Comprehensive Analysis of the Histone Methyltransferase Gene Family and their Expression Profiling Study on Yield in Two-line Hybrid Wheat

Shiqing Gao (✉ gshiq@126.com)

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Renwei Sun

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Jie Gong

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Yongjie Liu

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Zhaobo Chen

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Xianchao Chen

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Fengting Zhang

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Shengquan Zhang

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Jiangang Gao

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Junmei Cao

Xinjiang Academy of Agricultural Sciences ☐ Urumqi

Changping Zhao

Beijing Academy of Agriculture and Forestry Sciences (BAAFS)

Research Article

Keywords: hybrid wheat, yield, phylogenetic analysis, histone methyltransferase, heterosis, lncRNA

Posted Date: March 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1402181/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Histone methylation usually play important roles in plant development through post-translational regulation and may provide a new visual field for heterosis. The histone methyltransferase gene family has been identified in various plants, but its members and functions in hybrid wheat related in heterosis is poorly studied.

Results: In this study, 175 histone methyltransferase (*HMT*) genes were identified, including 152 histone lysine methyltransferase (*HKMT*) genes and 23 protein arginine methyltransferase (*PRMT*) genes. Gene structure analysis, physicochemical properties and subcellular localization predictions of the proteins, exhibited the adequate complexity of this gene family. HKMTs in angiosperms were classified into seven different orthologous groups, while four orthologous groups for PRMTs by phylogenetic analysis. As an allohexaploid species, the number of the genes in wheat was about three times than those in diploids and showed a relatively high degree of conservation, while only a small number of subfamilies such as ASH-like and Su-(var) subfamilies have expanded their members. Fifteen potential miRNA target sites in 11 different *TaHMT* genes were identified, and the expression of these *TaHMT* genes regulated by miRNAs were mainly through post-transcriptional cleavage manners. Transcriptome analysis showed that *HMT* genes were mainly expressed in the reproductive organs at distinct growth stages. Expression level of *TaHMT* genes exhibited different trends in various hybrid combinations; lncRNAs that are highly correlated with these *HMT*s were shown to have extremely similar expression trends in these materials. Pearson correlation analysis of gene expression and yield traits indicated that four *HMT* genes may participate in the yield heterosis of two-line hybrid wheat.

Conclusions: Our study identified characteristics of the histone methyltransferase gene family and enhances the understanding of the evolution and function of these members in allohexaploid wheat.

Background

Common wheat (*Triticum aestivum* L.), which is among the most widely consumed foods globally, is grown all over the world, second only to rice (*Oryza sativa* L.). It is generally believed that wheat was domesticated in Western Asia about 8000 years ago and consists of allohexaploid (AABBDD) in three sets of genomes: *Triticum urartu* (A-genome donor), an *Aegilops speltoides*-related grass (B-genome donor), and *Aegilops tauschii* (D-genome donor) [1, 2]. The huge genome and complex structure of wheat kept its genetic information a mystery for a long time. After years of unremitting scientific effort, some high-quality genome annotation data have been released, and multiple genome assemblies are available, which provide a powerful basis for the study of functional genomics [3–7]. A large amount of genetic diversity of cultivated crop species reside in its wild relatives. To harness that diversity and to expand the genetic base of the cultivated species, it is important to understand the evolutionary relationships among them and to know which genomic and phenomic tools would be appropriate.

Epigenetics, including DNA methylation, chromatin modification, and RNA editing, is a branch of genetics that studies the heritable changes of gene expression without changing the nucleotide sequence of genes [8]. In eukaryotes, nucleosomes that contain four groups of proteins (H2A, H2B, H3, and H4) are the main structural

components of chromatin. The four groups of proteins are two H3:H4 dimers and two H2A:H2B dimers, as well as a mixture of 146 bp DNA and a double helix. The N-end trailing of each histone conserved *in vitro* is the point of many targeted signal transduction pathways, resulting in post-translational modifications [9, 10]. Histone methylation is one of most important epigenetic modifications, and it plays a fundamental role in a range of biological processes, from transcriptional regulation to heterochromatin formation [11]. Methylation sites are usually located on the lysine (K) and arginine (R) residues of H3 and H4 and thus determine the two methylation patterns of histone lysine (Lys, K) methylation (HKMT) and histone arginine (Arg, R) methylation (PRMT). Methylation occurred in the arginine and lysine residues of histones, including seven sites in arginine and 17 sites in lysine [12]. These complex modification processes, one earlier and one more advanced, are deeply associated with gene expression. Studies have shown that histone methylation, catalyzed by HKMT and PRMT proteins, plays an important role in plant developmental processes, such as floral organogenesis, seed development, and plant senescence and defense [13–17].

The two-line method is a preferred way of realizing the heterosis of wheat, but the results of research on the molecular mechanism of the heterosis remain relatively weak. Generally, there is a strong correlation between heterosis and the genetic distance of the parents [18], but the study has shown that genetic distance does not fully explain heterosis [19]. The allelic variation caused by epigenetic modification provides a novel way of explaining heterosis. Previous study has shown that histone modification is closely related to the differential expression of genes in hybrid maize [20]. The methylation of H3K4me₂ can reduce the expression of two genes: *CCA* and *LHY*, which are the cores of heterosis in *Arabidopsis*, resulting in the improvement of photosynthetic efficiency of F₁ and the emergence of obvious hybrids [21]. Generally, H3K9ac and H3K4me₃ occur in regions of normal chromatin and activate the expression of genes, whereas H3K9me₂ and H3K27me₃ tend to occur in the heterochromatin regions and usually inhibit gene expression [22, 23]. The levels of H3K4me₃ and H3K27me₃ vary greatly among ecotypes and subspecies, but they showed additive patterns in *Arabidopsis* hybrids, whereas in rice they showed non-additive ones [20, 24]. Proteins in hybrid F₁, such as H3K9ac, H3K4me₂, and H3K4me₃, may be related to heterosis through the modification of different types of histone proteins to regulate genes related to photosynthesis and the circadian clock. However, due to the various methods of modification, the regulation mechanisms of histone modification and heterosis require further exploration. lncRNA-mediated deposition of H3K27me₃ is critical for chromatin modification and transcriptional regulation [25, 26]. lncRNAs participate in histone modification to regulate a series of important biological processes, such as vernalization and flowering, which can affect plant yield and quality [27, 28]. MicroRNAs (miRNAs) are a class of endogenous small noncoding RNAs which are involved in the regulation of gene expression at the posttranscriptional level by degrading their target mRNAs and/or inhibiting their translation. MicroRNA-mediated regulation is a key component in a wide range of biological processes such as plant developmental plasticity, abiotic/biotic responses, and symbiotic/parasitic interactions [29, 30]. *HMT* genes have been identified and studied in many plant systems and are widely involved in the regulation of plant growth and development [14, 31–37]. Although it is a fundamental process in an important cereal crop, histone methylation remains mysterious, and the role of regulatory factors such as miRNAs and lncRNAs in regulating histone methylation is largely unknown.

In this study, we provided a detailed overview of the phylogeny and expression of *HMT* genes in different hybrid bread wheats. We aim to explore the distribution and evolution of HMT members that catalyze histone

methylation in wheat, and search for some specific HMTs which may participate in the construction of Heterosis in hybrid wheat. All our efforts enhance the understanding of the evolution and function of the histone methylation family and sheds light on the histone methylation involved in regulating wheat yield.

Results

Identification of TaHMT Genes from the Whole Wheat Genome

Hmmsearch and the BLASTP program obtained 152 *HKMT* (histone lysine methyltransferase) genes and 23 *PRMT* (histone arginine methyltransferase) genes, which were named *TaHMT1–TaHMT152*, coinciding with *TaHKMT1–TaHKMT152*, *TaHMT153–TaHMT175*, and *TaPRMT1-23*, following the nomenclature 1A to 1D, 2A to 2D, up to 7A to 7D, based on their position on the chromosome. In total, we identified 175 genes that catalyze two kinds of histone methyltransferases (Fig. 1, Table 1, Table S1 and Table S2). Chromosome distribution showed that the *HKMT* gene family of wheat was distributed across all chromosomes and showed a trend of cluster distribution on chr2 and chr3. The small number of *PRMT* genes entails that they are orphans that exist on some chromosomes, such as chr4A (Figure S1).

Table 1
Numbers of *HMT* genes in wheat and related species

species	Number of <i>HMTs</i>	Number of <i>HKMTs</i>	Ratio to wheat	Number of <i>PRMTs</i>	Ratio to wheat
Wheat	175	152	–	23	–
<i>Arabidopsis</i>	57	48	3.17	9	2.56
Rice	47	39	3.90	8	2.88
<i>Brachypodium</i>	56	48	3.17	8	2.88
<i>Sorghum</i>	54	46	3.30	8	2.88
<i>Hordeum</i>	49	42	3.62	7	3.29

HMT Gene Family of Wheat Has Expanded with More In-paralogs Genes

Sorghum bicolor, *Hordeum vulgare*, and *Brachypodium distachyon* are usually considered to be relatives of wheat due to the whole-genome duplication events that occurred at about 45–60 million years ago [38]. In this study, the *HMT* gene families in *Arabidopsis thaliana*, *Oryza sativa*, *Hordeum vulgare*, *Sorghum bicolor*, and *Brachypodium distachyon* were also comprehensively identified in Table 1 using the same method as that employed for wheat. According to previous results, the number of *HMT* genes family in diploid species is between 40 and 60. We found that there were 47 and 57 *HMT* genes in rice (*Os*) and *Arabidopsis* (*At*),

respectively, which is close to the results previously obtained. In addition, 54, 56, and 49 *HMT* genes were obtained from the *Sorghum bicolor* (*Sb*), *Brachypodium distachyon* (*Bd*), and *Hordeum vulgare* (*Hv*) genome data, respectively. The proportion of gene quantity between wheat and other species was 3.07 (*At*), 3.72 (*Os*), 3.24 (*Sb*), 3.13 (*Bd*), and 3.57 (*Hv*), close to a ratio of 3:1, as expected (χ^2 test, $P > 0.05$).

To further explore the evolutionary relationship between wheat and the other five species, the orthologs and in-paralogs between each pair of these species were calculated. The number of orthologs in each species with one other genome is shown in Table 2. The result shows that wheat had a high average ortholog group size of 2.32, which also meant that every *HMT* gene had an average of 1.32 paralogs. The group sizes of other five species were significantly smaller than those of wheat, indicating that the genome ploidy of wheat enriches the number of paralogous genes for the *HMT* gene family. Significantly, the fact that wheat and *Brachypodium distachyon* clearly have more orthologous genes proves a later divergence between the two species, which is also a reflection of the previous conclusion that *Brachypodium distachyon* has a common ancestor with wheat.

Table 2
Total number of orthologs identified by In-paranoid 4. 1

species	Wheat	<i>Arabidopsis</i>	Rice	<i>Sorghum</i>	<i>Brachypodium</i>	<i>Hordeum</i>	Average in-paralogs
Wheat	–	137/401	121/43	132/46	141/50	111/44	2.32 ¹
<i>Arabidopsis</i>	48/40	–	42/34	42/34	45/36	42/34	1.23
Rice	44/43	38/34	–	40/39	38/37	36/36	1.04
<i>Sorghum</i>	47/46	39/34	40/39	–	44/44	40/39	1.04
<i>Brachypodium</i>	50/50	44/36	38/37	47/44	–	40/40	1.06
<i>Hordeum</i>	44/44	38/34	37/36	40/39	40/40	–	1.03

¹ The number of orthologs in an organism when clustered with another genome was shown on the left of the slash while the number on the right of the slash referred to the number of ortholog groups between two species. For wheat, average size of ortholog group: 2.32 = (137/40 + 121/43 + 132/46 + 141/50 + 111/44)/5. Notably, Table 2 was not a symmetrical table, since gene duplication frequency in organism ‘A’ generally differed from that in organism ‘B’ since speciation of organism ‘A’ and ‘B’ and thus the number of ‘A’ genes having orthologs in organism ‘B’ was unequal to the number of ‘B’ genes having orthologs in organism ‘A’.

According to the published wheat genome data, there are about in homologous groups of three, termed triads (1:1:1; 35.8% of genes) (IWGSC, 2018) [39]. Our results revealed a higher proportion 92.76% of *HKMTs* were in triads. In *PRMTs*, the ratio is 91.3% (Table 3). The high homology retention rate indicates that the *HMT* gene of three sets of wheat chromosome donors exhibited a certain conservation; in other words, there was almost no lack of homologous genes. With the improvement of accuracy in the knowledge of the wheat genome, it might be a higher proportion. This kind of conservative combination of triplets and high homology retention rate can also explain the richness of some gene families in wheat, such as the *HMT* genes. In addition to this 1:1:1 correspondence, some genes have expanded in breadth along the chromosomes. These members do not strictly follow the three homologies, like *HKMT1* and *PRMT1*. In addition to a correspondence to genes on

chromosomes with A/B/D homology, they also show homology with members of other chromosomes to a certain extent. The homology of these genes may have been established before the formation of wheat polyploidy.

Table 3
Groups of homoeologous *HMT* genes in wheat

Homoeologous group (A:B:D)	Number of groups	Number in total	Ratio in the gene family
<i>HKMT</i>			
1:1:1	47	141	92.76%
Paired	3	6	3.94%
Orphans/singletons	5	5	3.30%
<i>PRMT</i>			
1:1:1	7	21	91.30%
Paired	1	2	8.70%
Orphans/singletons	0	0	0.00%

Structure, Phylogeny, and Domain Analysis of HMT Genes in Wheat

The analysis of the physical and chemical properties of the protein sequences encoded by 152 *TaHKMT* genes revealed that the amino acid content of different HKMT proteins varied significantly, and the molecular weight of the encoded protein was significantly different from 24.76 kD (*TaHMT145*) to 221.14 kD (*TaHMT118*). Compared to the *HKMT* gene family, the protein encoded by the *PRMT* gene family seems not to show a significant difference in physicochemical properties. In the HKMT protein, the differences in the physicochemical properties of the protein were significantly higher than those of the histone arginine methyltransferase. With the exception of *TaHMT154*, the isoelectric point (pI) of the proteins encoded by the histone arginine methyltransferase gene was less than 7.0. This shows a more consistent acidity that tends to play a role in the acidic subcellular environment, which may be due to the coding of weakly acidic protein by the *PRMT* genes. Only 5.26% (8 /152) of HKMT proteins showed high stability, while 34.78% (8/23) of PRMT members were proteins with high stability. These findings suggest that HKMT may have wider catalytic functions and may be involved in more biological processes.

Interestingly, the subcellular localization prediction analysis told us that most HKMT proteins are located in the nucleus, while most PRMT proteins are located in the cytoplasm (Table S1). This difference in localization also implies that they may have different physiological functions. The structural analysis of all *TaHMT* genes indicated that all of the members contained exons, but the number of them was quite different. The *TaHMT*

gene family has a large number of introns and exons, most of which have more than four exons, and their genetic structure is complex and diverse. Interestingly, the number of exons in the *HKMT* gene family presents a huge difference from 25 (*TaHMT113*) to 1 (*TaHMT13.etc*). A group of 11 genes, including *TaHMT13* and *TaHMT29*, contain only one exon, which may have been produced late (Fig. 2, Table S1). We presume that their functions are relatively specific, and that *TaHMT113* has the largest number of exons (25), leading to the speculation that intron loss may have occurred in the evolution of the *TaHMT* gene family.

All 152 *TaHKMT* proteins were divided into seven classes according to the classification of the SET gene family in *Arabidopsis* [40]. The seven classes have fully different domain architectures and motif compositions (Fig. 2 and Fig. 3). Class I, namely, Enhancer of zeste [E(Z)] homologs (H3K27), has unique domains, such as SWI3, ADA2, N-CoR, and TFIIB (SANT) DNA-binding domains and DNA binding protein CXC domain. These domains may facilitate the ability of the SET domain to modify histones. Class II consists of 16 SET proteins, which are homologous with *Drosophila* ortholog ASH1 and may lead to H3K4 and H3K36 modifications. This classification includes the typical associated with SET (AWS) domain, followed by a cysteine-rich post-SET domain, which may play roles in the stability of SET protein. Class III is believed to be the SET protein most responsible for the active mark H3K4me1/2/3, similar to the *Drosophila* ortholog TRX, and there are various protein domains in Class III, including conserved protein domains PWWP, FYRN/C, and the PHD domain.

Class IV is the least of all of the classifications, which is highly conservative, and it is mainly characterized by a PHD finger and a C-terminally located SET domain. It is suggested that the members of this group may be involved in the methylation of H3K4. Class V is the largest population, consisting of 63 *SET* genes homologous to *Drosophila* SU(VAR)3–9, which is divided into two different subgroups. Subgroup I contains the WIYLD domain, pre-SET domains, SET domains, and post-SET domains, while subgroup II contains a typical SRA domain at the N-terminus, playing an essential role in the establishment of heterochromatic mark H3K9me1/2/3, as well as H4K20me and H3K27me2. Classes VI and VII contain members with a SET domain whose functions are to be determined. Classes VI and VII contain 44 members, and they lack proteins specific to other classifications. Among these, class VI contains an interrupted SET domain, and class VII mostly contains Rubis-sub-bind.

The existence of SET protein contains the activity of histone methylation, and the existence of Rubis-subunit may grant it the ability to modify other proteins. The 23 *TaPRMT* proteins were divided into four classes, following previous opinion (Bedford 2007). Classes I-1 to I-3 contain 20 members. Most of them include AdoMet-MTases with arginine methyl-transfer activity, which catalyzes the formation of mono-methylarginine and asymmetric dimethylarginine. *TaPRMT18* to 20 belong to class II, which owns the core domain PRMT5, which catalyzes mono-methylarginine and symmetric dimethylarginine (Fig. 2C and Fig. 3B). Ultimately, we counted the number of genes in each subfamily. The main reason that the *HMT* gene family has a large number of members is that the class I/II/V *HKMT* subfamily is expanding its members during evolution. The ratio of their number in wheat to that in rice and *Arabidopsis* is higher than the average of 3 (χ^2 test, $P < 0.05$) (Fig. 4, Table S3).

Expression characterization and transcription regulation analysis of TaHMTs

At present, the expression level of *HMTs* in various wheat tissues without any stress pressure is still unknown. In our study, the expression of *TaHMT* genes in the root, young leaf/shoot, and the young spikes during the vegetative growth stage were significantly higher than those of other tissues at other stages, and they may be widely involved in the physiological activities of wheat growth and development. Spike, as the reproductive organ, exhibits all-round expression for the genes during vegetative and reproductive stages. This result also indicates that the gene expression levels were significantly different in different organs and tissues, which may be biased. For example, *TaHMT6* and *TaHMT18* showed obvious tissue specificity, which are characterized by high expression only in the leaves/shoots during the seedling stage, and a relatively low relative expression in young spikes (Fig. 5), suggesting that they have a special role to play in the development and morphogenesis for the plants. There are many inert genes whose expression level is always very low, such as *TaHMT15* and *TaHMT41*.

Cis-regulatory elements analysis showed that light responsiveness related elements were the most extensive elements in the promoter regions of *TaHMT* genes. Except for *TaHMT151*, all genes contained light responsiveness related elements SP1 (GGGCGG) or GT1 motif (GGTTAA), of which 29 *TaHMTs* also contained circadian elements. These elements provided candidate binding sites for light responsive factors indicating that *TaHMT* genes were related to some photoperiod biological processes, such as flowering, photosynthesis, nutrient absorption and so on. In addition to light response elements, most *TaHMT* genes (173/175) contained stress resistance related elements, such as MBS motif related to drought-inducibility, ABRE motif to abscisic acid responsive, LTR motif to low-temperature responsiveness. Some of *TaHMT* members (142/175) also contained plant hormone responsive elements, such as TGA element and AuxRR core in response to auxin, P-box and GARE motif in response to gibberellin. These findings indicated that *TaHMT* genes might have a wide range of functions in wheat (Figure S2).

MicroRNAs (miRNAs) are a class of endogenous small noncoding RNAs which are involved in the regulation of gene expression at the posttranscriptional level by degrading their target mRNAs and/or inhibiting their translation. To further explore the relationship of miRNAs-regulated functions and *TaHMT* genes expression, we used psRNATarget server to predict *HMT* genes targeting annotated microRNAs in wheat. We found that 11 different miRNAs had 15 targeting sites in body parts of 11 *TaHMT* genes (Table S4, Figure S6). Some miRNAs can act on multiple *TaHMT* genes, and one *TaHMT* gene can also interact with multiple miRNAs. The expression level of most *TaHMTs* regulated by miRNAs were apparently inhibited which may regulated by post-transcriptional cleavage manners, except the two miRNAs, which were found to may inhibit expression of the target genes at the translational level: *tae-miR5049-3p* inhibiting gene *TaHMT40* and *tae-miR1134* inhibiting gene *TaHMT172*.

Expression Analysis of HMT Genes between Parents and Hybrids

Based on our transcriptome sequencing results of JM8 (BS366×TY806), we obtained four genes whose relative expression level was significantly different ($|\text{fold change}| > 2$, $p < 0.05$) from that of their parents in spikes (Fig. 6B). These parts are *TaHMT39*, *TaHMT49*, *TaHMT149*, and *TaHMT152*, which may be involved in heterosis during the period of wheat reproductive growth in spikes. However, there are no significant DEGs screened in the tillering tissue.

We conducted a profound analysis of the transcriptome sequencing data and the expression data downloaded from wheat-expression.com. We defined the transcriptome sequencing data as shoot-TR and spike-TR, respectively, and the data obtained from the website database was defined as shoot-WE and spike-WE, representing two organizations in two periods. The two parts showed the matching degrees of 56.67% in shoot tissue, with 43.33% in spikes (Fig. 6A, Table S5). Our expectation of seven active genes at high levels of expression in the four groups was met. Finally, among them, 12 genes were screened in shoots, and 15 were screened in spikes for further analysis.

TaHMT Genes Showed Multiple Gene Expression Patterns in Hybrid Combinations

Together with the 1000 seed weight (TSW) and effective tillers per plant in wheat, four *TaHMT* DEGs and 27 genes obtained by the previous methods were analyzed by qRT-PCR. WGCNA was performed on the expression results. The relative expression level of each maternal plant was standardized as relative 1, and all of the expression results were shown in Figure S3.

Based on the trait data of TSW, the expression of 19 genes in spikes showed low correlations on the whole. The results were divided into six modules, in which three DEGs (green modules), *TaHMT49*, *TaHMT149*, and *TaHMT152* showed a certain correlation ($R = 0.32$, $p = 0.04$) (Fig. 6C). In general, their expression levels showed a positive correlation with the TSW in the samples. These three genes, on the whole, showed under-dominant expression patterns. In addition, their expression in transgressive materials was somehow higher than that of the low-parent materials corresponding to transcriptome sequencing data (Fig. 7). In the HP combination, the gene expression in F_1 were up-regulated by about 20 times at most (*TaHMT49* in Combi201), while in some LP combinations, it was down regulated by more than 30 times (*TaHMT49* in Combi199), compared with one parent. It is worth noting that this expression trend is not consistent in all materials. For example, *TaHMT149* and *TaHMT152* in Combi7 has low parental expression. This seems to be a manifestation of the complexity of heterosis. Interestingly, another DEG, *TaHMT39*, showed a low-parent dominant expression pattern in the transgressive materials (HP) and high-parent dominant expression patterns in the low-parent materials (LP) (Fig. 8A). For example, in the Combi157 (LP), the TSW of F_1 was 0.8 times than that of the maternal parent, while the expression of the gene in F_1 was over six times that of the female parent (Fig. 8A).

The high-expression genes in spikes showed multiple expression patterns in all combinations. However, no consistent rule indicated that we could not obtain a more consistent conclusion. *TaHMT71* in the M5 combination, *TaHMT118* in the M6 combination, and *TaHMT162* in the M1 combination had negative correlation expression patterns. In HP, the expression in F_1 was up-regulated by about 6 times, while in LP, the

expression was down regulated by over 25 times at most (Fig. 8B). This means that in the HP materials, the genes presented a trend of low expression and a trend of high expression in the LP material. The expression patterns of *TaHMT83*, *TaHMT84*, and *TaHMT148* were positively correlated in the M4 combination, and their expression trends were consistent with the TSW, showing an over-dominant expression pattern (Fig. 8C). Conversely, *TaHMT86*, *TaHMT89*, and *TaHMT93* showed under-dominant patterns in some combinations. Their expression in F₁ was down regulated about 5 times compared with the parents (Fig. 8D).

The expression of the 12 genes in the tillers was analyzed as follows: as a whole, the expression of the 12 genes had a negative correlation with the number of tillers, but the trend was nevertheless not very significant. No significant correlation was seen between the first two modules and the number of tillers. Therefore, the gene may have no relation to the number of tillers. The expression of the six genes in the third modules (green part) was negatively correlated with the number of tillers ($R = -0.34$, $p = 0.03$). From the analysis of the expression of each combination, the *TaHMT* genes mainly showed a negative correlation (Fig. 9A). That is, the expression of F₁ was significantly down-regulated in HP and significantly up-regulated in LP. For example, *TaHMT50* was down-regulated by 3–5 times in HP in M6 and M4, and up-regulated by 3 times in LP in M4. In particular, the expression of *TaHMT122* in maternal M4 and *TaHMT50* in M6 had a strong negative correlation ($R > -0.6$, $p < 0.05$) with the tiller number (Fig. 9B).

Discussion

Characteristics Comparison of HMT Genes in Wheat and Other Diploid Species

Allopolyploidization is the key event in the formation of common wheat as one of the main reasons for environmental adaptability [41]. Histone methylation modifying genes are widely seen in animals and plants. The increased multi-function of the SET gene evolution may ensure plants to be more adaptive to the complex natural environment [42]. In this study, the number of *TaHMT* genes identified was about three times than other five diploids and shows that the gene family has expanded. We found that 92.76% of the HKMT genes were triplet genes which is a typical characteristic of hexaploid wheat. Comparing the number of genes in each subfamily in wheat, *Arabidopsis*, and rice, the main reason is the expansion of gene number in class I E(z)-like and class V Su(var) 3-9-like, while PRMTs is relatively conservative. our study is consistent with previous studies which have shown that the SUV subfamily shows the rapid evolution [37]. We also found that ASH-like also has a high evolution speed. The ratio of this subfamily in wheat to its amount in rice and *Arabidopsis* was significantly higher than the average.

Wheat has more orthologous genes and paralogous genes than other diploid species, which is a typical characteristics of a wheat heteropolyploid. According to the homology analysis, the ortholog group of wheat achieved a higher average size of 2.32, and therefore, every *TaHMT* gene had an average of 1.32 paralogs. This number is higher than other species, which supports the previous results in wheat [43]. In wheat, 67.11% of SET gene members belong to Suv, Ash, Trx and E(z) families, and most of them have colinearity support, showing that the diversity of SET gene is mainly caused by genome duplication. We also observed that, in addition to the SET protein domain, HKMT has evolved more new protein structures such as pre-SET, SRA,

AWS, WIYLD, and SANT, among which SRA and WIYLD are structures plant-specific [44]. In fact, the presence of more auxiliary proteins such as pre-SET and post-SET may make the function of the SET protein more stable.

Conservation and Complexity of TaHMT Genes in wheat

Histone methylation can activate or repress gene expression according to different modified types. The modification types of each subfamily of HMT protein have been widely studied. For example, the activated sites (H3K36 and H3K4) include subfamilies of class II, class III, class IV and class VI, while the exercise inhibited sites (H3K9 and H3K27) contains the subfamilies of class I and class V. Unlike HKMTs, PRMTs are relatively conservative. The role of PRMTs in plants is far less clear than that of HKMTs [45]. Our results show that PRMTs are mainly located in the cytoplasm, unlike the nucleus location of HKMTs, and their physical and chemical properties tend to be consistent. According to the previous studies, PRMTs is mainly divided into three types of arginine methylation: ω -NG monhyllargine (MMA), ω -NG, NG-asymmetric dimethylargine (aDMA), and ω -NG, N'-G-symmetric dimethylargine (sDMA) [46]. One type of PRMT can often catalyze two methylation modes.

The same arginine measurement plays a major role in general regulation because of the capability of the PRMTs to deposit key activation functions (H4R3me2a, H3R2me2s, H3R17me2a and H3R26me2a) or repressive functions (H3R2me2a, H3R8me2a, H3R8me2a and H3R8me2s) [47]. The phylogenetic tree shows that wheat contains four types of PRMTs, and their catalytic sites and functions need further study. In summary, HKMTs and PRMTs modify two types of histones, and they regulate the expression of post-transcriptional genes through their multiple sites and poly-methylation. This greatly deepens the complexity of gene expression and enriches the genetic information base.

Regulation of Histone Methylation by Cis-regulatory element and miRNAs May be Involved in Regulating Yield Heterosis in Two-line Hybrid Wheat

The utilization of heterosis is the main mean of improving crop yield. Heterosis utilization can be accelerated by using male sterile lines. In wheat, TGMS lines (such as BS366) are the main means of heterosis utilization. Many excellent two-line hybrid wheat varieties with good yield performance were developed by TGMS as maternal materials [48]. The male sterility of TGMS can be inherited and has some epigenetic characteristics. Almost all wheat *HMTs* are expressed using Chinese Spring (CS) wheat expression data. However, there are also inactive genes with low expression in various tissues. A previous study has shown that *HMT* (*HKMT*) genes is significantly responsive to stress resistance [34]. We found the same results that there were a large number of stress resistance related elements in the promoter regions of *TaHMT* genes, such as MBS related to drought-inducibility, ABRE to abscisic acid responsiveness, LTR to low-temperature responsiveness. Furthermore, the most extensive number of elements we found is the element of light responsiveness. Previous studies have shown that H3K4me2/3 methylation in hybrids can regulate genes related to photosynthesis and the circadian clock, so that F1 shows yield-heterosis [21]. Our findings indicated that the

promoter region of histone methylation genes provided opportunities for combining many genes related to light response such as photosynthesis and flowering time regulation, and finally affected the yield. Besides, histone methylation has been widely reported to be involved in the regulation of flowering time in plants [49, 50]. Therefore, the discovery of the target genes of these histone methyltransferases is worthy of further exploration. In addition, a wide variety of functional elements also implied the universality of histone methylation functions. Through the analysis of tissue-specific expression, we found that spike was the concentrated expression tissue, and its expression was concentrated in the vegetative and reproductive period. This is consistent with the previous work and above findings, which show that *HMTs* are closely related to the development of flower organs and flowering [51]. In addition, the same gene subfamily shows similar expression patterns, such as *TaHMT86*, *TaHMT89* and *TaHMT93*, which are three homologous genes of Class-I. Their expression patterns are quite similar: with the exception of the expression in leaves at seedling stage, they are highly expressed in different periods of other tissues, which indicates that their functions are not redundant but are all expressed, taking important part in the growth and development of wheat.

The transcriptome data of two-line hybrid wheat is to a certain degree consistent with CS expression data. They share a number of genes that are highly expressed in both spike and shoot. We are interested in these genes and have conducted expression analysis of hybrid combinations with different yield performance. The correlation between the relative expression of 31 genes and their yield traits is relatively low. This may be due to the lower gene expression level in the experimental context, the complexity of quantitative traits controlled by multiple genes, and the uncertainty of apparent modification [52, 53]. Among the known modes of microRNA regulating target genes, the two main ways are miRNAs direct target mRNA cleavage and translational repression [54]. Of the 15 target sites we obtained, 13 of them acted in a cleavage mode indicating a main way for miRNA to regulate *HMT* genes in wheat. In rice, some miRNAs have been identified to be involved in controlling rice grain size and panicle branching by targeting severe SPL genes, including *OsSPL14* and *OsSPL16* [55–57]. In our research, the expression of *TaHMT162* in M1 combination, *TaHMT71* in M5 combination showed a significant negative correlation with TSW, while the gene expression of *TaHMT50* in M1 and M6 combination also showed a significant negative correlation with the number of effective tillers. In the regulation of miRNA, there are cleavage-regulation modes of *tae-miR1122b-3p-TaHMT162* and *tae-miR9677a-TaHMT50/71*. We speculate that the histone methylation catalyzed by these three *HMT* genes may negatively regulate the expression of related yield genes, and the regulation of miRNA reduces their mRNAs, resulting in low expression in HP but high expression in LP. Whether their roles and regulation patterns are similar to those in rice deserve in-depth discussion.

It was also found that four DEGs, including *TaHMT39*, *TaHMT49*, *TaHMT149*, and *TaHMT152*, had significant expression differences in JM8 hybrid combinations, which were also reflected in other two-line hybrid wheat combinations. Heterosis is a complex biological phenomenon. The interaction network between histone methylation gene and yield control characters and the way to participate in heterosis require further experimental verification. In general, histone methylation sites are closely related to gene expression. The methylation of lysine H3K4 and H3K36 occurs in the promoter region, which generally functions as activating transcription. However, the methylation of H3K9, H3K27, H3K79, and H3K20 are mainly distributed in the body part of the gene, which usually suppresses gene expression [58–60].

In wheat, *TaHMT39* has a high degree of homology with *AT1G77300* of *Arabidopsis*, which belongs to the ASH1-like family. *AT1G77300* can catalyze the methylation of histone H3K4 and H3K36 which play an active role in gene expression [61, 62]. Methylation of H3K4me3 in target genes catalyzed by *TaHMT39* maybe activate the expression of related genes. *TaHMT39* may regulate yield through a more complex network, which requires further experimental evidence. PRMT usually activates gene expression in high plants [11]. *TaHMT149* and *TaHMT152* are two arginine methyltransferases. *TaHMT149* (*TaPRMT18*) is the homologous gene of *AT4G31120* (*ATPRMT5*), which can catalyze the symmetrical dimethylH4R3 (sDMA) of arginine and play an active role in the vernalization, flowering time, and growth in *Arabidopsis* [63]. This work gives us an important clue and the function of *TaHMT149* requires further experimental evidence.

Regulation of Histone Methylation by lncRNA May be Involved in Regulating Yield Heterosis in Two-line Hybrid Wheat

Studies have shown that the lncRNAs are important regulation factors involved in the methylation of histone [64, 65]. So far, a large number of lncRNAs have been found and studied in *Arabidopsis*, rice, and maize, which are obviously related to the development and reproduction of panicles [66, 67]. We obtained four lncRNAs using the from the gene-screening databases. Among the four DEGs, *TaHMT39* and *TaHMT49* showed a strong correlation (Fig. 10, Table 4). They showed a similar expression trend with *HMT* gene. In JM8 hybrid combination, the abundance in F₁ was 2–20 times that of female parent, while they were down regulated by at most two times compared with male parent. Among the four DEGs, *TaHMT39* and *TaHMT49* showed a strong correlation. The expression of these four lncRNAs in the two-line hybrid wheat combinations of JM6(BS366×GLDS and JM8(BS366×TY806) were higher than those in female parents, but lower than that of male parents, showing very similar expression patterns with their possible target genes *TaHMT39* and *TaHMT49*. The expression trends of lncRNAs are highly positively correlated with mRNA of the target *HMTs*. This result also further indicated that these *HMTs* DEGs may function positively in the process of heterosis cooperating with lncRNAs for two reasons: (1) lncRNAs play an auxiliary role in the stability of mRNA; (2) lncRNAs act as enhancers to promote gene expression [68]. These provide a theoretical foundation for further study on the methylation of candidate genes and the influence of non-coding RNA on plant growth and development. The cooperative relationship between miRNA- lncRNA-HMT gene will be a promising study for the analysis of heterosis in wheat.

Table 4
Correlation analysis of *TaHMT* genes and *lncRNAs*

mRNA	lncRNA (Genome Version: TGACv1)	R	p-value	mode
<i>TaHMT39</i>	<i>TRIAE_CS42_2DS_TGACv1_177271_AA0571600 lncRNA1</i> ¹	0.9504	1.73E-08	cis
	<i>TRIAE_CS42_2DS_TGACv1_178127_AA0591260 lncRNA2</i>	0.9541	1.02E-08	cis
	<i>TRIAE_CS42_7DS_TGACv1_622339_AA2037900 lncRNA3</i>	0.9639	1.93E-09	trans
<i>TaHMT49</i>	<i>TRIAE_CS42_5BL_TGACv1_404911_AA1314590 lncRNA4</i>	0.9517	1.44E-08	trans
¹ Numbered in the order of the tables, R and mode of action are detailed in Materials and Methods.				

Conclusion

In this study, two types of histone methylation genes, histone lysine methylation gene and histone arginine methylation gene, were comprehensively identified in wheat and its relative species, and the functions of the genes in terms of structure, evolution and expression regulation by cis-elements and miRNAs were deeply explored. Through transcriptome sequencing and analysis of expression in different yield performance hybrid combinations, we found some potential DEGs that may be involved in the construction of yield-heterosis. Furthermore, the identification and expression analysis of lncRNAs with strong correlation to the DEGs were conducted to elucidate their roles for yield-heterosis. Our results enhance the understanding of histone methyltransferase gene family at the beginning, accumulated the original data for the cloning of functional genes in the future, and laid a certain data foundation for in-depth research on the growth and development of wheat and the improvement of yield.

Materials And Methods

Identification of HMT Genes in Wheat and Its Relatives

To identify the putative *HMT* genes, the wheat genome sequence, gene annotation files, and protein sequences were downloaded from the ENSEMBL database (http://plants.ensembl.org/Triticum_aestivum/Info/Index: version: release-46, as well as <http://oct2017-plants.ensembl.org/index.html>: TGACv1). The genome and protein data for *Sorghum bicolor*, *Hordeum vulgare*, and *Brachypodium distachyon* were downloaded simultaneously. The Hidden Markov Model (HMM version 33.0) was conducted using the profiles of the SET domain (Pfam accession No: PF00856) as queries with the hmmsearch program of all annotated proteins of these species ($E < 1e^{-10}$). BLAT was used to compare the protein sequences of each species with the Pfam sequence of the gene family (the following analysis software used the default parameters unless otherwise specified). The modified DOT1L (Pfam

accession No: PF08123), which participates in H3K79 methylation, was the only lysine methyltransferase that has so far been found without the SET-conserved domain [69]. We did not find genes containing the DOT1L conserved domain in wheat.

PRMT5 is the most important type II protein arginine methyltransferase. It catalyzes the formation of symmetrical arginine demethylation [70]. Consequently, the Hidden Markov Model was also conducted using the HMM profiles of PRMT5 domain (Pfam accession No: PF05185). The nine reported *Arabidopsis* PRMTs and the eight reported rice PRMTs were used to conduct a BLASTP program (Version: blast-2.10) with a wheat protein database. The standard of the BLASTP is as follows: (1) e-value < $1e^{-10}$; (2) bit-score > 100 and global alignment of identity > 50% overlap, that is, the length of the matching length must exceed 50% of the longer sequence. All of the gene loci were numbered and mapped onto corresponding chromosomes, and the non-chromosome sequences scattered across the genome were merged into unfound-chromosomes (U). The name of each gene begins with Ta, an abbreviation for *Triticum aestivum* L., and the gene loci were re-numbered based on the order of A/B/D subgroups.

Analysis of the Characteristics of TaHMT Genes

The CD searches (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and ORF finder (<https://www.ncbi.nlm.nih.gov/orffinder>) functions of NCBI were used to find the location of each gene and the location of ORF reading frame, and ProtParam (<http://web.expasy.org/protparam/>) was used to predict the molecular weight, isoelectric point, and the instability index of all HMT protein [71, 72]. It is considered to be an unstable protein when the instability index is greater than 40. Subcellular localization was analyzed using cell-ploc2.0 (<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/>). All genes' exon–intron structures were depicted by the online tool GSDS (<http://gsds.cbi.pku.edu.cn>) [73]. Protein organization sketch maps were drawn using Dog2.0 (<http://dog.biocuckoo.org/>) [74]. The 2kb sequence before the CDS of each HMT gene was extracted as the promoter region and submitted to PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) for cis-regulatory response elements prediction [75], and the results were visualized by TBtools [76]. The miRNAs annotated in wheat and their targets to *HMTs* as well as 2kb promoter sequences were predicted using the psRNATarget server (<https://plantgrn.noble.org/psRNATarget/>) with default parameters; the mismatch expectation was 0–3 [77]. The Cytoscape tool was used to visualize the interaction networks of miRNAs and *HMT* targets [78].

Evolutionary Analysis of Wheat TaHMT Gene Family

Multiple sequence alignment and the phylogenetic analysis of the *HMT* gene family were performed using MUSCLE. InParanoid 4.1 was adopted to analyze the homologous protein (OG, orthologous groups) for every two species, and Multi-Paranoid was used for analyzing multiple species OG [79]. All *HMT* genes were mapped to their respective locus in the wheat genome in a circular diagram using shinyCircos [80]. The *HMT* proteins for wheat, rice, and *Arabidopsis* were introduced into MEGA-X software to construct a phylogenetic tree using the Neighbor-Joining method with the Poisson model for substitution, and a bootstrap test of 1000

replicates was adopted for internal branch reliability [81]. In order to obtain more reliable evolutionary analysis, we constructed the phylogenetic trees of HKMT and PRMT of six species at the same time.

Tissue-Specific Expression Analysis of TaHMT Gene Family

From the expVIP-Wheat Browser (<http://wheat-expression.com/> version: TGACv1), the expression data of root, shoot, leaf, spike, and grain of all *TaHMTs* in seedling, vegetative, and reproductive growth stages were downloaded (Ramirez-Gonzalez *et al.*, 2018). MeV4.9 was used as the heatmap for gene expression between tissues, and the cluster distance was measured as a Euclidean distance. The 28 *TaHMT* genes with the highest expression of shoots at the seedling stage and spikes at the reproductive growth stage were screened. In screening for differentially expressed genes (DEGs) from the parents, the two-line hybrid wheat JM8 (Jing Mai 8), cultivated by Beijing Hybrid Wheat Engineering Technology Research Center, was used for RNA sequencing. The transcriptome data can be found in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession number PRJNA398700 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA398700>) from JM8 and parental ones (BS366×TY806) [82]. The DEGs in each comparison from among all *TaHMT* genes were identified by EdgeR, using a significance threshold of $P < 0.05$ and a fold change ≥ 2 . The first 30 *TaHMT* genes with the highest expression were also screened. To clarify the relationship between the two groups' data, we used a Venn Diagram (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to analyze all of the highest-expression genes.

Plant Materials and Treatments

All plant materials used in this article were provided by Beijing Engineering Research Center for Hybrid Wheat. Two-line hybrid wheat JM6, JM8 and other materials have been approved and widely planted in China. A total of 201 two-line hybrid wheat hybridized combinations were prepared with six female parents (M1, M2, M3, M4, M5, and M6) and 34 male parents. These materials, grown at the experimental farm in Beijing Haidian District (China, N39°54', E116°18'), were regularly watered and fertilized. Each material was planted in three rows as three biological replicates. The corresponding next generations were numbered Combi1 to Combi201 (also 1 to 201 in figures), and the male parents were identified with a prefixed P to the label, such as in P5 or P13.

After vernalization, we selected plants in the four-leaf stage, the wheat seedling stage. Tillers from root base to stem about 0.5 cm in length were obtained, which contained the apical meristem. As the wheat grew into the anther separation stage, we sampled the spikes with the awn removed. All plant materials used for the above studies were frozen immediately in liquid nitrogen and stored at -80°C for RNA extraction.

Collection of Yield Phenotypes of Multiple Two-line Hybrids

When the wheat was fully matured, all of the samples were collected according to the names of the materials. The number of tillers and spikes with well-grown individuals was investigated. We selected five wheat plants randomly and counted tillers with mature grains. The mature grains were dried and threshed, and 1000 seeds were weighed three times. The yield character data obtained above were recorded, and finally, the dominant

differences in F_1 and parents were evaluated by Student's t -test, $p < 0.05$. All phenotypes and data statistics were listed in Supplementary parts.

Data Analysis

The expression of the genes was identified as over-dominance or under-dominance if the expression level of a gene in F_1 was significantly higher or lower than, respectively, the expression level in both parents. The high-parent dominant genes and low-parent dominant genes were defined such that the expression level in F_1 genotype was significantly different from that in only one parent. Phenotypic data were processed in a similar way to the genotype expression. The materials were divided into high-parent material (HP), middle-parent material (MP), and low-parent material (LP). The yield data of five maternal parents in the hybrid combinations were homogenized. Pearson analysis was conducted to calculate the correlation between yield-character data and gene-expression data after homogenization using SPSS27 software. The Pearson correlation coefficient (R) was calculated, and the p -value was considered statistically significant when it was below 0.05.

To identify discrete groups of co-expressed genes showing their relationship to multiple traits, we applied TKW/Tillers-*HMT* weighted gene co-expression network analysis (WGCNA) to integrate expression differences into a higher order. The network was built using 42 wheat samples. Eigengenes (the normalized linear combination of genes with the largest variance in a population) carry the representative gene expression pattern for each module. The association between co-expressed gene modules and phenotypic traits was further assessed using Pearson correlations. To analyze the correlation between yield traits and lncRNAs and target gene expression levels, we calculated a correlation between yield traits and lncRNAs and target-gene expression. The t -test was used for comparison among groups.

RNA Extraction and qRT-PCR

For the expression analysis of *HMT* genes, total RNA was extracted from plant tissues using TRIzol reagent (Ambion, USA) according to the manufacturer's instructions. First-strand cDNA synthesis was performed using a TaKaRa PrimeScript™ RT Reagent Kit with gDNA Eraser (TaKaRa, Dalian, China). qRT-PCR analysis was conducted using an CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) with Takara SYBR® Premix Ex Taq™ (Tli RNaseH Plus). The primers for all genes were designed using the Primer3plus online tools (<http://www.primer3plus.com/cgi-bin/dev/primer3plus.cgi>). Each reaction was performed in triplicate in a reaction volume of 10 μ L. The qRT-PCR parameters were as follows: 95°C for 2 min, 35 cycles of 95°C for 30 s and 55°C for 15 s. For the melting curve analysis, a program of 95°C for 15 s followed by a constant increase from 55°C to 95°C was included after the PCR cycles. Expression levels of genes in samples were normalized using endogenous wheat 18s gene; the relative expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method [83].

Abbreviations

TaHMT: Wheat Histone Methyltransferase; H3K4me3: Histone H3 lysine 4 tri-methylation; HKMT: Histone lysine methyltransferase; PRMT: Protein arginine methyltransferase; CS wheat: Chinese Spring wheat; TSW: thousand seeds weight; RNA-seq: RNA sequencing

Declarations

Acknowledgments

We thank Dr. Zhang Ning from Beijing Normal University, Dr. Sun Wei-fen and Jiang Teng-fei for the guidance to some graphs. We also thank the English language editors and Dr. Pan Zhuo-xi from Waseda University in Japan for the language editing guidance.

Authors' Contributions

SRW, GJ analyzed the data and drafted the manuscript. SRW, GJ and LYJ conducted most of the experiments; GSQ, ZSQ and ZCP conceived and designed researches. GSQ and ZCP revised the manuscript. ZCP, CZB, ZFT, GJG, CJM and CXC prepared different wheat hybrid combinations and contributed all the seeds. SRW, GJ and LYJ gathered the phenotype data. All authors read and approved the final version of this manuscript.

Funding

This work was financially supported by the Important Crops Genetically Modified New Germplasm Creation (KJCX20200205) Beijing Science and Technology Plan (Z191100004019014) Beijing Academy of Agriculture and Forestry Sciences Program (QNJJ202008) and Basic scientific research funding projects of public welfare scientific research institutes in Xinjiang Uygur Autonomous Region (KY2019009).

Availability of data and materials

The genome databases of wheat and other plants used are derived from <http://plants.ensembl.org/index.html>.

The expression data of wheat comes from <http://wheat-expression.com/>.

The transcriptome data of Two-line Hybrid Wheat (JM8) were derived from <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA398700>.

We declare that the datasets required to reproduce the results of this article are included in the article and additional files available.

Ethics approval and consent to participate

All plant materials used in this article were provided by Beijing Engineering Research Center for Hybrid Wheat. All experimental procedures including use of plant or collection of the plant material comply with relevant guidelines and regulation. This article does not contain any studies with human participants or animals and did not involve any endangered or protected species.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Shewry PR: Wheat. *J Exp Bot* 2009, 60(6):1537-1553.
2. Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D et al: Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 2012, 491(7426):705-710.
3. Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K et al: Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 2017, 357(6346):93-97.
4. He F, Pasam R, Shi F, Kant S, Keeble-Gagnere G, Kay P, Forrest K, Fritz A, Hucl P, Wiebe K et al: Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat Genet* 2019, 51(5):896-904.
5. International Wheat Genome Sequencing C, investigators IRp: Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 2018, 361(6403).
6. Ramirez-Gonzalez RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A et al: The transcriptional landscape of polyploid wheat. *Science* 2018, 361(6403).
7. Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej MC, Delorean E, Thambugala D et al: Multiple wheat genomes reveal global variation in modern breeding. *Nature* 2020, 588(7837):277-283.
8. Bird A: Perceptions of epigenetics. *Nature* 2007, 447(7143):396-398.
9. Egger G, Liang G, Aparicio A, Jones PA: Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004, 429(6990):457-463.
10. Holliday R: DNA methylation and epigenetic defects in carcinogenesis. *Mutat Res* 1987, 181(2):215-217.
11. Liu C, Lu F, Cui X, Cao X: Histone methylation in higher plants. *Annu Rev Plant Biol* 2010, 61:395-420.
12. Yi X, Jiang XJ, Li XY, Jiang DS: Histone methyltransferases: novel targets for tumor and developmental defects. *Am J Transl Res* 2015, 7(11):2159-2175.
13. Ay N, Janack B, Humbeck K: Epigenetic control of plant senescence and linked processes. *J Exp Bot* 2014, 65(14):3875-3887.
14. Huang Y, Mo Y, Chen P, Yuan X, Meng F, Zhu S, Liu Z: Identification of SET Domain-Containing Proteins in *Gossypium raimondii* and Their Response to High Temperature Stress. *Sci Rep* 2016, 6:32729.

15. Niu L, Lu F, Zhao T, Liu C, Cao X: The enzymatic activity of Arabidopsis protein arginine methyltransferase 10 is essential for flowering time regulation. *Protein Cell* 2012, 3(6):450-459.
16. Thorstensen T, Grini PE, Aalen RB: SET domain proteins in plant development. *Biochim Biophys Acta* 2011, 1809(8):407-420.
17. Wang X, Zhang Y, Ma Q, Zhang Z, Xue Y, Bao S, Chong K: SKB1-mediated symmetric dimethylation of histone H4R3 controls flowering time in Arabidopsis. *EMBO J* 2007, 26(7):1934-1941.
18. Comings DE, MacMurray JP: Molecular heterosis: a review. *Mol Genet Metab* 2000, 71(1-2):19-31.
19. Meyer RC, Torjek O, Becher M, Altmann T: Heterosis of biomass production in Arabidopsis. Establishment during early development. *Plant Physiol* 2004, 134(4):1813-1823.
20. He G, Chen B, Wang X, Li X, Li J, He H, Yang M, Lu L, Qi Y, Wang X et al: Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biol* 2013, 14(6):R57.
21. Ni Z, Kim ED, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ: Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 2009, 457(7227):327-331.
22. Jenuwein T, Allis CD: Translating the histone code. *Science* 2001, 293(5532):1074-1080.
23. Li B, Carey M, Workman JL: The role of chromatin during transcription. *Cell* 2007, 128(4):707-719.
24. Moghaddam AM, Roudier F, Seifert M, Berard C, Magniette ML, Ashtiyani RK, Houben A, Colot V, Mette MF: Additive inheritance of histone modifications in Arabidopsis thaliana intra-specific hybrids. *Plant J* 2011, 67(4):691-700.
25. Kim DH, Sung S: Environmentally coordinated epigenetic silencing of FLC by protein and long noncoding RNA components. *Curr Opin Plant Biol* 2012, 15(1):51-56.
26. Tian Y, Zheng H, Zhang F, Wang S, Ji X, Xu C, He Y, Ding Y: PRC2 recruitment and H3K27me3 deposition at FLC require FCA binding of COOLAIR. *Sci Adv* 2019, 5(4):eaau7246.
27. Heo JB, Sung S: Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 2011, 331(6013):76-79.
28. Swiezewski S, Liu F, Magusin A, Dean C: Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* 2009, 462(7274):799-802.
29. Rubio-Somoza I, Weigel D: MicroRNA networks and developmental plasticity in plants. *Trends Plant Sci* 2011, 16(5):258-264.
30. Tang J, Chu C: MicroRNAs in crop improvement: fine-tuners for complex traits. *Nat Plants* 2017, 3:17077.
31. Ahmad A, Dong Y, Cao X: Characterization of the PRMT gene family in rice reveals conservation of arginine methylation. *PLoS One* 2011, 6(8):e22664.
32. Aiese Cigliano R, Sanseverino W, Cremona G, Ercolano MR, Conicella C, Consiglio FM: Genome-wide analysis of histone modifiers in tomato: gaining an insight into their developmental roles. *BMC Genomics* 2013, 14:57.
33. Fan S, Wang J, Lei C, Gao C, Yang Y, Li Y, An N, Zhang D, Han M: Identification and characterization of histone modification gene family reveal their critical responses to flower induction in apple. *Bmc Plant Biol* 2018, 18(1):173.

34. Batra R, Gautam T, Pal S, Chaturvedi D, Rakhi, Jan I, Balyan HS, Gupta PK: Identification and characterization of SET domain family genes in bread wheat (*Triticum aestivum* L.). *Sci Rep* 2020, 10(1):14624.
35. Qian Y, Xi Y, Cheng B, Zhu S, Kan X: Identification and characterization of the SET domain gene family in maize. *Mol Biol Rep* 2014, 41(3):1341-1354.
36. Xu J, Xu H, Liu Y, Wang X, Xu Q, Deng X: Genome-wide identification of sweet orange (*Citrus sinensis*) histone modification gene families and their expression analysis during the fruit development and fruit-blue mold infection process. *Front Plant Sci* 2015, 6:607.
37. Zhang LS, Ma CR, Ji Q, Wang YF: [Genome-wide identification, classification and expression analyses of SET domain gene family in *Arabidopsis* and rice]. *Yi Chuan* 2009, 31(2):186-198.
38. International Brachypodium I: Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 2010, 463(7282):763-768.
39. Schilling S, Kennedy A, Pan S, Jermini LS, Melzer R: Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol* 2020, 225(1):511-529.
40. Ng DW, Wang T, Chandrasekharan MB, Aramayo R, Kertbundit S, Hall TC: Plant SET domain-containing proteins: structure, function and regulation. *Biochim Biophys Acta* 2007, 1769(5-6):316-329.
41. Feldman M, Levy AA: Genome evolution due to allopolyploidization in wheat. *Genetics* 2012, 192(3):763-774.
42. Fawcett JA, Maere S, Van de Peer Y: Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc Natl Acad Sci U S A* 2009, 106(14):5737-5742.
43. Li X, Gao S, Tang Y, Li L, Zhang F, Feng B, Fang Z, Ma L, Zhao C: Genome-wide identification and evolutionary analyses of bZIP transcription factors in wheat and its relatives and expression profiles of anther development related TabZIP genes. *BMC Genomics* 2015, 16:976.
44. Freiman RN, Tjian R: Regulating the regulators: lysine modifications make their mark. *Cell* 2003, 112(1):11-17.
45. Paik WK, Paik DC, Kim S: Historical review: the field of protein methylation. *Trends Biochem Sci* 2007, 32(3):146-152.
46. Blanc RS, Richard S: Arginine Methylation: The Coming of Age. *Mol Cell* 2017, 65(1):8-24.
47. Auclair Y, Richard S: The role of arginine methylation in the DNA damage response. *DNA Repair (Amst)* 2013, 12(7):459-465.
48. Tang Z, Zhang L, Xu C, Yuan S, Zhang F, Zheng Y, Zhao C: Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing. *Plant Physiol* 2012, 159(2):721-738.
49. Lu F, Cui X, Zhang S, Jenuwein T, Cao X: *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. *Nat Genet* 2011, 43(7):715-719.
50. Sun C, Fang J, Zhao T, Xu B, Zhang F, Liu L, Tang J, Zhang G, Deng X, Chen F et al: The histone methyltransferase SDG724 mediates H3K36me_{2/3} deposition at MADS50 and RFT1 and promotes flowering in rice. *Plant Cell* 2012, 24(8):3235-3247.

51. Schubert D, Primavesi L, Bishopp A, Roberts G, Doonan J, Jenuwein T, Goodrich J: Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J* 2006, 25(19):4638-4649.
52. Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS: All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proc Natl Acad Sci U S A* 2006, 103(18):6805-6810.
53. Zhang HY, He H, Chen LB, Li L, Liang MZ, Wang XF, Liu XG, He GM, Chen RS, Ma LG et al: A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. *Mol Plant* 2008, 1(5):720-731.
54. Song X, Li Y, Cao X, Qi Y: MicroRNAs and Their Regulatory Roles in Plant-Environment Interactions. *Annu Rev Plant Biol* 2019, 70:489-525.
55. Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X et al: Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat Genet* 2010, 42(6):541-544.
56. Miura K, Ikeda M, Matsubara A, Song XJ, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M: OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat Genet* 2010, 42(6):545-549.
57. Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, Zeng R, Zhu H, Dong G, Qian Q et al: Control of grain size, shape and quality by OsSPL16 in rice. *Nat Genet* 2012, 44(8):950-954.
58. Fischle W, Wang Y, Allis CD: Histone and chromatin cross-talk. *Curr Opin Cell Biol* 2003, 15(2):172-183.
59. Jaskelioff M, Peterson CL: Chromatin and transcription: histones continue to make their marks. *Nat Cell Biol* 2003, 5(5):395-399.
60. Pray-Grant MG, Daniel JA, Schieltz D, Yates JR, 3rd, Grant PA: Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* 2005, 433(7024):434-438.
61. Dorighi KM, Tamkun JW: The trithorax group proteins Kismet and ASH1 promote H3K36 dimethylation to counteract Polycomb group repression in *Drosophila*. *Development* 2013, 140(20):4182-4192.
62. Grini PE, Thorstensen T, Alm V, Vizcay-Barrena G, Windju SS, Jorstad TS, Wilson ZA, Aalen RB: The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in *Arabidopsis*. *PLoS One* 2009, 4(11):e7817.
63. Pei Y, Niu L, Lu F, Liu C, Zhai J, Kong X, Cao X: Mutations in the Type II protein arginine methyltransferase AtPRMT5 result in pleiotropic developmental defects in *Arabidopsis*. *Plant Physiol* 2007, 144(4):1913-1923.
64. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY: Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 2010, 329(5992):689-693.
65. Magistri M, Faghihi MA, St Laurent G, 3rd, Wahlestedt C: Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends Genet* 2012, 28(8):389-396.
66. Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chetoor AM, Givan SA, Cole RA, Fowler JE et al: Correction to: Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol* 2018, 19(1):122.
67. Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, Qu LH, Shu WS, Chen YQ: Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction

- of rice. *Genome Biol* 2014, 15(12):512.
68. Zhu QH, Wang MB: Molecular Functions of Long Non-Coding RNAs in Plants. *Genes (Basel)* 2012, 3(1):176-190.
 69. Gao WL, Liu HL: [DOT1: a distinct class of histone lysine methyltransferase]. *Yi Chuan* 2007, 29(12):1449-1454.
 70. Peterson CL, Laniel MA: Histones and histone modifications. *Curr Biol* 2004, 14(14):R546-551.
 71. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI et al: CDD: NCBI's conserved domain database. *Nucleic Acids Res* 2015, 43(Database issue):D222-226.
 72. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF: Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 1999, 112:531-552.
 73. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G: GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 2015, 31(8):1296-1297.
 74. Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X: DOG 1.0: illustrator of protein domain structures. *Cell Res* 2009, 19(2):271-273.
 75. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S: PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* 2002, 30(1):325-327.
 76. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R: TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant* 2020, 13(8):1194-1202.
 77. Dai X, Zhuang Z, Zhao PX: psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res* 2018, 46(W1):W49-W54.
 78. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003, 13(11):2498-2504.
 79. O'Brien KP, Remm M, Sonnhammer EL: Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res* 2005, 33(Database issue):D476-480.
 80. Yu Y, Ouyang Y, Yao W: shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* 2018, 34(7):1229-1231.
 81. Kumar S, Stecher G, Li M, Knyaz C, Tamura K: MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 2018, 35(6):1547-1549.
 82. Liu YJ, Gao SQ, Tang YM, Gong J, Zhang X, Wang YB, Zhang LP, Sun RW, Zhang Q, Chen ZB et al: Transcriptome analysis of wheat seedling and spike tissues in the hybrid Jingmai 8 uncovered genes involved in heterosis. *Planta* 2018, 247(6):1307-1321.
 83. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* 2001, 25(4):402-408.

Figures

Figure 1

Collinearity analyses of all *TaHMT* genes. All 175 genes were mapped to their respective locus in the wheat genome in a circular diagram. Sub-genomes are indicated by different shades of blue (outer track). Homoeologous genes were inferred by phylogeny (shown in the Materials and Methods section).

Figure 2

Analysis of phylogeny, structures, and protein motifs of all 175 TaHMTs in wheat. **(A)** and **(B)** Gene phylogeny, structures, and protein motifs of 152 *HKMT* genes. **(C)** Gene phylogeny, structures, and protein motifs of 23 *PRMT* genes. The full-length coding sequence (CDS) were aligned with Clustal X, and the phylogenetic tree was constructed using the neighbor-joining method by MEGA-X. In the gene structures, the purple boxes represent exons, and the green lines represent introns. The untranslated regions (UTRs) are indicated by red

boxes. The sizes of the exons and introns can be estimated using the scale at the bottom. The protein motifs were drawn by different colored boxes as shown in the figure using Dog 2.0.

Figure 3

The neighbor-joining phylogeny of HKMT **(A)** and PRMT **(B)** proteins from wheat, rice, and *Arabidopsis*. Wheat proteins are colored red and are subclade-specific, whereas rice and *Arabidopsis* genes are in blue and green, respectively. The HKMT proteins were divided into seven subfamilies and PRMT were divided into four subfamilies.

Figure 4

The number of *HMT* genes identified per subfamily in **(A)** *Arabidopsis*, **(B)** rice, and **(C)** wheat. **(D)** The ratio of total *HMT* gene numbers to those in all subfamilies is shown for wheat: rice (red) and wheat: *Arabidopsis* (purple). In (D) the expected ratio (3:1) is indicated by a black line, and asterisks mark a significant deviation from the expected value using χ^2 test, $P < 0.05$.

Figure 5

Tissue and organ expression of all identified *HMT* genes. The tpm values of all gene expression data are retrieved from downloaded from wheat-expression.com. The letters A/B/C/D/E/F/G/H/I in the figure represent seedling-roots/seedling-leaves, shoots/vegetative-roots/vegetative-leaves, shoots/vegetative-spike/reproductive-roots/reproductive-leaves, and shoots/reproductive-spike, respectively. The heatmap for gene expression patterns was generated with the software MultiExperiment Viewer (MeV) and the cluster distance was measured as Euclidean distances. Green modules indicate low expression level and red

modules indicate high expression level. The concentrate of expression was in spike both at vegetative growth and reproductive growth.

Figure 6

Screening and expression analysis of differentially expressed genes. **(A)** Pooled analysis of high expression level genes from two databases using Venn diagrams. **(B)** The mRNA abundance of four genes in JM8 combinations from transcriptome data, presented a significant difference between any one parent and F1. The relative expression levels were significantly different, which met the $|\text{fold change}| > 2$ and $p < 0.05$ in spikes. Values are given as means $\pm$ SDs ($n = 3$). **(C)** WGCNA analysis of 19 genes in spikes, wherein only three genes show correlations with TSW.

Figure 7

Analysis of three differentially expressed genes in five different two-line hybrid combinations. **(A)** Combi175, a representative of HP material. Combi99, a representative of LP material. The length of the red ruler represents 1 cm. The specific data were listed in the Additional file 2: Table S4. **(B)** Expression levels of three differentially expressed genes in five two-line hybrid combinations. Values are given as means $\pm$ SDs ($n = 3$).

Figure 8

Multiple expression patterns in all combinations. **(A)** The expression results *TaHMT39* in HP and LP materials. **(B)** Negative correlation expression patterns of *TaHMT71*, *TaHMT118*, and *TaHMT162*. **(C)** Positive correlation expression patterns of *TaHMT83*, *TaHMT84*, and *TaHMT148*. **(D)** *TaHMT86*, *TaHMT89*, and *TaHMT93* showed under-dominance patterns in M1 and M3 combinations. The relative expression of the female parent of each combination was relative 1, which is not shown in Figure 8A but appears in the blue-colored box in Figure 8B–D, marked as a dotted line. Red represents the relative expression of F1, and the yellow adjacent to the right represents the relative expression of its male paternity. ++ marks applicability of all HP combinations, and – marks LP combinations. Values are given as means $\pm$ SDs ($n = 3$).

Figure 9

Expression analysis 12 *TaHMT* genes in tillers. **(A)** WGCNA analysis of 12 genes in tillers. **(B)** The expression of *TaHMT50* and *TaHMT122* in specific combinations showed a strong negative correlation with effective tiller number. The relative expression of the female parent of each combination was relative 1, which is shown in the blue-colored and marked as a dotted line. Red represented the relative expression of F_1 , and the yellow adjacent to the right represents the relative expression of its male paternity. ++ marks applicability for all HP combinations, and - indicates LP combinations. R represents the Pearson correlation coefficient. Values are means $\pm$ SDs ($n = 3$).

Figure 10

Abundance of four *lncRNAs* in JM8 combination **(A)** and JM6 combinations **(B)**. The relative expression levels were significantly those that which met the $|\text{fold change}| > 2$ and $p < 0.05$ in spikes. **indicates a significant difference from the one parent at $p < 0.01$ according to Student's *t*-test.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1Basiccharacteristicfeaturesof175HMTgenesidentifiedinwheat.xlsx](#)
- [TableS2LocationandHomoeologousofHMTgenesinwheat.xlsx](#)
- [TableS3DistributionofHMTfamilymembersinsubgroupsinwheatArabidopsisandrice.xlsx](#)
- [TableS4TargetinganalysisofmiRNAsandHMTgenesinwheat.xlsx](#)
- [TableS5mRNAabundanceoftop30highlyexpressedgenes.xlsx](#)
- [TableS6Phenotypicstatistics.xlsx](#)
- [TableS7PrimersusedforqRT-PCR.xlsx](#)
- [Additionalfile1.pdf](#)