have a single annotated 5'UTR, over 22 thousand genes have multiple 5'UTR annotations. Since we do not know which of these alternative 5'UTRs correspond to the most frequent TSS in our samples, we excluded those genes from the correlation analysis. This had only a marginal effect on r^2 values ($+0.001$ and -0.0007 in the CpG and CHG contexts, respectively), suggesting that either the hypothesis of TSS switching by methylation is rejected, or that the 5'UTR annotations are not accurate enough to resolve the question. We therefore checked the quality of 5'UTR annotations by inspecting RNA-seq read alignments (bam files) for one of our samples (Recital) and a set of 144 highly-methylated/highly-expressed genes (above 95th percentile in both distributions, simultaneously). In most of these cases (131), 5'UTR annotations is approximately correct, that is, transcription seems to start mostly at the beginning of the annotated 5'UTR (Supplementary Fig. 6a) or in its close vicinity (within the region targeted by the probes; Supplementary Fig. 6b). Only for 13 (9%) of these genes, the most frequent TSS is hundreds of bp away from the annotated 5'UTR start (Supplementary Fig. 6c), sometimes associated with additional exons, or TSS is unclear, probably due to alignment problems (Supplementary Fig. 6d). Although our data does not allow us to conduct a comprehensive evaluation of the accuracy of 5'UTR annotations, our visual examination of RNA-seq alignments confirms that most of the highly-expressed genes with highly-methylated TSS appear to have a correct 5'UTR annotation.

We further explored the possibility that the 250 degree-day grain analyzed here is an outlier tissue or developmental stage, and the relationship between TSS

methylation and transcription in wheat is stronger in other tissues, e.g. leaves, as shown in other species [10]. We analyzed public data for wheat transcriptome and methylome sampled from a two-weeks old leaf of cv. Chinese Spring [19]. In a comparison of Chinese spring leaves and Recital grain, transcriptomes differ substantially ($r^2=0.33$), while TSS/SC methylation (CpG and CHG) are rather similar ($r^2=0.68$). After excluding genes with multiple 5'UTR annotations, we found that the correlation between TSS methylation and transcript abundance in young leaves is weak (Supplementary Fig. 7). The proportion of shared variance (r^2) is 0.011, 0.003 and 0.002 in the CpG, CHG and CHH contexts, respectively; not much different from the mean values across our grain samples (0.014, 0.003 and 0.0002, respectively). We conclude that the absence of a strong genome-wide relationship between TSS methylation and transcription can probably be generalized for various tissues in wheat.

Fine-scale methylation around TSS/SC

Additionally, we explored the relationship between gene expression and DNA methylation on a finer spatial scale, summarizing DNA methylation in 10 bp intervals around TSS/SC. We separated expressed (TPM > 0.1 in at least two libraries) and silent genes (TPM > 0.1 in less than two libraries), creating four categories: (i) expressed genes with TSS-centered methylation data; (ii) silent genes with TSS-centered methylation data; (iii) expressed genes with SC-centered methylation data; (iv) silent genes with SC-centered methylation data. Subsequently, we averaged the 10 bp windows across all genes in a sample, and further averaged multiple samples to visualize a single pattern per category (Fig. 4a).

Methylation levels around TSS are generally lower compared to SC. Within each dataset, expressed genes consistently show lower methylation levels than their silent counterparts (Fig. 4a). In general, DNA methylation levels are the lowest around TSS of expressed genes. Methylation progressively decreases as it approaches TSS and then remains low until 130 bp downstream. Compared to expressed genes, silent genes have their TSS methylation ~6% and ~2% higher in the CpG and CHG contexts, respectively. The methylation differences between expressed and silent genes are even more pronounced in the SC dataset, reaching ~23% and ~16% in the CpG and CHG contexts, respectively. This high difference is driven by a markedly higher SC methylation of silent genes, averaging around 27%-36% and 13%-21% in the CpG and CHG contexts, respectively. In other words, compared to TSS, local methylation patterns around SCs show a clearer methylation difference in relation to the gene expression status. While the pattern is consistent for the CpG and CHG context and the two groups of

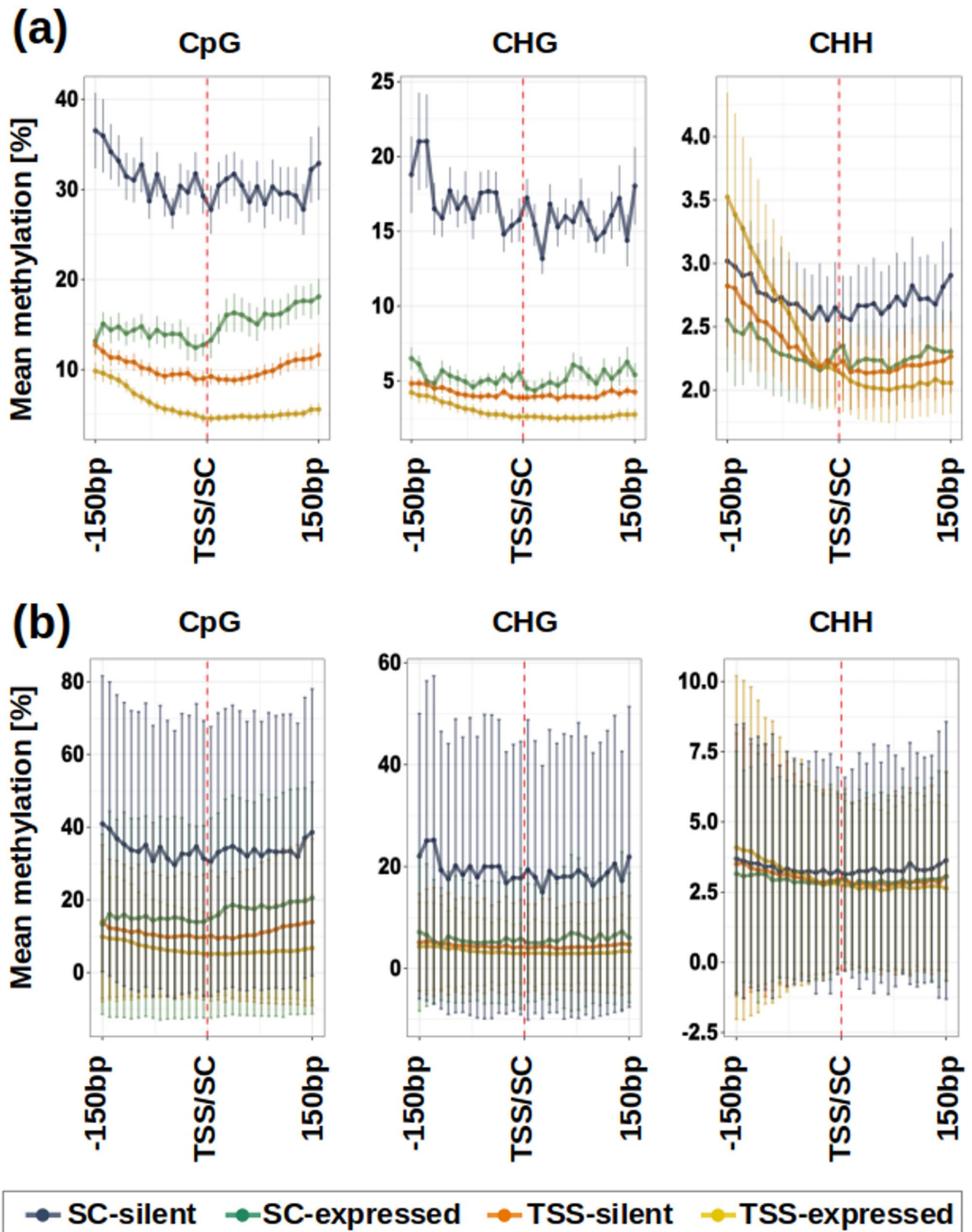


Fig. 4 Mean methylation levels at 10 bp resolutions around TSS (genes with multiple 5'UTR annotations removed) and SC of all analyzed genes. **a** Values averaged across all samples are shown, with error bars representing standard deviations across samples. **b** Mean methylation levels for Recital only, with error bars representing standard deviation across genes

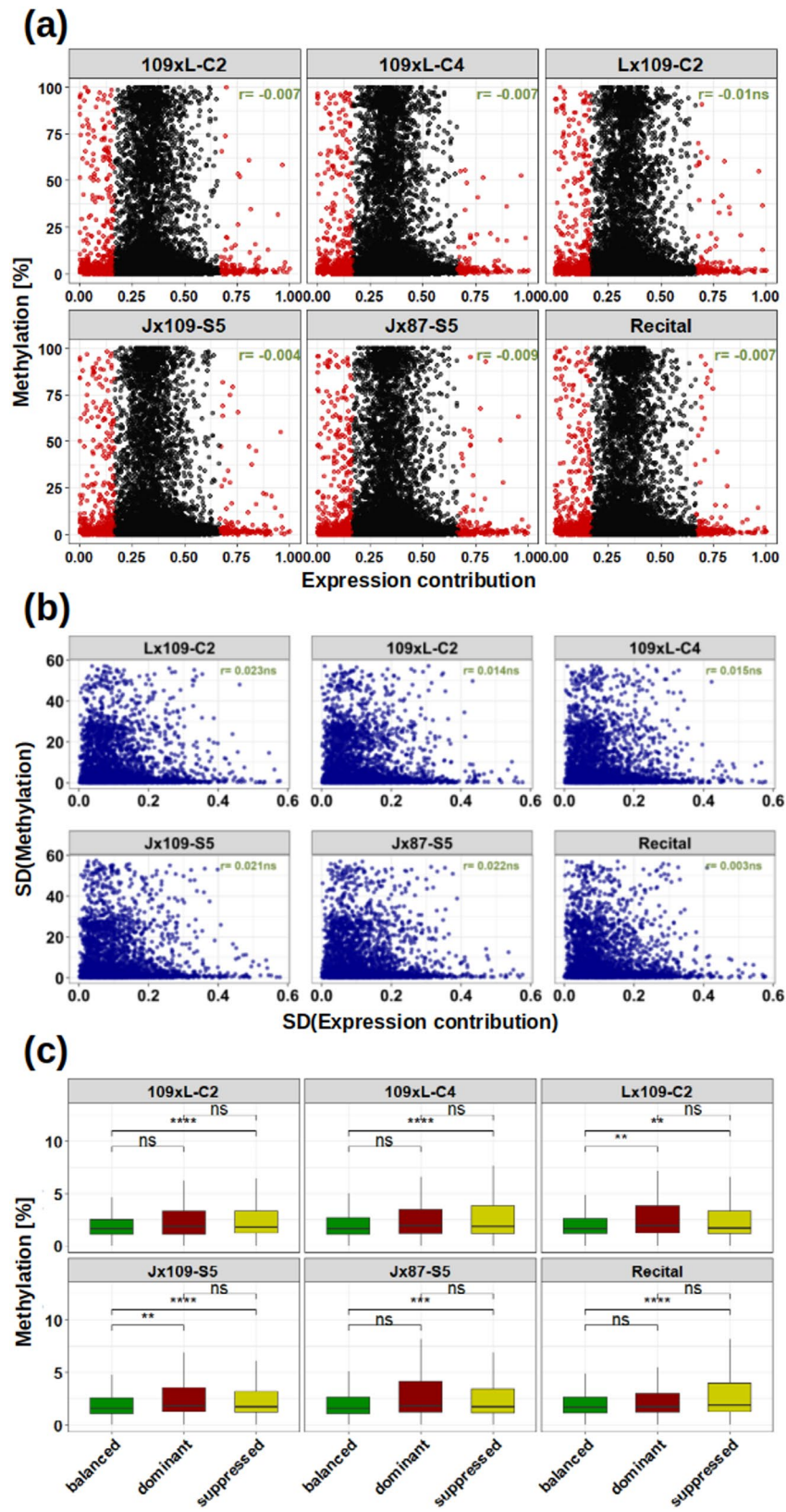


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 DNA methylation and homeolog expression bias. **a** Scatter plots with each data point representing a single homeolog, with its contribution to triad expression on the x-axis and CpG methylation level on the y-axis. Suppressed and dominant homeologs are shown in red. **b** Scatter plots with each data point representing a triad, with variance (SD) in the expression contribution on the x-axis, and variance of the CpG methylation level on the y-axis. Correlation coefficients are shown in green. **c** Boxplots showing CpG methylation across balanced, suppressed and dominant genes. Asterisks indicate statistical significance (Wilcoxon test; ** $p < 0.01$; *** $p < 0.001$; ns, not significant at $\alpha = 0.05$)

genes, the CHH context is more ambiguous, due to generally low methylation levels.

It is important to emphasize that Fig. 4a shows double-averaged data sets (all genes are averaged per sample before averaging all samples), with standard deviations relating to the across-sample variation (not across-gene variation). Although the graphs appear to suggest that the expression category of individual genes can be confidently determined by checking the methylation status around TSS and SC, this is far from the truth, because across-gene variability in each of the four categories is much greater than the across-sample variability, as we show for Recital (Fig. 4b). Just as suggested by the scatter plots on Fig. 3, each category contains genes with very high and very low DNA methylation.

Homeolog expression bias is not caused by TSS/SC methylation

In the following three sections, we investigated TSS/SC methylation in the polyploid context, analyzing methylation and transcription of homeologs; recording methylation changes after polyploidization; and comparing differential gene expression and differential methylation between synthetics and their parents.

This part of our study uses the concept of a triad and quantifies the contributions of individual homeologs to the total transcription of a triad on the basis of a 'ternary plot' [19], as described in Banouh et al. [26]. These individual homeolog contributions are compared with the weighted mean methylation level of their TSS/SC regions. In accordance with the ternary plot concept, we considered a homeolog to have a 'balanced' contribution to the triad expression if its transcripts make up between 16.7% and 66.7% of the triad total. Below or above this range, the homeolog contribution is considered suppressed or dominant, respectively. When individual homeolog contributions are analyzed in the form of a continuous variable, very low and mostly insignificant correlation with DNA methylation is observed for all samples and sequence contexts (Fig. 5a; Supplementary Fig. 8).

Homeologs span the full range of methylation values for almost any level of expression contribution. It appears that with increasing dominance, homeologs are associated with lower methylation levels. However, it should be noted that very few homeologs show strong dominance, therefore the apparent pattern can be due to decreasing number of data points. The lack of correlation suggests

that methylation levels do not determine the level of contribution to the overall triad expression.

Since the scatter plots with individual homeologs dissolve the triad information (i.e. homeologous groups are not identifiable on the scatter plot), we further checked whether the variance of methylation within triads correlates with the variance of transcription. Such a comparison tests the hypothesis that imbalanced triads are associated with high methylation differences while uniform triad methylation leads to a balanced homeolog expression, and is therefore independent from the direction of the relationship (i.e. it does not matter whether high methylation causes low transcription in one triad, but high transcription in another triad). This comparison did not detect significant correlations in the CpG and CHH contexts, and found significant but very weak positive correlations in the CHG context (proportion of shared variation $r^2 \sim 0.003$) (Fig. 5b; Supplementary Fig. 9). This does not suggest that DNA methylation plays an important role in homeolog expression differences. The relationship is not clearer when individual homeolog contributions are analyzed in the form of categorical data, creating balanced, dominant and suppressed homeolog groups (Fig. 5c). While the mean CpG methylation of suppressed homeologs is consistently distinguished from the mean of balanced homeologs (t -test; $\alpha = 0.05$), dominant and suppressed homeologs are not statistically different in any of the allohexaploids. This pattern is repeated in the CHG and CHH contexts (Supplementary Fig. 10). The result showing that suppressed homeologs are statistically different from balanced homeologs but not from the dominant ones is perplexing, as the dominant and suppressed categories are expected to be the most distinct.

DNA methylation changes in nascent polyploids

We searched for DMRs (differentially methylated regions) between synthetics and their parents, analyzing all covered genes and each cytosine context separately, with two different approaches defining the 'methylation regions'. The first approach uses the entire target regions (300 bp around TSS/SC) as fixed-length windows. The second approach is based on a change point segmentation calculated with the methseg function in MethyKit. It determines segments of varying sizes defined by the homogeneity of methylation levels across consecutive cytosines. The segments were identified from pooled parental data, and were set to contain a minimum of

three CpG sites with $\geq 10\times$ coverage. These segments were then used to track methylation changes across the polyploidization event.

The change point segmentation approach identified between 34 and 47 DMRs within sets of 44,944–45,835 genes analyzed (Fig. 6a). The CpG context recorded the highest number of TSS/SC DMRs, with counts ranging from 21 to 32 per sample, followed by the CHG (8–15) and the CHH context (1–2). Across samples, 109xL-C2 recorded the highest number of DMRs (1–32 depending on the context). In respect to subgenomes, most changes occurred on D, with 17–26 DMRs across samples. This is in agreement with the comparison of general methylation patterns along chromosomes (Fig. 3). The ratios of hypermethylation and hypomethylation were not consistent across the different synthetics. In Lx109-C2, hypomethylation was more common than hypermethylation

for all subgenomes. Hypomethylation was also more prevalent in 109xL-C2, but this pattern did not persist in the subsequent generation (109xL-C4). Some of the identified DMRs occur repeatedly in multiple synthetics (involving 3–7 genes per pairwise comparison; Fig. 6b). Such overlaps are statistically significant according to the hypergeometric test, which predicts < 2 genes to overlap by chance in sets of the given sizes. This indicates that some nonrandom and reproducible TSS/SC methylation changes occur in the synthetics. Nonetheless, it should be noted that significant methylation changes in parent-synthetic comparisons were detected in a very small fraction of the analyzed genes (0.08%–0.1%).

When the entire target windows of 300 bp around TSS/SC were used as regions in the DMR analysis, slightly more DMRs were identified (91–110; Supplementary Fig. 11), corresponding to 0.09%–0.1% of the total genes

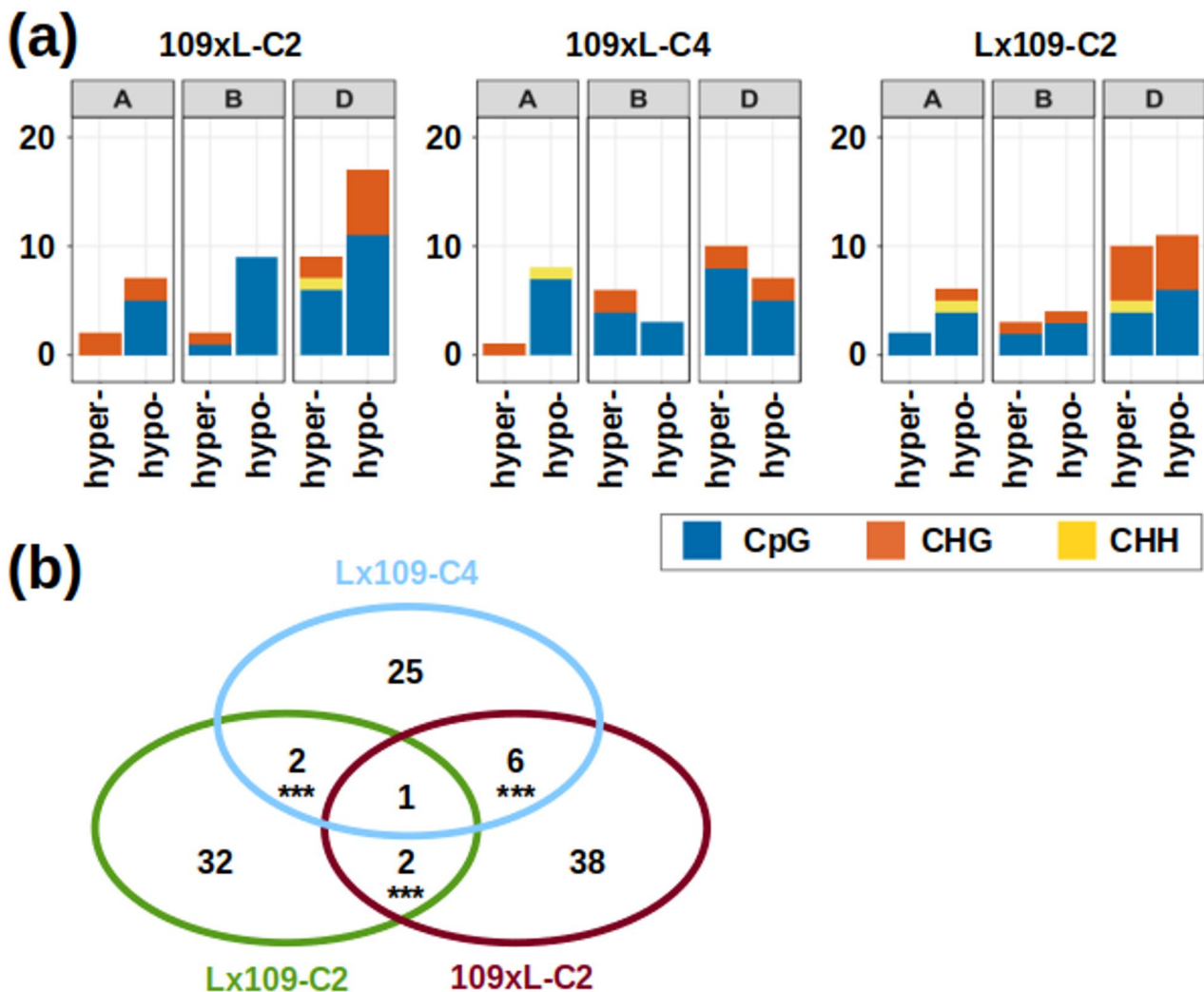


Fig. 6 DMR identification using the change point segmentation approach. **a** Counts of hyper/hypomethylated DMRs in different hexaploid synthetics across different subgenomes and different cytosine contexts. **b** Venn diagram showing the overlap of DMRs (considering all contexts) detected in different synthetic samples. Asterisks indicate the significance of the overlaps (***) $p < 0.001$)

analyzed. While the patterns seen in the change point segmentation were preserved here (i.e. the excess of changes on the D subgenome), the two methods detected generally different sets of genes (Supplementary Fig. 12). This is likely due to the different sizes of DMRs in the two methods. The change point segmentation focuses on segments with homogeneous methylation that can contain as few as three cytosines. This can lead to the detection of short DMRs that are frequently missed by the fixed window approach. On the other hand, 300 bp windows can have a more robust statistical signal that is missed in shorter segments. Nevertheless, it is difficult to assess the sensitivity and specificity (the false negative and false positive rates, respectively) of DMR detection methods, since it is unclear what a true positive is. A real DMR could be defined as a DMR with biological consequences (e.g. transcriptional change, chromatin structure, ...); however, we cannot assess biological consequences of DMRs prior to their identification, which creates a circular problem. For the moment, any method of DMR detection that identifies statistically significant differences may be considered valid, but we have to accept the possibility that some or all of the detected DMRs may have no biological consequences.

DEGs and differentially methylated TSS/SCs are unrelated in nascent wheat polyploids

In our previous study [26], we identified sets of DEGs triggered by polyploidization in wheat synthetics. With DNA from the same nucleic acid extracts that had been used for the transcriptomic analysis, here we searched for DNA methylation changes and tested whether the changes in transcription can be explained by mC modifications. Only DEGs with sufficient coverage in the BSseq analysis (10x) were included in the comparison. The association between DEGs and DMRs was assessed with Venn diagrams (Fig. 7a) and the exact hypergeometric test. Across all samples and either of the DMR detection methods, only four genes simultaneously change their methylation and transcription after polyploidization: TraesCS2D02G309500 (12-oxophytodienoate reductase2) downregulated with hypermethylated TSS; TraesCS5D02G553000 (beta-fructofuranosidase) downregulated with hypomethylated SC; TraesCS6D02G026500 (acid beta-fructofuranosidase) downregulated with hypermethylated TSS (all three in Lx109-C2); and TraesCS3B02G234600 (Vesicle transport v-SNARE 13) downregulated with hypermethylated SC (in 109xL-C4). However, it is unclear whether the differential expression and methylation in these genes are causally related. According to the hypergeometric test, such overlaps are not significantly different from random, given the sizes of the DMR and DEG subsets. Therefore,

we cannot reject the null hypothesis regarding the independence of DEGs and DMRs.

In addition to the DMR analysis, we also investigated the proportion of DMCs (differentially methylated cytosines) in DEGs and non-DEGs. A cytosine with $\geq 10x$ coverage was considered a DMC if the methylation level difference between parents and synthetics was $\geq 25\%$ and $p < 0.01$. We detected 0.43 and 0.7 DMCs per 1,000 DEG-associated cytosines in Lx109-C2 and 109xL-C2, respectively, which is not statistically different from the frequency of DMCs in non-DEGs (Fig. 7b). In 109xL-C4, we found that the frequency of DMCs was significantly higher in DEGs (compared to non-DEGs; chi-square test and hypergeometric test). This was mainly due to 19 hypermethylated DMCs around the SC of TraesCS3B02G234600 (Vesicle transport v-SNARE 13). This gene is downregulated in 109xL-C4, where it was also detected as a DMR (300 bp regions method). It is therefore possible that the hypermethylation and downregulation are causally related in this case. Overall, our results show that methylation changes around TSS/SC are rarely induced by allohexaploidization in AABBDD-type wheat synthetics, and perhaps with a single exception are not statistically associated with changes in gene expression.

Discussion

The unclear role of DNA methylation in gene expression and polyploid formation

It has long been suggested that the evolution of DNA methylation in eukaryotes is tied to the need for TE silencing [54]. But the role of DNA methylation expands far beyond the prevention of TE-related disturbances, appearing as one of the crucial mechanisms of epigenetic regulation of gene expression. In mammals, DNA hypomethylation is a hallmark of cancer cells, and appears to promote oncogenesis [54]. On the other hand, hypermethylation of CpG islands in some promoters causes transcriptional silencing, which can also lead to cancer development [55, 56]. Hypermethylation is also involved in the formation of facultative heterochromatin, best exemplified by the X-chromosome inactivation, where DNA methylation is employed to maintain (though not to initiate) stable silencing that compensates the gene dosage imbalance in mammalian females [57]. Due to its evolutionary conservation and taxonomic ubiquity, DNA methylation is assumed to play similar roles in plants. Nonetheless, particular features of plant DNA methylation, namely the occurrence of mC in any sequence context (symmetrical CpG and CHG, or asymmetrical CHH) and the scarcity of CpG islands, make the study of DNA methylation in plants more difficult, and its role in epigenetic regulation more complex.

DNA methylation has long been hypothesized to play an important role in polyploidization, which is a frequent

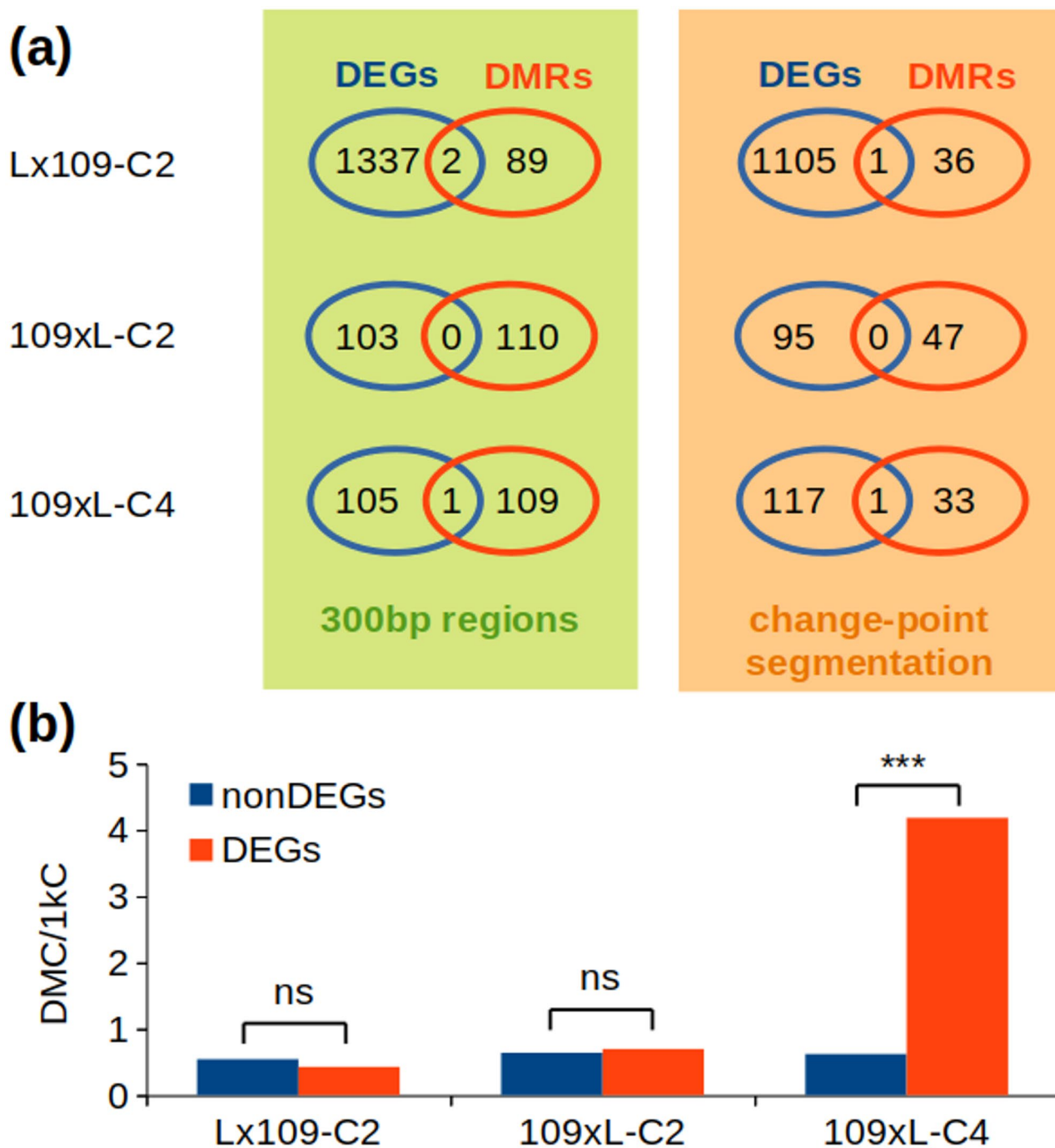


Fig. 7 Differential expression and differential methylation. **a** Venn diagrams representing per-sample overlaps between DEGs and DMRs obtained with either of the two methods used. Numbers of DEGs differ between the two DMR detection methods because only DEGs with sufficient DNA methylation data are included in the statistical comparison. None of the overlaps is statistically over-represented according to the exact hypergeometric test with normal approximation. **b** Occurrence of DMCs in DEGs and non-DEGs, expressed as the proportion of DMCs within 1,000 covered cytosines. Statistical differences between DEGs and non-DEGs were tested with the hypergeometric test and the chi-square test; ns, not significant at $\alpha = 0.05$; *** $p < 0.001$

phenomenon in the plant kingdom. Although polyploidization does not create a gene dosage imbalance, as copy numbers of all genes are multiplied equally, it is often assumed that nascent polyploids need to fine-tune their gene expression in the new genomic environment, and

that epigenetic changes are involved in this process. It has been speculated [31] that the RNA-directed DNA methylation (RdDM) pathway could be involved in the 'diploidization' of polyploid genomes, i.e. a gradual evolutionary process of sub-/neo-functionalization and loss of

redundant gene copies. Conceivably, RdDM triggered by the ‘shock of genome doubling’ could selectively inactivate redundant gene copies, leading to heterochromatin formation, and subsequently to sequence diversification or gene loss [31]. Following similar hypotheses, DNA methylation patterns have been investigated in a number of plant species with whole-genome doubling in their ancient history, including rice [58], maize [59, 60], soybean [60], pear [32], and *Brassicaceae* [61, 62], usually finding little evidence of epigenetic differentiation of subgenomes, and an unclear relationship between DNA methylation and transcriptional activity of the duplicated genes [63].

The study of methylomic changes after polyploidization presupposes a more general understanding of the role of DNA methylation in plants. The impact of cytosine methylation on gene expression is of particular interest, but the relationship is still shrouded by uncertainty. Given the scarcity of distinct CpG islands in plant promoters and plant genomes in general, there is typically no prior information on the position of elements whose methylation has consequences for the transcription of adjacent genes. DNA methylation is therefore typically measured in the transcribed gene partitions (gene bodies) and loosely-defined promoters. However, gene body methylation is persistently difficult to relate to transcription, not only in the polyploid context [59, 61, 62], but also in diploids. After detailed analyses of gene expression and methylomes in *Arabidopsis thaliana* epigenetic recombinant inbred lines and in *Eutrema salsugineum* that lacks gbM, Bewick et al. [12] found no evidence supporting gbM involvement in the regulation of transcription. Nonetheless, gbM could be involved in the prevention of aberrant and spurious transcripts that initiate outside of proper TSS, as it is described in mammals [64, 65]. This idea (together with the efficient silencing of TEs by DNA methylation) is consistent with the hypothesis that the initiation of transcription requires unmethylated DNA and is prevented in methylated regions. This is in agreement with observations of low methylation around TSS across multiple plant species [10]. However, it remains unclear whether TSS methylation can be employed by plants as an epigenetic on-off switch to fine-tune gene expression under internal (genomic) or external (environmental) stress.

TSS methylation is only weakly associated with gene silencing

We hypothesized that the region around TSS, which is the place of transcriptional initiation and includes the ‘core promoter’, could be crucial for the epigenetic control of gene expression. Focusing on 300 bp windows centered around TSS, we observed a wide range of CpG methylation for virtually any gene expression level in the

developing grain of wheat (Fig. 3; Supplementary Fig. 5). While most genes have mean CpG methylation of their TSS below 10%, there are typically >1,000 outliers with TSS methylation 25–100% and transcript abundance in the range 1×10^0 – 1×10^4 TPM. Although CpG methylation and gene expression are negatively correlated with statistical significance in all samples, the proportion of shared variation (r^2) is extremely low (0.009–0.021) and hardly indicative of a causal relationship. The situation is analogical in the CHG context, and for the set of genes with SC methylation data. Moreover, this weak relationship between TSS methylation and expression was also observed in young leaves of Chinese spring, and can be probably generalized across tissues. Overall, transcript abundance is as hard to relate to DNA methylation around TSS, as it is to gene body methylation reported in other species [61, 66]. On a finer spatial resolution (Fig. 4), across-gene means of CpG and CHG methylation can distinguish expressed (TPM > 0.1) and silent states, but the variance in these categories blurs the differences.

These results seem to be in contrast to those reported in Niederhuth et al. [10], who quantified for multiple species transcript abundance in mCG-TSS genes, i.e. genes where TSS overlaps with mCG regions (defined on the basis of a binomial test). For most examined species including *Brachypodium distachyon* and *Zea mays*, transcript abundance of mCG-TSS genes is virtually zero, although it should be noted that the boxplots (Supplementary Fig. 19 in [10]) do not show outlier data points. Our results indicate that wheat is probably more similar to *Oryza sativa* or *Solanum lycopersicum* [10], where CpG methylation of TSS permits moderate levels of transcription. The poor correlation between TSS methylation and gene expression in diploid, tetraploid and hexaploid wheats observed here, and the variability of this relationship across genera, do not suggest that TSS methylation has universally strong consequences for the level of gene expression.

This conclusion adds further ambiguity to the presumed link between transcription and DNA methylation. Below, we consider several explanations for this persistently obscure relationship, possibly working in combination.

- (i) DNA methylation controls the expression in only a fraction of genes and has no impact in most cases. There is strong evidence to support this possibility. In human cells, methylation of only 16.6% of CpG sites shows a significant correlation with the transcription at neighboring TSS [67]. The authors concluded that ‘*direct and selective methylation of certain TF binding sites that prevents TF binding is restricted to special cases and cannot be considered as a general regulatory mechanism of transcription.*’

In plants, the global effect of DNA methylation on transcription can be assessed with mutant lines of *A. thaliana* defective in DNA methylation maintenance. In DNA methylation-free quintuple mutant with all five known DNA methyltransferases (MET1, DRM1, DRM2, CMT3, and CMT2) knocked out, extreme growth retardation and small size was observed, together with a suite of severe developmental defects [68]. Nonetheless, the complete removal of DNA methylation caused only 3,738 DEGs (fold change > 2; FDR < 0.01), which is 13.5% of all annotated protein-coding genes. Much fewer DEGs (310; 1.1% of all annotated protein-coding genes) were observed in the quadruple mutant (*drm1 drm2 cmt3 cmt2*) retaining CpG methylation but defective in CHG and CHH methylation maintenance [68]. Since the complete removal of DNA methylation modifies the transcriptional regulation of only a fraction of genes, local and lower-magnitude changes of DNA methylation are likely to have even smaller impacts on transcription.

- (ii) The effect of DNA methylation on transcription levels is mild, frequently escaping detection in low-powered studies. It is possible that DNA methylation affects transcription on a very fine scale, with meaningful impacts only for lowly-expressed genes. Such small changes could be undetectable in studies with low sensitivity due to a small number of biological replicates, or the signal could be overwhelmed by highly-expressed genes where DNA methylation has no substantial effect.
- (iii) The role of DNA methylation in the regulation of gene expression is complex, masking clear genome-wide patterns. The complexity can be manifold. DNA methylation of a regulatory element can prevent the binding of both transcriptional activators and repressors, potentially leading to silencing and upregulation, respectively [3]. Moreover, some TFs show affinity to methylated binding sites [9]. Consequently, high DNA methylation can be conducive to high levels of transcription in some genes, but low levels in others, and the same can be true for low DNA methylation. This is consistent with the DEG analysis of the Arabidopsis quintuple mutant, where demethylation causes nearly as many down-regulations as up-regulations [68]. However, the down-regulations (and up-regulations alike) could have a different proximal cause and could be associated with demethylation only indirectly. Another study of the Arabidopsis quintuple mutant [69] demonstrated that the complete loss of DNA methylation causes extensive switches of chromatin states. Since DNA methylation is interconnected with histone modifications [70, 71], which in turn

determine repressive and active chromatin states, the differential expression in the *mddcc* mutant could result from complex changes of histone modifications rather than directly from DNA demethylation.

- (iv) DNA methylation drives transcriptional changes dynamically in response to challenge (genomic or environmental stress), and the effect is invisible in a static sample. In this scenario, basic gene expression levels are generally determined by non-epigenetic regulatory networks, but are responsive to DNA methylation changes under stress conditions. This type of association can be explored by comparing sets of DEGs and DMRs in multiple-sample comparisons. We explored this type of interaction here by investigating whether polyploidization (presumably causing genomic stress) induces TSS methylation changes that are associated with transcriptional changes. However, we detected very few DMRs, and these were statistically unrelated to the sets of DEGs. A recent study [72] examined the question of DEG-DMR overlaps by assessing results of dozens of studies on a statistical basis. The comparisons included various experimental conditions, different tissues, developmental stages and genotypes. An expectation implicit in an integrated analysis of methylome and transcriptome is that an overlap of the DEG and DMR sets would reveal the genes that are transcriptionally regulated by DNA methylation. However, contrary to the expectation, the DEG-DMR overlaps in the majority of cases are statistically random, or even underrepresented. This analysis suggests that the role of DNA methylation in a dynamic regulation of gene expression (in response to developmental or environmental stimuli) is minor, and quite possibly misunderstood [72].

Changes in gene expression and TSS/SC methylation are independent in nascent wheat allohexaploids

Polyploidization in wheat triggers very few DNA methylation changes around TSS and SC. When comparing synthetics (including reciprocal crosses and two different generations) to their parents, we detected only 91–110 DMRs among ~ 93,000 genes with sufficient methylation data (considering 300 bp TSS/SC windows as regions). Most of the DMRs and DMCs were consistently found on the D subgenome, which is similar to the distribution of DEGs detected in the same samples [26]. This indicates that the D subgenome of nascent wheat allohexaploids is more vulnerable to both transcriptomic and methylomic changes. However, the DMRs and DEGs are not statistically related, since we found only 0–2 genes per synthetic

where differential expression is accompanied by differential methylation.

Homeolog expression differences in wheat are not driven by DNA methylation

Among the three subgenomes of natural and synthetic allohexaploid wheat, B and D are the most and the least methylated, respectively, with statistical significance. It is tempting to relate this observation to the subtle yet significant D-dominance in transcript abundance observed in the same samples [26], as well as the Chinese Spring reference genome [19]. Moreover, when assessing homeolog expression bias in Banouh et al. [26], we observed a statistically strong under-representation of D-suppressed triads, sometimes accompanied by an excess of B-suppressed triads. We have demonstrated that homeolog expression imbalance is mainly due to parental legacy, i.e. it is not a consequence of polyploidization but merely a perpetuation of pre-existing parental expression differences. Similarly, here we demonstrate that the subgenome methylation differences are not a consequence of polyploidization, since the A, B and D subgenomes are also methylated at different levels in the parental genotypes. It is interesting that the methylomes of the highly syntenic subgenomes were not homogenized after ~10,000 years of coexistence in natural bread wheat, or even after 0.5 million years in the case of the A and B methylomes in tetraploid wheat. It indicates a very faithful long-term maintenance of DNA methylation patterns. This observation, together with the below-average D-methylation, the subtle D-dominance and the dearth of D-suppressed triads suggest that homeolog expression bias could be driven by DNA methylation differences.

However, here we show that the relationship between DNA methylation and homeolog contribution to triad expression is either absent, or only weakly significant (Fig. 5a; Supplementary Fig. 8). When the homeolog contributions are regarded as a qualitative variable, the dominant and suppressed copies are statistically indistinguishable on the basis of TSS/SC methylation in most cases (Fig. 5c; Supplementary Fig. 10). Considering the duality of the possible effects of DNA methylation on transcription (leading to up- and down-regulation in different cases), we also checked the correlation between methylation and expression variance within triads (Fig. 5b; Supplementary Fig. 9). This comparison does not assume a dominant effect of methylation (be it gene activation or silencing), it merely checks whether triads with unequal homeolog contributions are also characterized with high variability of DNA methylation. This is certainly not the case for CpG methylation nor for CHH methylation. We detected significant, but very low positive correlation in the CHG context (proportion

of shared variance $r^2 \sim 0.003$), not indicative of a causal relationship.

Our conclusion that homeolog expression bias in bread wheat and analogical synthetics is not driven by DNA methylation is consistent with observations in more ancient polyploids. Zhao et al. [59] studied homeologs in the soybean genome that went through a polyploidization 5–13 million years (Mya) ago, and found that gene pairs with >2 fold expression differences are not significantly different in their patterns of DNA methylation, or the distribution and abundance of 24-nucleotide siRNAs. Furthermore, Parkin et al. [61] found that the patterns of cytosine methylation in paralogs on different subgenomes of *Brassica oleracea* (hexaploidization 23 Mya) do not correlate with individual gene dominance. In a more recent polyploid *Brassica napus* (allotetraploidization 7,500 years ago), homeologs with differential UTR or gene body methylation are equally expressed in over half of the cases [62]. In this species, homeolog expression and gene body methylation in the CpG and CHG context are not independent, with highly-methylated/lowly-expressed and lowly-methylated/highly-expressed homeologs in excess. However, the violation of independence is much weaker for UTR methylation (this was not shown in the original study, but a chi-square test can be performed on data in Supplementary Table 45 of Chalhoub et al. [62]). In the maize genome (polyploidization ~10 Mya), the more highly expressed gene in the homeolog pair typically exhibits a lower proportion of mCs up-stream of TSS and in gene bodies, and a more open chromatin [60]. Nonetheless, the authors emphasize that the attribution of cause and effect was impossible. On the other hand, a causal link between gbM and expression of homeologs is claimed by Wang et al. [73]. The authors studied a link between gbM divergence and expression difference of paralogs in rice (polyploidization 70 Mya), using *OsMet1-2* null mutant defective in CpG methylation maintenance. They reported that duplicated gene copies correlate with each other in their expression levels more strongly once CpG methylation is removed (a change from $r=0.43$ to $r=0.51$). However, while Wang et al. [73] focus on gbM, which they claim to be the cause of the differences in duplicate gene expression, CpG methylation in the *OsMet1-2* null mutant is lost globally, not just in gene bodies. It is therefore unclear whether the observed small reduction in expression differences is connected to demethylation of gene bodies, regulatory elements, TEs, or even to changes in chromatin states that are triggered by the global demethylation (as shown by Zhao et al. [69]).

Our results and the overview of different polyploid genomes suggest that the association of homeolog expression differences with DNA methylation is sometimes detectable globally, but is so weak that the

transcriptional divergence of most homeologs cannot be attributed to DNA methylation differences. What then is the main cause of homeolog expression bias is a curious question. Homeologs, especially those in allohexaploid wheat, have high sequence similarity, which suggests that they share the same regulatory networks in most cases. However, unbalanced triads indicate that some homeologs either do not share the same regulatory pathways, or their expression is modulated by other factors. Since we have shown that DNA methylation does not explain the imbalance in homeolog expression, the real cause may involve histone modifications, 3D-positioning of chromatin within the nucleus, or other factors yet to be determined.

Wheat genome unfazed by polyploidization

Polyploidizations are commonly believed to cause a 'genome shock', i.e. a widespread disturbance of gene regulation and genome integrity related to changes in gene dosage or parental incompatibilities [30, 74]. Accordingly, allopolyploidization in wheat has been associated with rampant transcriptomic changes [75–77], sequence loss [78, 79], activation of TEs [80] and epigenetic changes [36, 37]. However, a genome-wide analysis of TEs in the bread wheat genome found no signs of transpositional bursts after allohexaploidization [81]. Moreover, we have previously reported that only about 1% of genes significantly change transcription in nascent AABBDD allohexaploids, without a statistical relationship to adjacent TEs [26]. Here we reported that DNA methylation differences around TSS and SC are extremely rare (~0.1% of genes) in comparisons of early generations of wheat allohexaploids and their parents. These observations do not support the notion of a genome shock in polyploid wheat, and we consider several reasons why our results contradict earlier studies.

- (i) First of all, large over-estimations of transcriptomic response to polyploidization are often caused by technical difficulties associated with interploidy comparisons (see Supplementary Note in Banouh et al. [26]).
- (ii) Secondly, studies conducted prior to the assembly of the wheat genome provided only a partial view of the changes, without a genome-wide perspective. A hunt for changes overlooks the fact that the vast majority of the genome remains unchanged.
- (iii) Finally, it is very likely that the magnitude of changes depends on parental genotypes. Since there is some evidence that allopolyploidizations trigger much stronger genomic responses than autopolyploidizations [30], more diverged parental combinations can be expected to cause more changes than the less diverged ones. It may be

possible to find a particular combination of *Ae. tauschii* and tetraploid wheat (e.g. *T. dicoccoides*) that leads to a massive genomic reorganization, or one that is entirely incompatible. However, if the goal is to find out what actually happened in the genomic history of bread wheat, there is no reason to assume such parental incompatibilities. In our assessment, changes caused by allohexaploidization in wheat may not be qualitatively or quantitatively different from those triggered by intra-specific hybridizations.

Conclusions

In developing grain of natural and synthetic wheat allohexaploids, TSS methylation is only weakly associated with gene silencing. Differences in the expression levels of homeologs from the A, B and D subgenome are not driven by divergence in TSS/SC methylation. DNA methylation around TSS/SC is very stable across polyploidization events, and generally does not explain transcriptional differences observed between nascent wheat allohexaploids and their parents. The relevance of these findings is twofold. (i) The relationship between transcription and DNA methylation around TSSs remains unclear, and might be indirect, or affect only a minority of genes. (ii) Polyploidization in wheat does not have a strong effect on DNA methylation patterns, at least not in the vicinity of TSSs. The merger of *Triticum/Aegilops* genomes does not necessarily lead to extraordinary changes in gene regulation.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-026-08607-6>.

Additional file 1. Supplementary Tables 1-6 (xlsx format).

Additional file 2. Supplementary Figures 1-12 (pdf format).

Acknowledgements

We are thankful to Annaig Bouguennec for the production of wheat synthetics, and to Cecile Huneau for nucleic acid extractions. A part of the experiments (capture-BSseq library preparation) was performed at the PGTB (doi:10.15454/1.5572396583599417E12), with the help of Christophe Boury and Erwan Guichoux. We are grateful to the Mésocentre Clermont-Auvergne and AuBi (plateforme Auvergne Bioinformatique) platform of the Université Clermont Auvergne for their support, computing and storage resources.

Authors' contributions

PC, JS and CP conceived the project and secured funding. MB performed data analyses and wrote an initial draft of the manuscript. MDS and MAS assisted with bioinformatics. PC supervised the analyses. All authors read and finalized the manuscript.

Funding

The authors acknowledge the support received from the International Research Center on Sustainable AgroEcosystems (SITE CAP20-25), and from the INRAE BAP department. This work is part of project ADAAPT of the Agroecology and Digital Technologies research program and received government funding managed by the Agence Nationale de la Recherche under the France 2030 program, reference ANR-24-PEAE-0001.

Data availability

The capture-BSeq data produced and analyzed for the purpose of this study are available at the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under the BioProject number PRJNA1247774.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹INRAE/UCA UMR 1095. 5 Chemin de Beaulieu, Clermont Ferrand 63100, France

²Bundeswehr Institute of Microbiology (IMB), Munich 80937, Germany

Received: 16 December 2025 / Accepted: 17 March 2026

Published online: 20 March 2026

References

- Galindo-González L, Sarmiento F, Quimbaya MA. Shaping Plant Adaptability, Genome Structure and Gene Expression through Transposable Element Epigenetic Control: Focus on Methylation. *Agronomy*. 2018;8:9.
- Moore LD, Le T, Fan G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology*. 2013;38:23–38.
- Zhang H, Zhu J-K. Epigenetic gene regulation in plants and its potential applications in crop improvement. *Nat Rev Mol Cell Biol*. 2024;26:51–67.
- Gu C, Pei MS, Guo ZH, Wu L, Qi KJ, Wang XP, et al. Multi-omics provide insights into the regulation of DNA methylation in pear fruit metabolism. *Genome Biol*. 2024;25:70.
- Downen RH, Pellizzola M, Schmitz RJ, Lister R, Downen JM, Nery JR et al. Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences*. 2012;109:E2183–E2191.
- Zheng G, Dong X, Wei J, Liu Z, Aslam A, Cui J, et al. Integrated methylome and transcriptome analysis unravel the cold tolerance mechanism in winter rapeseed (*Brassica napus* L). *BMC Plant Biol*. 2022;22:414.
- Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, et al. The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell*. 2014;26:2351–66.
- Jamge B, Lorković ZJ, Axelsson E, Osakabe A, Shukla V, Yelagandula R, et al. Histone variants shape chromatin states in Arabidopsis. *Elife*. 2023;12:RP87714.
- O'Malley RC, Huang SSC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. Cistrome and episcistrome features shape the regulatory DNA landscape. *Cell*. 2016;165:1280–92.
- Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, et al. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol*. 2016;17:1–19.
- Zeng Y, Somers J, Bell HS, Vejlupekova Z, Kelly Dawe R, Fowler JE, et al. Potent pollen gene regulation by DNA glycosylases in maize. *Nat Commun*. 2024;15:8352.
- Bewick AJ, Ji L, Niederhuth CE, Willing EM, Hofmeister BT, Shi X, et al. On the origin and evolutionary consequences of gene body DNA methylation. *Proc Natl Acad Sci*. 2016;113:9111–16.
- Muyle A, Gaut BS. Loss of Gene Body Methylation in *Eutrema salsugineum* Is Associated with Reduced Gene Expression. *Mol Biol Evol*. 2019;36:155–8.
- Zhu T, Wang L, Rimbart H, Rodríguez JC, Deal KR, De Oliveira R, et al. Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *Plant J*. 2021;107:303–14.
- Mcfadden ES, Sears ER. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J Hered*. 1946;37:81–9.
- Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, International Wheat Genome Sequencing Consortium. Ancient hybridizations among the ancestral genomes of bread wheat. *Science*. 2014;345:1250092.
- Alaux M, Rogers J, Letellier T, Flores R, Alfama F, Pommier C, et al. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biol*. 2018;19:1–10.
- Gardiner L-J, Quinton-Tulloch M, Olohan L, Price J, Hall N, Hall A. A genome-wide survey of DNA methylation in hexaploid wheat. *Genome Biol*. 2015;16:273.
- Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, et al. The transcriptional landscape of polyploid wheat. *Science*. 2018;361:eaar6089.
- Li Z, Wang M, Lin K, Xie Y, Guo J, Ye L, et al. The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. *Genome Biol*. 2019;20:139.
- Kawakatsu T, Ecker JR. Diversity and dynamics of DNA methylation: epigenomic resources and tools for crop breeding. *Breed Sci*. 2019;69:191–204.
- Varotto S, Krugman T, Aiese Cigliano R, Kashkush K, Kondić-Špika A, Aravanopoulos FA, et al. Exploitation of epigenetic variation of crop wild relatives for crop improvement and agrobiodiversity preservation. *Theor Appl Genet*. 2022;135:3987–4003.
- Singh D, Chaudhary P, Taunk J, Kumar Singh C, Sharma S, Singh VJ, et al. Plant epigenomics for extenuation of abiotic stresses: challenges and future perspectives. *J Exp Bot*. 2021;72:6836–55.
- Mutti JS, Bhullar RK, Gill KS. Evolution of Gene Expression Balance Among Homeologs of Natural Polyploids. *G3 Genes|Genomes|Genetics*. 2017;7:1225–37.
- Takahagi K, Inoue K, Mochida K. Gene Co-expression Network Analysis Suggests the Existence of Transcriptional Modules Containing a High Proportion of Transcriptionally Differentiated Homoeologs in Hexaploid Wheat. *Front Plant Sci*. 2018;9:1163.
- Banouh M, Armisen D, Bouguennec A, Huneau C, Sow MD, Pont C, et al. Low impact of polyploidization on the transcriptome of synthetic allohexaploid wheat. *BMC Genomics*. 2023;24:255.
- Chen ZJ. Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci*. 2010;15:57–71.
- Takahagi K, Inoue K, Mochida K. Gene co-expression network analysis suggests the existence of transcriptional modules containing a high proportion of transcriptionally differentiated homoeologs in hexaploid wheat. *Front Plant Sci*. 2018;9:1163.
- Qiu T, Liu Z, Liu B. The effects of hybridization and genome doubling in plant evolution via allopolyploidy. *Mol Biol Rep*. 2020;47:5549–58.
- Tayalé A, Parisod C. Natural Pathways to Polyploidy in Plants and Consequences for Genome Reorganization. *Cytogenet Genome Res*. 2013;140:79–96.
- Matzke MA, Kanno T, Matzke AJM. RNA-Directed DNA Methylation. The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annu Rev Plant Biol*. 2015;66:243–67.
- Li Q, Qiao X, Yin H, Zhou Y, Dong H, Qi K, et al. Unbiased subgenome evolution following a recent whole-genome duplication in pear (*Pyrus bretschneideri* Rehder). *Hortic Res*. 2019;6:34.
- Rao X, Ren J, Wang W, Chen R, Xie Q, Xu Y, et al. Comparative DNA-methylome and transcriptome analysis reveals heterosis- and polyploidy-associated epigenetic changes in rice. *Crop J*. 2023;11:427–37.
- Jiang X, Song Q, Ye W, Chen ZJ. Concerted genomic and epigenomic changes accompany stabilization of Arabidopsis allopolyploids. *Nat Ecol Evol*. 2021;5:1382–93.
- Xiang X, Gao Y, Cui J, Ren G, Yin C, Chang J. Methylome and transcriptome analysis of alters leaf phenotype with autotetraploid in grape. *Sci Hort*. 2023;307:111534.
- Yaakov B, Kashkush K. Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. *Genome*. 2011;54:42–9.
- Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K. Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. *Genetics*. 2010;186:801–12.
- Yuan J, Jiao W, Liu Y, Ye W, Wang X, Liu B, et al. Dynamic and reversible DNA methylation changes induced by genome separation and merger of polyploid wheat. *BMC Biol*. 2020;18:1–13.

39. Miao L, Xu W, Liu Y, Huang X, Chen Z, Wang H, et al. Reshaped DNA methylation cooperating with homoeolog-divergent expression promotes improved root traits in synthesized tetraploid wheat. *New Phytol.* 2024;242:507–23.
40. Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature.* 2020;588:277–83.
41. Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, et al. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet.* 2019;51:885–95.
42. Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science.* 2017;357:93–7.
43. Luo M-C, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, et al. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature.* 2017;551:498–502.
44. Zhao G, Zou C, Li K, Wang K, Li T, Gao L, et al. The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat Plants.* 2017;3:946–55.
45. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013. <https://doi.org/10.48550/arXiv.1303.3997>.
46. Geneious. 6.1. <https://www.geneious.com>
47. Chen Y, Song W, Xie X, Wang Z, Guan P, Peng H, et al. A Collinearity-Incorporating Homology Inference Strategy for Connecting Emerging Assemblies in the Triticeae Tribe as a Pilot Practice in the Plant Pangenomic Era. *Mol Plant.* 2020;13:1694–708.
48. Triticeae-Gene Tribe. <http://wheat.cau.edu.cn/TGT/>. Accessed 2024.
49. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2019;27:1571–2.
50. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 2012;13:1–9.
51. Schultz MD, Schmitz RJ, Ecker JR. Leveling the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet.* 2012;28:583–5.
52. Akalin A, Franke V, Vlahoviček K, Mason CE, Schübeler D. Genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics.* 2015;31:1127–9.
53. Zhao J, Xie Y, Kong C, Lu Z, Jia H, Ma Z, et al. Centromere repositioning and shifts in wheat evolution. *Plant Commun.* 2023;4:100556.
54. Deniz Ö, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. *Nat Rev Genet.* 2019;20:417–31.
55. Das PM, Singal R. DNA Methylation and Cancer. *J Clin Oncol.* 2004;22:4632–42.
56. Bernstein C. Epigenetic field defects in progression to cancer. *World J Gastrointest Oncol.* 2013;5:43.
57. Chow JC, Yen Z, Ziesche SM, Brown CJ. Silencing of the mammalian X chromosome. *Annu Rev Genom Hum Genet.* 2005;6:69–92.
58. Wang Y, Wang X, Lee TH, Mansoor S, Paterson AH. Gene body methylation shows distinct patterns associated with different gene origins and duplication modes and has a heterogeneous relationship with gene expression in *Oryza sativa* (rice). *New Phytol.* 2013;198:274–83.
59. Zhao M, Zhang B, Lisch D, Ma J. Patterns and Consequences of Subgenome Differentiation Provide Insights into the Nature of Paleopolyploidy in Plants. *Plant Cell.* 2017;29:2974–94.
60. Renny-Byfield S, Rodgers-Melnick E, Ross-Ibarra J. Gene fractionation and function in the ancient subgenomes of maize. *Mol Biol Evol.* 2017;34:1825–32.
61. Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, et al. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol.* 2014;15:1–18.
62. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science.* 2014;345:950–3.
63. Bellec A, Sow MD, Pont C, Civan P, Mardoc E, Duchemin W, et al. Tracing 100 million years of grass genome evolutionary plasticity. *Plant J.* 2023;114:1243–66.
64. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature.* 2017;543:72–7.
65. Wu F, Li X, Looso M, Liu H, Ding D, Günther S, et al. Spurious transcription causing innate immune responses is prevented by 5-hydroxymethylcytosine. *Nat Genet.* 2022;55:100–11.
66. Liang L, Chang Y, Lu J, Wu X, Liu Q, Zhang W, et al. Global Methylomic and Transcriptomic Analyses Reveal the Broad Participation of DNA Methylation in Daily Gene Expression Regulation of *Populus trichocarpa*. *Front Plant Sci.* 2019;10:243.
67. Medvedeva YA, Khamis AM, Kulakovskiy IV, Ba-Alawi W, Bhuyan MSI, Kawaji H, et al. Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics.* 2014;15:1–12.
68. He L, Huang H, Bradai M, Zhao C, You Y, Ma J, et al. DNA methylation-free *Arabidopsis* reveals crucial roles of DNA methylation in regulating gene expression and development. *Nat Commun.* 2022;13:1335.
69. Zhao L, Zhou Q, He L, Deng L, Lozano-Duran R, Li G, et al. DNA methylation underpins the epigenomic landscape regulating genome transcription in *Arabidopsis*. *Genome Biol.* 2022;23:197.
70. Wang J, Meng X, Yuan C, Harrison AP, Chen M. The roles of cross-talk epigenetic patterns in *Arabidopsis thaliana*. *Brief Funct Genomics.* 2016;15:278–87.
71. Nakamura K, Yamaguchi N, Ito T. The histone crosstalk code in plants: Deciphering epigenetic complexity. *Curr Opin Plant Biol.* 2025;87:102763.
72. Civan P, Sammarco I, Banouh M. The mirage of DNA methylation in transcriptional regulation of plants. *Plant Genome.* 2026;19:e70208.
73. Wang X, Zhang Z, Fu T, Hu L, Xu C, Gong L, et al. Gene-body CG methylation and divergent expression of duplicate genes in rice. *Sci Rep.* 2017;7:2675.
74. McClintock B. The Significance of Responses of the Genome to Challenge. *Science.* 1984;226:792–801.
75. Zhang H, Gou X, Zhang A, Wang X, Zhao N, Dong Y, et al. Transcriptome shock invokes disruption of parental expression-conserved genes in tetraploid wheat. *Sci Rep.* 2016;6:26363.
76. Hao M, Li A, Shi T, Luo J, Zhang L, Zhang X, et al. The abundance of homoeologue transcripts is disrupted by hybridization and is partially restored by genome doubling in synthetic hexaploid wheat. *BMC Genomics.* 2017;18:149.
77. Vasudevan A, Lévesque-Lemay M, Edwards T, Cloutier S. Global transcriptome analysis of allopolyploidization reveals large-scale repression of the D-subgenome in synthetic hexaploid wheat. *Commun Biology.* 2023;6:426.
78. Ozkan H, Levy AA, Feldman M. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell.* 2001;13:1735–47.
79. Ozkan H. Nonadditive changes in genome size during allopolyploidization in the wheat (*Aegilops-Triticum*) group. *J Hered.* 2003;94:260–4.
80. Kashkush K, Feldman M, Levy AA. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genet.* 2002;33:102–6.
81. Papon N, Lasserre-Zuber P, Rimbert H, De Oliveira R, Paux E, Choulet F. All families of transposable elements were active in the recent wheat genome evolution and polyploidy had no impact on their activity. *Plant Genome.* 2023;16:e20347.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.