HAL open science

New insights into homoeologous copy number variations in the hexaploid wheat genome

Caroline Juery, Lorenzo Concia, Romain de Oliveira, Nathan Papon, Ricardo Ramírez-González, Moussa Benhamed, Cristobal Uauy, Frédéric Choulet, Etienne Paux

To cite this version:

Caroline Juery, Lorenzo Concia, Romain de Oliveira, Nathan Papon, Ricardo Ramírez-González, et al.. New insights into homoeologous copy number variations in the hexaploid wheat genome. Plant Genome, 2021, 14 (1), pp.e20069. 10.1002/tpg2.20069 . hal-03084292

HAL Id: hal-03084292

https://hal.inrae.fr/hal-03084292v1

Submitted on 4 Jan 2021

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 Attribution 4.0 International License

Received: 9 September 2020 | Accepted: 11 October 2020

DOI: 10.1002/tpg2.20069

The Plant Genome

ORIGINAL RESEARCH

New insights into homoeologous copy number variations in the hexaploid wheat genome

Caroline Juery[1] | **Lorenzo Concia**[2,4] | **Romain De Oliveira**[1] | **Nathan Papon**[1] | **Ricardo Ramírez-González**[3] | **Moussa Benhamed**[2] | **Cristobal Uauy**[3] | **Frédéric Choulet**[1] | **Etienne Paux**[1]

[1] Université Clermont Auvergne, INRAE, GDEC, Clermont-Ferrand 63000, France

[2] Institute of Plant Sciences Paris-Saclay (IPS2), UMR 9213/UMR1403, CNRS, INRA, Université Paris-Sud, Université d'Evry, Université Paris-Diderot, Sorbonne Paris-Cité, Orsay 91405, France

[3] John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

[4] Current address: Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, Université PSL, Paris 75005, France

Correspondence
Etienne Paux, Université Clermont Auvergne, INRAE, GDEC, 63000 Clermont-Ferrand, France.
Email: etienne.paux@inrae.fr

Funding information
European Regional Development Fund, Grant/Award Number: SRESRI 2016

Abstract

Bread wheat is an allohexaploid species originating from two successive and recent rounds of hybridization between three diploid species that were very similar in terms of chromosome number, genome size, TE content, gene content and synteny. As a result, it has long been considered that most of the genes were in three pairs of homoeologous copies. However, these so-called triads represent only one half of wheat genes, while the remaining half belong to homoeologous groups with various number of copies across subgenomes. In this study, we examined and compared the distribution, conservation, function, expression and epigenetic profiles of triads with homoeologous groups having undergone a deletion (dyads) or a duplication (tetrads) in one subgenome. We show that dyads and tetrads are mostly located in distal regions and have lower expression level and breadth than triads. Moreover, they are enriched in functions related to adaptation and more associated with the repressive H3K27me3 modification. Altogether, these results suggest that triads mainly correspond to housekeeping genes and are part of the core genome, while dyads and tetrads belong to the *Triticeae* dispensable genome. In addition, by comparing the different categories of dyads and tetrads, we hypothesize that, unlike most of the allopolyploid species, subgenome dominance and biased fractionation are absent in hexaploid wheat. Differences observed between the three subgenomes are more likely related to two successive and ongoing waves of post-polyploid diploidization, that had impacted A and B more significantly than D, as a result of the evolutionary history of hexaploid wheat.

1 | INTRODUCTION

Polyploidy, or whole genome duplication (WGD), has been observed in many living organisms, both prokaryotic and eukaryotic, and is widely recognized as a key driver of species evolution, diversification, as well as domestication

Abbreviations: GO, Gene ontology; HC, High confidence; HomoeoCNV, Homoeologous copy number variation; IWGSC, International Wheat Genome Sequencing Consortium; LC, Low confidence; PAV, Presence absence variation; PPD, Post-polyploid diploidization; TE, Transposable element; TPM, Transcripts per million; WGD, Whole genome duplication.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. *The Plant Genome* published by Wiley Periodicals LLC on behalf of Crop Science Society of America

(Van de Peer, Mizrachi, & Marchal, 2017; Wendel, 2015). Polyploid species can be classified into two different categories: autopolyploids, which arise from genome doubling within one species, and allopolyploids, which arise from genome doubling following hybridization between two distinct species (Glover, Redestig, & Dessimoz, 2016). About one half of angiosperms are recent polyploids, including numerous important crop species such as oilseed rape (*Brassica napus*, 7500 years old), coffee (*Coffea arabica*, 10,000–50,000 years old) and wheat (*Triticum aestivum*, 10,000 years old). In addition, due to recurrent polyploidization events that occured through time, including the ζ (zeta) WGD event 300–350 million years ago (MYA), all flowering plants are ancient polyploids or paleopolyploids (Jiao et al., 2011). Well-characterized examples include sorghum (*Sorghum bicolor*, 95–115 MYA), maize (*Zea mays*, 26 MYA) and soybean (*Glycine max*, 13 MYA) (Qiao et al., 2019).

Each WGD event results in a doubling of the gene content. Salman-Minkov, Sabath, and Mayrose (2016) demonstrated that polyploidization gives fitness advantages through increasing the amount of raw genetic material on which natural and artificial selection can happen. However, despite the successive episodes of WGDs that occurred through time, the number of genes in plants is quite similar (Michael & Jackson, 2013) and far less than that expected by the doubling process (Adams & Wendel, 2005). This leads to the paradox that while being an important evolutionary process, polyploidy also seems to be an evolutionary 'dead-end' (Van de Peer et al., 2017) as polyploids systematically tend to return to a diploid state after a few million years (Wendel, 2015). Hence, duplicated genes can either be pseudogenized, silenced, and eventually lost or, alternatively, retained because having evolved a new function (neofunctionalization) or having diverged in expression (subfunctionalization) (Flagel & Wendel, 2009).

Previous studies on homoeologous gene loss and retention, as well as relative expression contribution in various polyploid species, revealed species-specific patterns, suggesting an effect of the age of the polyploidization and diploid progenitor divergence (Bottani, Zabet, Wendel, & Veitia, 2018). For example, in the paleopolyploid maize genome, 14% of coding sequences were lost during the diploidization process, with a 25% of differential loss between the two genomes and a biased fractionation (loss of functioning DNA sequence) in favor of one subgenome that exhibits an overall higher expression and higher impact on phenotypic variability (Jiao et al., 2017; Renny-Byfield, Rodgers-Melnick, & Ross-Ibarra, 2017). Similarly, in the ancient allotetraploid cotton genome, a biased fractionation was observed, with the A genome showing more gene loss, a faster evolution rate, and an overall lower expression level that the D genome (Zhang et al., 2015). In contrast, while frequent homoeolog sequence exchanges have been reported, no significant bias toward either subgenome was observed in the recent allotetraploid oilseed rape (*Brassica napus*) (Chalhoub et al., 2014). Therefore, for closely related progenitor genomes, like in soybean, a dosage sensitive pattern of expression leads to stochastic differentiation of homoeologous pairs. For highly divergent progenitor genomes, like maize, the more favorable homoeologous genes set of a subgenome are selected, leading to an overall subgenome retention and a biased fractionation.

Core Ideas

- Only one half of hexaploid wheat genes are in triads, i.e. in a 1:1:1 ratio across subgenomes
- Triads are likely part of the core genome; dyads and tetrads belong to the dispensable genome
- Subgenome dominance and biased fractionation are absent in hexaploid wheat
- Subgenome differences are related to two successive waves of post-polyploid diploidization

Bread wheat (*Triticum aestivum* L.) is an allohexaploid species (2n = 6X = AABBDD) originating from two successive rounds of hybridization (IWGSC, 2014; Marcussen et al., 2014). The first hybridization event occurred ~800,000 years ago between *Triticum urartu* (AA-genome) and an unknown *Aegilops* species (BB-genome). The second event took place ~10,000 years ago between *Triticum turgidum* (AABB-genome) and *Aegilops tauschii* (DD-genome). The resulting hexaploid AABBDD-genome was estimated to carry 107,891 high confidence (HC) protein-coding genes, although 161,537 low confidence (LC) genes and 303,818 pseudogenes and gene fragments were also annotated (IWGSC, 2018). Using a phylogenomics approach on a filtered set of 181,036 genes, 21,603 triads, defined as homoeologous genes that had a strict 1:1:1 correspondence (one copy per subgenome A, B and D), were identified. These account for only 36% of the gene set (64,809 genes) while the remaining 64% have a more complex homoeologous relationships (1:1:N or 0:1:1 for example). Similar proportions of genes in different homoeology contexts were observed on each of the subgenome. Together with equal contribution of the three homoeologous genomes to the overall gene expression, this supported the hypothesis of the absence of biased fractionation and global subgenome dominance (IWGSC, 2018). However, a cell type- and stage-dependent local subgenome dominance was observed (Harper et al., 2016; Pfeifer et al., 2014). A recent study reported that the vast majority of triads displayed a balanced contribution of each copy to the overall expression of the homoeologous group (Ramirez-Gonzalez et al., 2018). For those showing either dominance or suppression of one homoeologous copy, differences were associated with epigenetic changes,

especially in H3K9ac and H3K27me3 patterns. Such differences in gene expression likely represent the first steps toward neo- or subfunctionalization of wheat homoeologs.

Previous studies focused mainly on 1:1:1 triads leaving two third of the wheat genes apart. However, triads likely correspond to highly conserved and evolutionary constrained genes. In this regard, they may not be fully representative of the entire gene set and may not illustrate the complexity of the evolutionary trajectories that occurred within the hexaploid wheat genome. Here, we report on this unexplored part of the wheat genome by integrating not only triads but also dyads and tetrads, i.e. homoeologous groups that have undergone a single gene loss or duplication event, respectively. By combining genomic, transcriptomic and epigenetic data, we show that these two latter categories differ from triads not only by their chromosomal distribution but also by their transcriptional and epigenetic patterns as well as their conservation in wheat and other plant genomes, suggesting different evolutionary fates depending on the copy number of homoeologous genes.

2 | MATERIALS AND METHODS

2.1 | Definition and distribution of homoeologous group

Dyad, triad and tetrad gene information were retrieved from IWGSC (2018). Groups containing both high-confidence (HC) and low-confidence (LC) genes were filtered out to keep only those with HC genes. These data included gene position on the IWGSC RefSeq v1.0 reference sequence, homoeologous group category (dyad, triad or tetrad) and ID, as well as orthologous relationships with *Arabidopsis thaliana*, *Zea mays*, *Sorghum bicolor*, *Oryza sativa*, *Brachypodium distachyon* and *Hordeum vulgare* genes. The chromosome distribution of genes was performed by calculating the proportion of dyad, triad and tetrad genes over the total number of genes from this study within each of the five chromosomal regions defined by Pingault et al. (2015). Duplicated genes separated by less than 10 genes and less than 1 Mb on chromosomes were considered as tandem duplications. The other ones were considered as dispersed duplications.

2.2 | Characterization of ancestral duplications/deletions and presence-absence variations

To assess if deletions in dyads occurred within diploids or upon polyploidization, we aligned the two remaining copies onto diploid and tetraploid ancestor genomes using the GMAP package v2019.03.04 (Wu & Watanabe, 2005) with 85% of sequence identity and 85% of sequence coverage as parameters. For AB-dyads, we mined the *Aegilops tauschii* genome (Luo et al., 2017) with A and B coding sequences. For AD-dyads, we mined the B-genome of *Triticum dicoccoides* (Avni et al., 2017) with A and D coding sequences. For BD dyads, we mined the *Triticum urartu* genome (Ling et al., 2018), as well as the A-genome of *Triticum dicoccoides* with B and D sequences. To take into account polymorphisms between individual sequences (divergence, presence/absence variations…), we corrected these numbers by dividing them by the number of genes that are still present in the hexaploid wheat genome and that were found in the ancestral genomes (*e.g.* D-copy from an AD dyad in the *Ae. tauschii* genome). To assess if duplications in tetrads occurred within diploids or upon polyploidization, we used GMAP to estimate the number of copies in the diploid and tetraploid ancestor genomes. Presence/absence variation (PAV) analysis was performed as described by De Oliveira et al. (2020). Briefly, sequencing reads from 16 wheat accessions (Montenegro et al., 2017) were mapped on the IWGSC RefSeq v1.0 using BWA-MEM v0.7.12 (Li & Durbin, 2010). Alignments were then filtered using samtools view (samtools view -F 2308 -q11; Li et al., 2009) and PCR duplicates were removed using samtools rmdup. Depth of coverage was assessed using bedtools coverage (v2.26; Quinlan & Hall, 2010). Genes were considered as putative PAVs when their coding sequence was covered over less than 10% of their length in at least two accessions.

2.3 | Gene ontology enrichment and functional analysis

Gene Ontology (GO) terms and functional annotation data were retrieved from IWGSC (2018). GO enrichment analysis was conducted using R package *topGO* (Alexan & Rahnenfuhrer, 2019).

2.4 | Gene expression and relative contribution analysis

Expression data from 15 samples representing five different organs (root, leaf, stem, spike and grain) at three developmental stages each in controlled non-stressed conditions were retrieved from Ramirez-Gonzalez et al. (2018). Genes with expression levels below 0.5 TPM were considered as non-expressed. Outlier identification was conducted in R (R Core Team, 2014) using the *boxplot* function of the *ggplot2* package (Villanueva & Chen, 2019). They were defined as genes outside 1.5 times the interquartile range (IQR) above the third quartile ($Q3 + 1.5 \times IQR$) and below the first quartile ($Q1 - 1.5 \times IQR$).

For relative contribution analyses, we used the calculation method described by Ramirez-Gonzalez et al. (2018).

Briefly, to standardize the relative expression of each homoeolog across a group, we normalized the absolute TPM for each gene within this group, as follows:

$$\text{Expression of A} - \text{copy in an AB dyad, expression}_A = \frac{\text{TPM (A)}}{\text{TPM (A)} + \text{TPM (B)}}$$

$$\text{Expression of A} - \text{copy in an ABD triad, expression}_A = \frac{\text{TPM (A)}}{\text{TPM (A)} + \text{TPM (B)} + \text{TPM(D)}}$$

$$\text{Expression of A} - \text{copy in an ABDD tetrad, expression}_A = \frac{\text{TPM (A)}}{\text{TPM (A)} + \text{TPM (B)} + \text{TPM}\left(\text{D}_1\right) + \text{TPM}\left(\text{D}_2\right)}$$

The normalized expression was calculated for the average across all expressed tissues as well as for each tissue individually. In order to assign theoretical expression bias categories to each group within triad, tetrad and dyad, we constructed theoretical matrix (Supplemental Table S1). We calculated the Euclidean distance with the *rdist* function from R from the observed normalized expression of each group to each of the ideal categories. We assigned the homoeolog expression bias category for each group by selecting the shortest distance between theoretical and observed relative contribution values. For binary organ expression, genes expressed in an organ ($\geq$ 0.5 TPM) were given a value of 1 and those not expressed (<0.5 TPM), 0. This resulted in 32 binary expression profiles (0-0-0-0-0, 0-0-0-0-1, 0-0-0-1-1…).

2.5 | Histone mark analysis

Wheat H3K9ac and H3K27me3 data from bread wheat cultivar Chinese Spring at three-leaf stage were retrieved from IWGSC (2018). Genes were assigned a histone mark category, either H3K9ac, H3K27me3, both or no mark. We calculated meta-gene profiles for each category by computing the read density of each histone mark over different categories using Deeptools (Ramirez et al., 2016) computeMatrix scale-regions and plotted it with plotProfile. Only reads mapping within gene bodies plus 1 kb upstream of the transcription start site and 1 kb downstream of the transcription end site were considered. To account for different gene size, we divided the read counts over each gene by its length. H3K9ac and H3K27me3 data from *Oryza sativa* and *Zea mays* were retrieved from the Plant Chromatin State Database (Liu et al., 2018).

3 | RESULTS

3.1 | Defining homoeologous groups

To decipher the impact of gene loss and duplication in the wheat genome, we focused our study on high confidence (HC) genes from the IWGSC RefSeq v1.0. These 107,891 genes were previously assigned an homoeologous group defined through an iterative phylogenomic approach and for all of these groups, the cardinality was determined based on the number of homoeologs identified on each sub-genome (IWGSC, 2018). Using these data, we found 55,170 homoeologs (51.1%) belonging to 18,390 triads (Table 1; Supplemental Table S2; Supplemental Data S1). It is worth noting that an additional 2,218 triads containing both HC and LC genes were found. However, since LC genes corresponded to partially supported gene models, we did not include these triads in our analysis. We also found 12,640 genes (11.7%) corresponding to 6,320 groups having undergone a single gene loss during the course of evolution since the divergence of the A, B, and D subgenomes. The corresponding groups will be hereafter referred to as AB, AD and BD dyads, i.e. groups of HC genes being in 1:1:0, 1:0:1 or 0:1:1: ratios across homoeolog genomes, respectively. Finally, we identified 3,008 genes (2.8%) belonging to 240 AABD, 315 ABBD and 197 ABDD tetrads, i.e. groups having undergone one single gene duplication thus being in 2:1:1, 1:2:1 or 1:1:2 ratios, respectively. Similar to triads, 2,085 and 658 additional dyads and tetrads containing both HC and LC genes were found but not selected for further analyses. While in the present study we will focus only on dyads, triads and tetrads (70,818 genes), it is worth noting that 37,073 HC genes (34.4%) depart from these ratios, corresponding to genes that have undergone more than one deletion or duplication or that were not clustered into a homoeolog group. The overall high proportion of genes that are not in a strict 1:1:1 ratio, hereafter referred to as genes affected by homoeologous copy number variations (HomoeoCNVs), represent the dynamic part of the wheat genome during the course of its evolution either in the ancestral diploid species or after polyploidization.

3.2 | Conservation of homoeoCNVs

To assess whether dyad genes were lost upon or after polyploidization or were already missing in the progenitor genomes, we searched for orthologs in diploid and tetraploid genomes: *Triticum urartu* (AA-genome), *Triticum dicoccoides* (AABB-genome) and *Aegilops tauschii* (DD-genome). For the A-missing copies (i.e. BD dyads), we estimated that approximately 40.7% and 19.0% were still present in diploid and tetraploid ancestor genomes, respectively (Supplemental

TABLE 1 Number of groups and genes in the dyad, triad and tetrad categories

	Dyads			Triads	Tetrads			Total
	AB	AD	BD	ABD	AABD	ABBD	ABDD	–
Number of groups	1,776	2,253	2,291	18,390	240	315	297	25,462
	6,320			18,390	752			25,642
Number of genes	3,552	4,506	4,582	55,170	960	1,260	788	70,818
	12,640			55,170	3,008			70,818

Data S1). For the B-missing (i.e. AD dyads) and D-missing ones (i.e. AB dyads), the estimates were of 20.6% and 27.7% still present in tetraploid and DD-diploid ancestor genomes, respectively. These results revealed that most of the genes of the dyad category were already absent from the diploid and tetraploid progenitors and that roughly 450 genes were lost on each subgenome at each step of polyploidization (498 A genes from *T. urartu* to *T. dicoccoides*; 434 A genes from *T. dicoccoides* to *T. aestivum*; 465 B genes from *T. dicoccoides* to *T. aestivum*; 493 D genes from *Ae. tauschii* to *T. aestivum*). Similarly, for tetrads, the majority of genes duplicated in the hexaploid wheat genome were also found in two copies in the diploid or tetraploid genomes. Indeed, 65.8% and 76.3% of A-duplicates were found in two copies in *T. urartu* and *T. dicoccoides*, respectively, 67.3% of B-duplicates in *T. dicoccoides* and 83.2% of D-duplicates in *Ae. tauschii*. Overall, we estimated that 172 genes were duplicated upon hexaploidization, 29.1% on the A-genome, 52.9% on the B-genome and 18.0% on the D-genome. However, one cannot exclude that the absence of a gene is due to an intraspecific polymorphism.

To investigate the conservation of dyad, triad and tetrad genes in other bread wheat accessions, we mined for presence-absence variations (PAVs) of genes in the genome of 16 resequenced wheat accessions (Montenegro et al., 2017). Out of the 70,818 genes, we identified 2,270 putative PAVs representing 3.2% of the dataset (Supplemental Data S1). Consistent with the percentage of genes duplicated upon hexaploidization, the B-subgenome appeared to be more subject to variations (47.0% of all PAVs) than the A- and D-subgenomes (30.9% and 22.1%, respectively). When analyzing each category of homoeogroups individually, only 1.8% of triad genes were affected by PAVs whereas they accounted for 7.9% and 9.7% of dyads and tetrads, respectively. Interestingly, for tetrads, 74.1% of PAVs appeared to affect one of the duplicated homoeologs.

Finally, to look at the conservation of dyad, triad and tetrad genes in other plants, we used the orthologous relationships with *Arabidopsis thaliana*, *Sorghum bicolor*, *Zea mays*, *Oryza sativa*, *Brachypodium distachyon* and *Hordeum vulgare* determined by the IWGSC (2018). The overall percentage of orthologs found for our 25,462 groups ranged from 52.8% in *A. thaliana* to 75.0% in *B. distachyon* (Supplemental Table S2). These proportions were consistent with the phylogenetic distance, the most distant species sharing the lowest number of orthologs, with the notable exception of barley, consistent with the a lower BUSCO score indicating the completeness of genome assembly, gene set and transcriptome calculated by the IWGSC (2018). When analyzing each category of homoeogroups separately, triad genes were found to be the most conserved (Figure 1). Indeed, the proportion of orthologous genes ranged from 58.4% in *A. thaliana* to 82.1% in *B. distachyon.* In tetrads, this proportion ranged from 37.6 to 56.4%. The least conserved genes were dyad ones with 32.0% of orthologs in *A. thaliana* and 48.3% in *B. distachyon.*

3.3 | Distribution of homoeoCNVs along wheat chromosomes

Previous studies revealed a partitioning of the wheat genome based on different structural and functional features, including the recombination rate, gene and transposable element (TE) densities, gene expression breadth, histone modifications, as well as gene and TE structural variation rate (Choulet et al., 2014; De Oliveira et al., 2020; IWGSC, 2018; Pingault et al., 2015). Consequently, chromosomes can be divided into five chromosomal compartments: the short arm distal R1, the short arm proximal R2a, the centromeric-pericentromeric C, the long arm proximal R2b and the long arm distal R3 regions. To investigate the distribution of HomoeoCNVs in the light of chromosome partitioning, we analyzed the proportions of each category (dyads, triads, and tetrads) in the proximal (R2 and C) and distal (R1 and R3) regions of the chromosomes (Figure 2a; Supplemental Table S2). We observed that triad homoeologs were more abundant in proximal than in distal regions: 64.3% *vs.* 35.7%, respectively. The opposite pattern was found for dyad genes with 62.2% located in distal regions and 37.8% in proximal regions. For tetrad genes, 57.6% were in distal and 42.4% in proximal regions.

At the chromosome scale, 95.6% of triads had their three genes located on homoeologous chromosomes (Figure 3). Out of the remaining 4.4%, 2.8% were found to have a mosaic distribution between chromosomes 4B, 4D and 5A or 4A, 5B and 5D, as a result of the structural evolution of the chromosomes 4A and 5A that have experienced inversions and translocations (Dvorak et al., 2018; Hernandez et al., 2012).

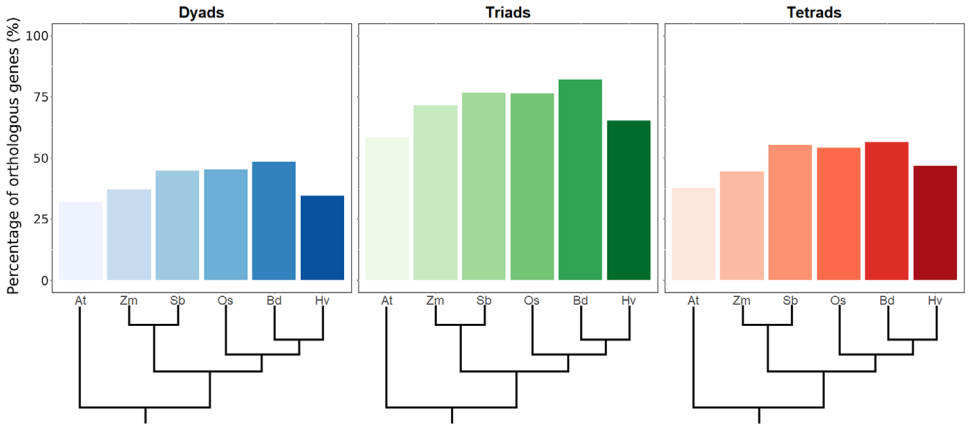

FIGURE 1 Conservation of genes in different plant genomes according to their category. For each category, the percentage of orthologs found in the *A. thaliana* (At), *Z. mays* (Zm), *S. bicolor* (Sb), *O. sativa* (Os), *B. distachyon* (Bd) and *H. vulgare* (Hv) genomes are given. Dyads are in blue, triads in green and tetrads in red

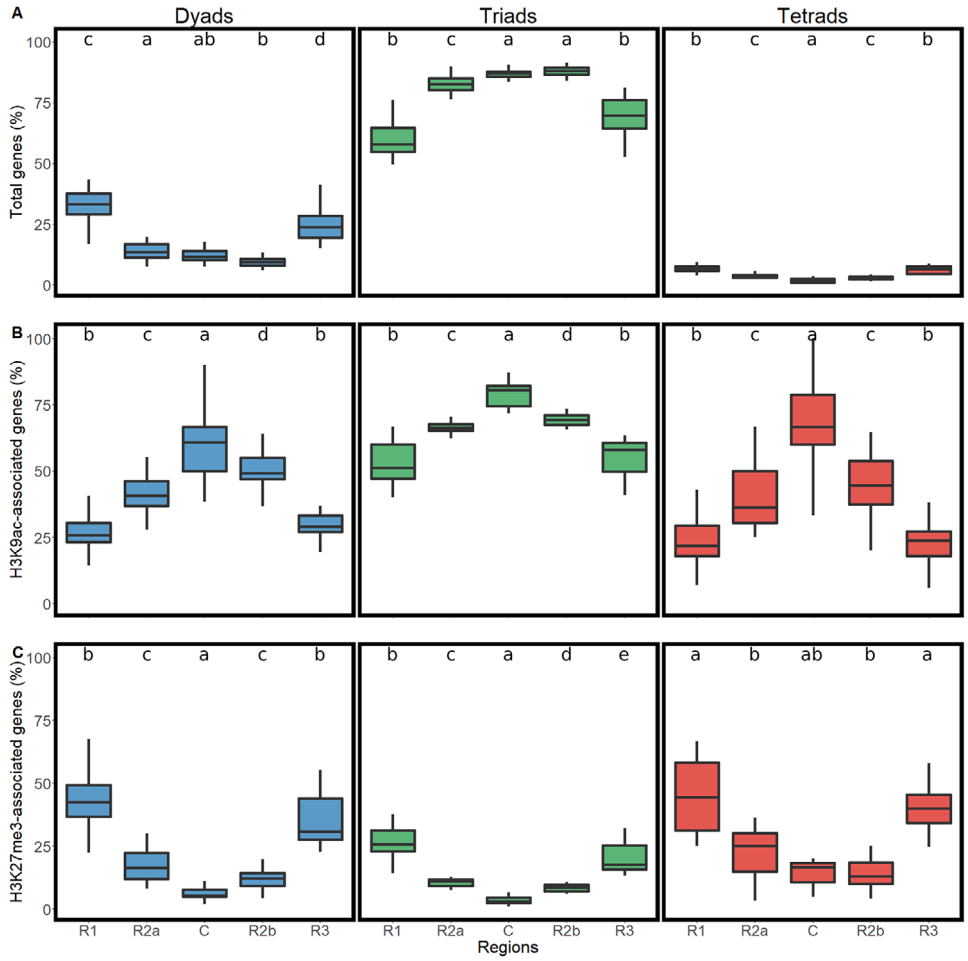

FIGURE 2 Distributions of genes of the different categories in the five regions of the wheat chromosomes. (a) Percentage of genes of a given category in a given region according to the total number of genes in this region. (b) Percentage of H3K9ac-associated genes according to the total number of genes from the same category in a given region. (B) Percentage of H3K27me3-associated genes according to the total number of genes from the same category in a given region. R1, R2a, C, R2b and R3 are the five chromosomal regions. Dyads are in blue, triads in green and tetrads in red. The boxplots depict the minimum without outliers, first quartile, median, third quartile and maximum without outliers. Different letters above the boxplots indicate significant differences ($P < .01$, Wilcoxon test)

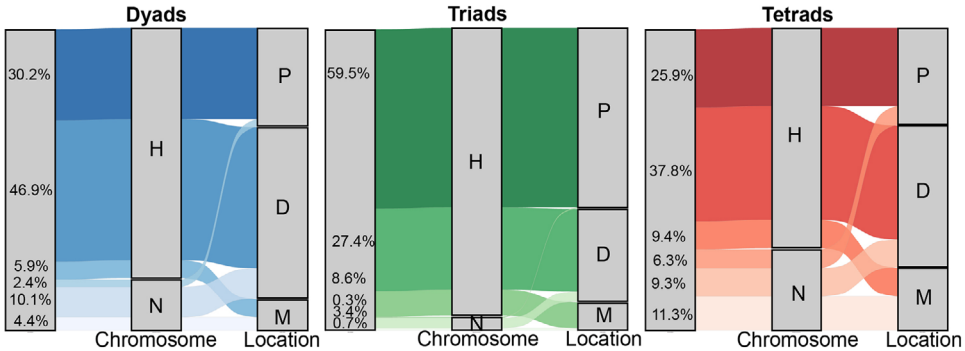

FIGURE 3 Alluvial plots of the different categories according to their location on chromosomes. For each category, the left-hand bar represents to whole set of genes, the central bar represents the position on homoeologous (H) or non-homoeologous (N) chromosomes; the right-hand bar represents the location, either proximal for all copies of a group (P), distal (D) or a mosaic of proximal and distal genes (M). The number indicated in the left-hand bar are the percentages of each class. Dyads are in blue, triads in green and tetrads in red

For dyads, 83.1% were found on homoeologous chromosomes and 3.3% showed a mosaic between chromosomes 4 and 5. For tetrads, the proportion of conserved homoeologous locations was much lower (73.1%) whereas that of mosaic distributions related to chromosome 4A evolutionary history was similar (3.2%).

As the boundaries of these regions are conserved between homoeologous chromosomes (IWGSC, 2018), we wondered to what extent the different gene copies of the same homoeologous group were located in the same regions (Figure 3). For triads, 90.7% of the genes belong to groups showing conserved locations for the three copies, with 30.8% being exclusively in distal regions and 59.9% being exclusively in proximal regions, confirming the high level of collinearity between A, B, and D. Only 9.4% of triad genes showed a variable location (mosaic distribution) between A, B, and D. For dyads, 57.0% of the homoeologs were exclusively located in distal regions, 32.7% were only in proximal and 10.3% were located in two different regions. For tetrads, 47.1% of the genes belong to groups having all their copies located exclusively in distal regions, 32.2% located exclusively in proximal regions and 20.7% with a mosaic distribution. The higher proportion of mosaic distribution for the tetrad category is explained by dispersed duplications that represented 36.6% of duplicated genes, of which 19.8% showed inter-chromosomal duplications.

3.4 | Functions of homoeoCNVs

Gene Ontology (GO) enrichment analysis revealed that dyads, triads and tetrads were involved in different biological processes. Indeed, triads were associated with basic cell processes such as transport, protein folding or DNA repair, replication and recombination. In contrast, tetrad and dyad genes were enriched in GO terms such as protein phosphorylation, oxidation-reduction processes, and response to fungus and oxidative stress (Supplemental Table S3). Analyzing the distal and proximal regions separately reached the same results, demonstrating that the GO enrichment was not only related to the preferential chromosomal location of the different categories (data not shown).

We expanded the analysis using the functional annotation of these genes to search for putative enrichment in protein functions (IWGSC, 2018) (Supplemental Data S1). F-box family proteins appeared to be the most abundant family in dyads and tetrads, comprising 7.9% and 6.0% of genes, respectively, while it represented 2.1% of triads. Similarly, consistent with GO enrichment analyses, disease resistance associated genes such as NLR, RLK, BTB/POZ-domain or ankyrin represented 9.7% of dyads and 7.8% of tetrads but only 2.9% of triads. Among other functions enriched in dyads and/or tetrads compared to triads were oxidation-reduction processes-associated proteins such as peroxidases, Cytochrome P450 and glutathione S-transferases.

3.5 | Expression of homoeoCNVs

To evaluate expression differences between the three categories of homoeologs, we used a gene expression atlas covering the whole plant development in controlled non-stressed conditions (Pingault et al., 2015; Ramirez-Gonzalez et al., 2018). We found detectable expression (TPM values >0.5) in at least one out of 15 tissues for 61,680 homoeologous genes from our dataset (87.1%) (Supplemental Table S2; Supplemental Data S1). The proportion of expressed genes was slightly higher on the D-genome genes (87.7%) than on the A- and B-genomes (86.6% and 86.9%, respectively; χ^2 p-value <.01). The percentage of expressed genes varied between categories too: 91.9% for triads (50,710 genes), 69.9% for dyads (8,838 genes) and 70.9% for tetrads (2,132 genes). In addition, we observed intra-category differences. For dyads, the AB-groups contained significantly

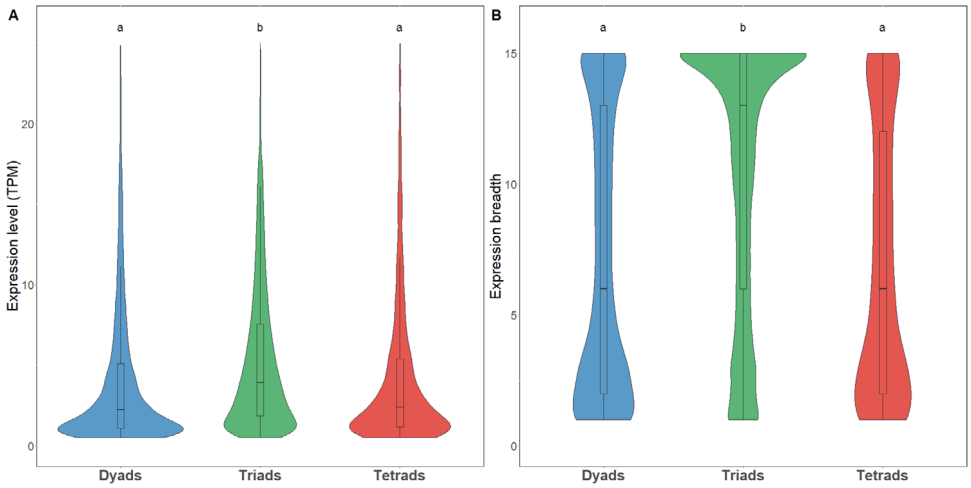

FIGURE 4 Expression level (a) and breadth (b) of the different categories. Expression level in TPM; expression breadth in number of conditions. Dyads in blue, triads, in green and tetrads in red. The boxplots depict the minimum without outliers, first quartile, median, third quartile and maximum without outliers. Different letters above the violin plots indicate significant differences ($P < .01$, Wilcoxon test)

fewer expressed genes (66.2%) than the BD- (70.5%) and AD-groups (72.2%) (χ^2 *p*-value <.01). For tetrads, at a χ^2 *p*-value of 1%, no significant differences were observed between groups. Interestingly, while in dyads and triads, the homoeologous genomes tended to have similar proportions of expressed copies, the duplicated-genome copies of tetrads displayed fewer expressed genes (χ^2 *p*-value <.01; Supplemental Table S2).

After discarding 7,179 outliers, i.e. genes with abnormally high expression level values (649 dyad, 6,375 triad and 155 tetrad genes), we investigated the mean expression level and expression breadth (i.e. the number of tissues in which genes were expressed) of 54,501 expressed genes: 8,189 dyad, 44,335 triad and 1,977 tetrad genes.

We found that triad genes were expressed at a higher level (mean = 5.9 TPM) and a higher breadth (10.4 tissues) than dyad (mean expression level = 5.2 TPM; mean expression breadth = 7.1) and tetrads (mean expression level = 5.2 TPM; mean expression breadth = 6.9) genes (Figure 4; Wilcoxon test *p*-value $<2.2\times10^{-16}$). To rule out the possibility that differences in expression level and breadth between dyads, triads and tetrads were only related to their chromosomal location, we divided the three categories in two sub-classes corresponding to their location, either proximal (R2 and C) or distal (R1 and R3). For all categories, genes located in proximal regions were expressed at significantly higher breadth than those in distal regions, confirming the impact of gene position on its expression. However, in general, triad genes were expressed at higher level and breadth than dyad and tetrad genes located in the same compartment (distal or proximal) (data not shown). This showed that the higher expression observed for triads may not only be due to their chromosomal location but to other factors. In tetrads, as for the proportion of expressed genes, the duplicated genome copies were less expressed than the non-duplicated ones, with lower expression breadth (6.6 vs. 7.1) and level (4.7 vs. 5.6; Wilcoxon test *p*-value <.01).

3.6 | Relative contribution of each copy to the expression of the overall homoeologous group

To go further on expression analysis of our three categories of homoeologs, we calculated the relative contribution of each homoeolog to the overall group expression, for groups having at least one gene expressed. We then assigned each group to an expression bias category, as defined by Ramirez-Gonzalez et al. (2018): the balanced category with similar relative abundance of transcripts from each of the homoeologs, and the homoeolog-dominant or homoeolog-suppressed categories, classified based on the higher or lower abundance of transcripts from a given homoeolog with respect to those from the other(s) (Supplemental Tables S2 and S3).

For dyads, 64.0% of the groups were balanced, while 36.0% were dominant/suppressed. AB dyads appeared to be less frequently balanced than AD and BD dyads (60.1%, 66.6% and 64.2%, respectively; χ^2 *p*-value <.05). The expression breadth of balanced dyads was higher than that of suppressed/dominant ones (8.0 and 4.9, respectively) (Figure 5).

For triads, an even higher proportion of balanced groups was observed (81.2%), while suppressed and dominant groups represented 14.0% and 4.8%, respectively. No difference was observed in the proportion of groups presenting a single-homoeolog dominance toward one sub-genome. Nevertheless, we observed a D-homoeolog suppression significantly less frequent (3.4%) than either A- or B-homoeolog suppression (5.3% and 5.2%, respectively; χ^2 *p*-value $<2.2\times10^{-16}$). As observed by Ramirez-Gonzalez and collaborators

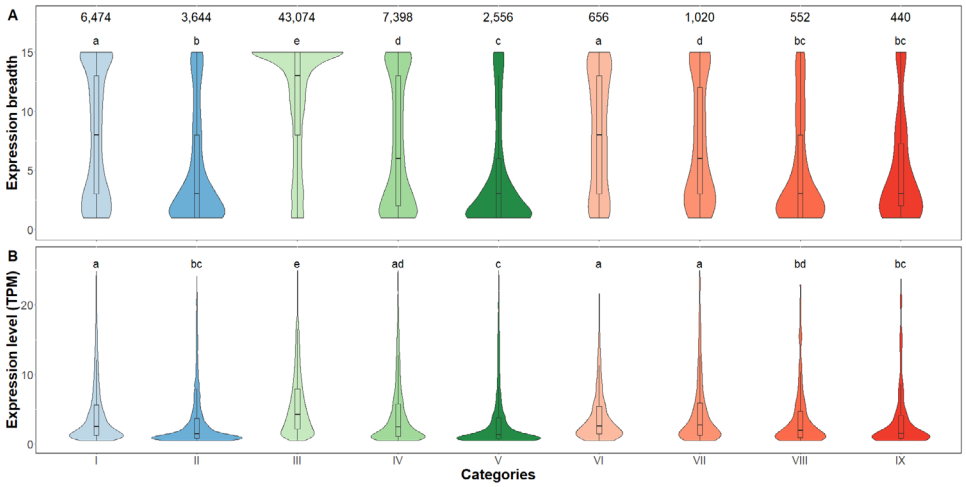

FIGURE 5 Expression breadth (a) and level (b) of the different categories according to their relative contribution classes. I: balanced dyads; II: Suppressed dyads; III: balanced triads, IV: Suppressed triads, V: Dominant triads, VI: balanced tetrads, VII: tetrads with one suppressed copy, VIII: tetrads with two suppressed copies, IX: tetrads with one dominant copy. The numbers above violin plots indicate the number of genes within the category. The boxplots depict the minimum without outliers, first quartile, median, third quartile and maximum without outliers. Different letters above the violin plots indicate significant differences ($P < .01$, Wilcoxon test)

(2018), expression breadth decreased from balanced to suppressed to dominant triads (11.1, 7.1 and 4.4, respectively) (Figure 5).

For tetrads, as expected from the greater number of gene copies, the pattern of relative contributions was much more complex. Balanced tetrads represented only 24.6% of the 667 groups having at least one expressed gene. The rest of the groups included 16.5% of groups with one copy dominant over the three others, 20.7% with two copies suppressed and 38.2% with one copy suppressed. It is worth noting that, for 74.5% of this latter, one of the two duplicates was suppressed. In addition, for ABDD tetrads, the duplication of a D-copy seems to have a similar impact on the suppression of A- and B-copies. By contrast, duplications of A-copies led to a slightly yet significantly higher proportion of B-copies than D-copies suppression (17.6% vs. 11.7%, respectively; χ^2 p-value <.01). A similar trend was observed in ABBD tetrads, where the B-copy duplication had greater impact on A-copy than D-copy suppression (15.7% *vs*. 11.2%, respectively; χ^2 p-value <.01). Interestingly, no significant difference was observed in terms of expression level between balanced tetrads and tetrads with one copy suppressed (5.2 TPM and 5.6 TPM, respectively) (Figure 5). The other tetrads displayed a significantly lower level (4.6 TPM for two suppressed copies and 4.6 TPM for one dominant copy; χ^2 p-value <.01).

We then explored whether the different categories retain their homoeologous expression bias category across the five organs (root, leaf, stem, spike and grain) (Supplemental Data S1). We found that 64.4% of balanced triads were also balanced (or not expressed) in the five organs, whereas, for dyads and tetrads, the proportions were 56.0% and 26.8%, respectively.

To complement this analysis, we investigated the divergence in spatial expression patterns. To this aim, we computed the binary expression (i.e. expressed or not) of each gene in the five different organs. This resulted in 32 binary expression clusters (0-0-0-0-0, 0-0-0-0-1, 0-0-0-1-1…). We then analyzed each group to see whether genes from a given group belong to the same or divergent binary expression groups (Supplemental Table S2).

For triads, 65.3% had their three copies in the same cluster, among which 85.8% were expressed in all five organs. When analyzing triads with one single divergent copy, we found a lower proportion of D-genome divergence, with 7.3% compared to 8.7% and 8.2% for the A and B-genomes, respectively (χ^2 p-value <.01).

For dyads and tetrads, the proportion of groups having all the genes in the same binary expression cluster dropped to 45.7% and 21.1%, respectively. Interestingly, 21.9% of tetrads had one single divergent copy and in 71.2% of the cases, the divergent copy was one of the duplicates. The proportion of D-divergent ABDD tetrads was found to be lower (63.9%) even though the difference was not significant, probably due to the small sample size.

Finally, when considering only balanced groups, the percentage of groups having all their copies in the same binary expression cluster raised to 64.8% for dyads, 75.4% for triads and 46.3% for tetrads.

3.7 | Epigenetic status of homoeoCNVs

Epigenetic marks, and especially H3K9ac and H3K27me3 histone modifications, have been shown to be associated

with differences in homoeolog expression patterns in triads (Ramirez-Gonzalez et al., 2018). These two marks have antagonist effects: H3K9ac is associated with open euchromatin and transcriptional activation whereas H3K27me3 is associated with facultative heterochromatin and transient transcriptional repression. To assess whether these marks may also be involved in the differences of expression patterns of dyads and tetrads, we analyzed the presence of these two marks on the 70,818 genes from our dataset. We found 44,954 genes associated with H3K9ac and 15,357 with H3K27me3 (Supplemental Data S1). After removing 3,809 genes that were associated with both marks, our dataset comprised 41,145 H3K9ac- and 11,548 H3K27me3-marked genes, i.e. 58.1% and 16.3% of all genes in the dataset, respectively (Supplemental Table S2). These proportions differed according to the chromosomal locations: distal and proximal regions were enriched in H3K27me3 and H3K9ac genes, respectively, as shown previously (IWGSC, 2018) (Figures 2b and 2c).

As expected, H3K27me3 genes tended to be more often repressed than H3K9ac, with 31.5% and 3.5% of genes never expressed across the 15 tissues, respectively. We also found a higher expression breadth for H3K9ac genes (12.1 tissues) compared to H3K27me3 genes (2.8 tissues). When analyzing gene expression in leaves at three-leaf stage (corresponding to ChIP-seq data), 85.1% of the H3K27me3 genes did not display any detectable expression while 83.2% of H3K9ac genes did.

When considering the three categories separately, we found differences in the proportion of the two marks. H3K27me3 was associated with 29.1% and 31.9% of dyad and tetrad genes, respectively, but only with 12.5% of triad ones (Figure 2b and 2c). By contrast, 64.7% of triad genes were marked by H3K9ac *vs.* 35.6% for dyads and 30.8% for tetrads.

Similar to what was observed previously, these proportions varied according to chromosomal regions (Figures 2b and 2c). However, triads were always more associated with H3K9ac and less with H3K27me3 than dyads and tetrads in the same chromosomal compartment. We also observed differences among tetrads. Indeed, the proportion of H3K9ac-associated genes increased from AABD to ABBD to ABDD (26.3%, 31.7% and 34.8%), while the opposite pattern was observed for H3K27me3 (34.7% for AABD, 31.7% for ABBD and 28.9% for ABDD). In addition, a significantly lower proportion of H3K9ac-associated genes was observed for the genome carrying the duplicated copies compared to the two others (27.7% vs. 33.9%; χ^2 *p*-value <.01). For H3K27me3, the proportion was slightly higher yet not significantly (33.1 vs. 30.8%; (χ^2 *p*-value >.01).

At the gene scale, balanced dyad, triad and tetrad genes had generally high H3K9ac and low H3K27me3 densities (Figure 6). For suppressed/non-suppressed and dominant/non-dominant groups, suppressed and non-dominant copies displayed higher H3K27me3 and lower H3K9ac than non-suppressed and dominant ones. Interestingly, the higher the number of suppressed copies in a group (from one to two in triads and from one to three in tetrads), the higher the H3K27me3 density, not only in the gene body but also into the upstream and downstream regions. This is consistent with the tight association of this mark with inactive promoters (Zhang et al., 2007).

We then analyzed whether genes from a group containing at least one gene associated with a mark tended to share the same epigenetic mark. For triads, 64.4% of the groups comprised three genes sharing the same mark, either H3K9ac (88.5%) or H3K27me3 (11.5%). This percentage was lower (55.9%) for dyads (57.1% H3K9ac and 42.9% H3K27me3). For tetrads, only 28.1% of groups comprised four genes sharing the same mark. Nevertheless, this percentage raised to 52.8% when including groups with three copies sharing the same mark.

Finally, we investigated the conservation of histone marks in two other species, *Zea mays* and *Oryza sativa*, for which H3K27me3 and H3K9ac data on a young leaf stage were available in the Plant Chromatin Database (Liu et al., 2018). Out of 12,995 and 12,006 groups containing at least one wheat H3K9ac-marked gene that had an ortholog in rice and maize, 10,541 (81.1%) and 7,604 (63.3%) were also marked by H3K9ac in these two species, respectively (Supplemental Table S2). For H3K27me3, the conservation was lower with 47.8% and 56.4% of the groups containing orthologs also targeted by this mark in rice and maize, respectively.

Surprisingly, while strong differences were observed at the sequence orthology level between dyads, triads and tetrads, the conservation of histone marks was not so different between the three categories. For example, while triads were much more conserved with rice than dyads (76.4% *vs.* 45.1%, respectively), quite similar proportions of groups containing conserved histone-marked genes were observed (48.4% of triads and 46.7% of dyads for H3K27me3, 81.5% of triads and 79.1% of dyads for H3K9ac).

4 | DISCUSSION

Wheat is an allohexaploid species originating from two successive and recent rounds of hybridization between three diploid species that were very similar in terms of chromosome number, genome size, TE content, gene content and synteny (IWGSC, 2018; Wicker et al., 2018). As a result, and considering that wheat is an autogamous homozygous species, it has long been considered that most of the genes were in three homoeologous copies. This perception started to change with the advent of the first draft assembly of the wheat genome sequence (IWGSC, 2014). The reference sequence of the hexaploid wheat genome confirmed that a significant fraction of genes departed from this 1:1:1 ratio and

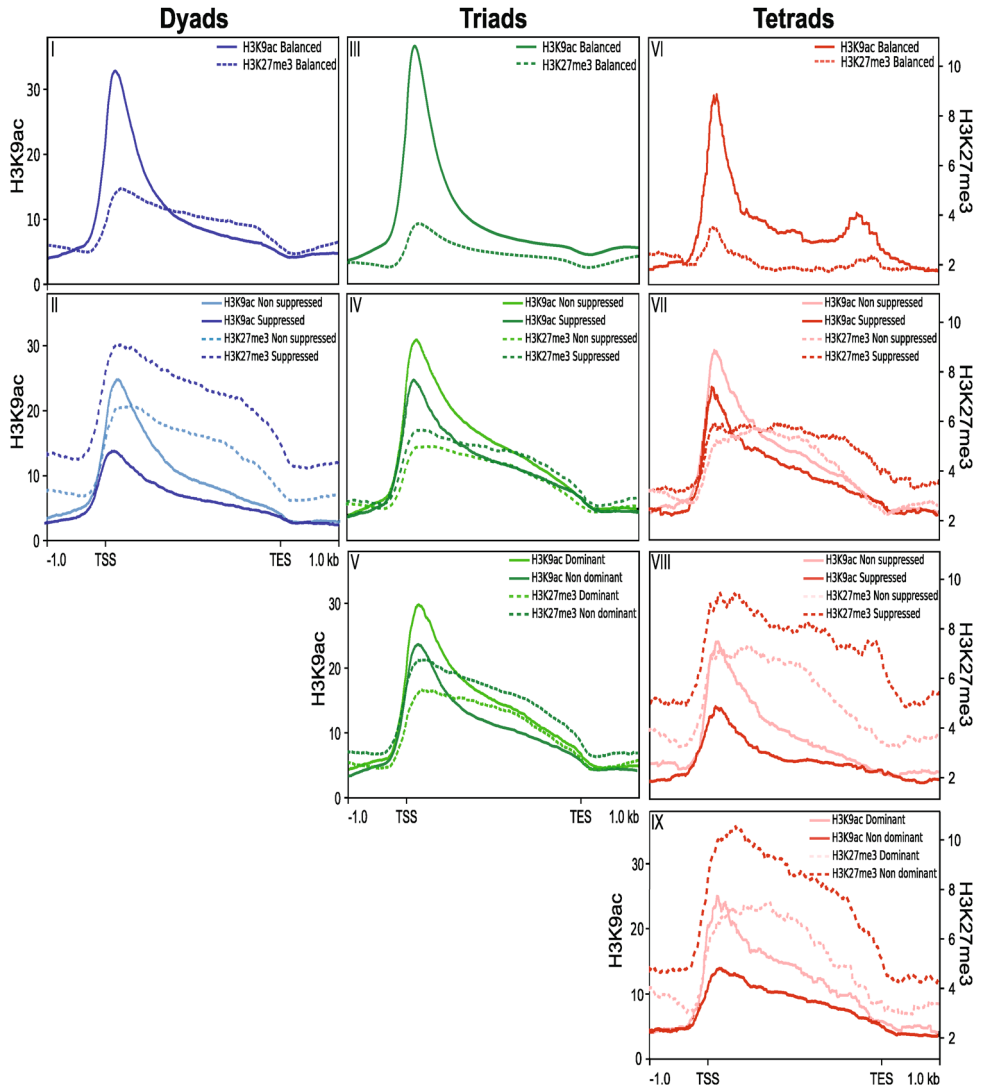

FIGURE 6 Epigenetic profiles of the different categories according to their relative contribution classes. Metagene profile for histone H3K9ac (solid lines) and H3K27me3 (dashed lines) marks from −1 kb upstream of the ATG to +1 kb downstream of the stop codon (normalized for gene length) for dyads (blue), triads (green) and tetrads (red) dominant/non-suppressed (light colors) and suppressed/non-dominant (dark colors) copies. I: balanced dyads; II: Suppressed dyads; III: balanced triads, IV: Suppressed triads, V: Dominant triads, VI: balanced tetrads, VII: tetrads with one suppressed copy, VIII: tetrads with two suppressed copies, IX: tetrads with one dominant copy

that these so-called 'triads' represent less than one half of all wheat genes.

In a recent work, Ramirez-Gonzalez et al. (2018) characterized the transcriptome atlas of wheat with a focus on these triads. This analysis provided new insights into the relative contribution of homoeologous copies to the overall group expression and the possible role of epigenetic marks in establishing this pattern.

In this study, we extended this analysis to genes departing from the 1:1:1 ratio, and more particularly the homoeologous groups having undergone a single gene loss or duplication event. These so-called dyads and tetrads, collectively referred to as HomoeoCNVs, represented 17.8% and 4.2% of our HC gene datasets whereas triads represented 77.9%. These proportions differed from those reported by the IWGSC (2018) as we focused our analysis on HC genes while the filtered dataset used by the IWGSC consisted of both HC and LC genes and took into account all categories, including genes in N:N:N ratio and not only dyads, triads and tetrads.

Because they have been kept in a strict 1:1:1 ratio through the course of evolution, triads are likely to correspond to highly conserved and evolutionary constrained genes. By contrast, dyads and tetrads have been either deleted or duplicated in the hexaploid wheat genome or in its diploid and tetraploid progenitors. We therefore suggest that triads are enriched in housekeeping genes and are part of the core genome, while dyads and tetrads belong to the dispensable genome of wheat. Several findings support this hypothesis.

First, we found that triads were more conserved in other plant genomes than HomoeoCNVs. By contrast, dyads and tetrads were found to be less conserved not only in distant plant genomes such as *A. thaliana*, *Z. mays*, *O. sativa*, *S. bicolor* or *B. distachyon* but also in the *Triticum/Aegilops* species, as most of these genes were already missing or duplicated in the wheat progenitors. In addition, HomoeoCNVs were also enriched in PAVs in a panel of 16 hexaploid wheat accessions. Previous studies in soybean, rice and *B. distachyon* demonstrated that core genes tend to have a higher percentage of homologs in other species than dispensable ones (Gordon et al., 2017; Li et al., 2014; Zhao et al., 2018). In wheat, this difference in terms of gene conservation is consistent with the genomic distribution of the different categories and the chromosome partitioning (Choulet et al., 2014; Daron et al., 2014; Darrier et al., 2017; Glover et al., 2015; IWGSC, 2018). Indeed, we showed that triads were more abundant in the low-recombination proximal regions. By contrast, dyads and tetrads were enriched in distal regions where differential TE content and recombination rate have likely driven gene duplications and deletions (Akhunov et al., 2003; Dvorak & Akhunov, 2005; Feldman, Levi, Fahima, & Korol, 2012; Reams & Roth, 2015; Zhang, 2003).

We also found that triads were expressed at higher level and breadth, while dyads and tetrads tend to be more specific to some tissues or developmental stages. In *B. distachyon*, core genes tend to be expressed at a higher level and more broadly than dispensable genes (Gordon et al., 2017). Choulet et al. (2014) and Pingault et al. (2015) reported on the physical partitioning of wheat genes, with highly and constitutively expressed genes being mainly located in proximal regions and genes expressed at lower level and breadth in distal ones. However, by analyzing distal and proximal regions separately, we showed that triads were always more expressed than dyads and tetrads whatever their position on the chromosome, which ruled out the possibility that the differences in expression patterns were only related to the chromosomal positions. Conversely, this difference can at least partly be explained by the epigenetic pattern of the categories of homoeologous genes. Indeed, triads were enriched in H3K9ac active euchromatin mark whereas dyads and tetrads were enriched in H3K27me3, a repressive mark related to facultative heterochromatin (Wiles & Selker, 2017). This differential association with active or repressive histones marks have already been reported in other species, such as potato where CNV frequency increased in genes lacking histone marks associated with permissive transcription (Hardigan et al., 2016). In wheat, we showed recently that genes affected by intra- and interspecific copy number variations were enriched in H3K27m3 (De Oliveira et al., 2020).

Finally, dyads and tetrads were enriched in functions associated with environmental and defence responses, a common feature of most plant dispensable genomes (Golicz et al., 2016; Gordon et al., 2017; Hurgobin et al., 2018; Li et al., 2014; McHale et al., 2012; Schatz et al., 2014). In particular, we found a higher proportion of genes associated to oxidation-reduction process that are known to be related to reactive oxygen species and putatively to biotic (pathogens) and abiotic (heavy metals, salt…) stress response mechanisms (Gullner, Komives, Király, & Schröder, 2018; Mir et al., 2015; Mittler, Vanderauwera, Gollery, & Van Breusegem, 2004; Veith & Moorthy, 2018). Disease resistance-associated families, such as NLRs, RLKs, ankyrin repeat or BTB/POZ domain-containing proteins were also found in higher proportions in dyads and tetrads than in triads (Sun, Zhu, Balint-Kurti, & Wang, 2020; Wang, Zou, Li, Lin, & Tang, 2020; Ye et al., 2017; Zhang et al., 2019).

We then examined intra-categories differences to investigate the possible impact of polyploidization on both core and dispensable genomes in wheat. Indeed, polyploidization is usually followed by a post-polyploid diploidization (PPD) process that tends to revert the polyploid genome into a quasi-diploid one (Mandáková & Lysak, 2018). PPD is accompanied by several mechanisms including gene neo/subfunctionalization, activation of transposable elements, epigenetic reprogramming and genome fractionation. Genome fractionation is a long-term process involving the loss of redundant genes and/or noncoding regulatory elements (Cheng et al., 2018). While it has been observed in several species and seems to be a common mechanism, differences have been observed according to the type of whole genome duplication (Garsmeur et al., 2013). In allopolyploids or paleo-allopolyploids such as *Arabidopsis thaliana*, maize (*Zea mays*), Chinese cabbage (*Brassica rapa*) and *Brassica oleracea,* duplicated genes are lost preferentially from one parental genome (biased fractionation) and the subgenome having retained the highest number of genes is more expressed (genome dominance) (Liu et al., 2014; Schnable, Springer, & Freeling, 2011; Wang et al., 2011). By contrast, in autopolyploids or paleo-autopolyploids, such as poplar (*Populus trichocarpa*) and pear (*Pyrus bretschneideri*), subgenome dominance is absent and genes tend to be evenly lost between the two subgenomes (Li et al., 2019; Liu et al., 2017). In wheat, some rapid changes following polyploidization have been reported, including chromosomal rearrangements, epigenetic changes or TE-related shift in centromere position (Badaeva, Dedkova, Pukhalskyi, & Zelenin, 2015; Dvorak et al., 2018; Jiao et al., 2018; Li et al., 2013; Liu et al., 2009; Shaked, Kashkush, Ozkan, Feldman, & Levy, 2001; Zhao et al., 2019). However, whether the wheat genome experiences subgenome dominance or biased fractionation is still a matter of debate. Different analyses reached contradictory results (El Baidouri et al., 2017; IWGSC, 2018; Pont & Salse, 2017).

Consistent with what was observed on a filtered set of 181,036 genes comprising both HC and LC genes (IWGSC, 2018), the number of genes analyzed in our study was highly

similar between subgenomes, with 23,411 on A, 23,524 on B and 23,883 on D. These similar proportions can be partly explained by the fact that the vast majority of these groups (75.9%) corresponded to triads, with one copy on each of the subgenomes. Such a high percentage of genes that are still present on the A, B and D-genomes demonstrate that no massive gene loss occurred upon polyploidization. Homoeologous groups that have lost one copy, i.e. dyads, represented 17.8% of our dataset. This category might reflect post-polyploidization gene loss. Interestingly, a lower number of AB-dyads (1,776) was observed compared to AD- or BD-ones (2,253 and 2,291, respectively). However, the analysis of the diploid and tetraploid ancestors suggested that only approximately 450 genes were lost on each subgenome at each step of polyploidization. This similar number of lost genes reveals the absence of a biased fractionation in wheat. Nevertheless, while no bias was found in terms of gene loss, we noticed subtle differences between genomes at the transcription and epigenetic levels.

The majority of triads displayed a balanced contribution of each copy to the overall group expression (81.2%). They also showed a high proportion of homoeologous genes having the same binary spatial expression (65.3%) and sharing the same histone mark (69.3%). However, we observed a slightly higher proportion of D-genome expressed genes compared to A and B, together with a lower proportion of D-suppressed triads, and a lower proportion of D divergent copies. This suggests a lower repression or subfunctionalization of D-genome homoeologs compared to their A and B counterparts. While these are very faint differences, they might explain the subtly yet significantly higher relative abundance of the D-subgenome transcripts (33.7%) compared to the A (33.1%) and B (33.3%) reported by Ramirez-Gonzalez et al. (2018).

Similarly, little differences were observed in dyads. They were mainly balanced (64.0%), with most of the homoeologs associated with the same epigenetic mark (65.1%). In addition, we found no bias in terms of dominance of one genome over the other. However, AB-dyads tended to be slightly less numerous and less expressed than the AD- and BD-ones, these two latter being similar on several aspects.

The pattern was much more complex for tetrads, with fewer balanced groups (24.6%) that might be explained by a higher proportion of mosaic patterns of epigenetic marks (71.9%). The overall higher proportion of H3K27me3- and lower proportion of H3K9ac-associated genes, especially on the genome carrying the duplicated copy, were likely related to subfunctionalization of retained paralogs as suggested by Makarevitch et al. (2013) and Berke, Sanchez-Perez, and Snel (2012). According to their mean expression level and breadth, as well as the epigenetic pattern, ABDD tetrads were more comparable to AB-dyads than to other tetrads. Indeed, the A and B copies of both ABDD tetrads and AB dyads were more associated with H3K9ac and less with H3K27me3, while the opposite pattern was found for those of AABD and ABBD tetrads. In addition, an extra copy on the B-genome (ABBD tetrads) appeared to have a much stronger impact on the A-genome than on the D-genome, while an extra D-copy had a similar impact on A and B in ABDD tetrads.

Based on these different results, we propose that the D-subgenome homoeologous genes are less repressed than the two others, and conversely, that their presence, either as single or duplicated genes, had a limited impact on the A- and B-copies. Differences observed between subgenomes are likely related to the D-genome more recent hybridization with the AABB tetraploid genome progenitor. This resulted in two successive PPD waves, that had impacted A and B more significantly since they spent more time together. However, unlike most of the allopolyploid species, subgenome dominance and biased fractionation are absent in hexaploid wheat. Indeed, while originating from the hybridization of three distinct species, the diploid donor genomes were very similar in terms of gene and TE contents prior to polyploidization. Consequently, individual genes, rather than subgenomes, experienced stochastic differences over longer periods of time, resulting in retention of the majority of WGD duplicates. In this regard, while being an allohexaploid species, wheat somehow resembles more to an autopolyploid in terms of evolutionary fate, as already observed in other paleo-allopolyploids such as soybean (*Glycine max*) and cucurbits (*Cucurbita maxima* and *Cucurbita moschata*) (Sun et al., 2017; Zhao, Zhang, Lisch, & Ma, 2017).

ACKNOWLEDGMENTS

We acknowledge Hélène Rimbert and Philippa Borrill for their assistance. C.J. was supported by Région Auvergne and the European Regional Development Fund (SRESRI 2016).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Lorenzo Concia https://orcid.org/0000-0002-7401-7214
Romain De Oliveira https://orcid.org/0000-0003-0017-6308
Etienne Paux https://orcid.org/0000-0002-3094-7129

REFERENCES

Adams, K. L., & Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, *8*, 135–141. https://doi.org/10.1016/j.pbi.2005.01.001

Akhunov, E. D., Akhunova, A. R., Linkiewicz, A. M., Dubcovsky, J., Hummel, D., Lazo, G., … Dvorak, J. (2003). Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 10836–10841. https://doi.org/10.1073/pnas.1934431100

Alexan, A., & Rahnenfuhrer, J. (2019). topGO: Enrichment analysis for gene ontology.

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., … Distelfeld, A. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*, *357*, 93–97. https://doi.org/10.1126/science.aan0032

Badaeva, E. D., Dedkova, O. S., Pukhalskyi, V. A., & Zelenin, A. V. (2015). Chromosomal changes over the course of polyploid wheat evolution and domestication. In Y. Ogihara, S. Takumi, & H. Handa (Eds.), *Advances in wheat genetics: From genome to field* (pp. 83–89). Tokyo: Springer.

Berke, L., Sanchez-Perez, G. F., & Snel, B. (2012). Contribution of the epigenetic mark H3K27me3 to functional divergence after whole genome duplication in *Arabidopsis*. *Genome Biology*, *13*, R94. https://doi.org/10.1186/gb-2012-13-10-r94

Bottani, S., Zabet, N. R., Wendel, J. F., & Veitia, R. A. (2018). Gene expression dominance in allopolyploids: Hypotheses and models. *Trends in Plant Science*, *23*, 393–402. https://doi.org/10.1016/j.tplants.2018.01.002

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., … Wincker, P. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, *345*, 950–953. https://doi.org/10.1126/science.1253435

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

Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., … Feuillet, C. (2014). Structural and functional partitioning of bread wheat chromosome 3B. *Science*, *345*, 1249721. https://doi.org/10.1126/science.1249721

Daron, J., Glover, N., Pingault, L., Theil, S., Jamilloux, V., Paux, E., … Choulet, F. (2014). Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biology*, *15*, 546. https://doi.org/10.1186/s13059-014-0546-4

Darrier, B., Rimbert, H., Balfourier, F., Pingault, L., Josselin, A.-A., Servin, B., … Sourdille, P. (2017). High-resolution mapping of crossover events in the hexaploid wheat genome suggests a universal recombination mechanism. *Genetics*, *206*, 1373–1388. https://doi.org/10.1534/genetics.116.196014

De Oliveira, R., Rimbert, H., Balfourier, F., Kitt, J., Dynomant, E., Vrána, J., … Choulet, F. (2020). Structural variations affecting genes and transposable elements of chromosome 3B in wheats. *Frontiers in Genetics*, *11*, 891. https://doi.org/10.3389/fgene.2020.00891

Dvorak, J., & Akhunov, E. D. (2005). Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *Aegilops-Triticum* alliance. *Genetics*, *171*, 323–332. https://doi.org/10.1534/genetics.105.041632

Dvorak, J., Wang, L., Zhu, T., Jorgensen, C. M., Luo, M.-C., Deal, K. R., … McGuire, P. E. (2018). Reassessment of the evolution of wheat chromosomes 4A, 5A, and 7B. *Theoretical and Applied Genetics*, *131*, 2451–2462. https://doi.org/10.1007/s00122-018-3165-8

El Baidouri, M., Murat, F., Veyssiere, M., Molinier, M., Flores, R., Burlot, L., … Salse, J. (2017). Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytologist*, *213*, 1477–1486. https://doi.org/10.1111/nph.14113

Feldman, M., Levi, A., Fahima, T., & Korol, A. (2012). Genomic asymmetry in allopolyploid plants: Wheat as a model. *Journal of Experimental Botany*, *63*, 5045–5059. https://doi.org/10.1093/jxb/ers192

Flagel, L. E., & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist*, *183*, 557–564. https://doi.org/10.1111/j.1469-8137.2009.02923.x

Garsmeur, O., Schnable, J. C., Almeida, A., Jourda, C., D'Hont, A., & Freeling, M. (2013). Two evolutionarily distinct classes of paleopolyploidy. *Molecular Biology and Evolution*, *31*, 448–454. https://doi.org/10.1093/molbev/mst230

Glover, N., Daron, J., Pingault, L., Vandepoele, K., Paux, E., Feuillet, C., & Choulet, F. (2015). Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B. *Genome Biology*, *16*, 188. https://doi.org/10.1186/s13059-015-0754-6

Glover, N. M., Redestig, H., & Dessimoz, C. (2016). Homoeologs: What are they and how do we infer them? *Trends in Plant Science*, *21*, 609–621. https://doi.org/10.1016/j.tplants.2016.02.005

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., … Edwards, D. (2016). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature Communications*, *7*, 13390.

Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., … Vogel, J. P. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications*, *8*, 2184. https://doi.org/10.1038/s41467-017-02292-8

Gullner, G., Komives, T., Király, L., & Schröder, P. (2018). Glutathione S-transferase enzymes in plant-pathogen interactions. *Frontiers in plant science*, *9*, 1836–1836. https://doi.org/10.3389/fpls.2018.01836

Hardigan, M. A., Crisovan, E., Hamilton, J. P., Kim, J., Laimbeer, P., Leisner, C. P., … Buell, C. R. (2016). Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. *The Plant Cell*, *28*, 388–405.

Harper, A. L., Trick, M., He, Z., Clissold, L., Fellgett, A., Griffiths, S., & Bancroft, I. (2016). Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat. *Plant Biotechnology Journal*, *14*, 1207–1214. https://doi.org/10.1111/pbi.12486

Hernandez, P., Martis, M., Dorado, G., Pfeifer, M., Galvez, S., Schaaf, S., … Mayer, K. F. (2012). Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *The Plant Journal*, *69*, 377–386. https://doi.org/10.1111/j.1365-313X.2011.04808.x

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

International Wheat Genome Sequencing Consortium (IWGSC) (2014). A chromosome-based draft sequence of the hexaploid bread wheat genome. *Science*, *345*, 1251788.

International Wheat Genome Sequencing Consortium (IWGSC) (2018). Shifting the limits in wheat research and breeding through a fully annotated and anchored reference genome sequence. *Science*, *361*, eaar7191.

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

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., … Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*, 524–527. https://doi.org/10.1038/nature22971

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., … dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, *473*, 97–100. https://doi.org/10.1038/nature09916

Li, B., Choulet, F., Heng, Y., Hao, W., Paux, E., Liu, Z., … Zhang, X. (2013). Wheat centromeric retrotransposons: The new ones take a major role in centromeric structure. *The Plant Journal*, *73*, 952–965. https://doi.org/10.1111/tpj.12086

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, H., & Durbin, R. (2010). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, Q., Qiao, X., Yin, H., Zhou, Y., Dong, H., Qi, K., … Zhang, S. (2019). Unbiased subgenome evolution following a recent whole-genome duplication in pear (*Pyrus bretschneideri* Rehd.). *Horticulture Research*, *6*, 34. https://doi.org/10.1038/s41438-018-0110-6

Li, Y.-H., Zhou, G., Ma, J., Jiang, W., Jin, L.-G., Zhang, Z., … Qiu, L.-J. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*, *32*, 1045–1052. https://doi.org/10.1038/nbt.2979

Ling, H.-Q., Ma, B., Shi, X., Liu, H., Dong, L., Sun, H., … Liang, C. (2018). Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature*, *557*, 424–428.

Liu, B., Xu, C., Zhao, N., Qi, B., Kimatu, J. N., Pang, J., & Han, F. (2009). Rapid genomic changes in polyploid wheat and related species: Implications for genome evolution and genetic improvement. *Journal of Genetics and Genomics*, *36*, 519–528. https://doi.org/10.1016/S1673-8527(08)60143-5

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A. P., … Paterson, A. H. (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications*, *5*, 3930. https://doi.org/10.1038/ncomms4930

Liu, Y., Tian, T., Zhang, K., You, Q., Yan, H., Zhao, N., … Su, Z. (2018). PCSD: A plant chromatin state database. *Nucleic Acids Research*, *46*, D1157–D1167. https://doi.org/10.1093/nar/gkx919

Liu, Y., Wang, J., Ge, W., Wang, Z., Li, Y., Yang, N., … Wang, X. (2017). Two highly similar poplar paleo-subgenomes suggest an autotetraploid ancestor of Salicaceae plants. *Frontiers in Plant Science*, *8*, 571–571.

Luo, M.-C., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., … Dvořák, J. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature*, *551*, 498.

Makarevitch, I., Eichten, S. R., Briskine, R., Waters, A. J., Danilevskaya, O. N., Meeley, R. B., … Springer, N. M. (2013). Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. *Plant Cell*, *25*, 780–793. https://doi.org/10.1105/tpc.112.106427

Mandáková, T., & Lysak, M. A. (2018). Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology*, *42*, 55–65. https://doi.org/10.1016/j.pbi.2018.03.001

Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., … Olsen, O. A. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, *345*, 1250092. https://doi.org/10.1126/science.1250092

McHale, L. K., Haun, W. J., Xu, W. W., Bhaskar, P. B., Anderson, J. E., Hyten, D. L., … Stupar, R. M. (2012). Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiology*, *159*, 1295–1308. https://doi.org/10.1104/pp.112.194605

Michael, T. P., & Jackson, S. (2013). The first 50 plant genomes. *The Plant Genome*, *6*, 1–7. https://doi.org/10.3835/plantgenome2013.03.0001in

Mir, A. A., Park, S.-Y., Sadat, M. A., Kim, S., Choi, J., Jeon, J., & Lee, Y.-H. (2015). Systematic characterization of the peroxidase gene family provides new insights into fungal pathogenicity in *Magnaporthe oryzae*. *Scientific Reports*, *5*, 11831.

Mittler, R., Vanderauwera, S., Gollery, M., & Van Breusegem, F. (2004). Reactive oxygen gene network of plants. *Trends in Plant Science*, *9*, 490–498. https://doi.org/10.1016/j.tplants.2004.08.009

Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C.-K. K., … Edwards, D. (2017). The pangenome of hexaploid bread wheat. *The Plant Journal*, *90*, 1007–1013. https://doi.org/10.1111/tpj.13515

Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B., Rudi, H., Hvidsten, T. R., … Olsen, O. A. (2014). Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science*, *345*, 1250091. https://doi.org/10.1126/science.1250091

Pingault, L., Choulet, F., Alberti, A., Glover, N., Wincker, P., Feuillet, C., & Paux, E. (2015). Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. *Genome Biology*, *16*, 29. https://doi.org/10.1186/s13059-015-0601-9

Pont, C., & Salse, J. (2017). Wheat paleohistory created asymmetrical genomic evolution. *Current Opinion in Plant Biology*, *36*, 29–37. https://doi.org/10.1016/j.pbi.2017.01.001

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., … Paterson, A. H. (2019). Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biology*, *20*, 38. https://doi.org/10.1186/s13059-019-1650-2

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: Retrieved from www.r-project.org.

Ramirez-Gonzalez, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., … Uauy, C. (2018). The transcriptional landscape of polyploid wheat. *Science*, *361*. https://doi.org/10.1126/science.aar6089

Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., … Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*, W160–165. https://doi.org/10.1093/nar/gkw257

Reams, A. B., & Roth, J. R. (2015). Mechanisms of gene duplication and amplification. *Cold Spring Harbor perspectives in biology*, *7*, a016592–a016592. https://doi.org/10.1101/cshperspect.a016592

Renny-Byfield, S., Rodgers-Melnick, E., & Ross-Ibarra, J. (2017). Gene fractionation and function in the ancient subgenomes of maize. *Molecular Biology and Evolution*, *34*, 1825–1832. https://doi.org/10.1093/molbev/msx121

Salman-Minkov, A., Sabath, N., & Mayrose, I. (2016). Whole-genome duplication as a key factor in crop domestication. *Nature Plants*, *2*, 16115. https://doi.org/10.1038/nplants.2016.115

Schatz, M. C., Maron, L. G., Stein, J. C., Wences, A. H., Gurtowski, J., Biggers, E., … McCombie, W. R. (2014). Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biology*, *15*, 506.

Schnable, J. C., Springer, N. M., & Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 4069–4074. https://doi.org/10.1073/pnas.1101368108

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., & Levy, A. A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell*, *13*, 1749–1759. https://doi.org/10.1105/TPC.010083

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., … Xu, Y. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. *Molecular Plant*, *10*, 1293–1306. https://doi.org/10.1016/j.molp.2017.09.003

Sun, Y., Zhu, Y. X., Balint-Kurti, P. J., & Wang, G. F. (2020). Fine-Tuning Immunity: Players and Regulators for Plant NLRs. *Trends in Plant Science*, *25*, 695–713. https://doi.org/10.1016/j.tplants.2020.02.008

Van de Peer, Y., Mizrachi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics*, *18*, 411–424. https://doi.org/10.1038/nrg.2017.26

Veith, A., & Moorthy, B. (2018). Role of cytochrome P450s in the generation and metabolism of reactive oxygen species. *Current opinion in toxicology*, *7*, 44–51. https://doi.org/10.1016/j.cotox.2017.10.003

Villanueva, R. A. M., & Chen, Z. J. (2019). ggplot2: Elegant graphics for data analysis, 2nd edition. *Measurement-Interdisciplinary Research and Perspectives*, *17*, 160–167. https://doi.org/10.1080/15366367.2019.1565254

Wang, H., Zou, S., Li, Y., Lin, F., & Tang, D. (2020). An ankyrin-repeat and WRKY-domain-containing immune receptor confers stripe rust resistance in wheat. *Nature Communications*, *11*, 1353. https://doi.org/10.1038/s41467-020-15139-6

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., … Zhang, Z. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics*, *43*, 1035–1039.

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *American Journal of Botany*, *102*, 1753–1756. https://doi.org/10.3732/ajb.1500320

Wicker, T., Gundlach, H., Spannagl, M., Uauy, C., Borrill, P., Ramirez-Gonzalez, R. H., … Choulet, F. (2018). Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biology*, *19*, 103. https://doi.org/10.1186/s13059-018-1479-0

Wiles, E. T., & Selker, E. U. (2017). H3K27 methylation: A promiscuous repressive chromatin mark. *Current Opinion in Genetics & Development*, *43*, 31–37.

Wu, T. D., & Watanabe, C. K. (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, *21*, 1859–1875. https://doi.org/10.1093/bioinformatics/bti310

Ye, Y., Ding, Y., Jiang, Q., Wang, F., Sun, J., & Zhu, C. (2017). The role of receptor-like protein kinases (RLKs) in abiotic stress response in plants. *Plant Cell Report*, *36*, 235–242. https://doi.org/10.1007/s00299-016-2084-x

Zhang, C., Gao, H., Li, R., Han, D., Wang, L., Wu, J., … Zhang, S. (2019). GmBTB/POZ, a novel BTB/POZ domain-containing nuclear protein, positively regulates the response of soybean to *Phytophthora sojae* infection. *Molecular Plant Pathology*, *20*, 78–91. https://doi.org/10.1111/mpp.12741

Zhang, J. (2003). Evolution by gene duplication: An update. *Trends in Ecology & Evolution*, *18*, 292–298.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., … Chen, Z. J. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology*, *33*, 531–537. https://doi.org/10.1038/nbt.3207

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V., Pellegrini, M., Goodrich, J., & Jacobsen, S. E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biology*, *5*, e129.

Zhao, J., Hao, W., Tang, C., Yao, H., Li, B., Zheng, Q., … Zhang, X. (2019). Plasticity in Triticeae centromere DNA sequences: A wheat × tall wheatgrass (decaploid) model. *The Plant Journal*, *100*, 314–327. https://doi.org/10.1111/tpj.14444

Zhao, M., Zhang, B., Lisch, D., & Ma, J. (2017). Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. *The Plant Cell*, *29*, 2974–2994. https://doi.org/10.1105/tpc.17.00595

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., … Huang, X. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics*, *50*, 278–284. https://doi.org/10.1038/s41588-018-0041-z

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Juery C, Concia L, De Oliveira R, et al. New insights into homoeologous copy number variations in the hexaploid wheat genome. *Plant Genome*. 2020;e20069. https://doi.org/10.1002/tpg2.20069