DR NING LI (Orcid ID : 0000-0002-7721-8222)

Article type : - Regular Manuscript

DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids

Ning Li[1], Chunming Xu[1], Ai Zhang[1], Ruili Lv[1], Xinchao Meng[1], Xiuyun Lin[1], Lei Gong[1,2], Jonathan F. Wendel[2] and Bao Liu[1]

[1]Key Laboratory of Molecular Epigenetics of the Ministry of Education (MOE), Northeast Normal University, Changchun 130024, P. R. China

[2]Department of Ecology, Evolution & Organismal Biology, Iowa State University, Ames, Iowa 50011, USA

Authors for correspondence:
Lei Gong
Tel: +86 135 0433 7440
Email: gongl100@nenu.edu.cn

Bao Liu
Tel: +86 0431 85099822
Email: baoliu@nenu.edu.cn

ORCID iD: http://orcid.org/0000-0001-5481-1675

Received: 5 January 2019

Accepted: 21 March 2019

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/nph.15820

Total word count (excluding summary, references and legends):	6314	No. of figures:	6 (all in color)
Summary:	190	No. of Tables:	0
Introduction:	1156	No. of Supporting Information files:	16 (Fig. S1-6; Table S1-9; Method S1)
Materials and Methods:	681		
Results:	3180		
Discussion:	1051		
Acknowledgements:	56		

Summary

- Allopolyploidization, which entails interspecific hybridization and whole genome duplication (WGD), is associated with emergent genetic and epigenetic instabilities that are thought to contribute to adaptation and evolution. One frequent genomic consequence of nascent allopolyploidization is homoeologous exchange (HE), which arises from compromised meiotic fidelity and generates genetically and phenotypically variable progenies.
- Here, we used a genetically tractable synthetic rice segmental allotetraploid system to interrogate genome-wide DNA methylation and gene expression responses and outcomes to the separate and combined effects of hybridization, WGD and HEs.
- Progenies of the tetraploid rice were genomically diverse due to genome-wide HEs that affected all chromosomes, yet they exhibited overall methylome stability. Nonetheless, regional variation of cytosine methylation states was widespread in the tetraploids. Transcriptome profiling revealed genome-wide alteration of gene expression, which at least in part associates with changes in DNA methylation. Intriguingly, changes of DNA methylation and gene expression could be decoupled from hybridity and sustained and amplified by HEs.
- Our results suggest that HEs, a prominent genetic consequence of nascent allopolyploidy, can exacerbate, diversify and perpetuate the effects of allopolyploidization on epigenetic and gene expression variation, and hence may contribute to allopolyploid evolution.

Key words: DNA methylation, homoeologous exchange, homolog copy number variation, gene expression, segmental allopolyploidy, sustained epigenetic diversity

Introduction

Polyploidization, or whole genome duplication (WGD), is a pervasive and evolutionarily creative force that has shaped the evolution of all higher plants (Wendel, 2000; Van de Peer *et al.*, 2009; Jiao *et al.*, 2011; Soltis & Soltis, 2012). Given sufficient time, each successful WGD episode is followed by a whole-genome (or nearly so) diploidization process (Freeling *et al.*, 2015; Van de Peer *et al.*, 2017; Zhao *et al.*, 2017; Cheng *et al.*, 2018; Wendel *et al.*, 2018). Thus, genomically, the evolutionary history of angiosperms entails a cyclic interplay of WGD and diploidization, processes thought to enhance evolvability by generating genetic variability and heritable phenotypic diversity (Soltis & Soltis, 2009; Schubert & Lysak, 2011; Han *et al.*, 2015; Wendel, 2015; Dodsworth *et al.*, 2016; Soltis & Soltis, 2016; Mandakova & Lysak, 2018). Relatively recent polyploids, i.e., neopolyploids and mesopolyploids, which contain cytogenetically recognizable duplication(s) of whole chromosome sets (Schubert & Lysak, 2011), account for about one-third of all extant vascular plant species (Mayrose *et al.*, 2011) and include many of our important crops (Renny-Byfield & Wendel, 2014), often contain largely intact duplicated chromosome sets.

There are two major types of neopolyploidy, i.e., autopolyploidy, which is WGD of a single species genome, and allopolyploidy, which frequently is WGD concomitant with interspecific hybridization (Stebbins, 1947). Under natural settings, however, there is a continuum of intermediates between these two poles, and many polyploids may be included under the umbrella of segmental allopolyploidy (Stebbins, 1947; Grant, 1981; Ramsey & Schemske, 2002; Wendel & Doyle, 2005; Spoelhof *et al.*, 2017). Notably, segmental allopolyploidy can fully recapitulate the genomic properties of allopolyploidy in the sense that they contain both homologs and homoeologs (Wendel & Doyle, 2005), and in fact sometimes the distinction between segmental allopolyploidy and allopolyploidy can be blurred, for example in young tetraploid species of *Tragopogon*, e.g., *T. miscellus* (Chester *et al.*, 2012).

Early generations of allopolyploids are often associated with emergent genetic and epigenetic instabilities (Wendel, 2000; Comai, 2005; Salmon *et al.*, 2005; Adams, 2007; Chen, 2007; Otto, 2007; Doyle *et al.*, 2008; Leitch & Leitch, 2008; Ainouche & Jenczewski, 2010; Feldman & Levy, 2012; Madlung & Wendel, 2013; Diez *et al.*, 2014; Song & Chen, 2015; Wendel *et al.*, 2016; Dion-Cote & Barbash, 2017; Ding & Chen, 2018). The most pervasive and immediate genetic consequence of nascent polyploidy is disruption of normal meiosis due to mismatches between the meiotic machinery of diploids that now must adapt to handle

the abruptly doubled chromosome set (Hollister, 2015; Mercier *et al.*, 2015; Bomblies *et al.*, 2016). Consequently, multivalents and univalents occur due to compromised pairing fidelity, resulting in homoeologous exchanges (HEs) and aneuploidy (Pecinka *et al.*, 2011; Higgins *et al.*, 2018; Lloyd *et al.*, 2018). Conceivably, while most aneuploidies that cause deficiency and/or chromosome-wide dosage imbalance will be rapidly purged due to lethality or lack of fitness, many progenies with HEs may remain and be transgenerationally persistent due to the frequent (Gou *et al.*, 2018), though not everpresent (Zhang *et al.*, 2011; Gong *et al.*, 2014; Lloyd *et al.*, 2018), mutual functional compensation of homoeologs (Xiong *et al.*, 2011; Chester *et al.*, 2012). HEs generate alterations of the otherwise 2:2 homoeolog ratio, and hence may impact epigenetic stabilities (e.g., DNA methylation and histone modifications) and gene expression due to inherent differences between homoeologs as well as *de novo* alterations that arise following genome merger and doubling.

DNA methylation in plants is a relatively stable and transgenerationally heritable epigenetic mark, yet it also is known to be dynamic in its genomic distribution, undergoing spontaneous epimutations and being subject to intrinsic and extrinsic perturbations (Zhang *et al.*, 2006; Schmitz *et al.*, 2013; Kawakatsu *et al.*, 2016; Niederhuth *et al.*, 2016; Quadrana & Colot, 2016; Takuno *et al.*, 2016). Because plants do not set aside an early, clearly defined germline (Grossniklaus, 2011; but see Lanfear, 2018), somatically acquired DNA methylation modifications can be transgenerationally inherited, potentially contributing to adaptation and evolution. Among the intrinsic causative factors that may drive DNA methylation evolution, interspecific hybridization and WGD are perhaps the most pervasive. Indeed, studies in diverse plant taxa, including *Arabidopsis* (Madlung *et al.*, 2002), *Spartina* (Salmon *et al.*, 2005), *Brassica* (Song *et al.*, 1995; Gaeta *et al.*, 2007; Xu *et al.*, 2009), *Tragopogon* (Sehrish *et al.*, 2014) and *Triticum-Aegilops* complex (Shaked *et al.*, 2001; Kenan-Eichler *et al.*, 2011; Zhao *et al.*, 2011) have shown that allopolyploidization causes extensive changes in both DNA methylation and gene expression (Adams, 2007; Song & Chen, 2015). However, the possible distinct effects of hybridization, WGD and HE during allopolyploidization have not hitherto been explored.

It was recently shown in allotetraploid rapeseed (*Brassica napus*) that significant and nonadditive gene expression alterations can be traced to HEs (Lloyd *et al.*, 2018). However, it remains unclear if and to what extent changes in DNA methylation are correlated with HEs and gene expression. A recent methylome study of a newly synthesized autotetraploid rice demonstrated the occurrence of genome-wide DNA methylation, with a prominent type being hypermethylation of class II transposable elements (TEs) involving CHG and CHH contexts

(Zhang *et al.*, 2015), suggesting that changes in DNA methylation in newly formed autopolyploids plays a role in genome stabilization (Zhang *et al.*, 2015). It is not clear, however, if DNA methylation changes contribute to enhanced evolvability in allopolyploids, that is, whether methylation alteration generates heritable variation (Otto, 2007).

The two subspecies of cultivated Asian rice (*Oryza sativa* L.), *indica* and *japonica*, share high homologous and syntenic genomes but also harbor substantial genetic and epigenetic differentiation due to their distinct domestication histories and human selection under different climatic and edaphic conditions (Civan *et al.*, 2016; Choi *et al.*, 2017; Choi & Purugganan, 2018). High-quality genome sequences are available for the two laboratory genotypes representing the two subspecies, cv. 9311 for subsp. *indica* and cv. Nipponbare for subsp. *japonica*. Our previous studies have shown that segmental allotetraploids can be readily produced using reciprocal F1 hybrids of the two cultivars (Xu *et al.*, 2014). These tetraploids at the immediate generations (S0 and S1) already manifested unique transcriptome profiles compared with those of both of their parental lines and the corresponding F1 hybrids, due to distinct *cis*-/*trans*-regulations (Xu *et al.*, 2014). Moreover, extensive expression rewiring occurred at a set of selected gene loci in later generations of the tetraploids at the population level, due to variation of homoeolog copy numbers among the individuals (Sun *et al.*, 2017). Together, this system provides a genetically tractable system to explore the immediate genetic, epigenetic, gene expression and phenotypic consequences of allopolyploidization, using advanced generation tetraploids derived from reciprocal F1 hybrids.

Here, we used this experimental system to interrogate genome-wide DNA methylation and gene expression in response to the separate and combined effects of hybridization, WGD and HEs. We demonstrate that the three factors have both counteracting and exacerbating effects on changes of DNA methylation and gene expression in progenies of the rice tetraploids, which acting together may contribute to adaptation, diversification and evolutionary success.

Materials and Methods

Plant materials

The rice segmental allotetraploids used in this study were the 5th selfed generation (S5) of colchicine-doubled tetraploids from reciprocal F1 hybrids (N9 and 9N), created by crossing rice cultivars Nipponbare (*Oryza sativa* ssp. *japonica*) and 9311 (*O. sativa* ssp. *indica*) (Xu *et al.*, 2014). The diploid Nipponbare, 9311, F1 hybrids (9N and N9) and four tetraploid individuals (99NN-3, 99NN-7, NN99-4 and NN99-8) were used for all experiments. Rice seeds were germinated in Petri dishes with distilled water at room temperature. After germination, seedlings were transplanted into soil in the greenhouse, under a 16/8 hour and 26/20 ^{0} C light/dark regime and a relative humidity of about 50%. Leaf tissues were collected from 6-week-old seedlings when the 4th-leaves were fully expanded. For diploids, leaves were harvested from 5 individuals as a pool. For parental lines and F1 hybrids, multiples individuals (>10) for each line were used, while for tetraploids, leaves were harvested on a per individual basis. Collected leaves were immediately frozen in liquid nitrogen. Genomic DNA was isolated using a modified CTAB method (Doyle & Doyle, 1987). Total RNA was isolated using Trizol (Invitrogen) according to the manufacturer's protocol.

Data collection, processing and analysis

Purified DNA and RNA samples were sent to Beijing Genomics Institute (BGI, Shenzhen, China) for whole genome resequencing, whole genome bisulfite sequencing (WGBS) and RNA-seq. The bisulfite treatment, library construction, cluster generation, and next-generation sequencing were carried out using standard protocols. Following sequencing (HiSeq-2000, Illumina), the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was used to evaluate sequence quality and filter out adaptors and low-quality reads (keeping reads with >80% of bases having a quality score > than 20). Clean data have been deposited at the SRA database (http://www.ncbi.nlm.nih.gov/sra/) with accession number PRJNA514100. Detailed experimental and analytical procedures were described in Methods S1.

Analyses of genomic composition

Whole-genome resequencing was used to analyze genomic compositions of the four tetraploid individuals. The total sequencing yield was 82-173 million reads (**Table S1**). For each sample, a minimum of 92% of the rice genome was covered by at least one read, corresponding to 21- to 41-fold sequencing depth (**Table S1**). First, we assessed genomic polymorphism between the parental cultivars, Nipponbare and 9311, representing the two *O.*

sativa subspecies, *indica* and *japonica*. Compared with the Nipponbare reference genome (MSU 7.0), 1,719,154 single nucleotide polymorphisms (SNPs) were detected in 9311, which translate into 0.46% divergence between Nipponbare and 9311 at the nucleotide sequence level (**Table S2**). Most single nucleotide substitutions are C/T and G/A transitions, which together accounted for 71.9% of all SNPs. These SNPs were used to distinguish the two subgenomes and estimate homoeologous exchanges (HEs) in the tetraploids.

Analyses of methylome

Whole-genome bisulfite sequencing (WGBS) was used to construct the methylomes, at single base resolution, for each of the four tetraploids, each of the two F1 hybrids and the two parents. A total of 49-66 million base pairs of sequencing yield was generated (**Table S3**). Before alignment, SNPs between 9311 and Nipponbare in the reference genome (MSU 7.0) were substituted with the corresponding degenerate bases. For each sample, a minimum of 83% of both strands of the rice genome were covered by at least one read, equivalent to 18- to 27-fold depth of the genome coverage (**Table S3**).

Calculations of differentially methylated region (DMR) and differentially methylated cytosine (DMC) were described in Methods S1. To explore whether the relative homoeolog compositions in tetraploids would associate with DNA methylation alterations, we calculated DMRs separately according to the proportional homoeolog compositions. This was done by comparing genomic regions representing each of the five types of homoeolog compositions separately, in each of the four tetraploids. Specifically, for regions of the two homozygous types, i.e., homo-9 or homo-N, the corresponding parents (9311 or Nipponbare) were used as controls to calculate DMRs, while for regions containing the three types of heterozygosity (N:9 = 1:3, 2:2 or 3:1), the corresponding *in silico* "hybrids" constructed by mixing the WGBS reads of Nipponbare and 9311 according to the corresponding homoeolog ratios were used as controls to calculate the expected DMRs. F1 hybrids for both crossing directions were also included in the comparisons.

Data availability

All raw reads of genome resequencing, methylomes and transcriptomes generated in this study have been deposited in the public database of National Center of Biotechnology Information under PRJNA514100 (https://www.ncbi.nlm.nih.gov/sra/).

Results

Genome composition of the segmental allotetraploids

Whole-genome resequencing revealed extensive homoeologous exchanges (HEs) in all four tetraploid individuals of the 5th selfed generation studied. These HEs were categorized into five types of *japonica-indica* homoeolog composition: (*i*) 4 copies of Nipponbare and 0 of 9311 (designated homo-N); (*ii*) 3 copies of Nipponbare and 1 of 9311; (*iii*) 2 copies each of Nipponbare and 9311; (*iv*) 1 copy of Nipponbare and 3 of 9311; and (*v*) 0 copy of Nipponbare and 4 of 9311 (designated homo-9). We found tetraploids of the reciprocal crosses, NN99-4 and NN99-8 vs 99NN-3 and 99NN-7, differed considerably in their particular combinations of homoeolog ratios on every chromosome (**Fig. 1a**); thus, the two sets of reciprocally generated synthetic allopolyploids are strikingly different in genomic composition by the 5th selfed generation, having experienced different histories of HEs (**Table S4**). Notably, however, this may simply be due to earlier segregation rather than parent-of-origin effect (e.g., maternal or cytoplasmic effect), as documented earlier (Xu *et al.*, 2014). In contrast, the two synthetics within each of the two sets of reciprocals were quite similar to each other, sharing the vast majority of their HEs, notwithstanding their evident distinctions (**Fig. 1b, c; Table S4**); these differences reflect either segregating heterozygosity and/or still ongoing HEs despite their sibling relationship. Consequently, there is only a small proportion (< 12%, 44.38 out of 372.63 Mb) of genomic regions where all four allotetraploids shared the same genomic composition, or history of HEs (**Fig. 1d**), emphasizing the genetic diversity that rapidly arises from this process; examples of shared fixation of HEs include the upper half of chromosome 2 and the middle of chromosome 6 (**Fig. 1a**). Notably, although there are few regions fixed for the same HEs among all four lines, the extent of homozygosity within each of the lines was high (**Fig. 1e; Table S4**); regional homozygosity (i.e., homo-N or homo-9) ranged from 48% to 63%, indicating that about half of these tetraploid genomes have been homogenized to either the Nipponbare or 9311 genome, with the remainder being still in a heterologous state. We also noted that the composition of the Nipponbare parent (homo-N) was approximately equivalent among lines (ranging only from 26-28%), whereas the genomic fraction homozygous for the 9311 parent (homo-9) was more variable (22-36%), thus yielding tetraploid lines that vary overall in their relative parental composition from approximately equal (both 99NN lines) to biased (both NN99 lines) (**Fig. 1e; Table S4**).

Patterns of localized DNA methylation changes in the segmental allotetraploids recapitulate epigenetic divergence between the parental subspecies

Whole-genome bisulfite sequencing (WGBS) identified ~14.6 - 25.8 million methylated cytosines (^{m}Cs) from the total WGBS-reads across the samples, of which > 40% are in the CG context, while 29% and 31% are in CHG and CHH contexts, respectively (**Fig. S1; Table S3**). This context-partitioning ratio is broadly in line with previous methylome profiling results in rice (Zemach *et al.*, 2010; Chodavarapu *et al.*, 2012; Li *et al.*, 2012; Hu *et al.*, 2014; Deng *et al.*, 2016). The overall levels of genome-wide methylation in all three contexts in the tetraploids are not markedly different from each other nor from those of the diploid F1 hybrids and parents (**Fig. 2a; Fig. S2, S3**). Statistically, however, methylation levels along the length of protein-coding genes, genes annotated as transposons or retrotransposons, i.e., TE-genes, class I TEs (retrotransposons) and class II TEs (transposons) are all significantly different in each of the three contexts in most of the pairwise comparisons (**Table S5 and S6**). This suggests that although both hyper- and hypo-methylation alterations occurred widely in the tetraploids in a locus-specific manner, this was insufficient to alter overall methylation levels, with hyper- and hypo-methylation largely offsetting each other.

Among lines, the largest numbers of differentially methylated cytosines (DMCs), including both hyper- and hypo-DMCs in all three contexts, were between the parental cultivars, 9311 and Nipponbare (**Fig. 2b; Table S7**), in keeping with their status as different subspecies (Chodavarapu *et al.*, 2012). Interestingly, the lowest numbers of DMCs (both hyper and hypo) in all contexts were in the two F1 hybrids in comparison with their parental average (**Fig. 2b; Table S7**). This suggests largely stable inheritance of parental ^{m}Cs to the F1 hybrids with only minor *de novo* repatterning in the hybrid genome, also in line with a previous study (Chodavarapu *et al.*, 2012). In contrast, large numbers of DMCs, again in all contexts, were detected in all four tetraploids compared with the parental averages (**Fig. 2b; Table S7**), suggesting extensive locus-specific remodeling of DNA methylation in the tetraploid genomic environment, albeit the overall conservation at the methylome scale.

Markedly more CG-DMCs in the tetraploids were mapped to protein-coding genes than to other regions, followed by those mapping to intergenic regions, with far fewer in TEs and especially class II TEs (**Fig. 2c; Fig. S4**). This result indicates that ^{m}CGs are largely stable in TEs of the tetraploids. For CHG-DMCs, all types of genomic regions except class II TEs showed similar numbers of DMCs (**Fig. 2c; Fig. S4**). Class II TEs showed a strikingly small number of CHG-DMCs (**Fig. 2c; Fig. S4**), indicating their ^{m}CHG has been particularly stable in the tetraploids. CHH-DMCs occurred uniformly in all genomic regions (**Fig. 2c; Fig. S4**).

Intriguingly, all patterns of differentially methylated cytosines (DMCs) in the tetraploids vs the parental mix (*in silico* "hybrids") with respect to their context-partitioning and variable abundance across the genomic features mirrored those seen between the parental subspecies (**Fig. 2c**). This suggests commonality between allopolyploidy-induced immediate DNA methylation alteration and that precipitated under different natural and artificial selections between the rice subspecies.

Impact of genomic composition on DNA methylation variation in the segmental allotetraploids

To explore whether relative homoeolog composition (**Fig. 1a**) might be associated with the localized alteration of DNA methylation in the tetraploids (**Fig. 2**), we separately tabulated differentially methylated regions (DMRs) for each of the three sequence contexts, CG, CHG and CHH, relative to the proportional homoeolog compositions. In all four tetraploids, we identified a large albeit variable number of DMRs, for all three sequence contexts, in genomic regions harboring each of the five types of homoeolog composition relative to the corresponding controls (**Table S8**). Remarkably, the two types of homozygous regions, i.e., homo-N and homo-9, were the most dramatically different in methylation relative to their parental counterparts, showing 2-5 times as many DMRs as other genomic combinations (**Fig. 3**). This was true for all three sequence contexts, indicating that genomic composition plays a major role in the genesis and/or maintenance of DMRs. F1 hybrids at these genomic regions also showed a higher number of DMRs than those at heterozygous regions; this is because their DMRs were also defined by comparison to one inbred parent only. Notwithstanding this escalating factor causing more DMRs in the hybrids, the DMR numbers of the tetraploids are still all substantially greater than those of the corresponding hybrids (**Fig. 3**). Heterozygous regions (N:9 = 1:3, 2:2 or 3:1) also manifested DMRs in all three contexts but with much smaller numbers than those of the homogenized regions (**Fig. 3**).

Contrasting inter-subgenomic DNA methylation interactions in hybrids and segmental allotetraploids

The foregoing analyses of DNA methylation in the tetraploids and F1 hybrids were based on total methylation levels without separating the two constituent subgenomes. Taking advantage of the diagnostic SNPs identified from our whole-genome resequencing data, the methylation levels of each subgenome in the hybrids and tetraploids could be separately quantified using the overall methylome data. Similar to the results presented in the foregoing sections, when compared with the corresponding diploid parents, subgenome DNA

methylation levels in hybrids and tetraploids did not show conspicuous methylation alterations from those of the corresponding parents, for all three contexts (**Fig. S5**).

We next investigated if and to which extent the original DMRs between the two parental cultivars were inherited to, and new DMRs were generated *de novo* due to subgenomic interactions in, the F1 hybrids and tetraploids. To avoid confounding factors, for this purpose we only analyzed genomic regions with a balanced (2:2) homoeolog constitution, which account for *ca.* 30% of the tetraploid genome (**Fig. 1a**). Thus, seven groups of inter-genomic (between parents) or inter-subgenomic (in hybrids and tetraploids) DMRs for each sequence context (CG, CHG or CHH) can be categorized: group 1 refers to DMRs between the parental genomes; group 2 refers to DMRs between the subgenomes in the F1 hybrids; group 3 refers to DMRs between the two subgenomes in the tetraploids; group 4 refers to DMRs common between groups 1 and 2; group 5 refers to DMRs common between groups 1 and 3; group 6 refers to DMRs common between groups 2 and 3; and group 7 refers to DMRs common between groups 1, 2 and 3.

We used a Venn diagram to illustrate one example of the relationships among, and proportions of, these seven groups of DMRs of each of the three sequence contexts (**Fig. 4a**), while the relative proportions of hypo- vs hyper-methylation of all three contexts (CG, CHG and CHH) in each of the seven groups of DMRs in the four tetraploids are illustrated in a heatmap (**Fig. 4b**). These analyses led to the following major observations: First, there are always more DMRs in group 1 than in group 2 in all three contexts (CG, CHG and CHH) and for both DMR directions (i.e., Nipponbare > 9311 and Nipponbare < 9311), indicating that hybridization eliminates a substantial number of DMRs of all types between the parental subspecies. Second, most types of DMRs in group 3, especially those of CG-DMRs in the Nipponbare > 9311 category and also those of CHG-DMRs in the Nipponbare > 9311 category, are markedly more numerous than those of the corresponding group 2, indicating that the reduced inter-subgenomic differences in DNA methylation in the F1 hybrids was re-augmented in the tetraploids, a conclusion further supported by the substantially reduced numbers of DMRs of all types in group 4. Third, although both groups 5 and 6 DMRs of all types contain smaller numbers than the other five categories, group 5 DMRs are uniformly higher in number than those in group 6, indicating different inter-subgenomic interactions in the F1 hybrids and tetraploids, as well as some common metastable genomic regions for DNA methylation repatterning. Some representative DMRs are illustrated as IGV (integrative genomic view) snapshots **(Fig. 4c**). Altogether, these results indicate that a major effect of genome merger is to attenuate the original DNA methylation divergence that evolved during

subspecies divergence, while a major role of WGD is to re-augment and exacerbate these differences in DNA methylation, with HEs further diversifying and entangling these genetic and epigenetic outcomes. This clearly points to contrasting inter-subgenomic DNA methylation interactions in the hybrids vs the tetraploids.

Changes of gene expression in the segmental allotetraploids due to hybridization, WGD and HEs, and their relationship to localized DNA methylation alteration

A recent study in allotetraploid rapeseed (*Brassica napus*) demonstrated that HEs cause significant changes in gene expression, proportional to homoeolog copy number (Lloyd *et al.*, 2018). Similarly, we reported that extensive homoeologous expression rewiring occurred in progenies of the segmental rice allotetraploids, and that this was also largely dependent on homoeolog copy numbers (Sun *et al.*, 2017). These prior results, together with the finding that differentially methylated genes are more likely to show differential expression than are equally methylated genes in rice hybrids (Chodavarapu *et al.*, 2012), prompted us to investigate changes of gene expression in relation to the separate and combined effects of hybridization, WGD and HEs in the tetraploids, and whether these changes are related to the localized alterations of DNA methylation.

To address these questions, we profiled genome-wide gene expression of the tetraploids, their parents and the corresponding F1 hybrids. We tabulated differentially expressed genes (DEGs) in genomic regions that had a N:9 homoeolog ratio of 2:2, 4:0 or 0:4 in the tetraploids, in comparison with expression levels of the genes in the corresponding parental controls. As noted (**Fig. 1b**), these three types of homoeolog compositions occupied similar proportions (each *ca.* 25-30%) in each of the four S5 tetraploids. The control for the analyses of genes in genomic regions with a 2:2 ratio (i.e., no HEs) is an *in silico* "hybrid" constructed by mixing the parental RNA-seq reads at a ratio of 1:1. Accordingly, DEGs are those that are expressed at significantly higher or lower levels than in the 1:1 *in silico* "hybrid". Controls for genes in genomic regions with a 4:0 or 0:4 ratio are the corresponding inbred parents, using normalized counts. To analyze the effects of hybridization alone vs hybridization coupled with WGD and/or HEs, DEGs were also tabulated in the F1 hybrids using the *in silico* "hybrid" with a 1:1 parental RNA-seq mixture as a control.

Our main results are summarized in **Fig. 5**, which showed that all three types of genomic compositions manifested from >300 up to 10,00 DEGs across the tetraploids and F1 hybrids. For genomic regions with a balanced homoeolog composition, more DEGs were found in each of the four tetraploids than in the corresponding F1 hybrids, an observation that is most

prominent for down-regulated DEGs (**Fig. 5a; Table S9**); this demonstrates that polyploidy, even in the absence of HEs, has a marked effect on generating DEGs. This is consistent with our previous result using the immediate generations (S0 and S1) of this tetraploid rice system, in which few HEs occurred but transcriptome profiles of tetraploids are significantly different from those of the F1 hybrids (Xu *et al.*, 2014). For genomic regions that have become fully homozygous (4:0 or 0:4) in the tetraploids, F1 hybrids showed much higher numbers of DEGs (**Fig. 5b; Table S9**) than did genomic regions with 2:2 constitutions in the tetraploids (**Fig. 5a**). This is expected because the controls used to define the DEGs are different for the two types of genomic regions: the *in silico* "hybrid" with 1:1 parental RNA-seq admixture is the control for the 2:2 regions while the inbred parents are controls for the 4:0 or 0:4 homogenized regions. Notwithstanding this escalating factor in the hybrids, the numbers of upregulated DEGs appeared less in the tetraploids than in the F1 hybrids (**Fig. 5b, c; Table S9**). Similar numbers of downregulated DEGs were detected in homo-N region, but more downregulated DEGs detected in homo-9 region in each of the tetraploid than those in hybrids (**Fig. 5 b, c; Table S9**). Notably, the absolute numbers of DEGs in these genomic regions were greater than those in the 2:2 genomic regions (**Fig. 5b and c vs a; Table S9**). Collectively, these results suggest that the extent of gene expression changes in the fully homogenized genomic regions in the tetraploids is no less than those remaining as balanced heterozygotes (2:2), suggesting that HEs can substitute for and perpetuate the role of heterozygosity in sustaining gene expression alteration in tetraploids, an observation that mirrors the finding in allotetraploid rapeseed (Lloyd *et al.*, 2018).

We further analyzed the relevance of genomic composition for a set of tetraploid-specific, expression-altered genes relative to both parents and hybrids. This set included 378 upregulated genes and 310 downregulated genes. The expression patterns of these genes are illustrated as a heatmap (**Fig. 5d**). We found that for both upregulated and downregulated genes, their distributions among the five genomic possibilities are not statistically different from the expected ratios (**Fig. 1**) in the four tetraploids (**Fig. 5e** prop.test, $p > 0.05$). This indicates that homoeolog proportion variation is not a prerequisite for the generation of polyploidy-specific gene expression patterns; we note, however, that in contrast to genes located in heterozygous regions, those in homogenized genomic compartments are more likely to perpetuate transgenerationally.

For genes with balanced homoeolog composition, lower proportions of Nipponbare expression (i.e., higher 9311 subgenome expression) contribute more to tetraploid-specific upregulation than the alternative situation **(Fig. 5f)**. We further analyzed some of the up- or down- regulated genes by locus-specific RNA pyrosequencing (pyro-seq). The highly similar expression patterns between RNA-seq and pyro-seq (**Fig. 5g**) confirmed this expression trend.

To explore the possible roles of DNA methylation alteration in differential gene expression, we tabulated the distribution of DMCs of all three sequence contexts (CG, CHG and CHH) in the DEGs vs those in non-DEGs. Given that TEs are major targets for DNA methylation modification in plant genomes (Kawakatsu *et al.*, 2016; Springer *et al.*, 2016; Song & Cao, 2017), and our observation that TE-genes and protein-coding genes manifested variable abundance of DMCs in the three difference sequence contexts (**Fig. 2c**), we separately analyzed the two groups of genes. Our major findings are illustrated in **Fig. 6a**, which indicate that: (*i*) for protein-coding genes, DEGs contained significantly more CG- and CHG-DMCs than those in non-DEGs across the gene, including gene bodies and their 5'- and 3'-adjacent 2 kb flanking regions in all four tetraploids, with the trend being most prominent for CHG-DMCs; (*ii*) for TE-genes, the same conclusion as for protein-coding genes can be drawn, but in this case, the trend for the differences in CHG-DMCs between DEGs and non-DEGs is even more conspicuous, although the differences in CG-DMCs are also clearly larger than those for protein-coding genes; (*iii*) for both gene categories, the difference in abundance of CHH-DMCs between DEGs and non-DEGs genes is not significantly different (**Fig. 6a**).

We next investigated whether the proportions of DEGs that are also differentially methylated (DMGs), designated as 1, are significantly greater than those that are not, designated as 2 (**Fig. 6b**). Again, we tabulated protein-coding genes and TE-genes separately. We found that: (*i*) for protein-coding genes, in both ^{m}CG and ^{m}CHG contexts, 1s are significantly higher than 2s in three of the four studied tetraploids, while in the ^{m}CHH context, the 1s are not significantly different from the 2s in any of the tetraploids (**Fig. 6c**); (*ii*) for TE-genes, the 1s are not significantly different from the 2s in any of the tetraploids in the ^{m}CG context, while the 1s are significantly different from the 2s in three tetraploids and one tetraploid in the ^{m}CHG and ^{m}CHH contexts, respectively (**Fig. 6d**).

Given that a large number of genes in homogenized regions (4:0 or 0:4) showed differential expression relative to those of the corresponding parents (**Fig. 5b, c**), we further tested whether these differentially expressed genes are enriched for DMGs in any sequence context. We found that differentially expressed protein-coding genes of homo-N vs Nip were significantly enriched for CG-DMGs and CHG-DMGs but not for CHH-DMGs. For the set of DEGs of homo-9 vs 9311, only CG-DMGs were significantly enriched. For TE-genes, CHG-DMG seems more significant for most comparisons (**Fig. S6**).

Overall, these results indicate that localized alteration of DNA methylation due to the combined effects of hybridization, WGD and HE is an important contributing factor to changes of gene expression in the tetraploids, and that this effect is greater for TE-genes than for protein-coding genes. Among the three sequence contexts, the effect of ^{m}CHGs is much stronger than that of ^{m}CGs for both protein-coding genes and TE-genes, but again, is more prominent for the latter. Finally, ^{m}CHHs apparently have no appreciable effect on changed expression of genes belonging to either gene category.

Discussion

Analogous to several salient cases of homoploid hybrid speciation (Salazar *et al.*, 2010; Leducq *et al.*, 2016; Meier *et al.*, 2017; Nieto Feliner *et al.*, 2017; Lamichhaney *et al.*, 2018), allopolyploidization, when accompanied with rampant homoeologous exchanges (HEs), may not only generate phenotypic novelty but also rapid diversity. This, together with the known presence of multiple pairs of interacting incompatible loci that evolved in the two rice subspecies (https://shigen.nig.ac.jp/rice/oryzabase/) and their different combinations in the tetraploid progenies, may lead to adaptive radiation (Schumer *et al.*, 2015). This is because HEs, being products of meiotic recombination, are generated largely at random and assort independently in progenies until every genomic region (but not the chromosome as a whole) becomes homogenized to either of the progenitor homoeologs in a given tetraploid individual (Gaeta & Chris Pires, 2010; Chester *et al.*, 2012; Han *et al.*, 2015; Wendel, 2015; Lloyd *et al.*, 2018). Conceivably, this feature of allopolyploidization should be most prominent when the two hybridizing progenitors are diverged to only moderate extent, such that the nascent allopolyploid genome represents a platform with a strong buffering capacity to allow homoeologous chromosomes to pair and recombine during meiosis. This process and

outcome are on full display in the present study, where the four segmental allotetraploid individuals at the 5th selfed generation (S5) derived from WGD of reciprocal F1 hybrids of the two rice subspecies, *indica* and *japonica* (Xu *et al.*, 2014), are each distinctly different from one another in genome composition, due to accumulated and still-ongoing HEs.

One key result of this study is the demonstration that in spite of the genome-wide mosaicism and allelic variance in the segmental allotetraploids, their overall methylomes are not dramatically different from those of the diploid parents or F1 hybrids. This conclusion is consistent with results of the prior study on autotetraploid rice (Zhang *et al.*, 2015), suggesting that there is built-in homeostatic mechanism that maintains overall methylation stability in plants.

A second key result of this study is that, notwithstanding the overall methylome stability among the genomically divergent and recombinant allopolyploid rice lines, locus-specific DNA methylation repatterning occurred extensively, and that this effect appears to be caused by the combination of WGD and HEs, rather than by hybridization *per se*. In this aspect, we note that several previous studies in a wide spectrum of plant taxa including *A. thaliana* (Greaves *et al.*, 2012; Yang *et al.*, 2016; Zhu *et al.*, 2017), rice (He *et al.*, 2010; Chodavarapu *et al.*, 2012; Zhao & Zhou, 2012; Chen & Zhou, 2013), and maize (Springer & Stupar, 2007; Zhao *et al.*, 2007) have shown that DNA methylation may undergo extensive repatterning in F1 hybrids. However, our results indicate that the effect of hybridization on DNA methylation alteration pales when compared with that of WGD and HEs. One surprising dimension of our results is that hybridization alone, to some extent, actually mitigates the original parental DNA methylation divergence; that is, the two subgenomes of F1 hybrids become more similar to each other in DNA methylation than are their parental subspecies. From a mechanistic perspective, this implies that there is strong common *trans* control of methylation in both subgenomes of the reciprocal rice diploid hybrids. In contrast to this quantitative stability at the diploid level, the combined effects of WGD and HEs augment DNA methylation differences between the subgenomes and generate genome-wide diversity in methylation footprints. Importantly, we found the changing features of DNA methylation in the tetraploids, including contextual partitioning of DMCs and their variable abundance between TEs and genes, largely mirror those between the parental subspecies that have evolved under differential natural and artificial selections (He *et al.*, 2011; Molina *et al.*, 2011; Huang *et al.*, 2012; Choi *et al.*, 2017).

We have previously shown that extensive homoeologous expression rewiring occurred in the segmental rice allotetraploids, largely due to the variable copy numbers of homologs resulting from HEs (Sun *et al.*, 2017). This accords with the more recent finding in allotetraploid rapeseed (*Brassica napus*), where it was also shown that HEs caused significant changes in gene expression (Lloyd *et al.*, 2018). The design of the present study, which includes the exact diploid parents and F1 hybrids, enabled us to precisely partition gene expression alteration as a function of the separate and combined effects of hybridization, WGD and HEs, and to explore if localized alteration of DNA methylation is involved in the process. One notable result of this analysis is that genomic regions with a 2:2 homolog/homoeolog constitution in the tetraploids manifested much greater numbers of differentially expressed genes (DEGs) than those in the corresponding F1 hybrids at the same genomic regions. This indicates that the number of DEGs due to hybridization is enhanced by WGD in the tetraploids even without invoking the effect of HEs, consistent with our previous results using the immediate generations (S0 or S1) of this tetraploid rice system when no HEs occurred (Xu *et al.*, 2014). More strikingly, the fully homogenized genomic regions (4:0 or 0:4) showed more or less the same numbers of DEGs as those at the 2:2 heterozygous regions, suggesting that genomic homozygosity and chromosome mosaicism due to HEs do not reduce the numbers of DEGs in the tetraploids.

We consider these results bear implications to further our understanding of allopolyploid diversification and adaptation, in that potentially adaptive phenotypic variants related to HE-generated DEGs can be fixed in each of the divergent progeny swarms derived from a single allopolyploid. This realization, together with the massive inter-individual genomic variation generated by HEs from the same allopolyploidization event, each with a patchwork of regional genomic homozygosity, may provide fresh insight into genesis of the genomic substrates for both natural and artificial selections. One additional, possibly relevant aspect of this process, also shown here, is that localized repatterning of DNA methylation is a factor contributing to the emergence of DGEs in the tetraploids. Finally, we demonstrate that transposable elements (TEs) plays a role in mediating the connection between altered DNA methylation (primarily in the CHG context) and gene expression. These results are consistent with the increasingly recognized importance of TEs in gene expression in the course of plant evolution, especially in plants with large, complex and TE-embedded genomes (Ma & Bennetzen, 2004; Lisch, 2013; Bennetzen & Wang, 2014; Springer *et al.*, 2016; Song & Cao, 2017; Vicient & Casacuberta, 2017).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC#31700187, NL), the National Recruitment Program of Youth Talents (LG), the Program for Introducing Talents to Universities (#B07017, BL), and the Plant Genome Research Program of the United States National Science Foundation (JFW). We are grateful to two anonymous reviewers for their constructive comments. The authors declare that they have no competing interests.

Author contributions

BL and JFW conceived and supervised the study. NL, CX, AZ, RL, XM and LG performed the data analyses and conducted the experiments. XL grew the plants and collected the samples. NL, GL and BL interrogated and verified the analyses. NL, BL and JFW wrote the manuscript with contributions from all coauthors.

References

Adams KL. 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. *Journal of Heredity* **98**(2): 136-141.

Ainouche ML, Jenczewski E. 2010. Focus on polyploidy. *New Phytologist* **186**(1): 1-4.

Bennetzen JL, Wang H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annual Review of Plant Biology* **65**: 505-530.

Bomblies K, Jones G, Franklin C, Zickler D, Kleckner N. 2016. The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role? *Chromosoma* **125**(2): 287-300.

Chen X, Zhou D-X. 2013. Rice epigenomics and epigenetics: challenges and opportunities. *Current Opinion in Plant Biology* **16**(2): 164-169.

Chen ZJ. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* **58**: 377-406.

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**(5): 258-268.

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proceedings of the National Academy of Sciences of the United States of America* **109**(4): 1176-1181.

Chodavarapu RK, Feng S, Ding B, Simon SA, Lopez D, Jia Y, Wang G-L, Meyers BC,

Jacobsen SE, Pellegrini M. 2012. Transcriptome and methylome interactions in rice hybrids. *Proceedings of the National Academy of Sciences of the United States of America* **109**(30): 12040-12045.

Choi JY, Platts AE, Fuller DQ, Hsing Y-I, Wing RA, Purugganan MD. 2017. The rice paradox: multiple origins but single domestication in Asian rice. *Molecular Biology and Evolution* **34**(4): 969-979.

Choi JY, Purugganan MD. 2018. Multiple origin but single domestication led to *Oryza sativa*. *G3: Genes | Genomes | Genetics* **8**(3): 797-803.

Civan P, Craig H, Cox CJ, Brown TA. 2016. Multiple domestications of Asian rice. *Nature Plants* **2**: 16037.

Comai L. 2005. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**(11): 836-846.

Deng X, Song X, Wei L, Liu C, Cao X. 2016. Epigenetic regulation and epigenomic landscape in rice. *National Science Review* **3**(3): 309-327.

Diez CM, Roessler K, Gaut BS. 2014. Epigenetics and plant genome evolution. *Current Opinion in Plant Biology* **18**: 1-8.

Ding M, Chen ZJ. 2018. Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. *Current Opinion in Plant Biology* **42**: 37-48.

Dion-Cote A-M, Barbash DA. 2017. Beyond speciation genes: an overview of genome stability in evolution and speciation. *Current Opinion in Genetics & Development* **47**: 17-23.

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

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**(1): 11-15.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443-461.

Feldman M, Levy AA. 2012. Genome evolution due to allopolyploidization in wheat. *Genetics* **192**(3): 763-774.

Freeling M, Scanlon MJ, Fowler JE. 2015. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Current Opinion in Genetics & Development* **35**: 110-118.

Gaeta RT, Chris Pires J. 2010. Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytologist* **186**(1): 18-28.

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* **19**(11): 3403-3417.

Gong L, Olson M, Wendel JF. 2014. Cytonuclear evolution of rubisco in four allopolyploid lineages. *Molecular Biology and Evolution* **31**(10): 2624-2636.

Gou X, Bian Y, Zhang A, Zhang H, Wang B, Lv R, Li J, Zhu B, Gong L, Liu B. 2018.

Transgenerationally precipitated meiotic chromosome instability fuels rapid karyotypic evolution and phenotypic diversity in an artificially constructed allotetraploid wheat (AADD). *Molecular Biology and Evolution* **35**(5): 1078-1091.

Grant V. 1981. *Plant Speciation*. New York: Columbia University Press.

Greaves IK, Groszmann M, Ying H, Taylor J, Peacock WJ, Dennis ES. 2012. Trans chromosomal methylation in Arabidopsis hybrids. *Proceedings of the National Academy of Sciences of the United States of America* **109**(9): 3570-3575.

Grossniklaus U. 2011. Plant germline development: a tale of cross-talk, signaling, and cellular interactions. *Sexual Plant Reproduction* **24**(2): 91-95.

Han T-S, Wu Q, Hou X-H, Li Z-W, Zou Y-P, Ge S, Guo Y-L. 2015. Frequent introgressions from diploid species contribute to the adaptation of the tetraploid Shepherd's purse (*Capsella bursa-pastoris*). *Molecular Plant* **8**(3): 427-438.

He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F, et al. 2010. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* **22**(1): 17-33.

He Z, Zhai W, Wen H, Tang T, Wang Y, Lu X, Greenberg AJ, Hudson RR, Wu C-I, Shi S. 2011. Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLOS Genetics* **7**(6): e1002100.

Higgins EE, Clarke WE, Howell EC, Armstrong SJ, Parkin IAP. 2018. Detecting *de novo* homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array. *G3: Genes | Genomes | Genetics* **8**(8): 2673-2683.

Hollister JD. 2015. Polyploidy: adaptation to the genomic environment. *New Phytologist* **205**(3): 1034-1039.

Hu L, Li N, Xu C, Zhong S, Lin X, Yang J, Zhou T, Yuliang A, Wu Y, Chen Y-R, et al. 2014. Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proceedings of the National Academy of Sciences of the United States of America* **111**(29): 10642-10647.

Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, et al. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**(7421): 497-501.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**(7345): 97-100.

Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, et al. 2016. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**(2): 492-505.

Kenan-Eichler M, Leshkowitz D, Tal L, Noor E, Melamed-Bessudo C, Feldman M, Levy AA. 2011. Wheat hybridization and polyploidization results in deregulation of small RNAs. *Genetics* **188**(2): 263-272.

Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2018. Rapid hybrid speciation in Darwin's finches. *Science* **359**(6372): 224-228.

Lanfear R. 2018. Do plants have a segregated germline? *PLOS Biology* **16**(5): e2005439.

Leducq J-B, Nielly-Thibault L, Charron G, Eberlein C, Verta J-P, Samani P, Sylvester K, Hittinger CT, Bell G, Landry CR. 2016. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature Microbiology* **1**: 15003.

Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. *Science* **320**(5875): 481-483.

Li X, Zhu J, Hu F, Ge S, Ye M, Xiang H, Zhang G, Zheng X, Zhang H, Zhang S, et al. 2012. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics* **13**: 300.

Lisch D. 2013. How important are transposons for plant evolution? *Nature Reviews Genetics* **14**(1): 49-61.

Lloyd A, Blary A, Charif D, Charpentier C, Tran J, Balzergue S, Delannoy E, Rigaill G, Jenczewski E. 2018. Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *New Phytologist* **217**(1): 367-377.

Ma J, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America* **101**(34): 12404-12410.

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. 2002. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiology* **129**(2): 733-746.

Madlung A, Wendel JF. 2013. Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenetic and Genome Research* **140**(2-4): 270-285.

Mandakova T, Lysak MA. 2018. Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology* **42**: 55-65.

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011. Recently formed polyploid plants diversify at lower rates. *Science* **333**(6047): 1257.

Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications* **8**: 14363.

Mercier R, Mezard C, Jenczewski E, Macaisne N, Grelon M. 2015. The molecular biology of meiosis in plants. *Annual Review of Plant Biology* **66**: 297-327.

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD, et al. 2011. Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences of the United States of America* **108**(20): 8351-8356.

Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, et al. 2016. Widespread natural variation of DNA methylation within angiosperms. *Genome Biology* **17**(1): 194.

Nieto Feliner G, Alvarez I, Fuertes-Aguilar J, Heuertz M, Marques I, Moharrek F, Pineiro R, Riina R, Rossello JA, Soltis PS, et al. 2017. Is homoploid hybrid

speciation that rare? An empiricist's view. *Heredity* **118**(6): 513-516.
Otto SP. 2007. The evolutionary consequences of polyploidy. *Cell* **131**(3): 452-462.
Pecinka A, Fang W, Rehmsmeier M, Levy AA, Mittelsten Scheid O. 2011. Polyploidization increases meiotic recombination frequency in *Arabidopsis*. *BMC Biology* **9**: 24.
Quadrana L, Colot V. 2016. Plant transgenerational epigenetics. *Annual Review of Genetics* **50**: 467-491.
Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* **33**(1): 589-639.
Renny-Byfield S, Wendel JF. 2014. Doubling down on genomes: polyploidy and crop plants. *American Journal of Botany* **101**(10): 1711-1725.
Salazar C, Baxter SW, Pardo-Diaz C, Wu G, Surridge A, Linares M, Bermingham E, Jiggins CD. 2010. Genetic evidence for hybrid trait speciation in *heliconius* butterflies. *PLoS Genetics* **6**(4): e1000930.
Salmon A, Ainouche ML, Wendel JF. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* **14**(4): 1163-1175.
Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen H, Schork NJ, et al. 2013. Patterns of population epigenomic diversity. *Nature* **495**(7440): 193-198.
Schubert I, Lysak MA. 2011. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends in Genetics* **27**(6): 207-216.
Schumer M, Cui R, Rosenthal GG, Andolfatto P. 2015. Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS Genetics* **11**(3): e1005041.
Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA. 2014. Gene silencing via DNA methylation in naturally occurring *Tragopogon miscellus* (Asteraceae) allopolyploids. *BMC Genomics* **15**: 701.
Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* **13**(8): 1749-1759.
Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* **60**: 561-588.
Soltis PS, Soltis DE. 2012. *Polyploidy and Genome Evolution*. Berlin Heidelberg: Springer-Verlag Berlin Heidelberg.
Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* **30**: 159-165.
Song K, Lu P, Tang K, Osborn TC. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**(17): 7719-7723.
Song Q, Chen ZJ. 2015. Epigenetic and developmental regulation in plant polyploids. *Current Opinion in Plant Biology* **24**: 101-109.
Song X, Cao X. 2017. Transposon-mediated epigenetic regulation contributes to phenotypic

diversity and environmental adaptation in rice. *Current Opinion in Plant Biology* **36**: 111-118.

Spoelhof JP, Soltis PS, Soltis DE. 2017. Pure polyploidy: Closing the gaps in autopolyploid research. *Journal of Systematics and Evolution* **55**(4): 340-352.

Springer NM, Lisch D, Li Q. 2016. Creating order from chaos: epigenome dynamics in plants with complex genomes. *Plant Cell* **28**(2): 314-325.

Springer NM, Stupar RM. 2007. Allelic variation and heterosis in maize: how do two halves make more than a whole? *Genome Research* **17**(3): 264-275.

Stebbins GL. 1947. Types of polyploids: their classification and significance. *Advances in Genetics* **1**: 403-429.

Sun Y, Wu Y, Yang C, Sun S, Lin X, Liu L, Xu C, Wendel JF, Gong L, Liu B. 2017. Segmental allotetraploidy generates extensive homoeologous expression rewiring and phenotypic diversity at the population level in rice. *Molecular Ecology* **26**(20): 5451-5466.

Takuno S, Ran J-H, Gaut BS. 2016. Evolutionary patterns of genic DNA methylation vary across land plants. *Nature Plants* **2**: 15222.

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K. 2009. The flowering world: a tale of duplications. *Trends in Plant Science* **14**(12): 680-688.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18**(7): 411-424.

Vicient CM, Casacuberta JM. 2017. Impact of transposable elements on polyploid plant genomes. *Annals of Botany* **120**(2): 195-207.

Wendel J, Doyle J 2005. Polyploidy and evolution in plants. In: Henry RJ ed. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. Oxon, UK.: CABI Publishing, 97-117.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* **42**(1): 225-249.

Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *American Journal of Botany* **102**(11): 1753-1756.

Wendel JF, Jackson SA, Meyers BC, Wing RA. 2016. Evolution of plant genome architecture. *Genome Biology* **17**: 37.

Wendel JF, Lisch D, Hu G, Mason AS. 2018. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Current Opinion in Genetics & Development* **49**: 1-7.

Xiong Z, Gaeta RT, Pires JC. 2011. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proceedings of the National Academy of Sciences of the United States of America* **108**(19): 7908-7913.

Xu C, Bai Y, Lin X, Zhao N, Hu L, Gong Z, Wendel JF, Liu B. 2014. Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. *Molecular Biology and Evolution* **31**(5): 1066-1076.

Xu Y, Zhong L, Wu X, Fang X, Wang J. 2009. Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta*

229(3): 471-483.

Yang M, Wang X, Huang H, Ren D, Su Ye, Zhu P, Zhu D, Fan L, Chen L, He G, et al. 2016. Natural variation of H3K27me3 modification in two *Arabidopsis* accessions and their hybrid. *Journal of Integrative Plant Biology* **58**(5): 466-474.

Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D. 2010. Local DNA hypomethylation activates genes in rice endosperm. *Proceedings of the National Academy of Sciences of the United States of America* **107**(43): 18729-18734.

Zhang J, Liu Y, Xia E-H, Yao Q-Y, Liu X-D, Gao L-Z. 2015. Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **112**(50): E7022-7029.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**(6): 1189-1201.

Zhang Z, Belcram H, Gornicki P, Charles M, Just J, Huneau C, Magdelenat G, Couloux A, Samain S, Gill BS, et al. 2011. Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. *Proceedings of the National Academy of Sciences of the United States of America* **108**(46): 18737-18742.

Zhao M, Zhang B, Lisch D, Ma J. 2017. Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. *Plant Cell* **29**(12): 2974-2994.

Zhao N, Zhu B, Li M, Wang L, Xu L, Zhang H, Zheng S, Qi B, Han F, Liu B. 2011. Extensive and heritable epigenetic remodeling and genetic stability accompany allohexaploidization of wheat. *Genetics* **188**(3): 499-510.

Zhao XX, Chai Y, Liu B. 2007. Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids. *Plant Science* **172**(5): 930-938.

Zhao Y, Zhou D-X. 2012. Epigenomic modification and epigenetic regulation in rice. *Journal of Genetics and Genomics* **39**(7): 307-315.

Zhu W, Hu B, Becker C, Dogan ES, Berendzen KW, Weigel D, Liu C. 2017. Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific *Arabidopsis* hybrid. *Genome Biology* **18**(1): 157.

Figure Legends

Fig. 1 Genome-wide distribution and proportion of the five types of *japonica-indica* homoeolog composition in each of the four euploid segmental allotetraploids derived from reciprocal F1 hybrids between *Oryza sativa* ssp. *japonica* (cv. Nipponbare) and ssp. *indica* (cv. 9311).

(**a**) A circos plot depicting chromosome compositions (homoeolog ratios) in the four tetraploids individuals, 99NN-3, 99NN-7, NN99-4 and NN99-8, along each of the 12 rice chromosomes. Recombined and non-recombined segments between homoeologous chromosomes are represented by Nipponbare (Nip) homoeolog copy numbers, which ranged from 0 to 4 (y-axis). The green lines denote Nipponbare copy numbers using a 5 kb sliding window; the black bold lines represent adjacent 5 kb windows with the same Nipponbare copy number. (**b-c**) Heatmaps depicting the segment length of common *japonica-indica* homoeolog compositions between the two tetraploids individuals of the same crossing direction. (**d**) Heatmaps depicting the segment length of common *japonica-indica* homoeolog compositions between two crossing directions. (**e**) Proportions of the five types of *japonica-indica* homoeolog composition in each of the four euploid allotetraploid individuals.

Fig. 2 DNA methylation differences among the diploid parents, reciprocal F1 hybrids and segmental allotetraploids, both overall and across different genomic features (Also see Figs. S1-S4).

(**a**) Overall methylation levels in diploid rice parents, reciprocal F1 hybrids and tetraploids. (**b**) Distribution of differentially methylation cytosines (DMCs) detected in the comparisons of Nipponbare (N) vs 9311 (9), and hybrids and tetraploids vs the *in silico* "hybrids" in CG, CHG and CHH sequence contexts, respectively. (**c**) Numbers of DMCs residing in different genomic features, including protein-coding genes, TE-genes (genes annotated as transposons or retrotransposons), transposons, retrotransposons and intergenic regions.

Fig. 3 Numbers of differentially methylated regions (DMRs) detected in each of the four euploid tetraploids and both F1 hybrids in comparison with either parent, Nipponbare (N) or 9311 (9), or with artificial mixtures of both parents (*in silico* "hybrids").

For a given tetraploid or F1 hybrid, the numbers of DMRs were scored and tabulated for each of the five types of genomic region defined by differential homoeolog compositions. Thus,

for genomic regions with a homoeolog ratio of N:9 = 4:0 (homo-N) or N:9 = 0:4 (homo-9), the DMRs were scored by comparing with the corresponding regions of either of the parents; for genomic regions with a homoeolog ratio of N:9 = 1:3 or 2:2 or 3:1, the DMRs were scored by comparing with *in silico* "hybrids" constructed by mixing the parental data in proportion to the homoeolog ratios. DMRs were calculated separately in the CG, CHG and CHH contexts.

Fig. 4 Parental inheritance and *de novo* generation of DMRs in subgenomes of F1 hybrids and tetraploids.

(**a**) Comparison of differentially methylated regions (DMRs) between diploid parents, subgenomes in hybrids, and subgenomes in tetraploids for the same set of genomic regions that had a balanced (2:2) homoeolog constitution in a given tetraploid which could be classified into seven groups (defined in main text). Numbers in the Venn diagram illustrate an example of the DMR counts for comparisons among parents, F1 hybrid 9N and tetraploid 99NN-3 in the cases where the Nipponbare parental genome had less methylation than that of 9311. (**b**) Counts of DMRs in all seven different groups as exemplified in (**a**). (**c**) Examples of subgenomic methylation distribution in parents, F1 hybrids and tetraploids by the integrative genomics viewer (IGV) snapshots. For each sample, there are two panels, "^{m}C" and "cov"; the "^{m}C" is methylation level with Y axes ranging from 0 to 1; the "cov" is read coverage with Y axes ranging from 0 to 50. For 9N and 99NN-3, the Nipponbare and 9311 subgenomes are displayed separately, for example 9N|Nip mC is the methylation distribution of Nipponbare homoeologs in 9N.

Fig. 5 Counts of differentially expressed genes (DEGs) and tetraploid-specific upregulated and downregulated genes.

(**a**) Numbers of non-additively expressed genes with N:9 = 2:2 homoeolog compositions between hybrids/tetraploids and the *in silico* "hybrids". (**b**) and (**c**) Numbers of DEGs located on homo-9 or homo-N regions in tetraploids, and DEGs from the respective regions in F1 hybrids, when compared to the corresponding regions of the inbred parents. (**d**) A heatmap depicting expression patterns of a set of tetraploid-specific upregulated and downregulated genes in comparison with parents and F1 hybrids. (**e**) Proportions of the five types of homoeolog composition for the tetraploid-specific upregulated and downregulated genes. (**f**) Box plots depicting subgenomic expression trends in tetraploids-specific up- or downregulated genes with the N:9 = 2:2 homoeolog compositions. The middle horizontal

lines represent the median, lower and upper edges of the box represent the 25th and 75th percentiles, lines extending vertically from the boxes (whiskers) indicating variability outside the upper and lower quartiles, and individual dots are outliers beyond the whiskers. (**g**) Validation of the RNA-seq data-based subgenomic expression ratios by locus-specific RNA-pyrosequencing.

Fig. 6 Relationship between differential methylation and differential expression in the four euploid tetraploids (Also see Fig. S5).

(**a**) Distribution of DMCs along DEGs and non-DEGs (vs *in silico* "hybrids") in the four tetraploids (in rows) for protein-coding genes and TE-genes (genes annotated as transposons or retrotransposons). (**b**) Expression and methylation classifications, where #1 (grey bar) = X1/(X1 + Y1) denotes the proportion of differentially expressed DMGs (X1) out of total expressed DMGs (X1 + Y1), and #2 (black bar) = X2/(X2 + Y2), denotes the proportion of differentially expressed non-DMGs out of total expressed non-DMGs. (**c-d**) Comparison of the percentages of DEGs (tetraploids vs *in silico* "hybrids") with and without differential methylation (DMGs and non-DMGs) in protein-coding genes and TE-genes, respectively. (*p values < 0.05 by prop.test)

Supporting Information

Fig. S1 Proportion of methylated cytosines for each sample in the CG, CHG and CHH contexts, respectively. (Refer to Fig. 2)

Fig. S2 Overall methylation levels of diploids and tetraploids along chromosome 1 in CG (red), CHG (blue) and CHH (green), respectively. (Refer to Fig. 2.; window size is 100kb)

Fig. S3 DNA methylation pattern along protein-coding genes, TE-genes and TEs in three sequence contexts, respectively. (Refer to Fig. 2)

Fig. S4 Profiles of differentially methylation cytosine (DMC). (Refer to Fig. 2)

Fig. S5 Subgenomic DNA methylation level. (Refer to Fig. 4)

Fig. S6 Comparison of differentially expressed DMGs (differentially methylated genes) and non-DMGs. (Refer to Fig. 6)

Table S1 General information regarding the DNA resequencing data.

Table S2 SNP profiles of the parental lines, Nipponbare and 9311.

Table S3 General information regarding the DNA methylome data.

Table S4 P values from pairwise comparison (prop.test) for five types of homoeolog composition in four tetraploid individuals.

Table S5 DNA methylation profiles of the diploid parents, reciprocal F1 hybrids and segmental allotetraploids.

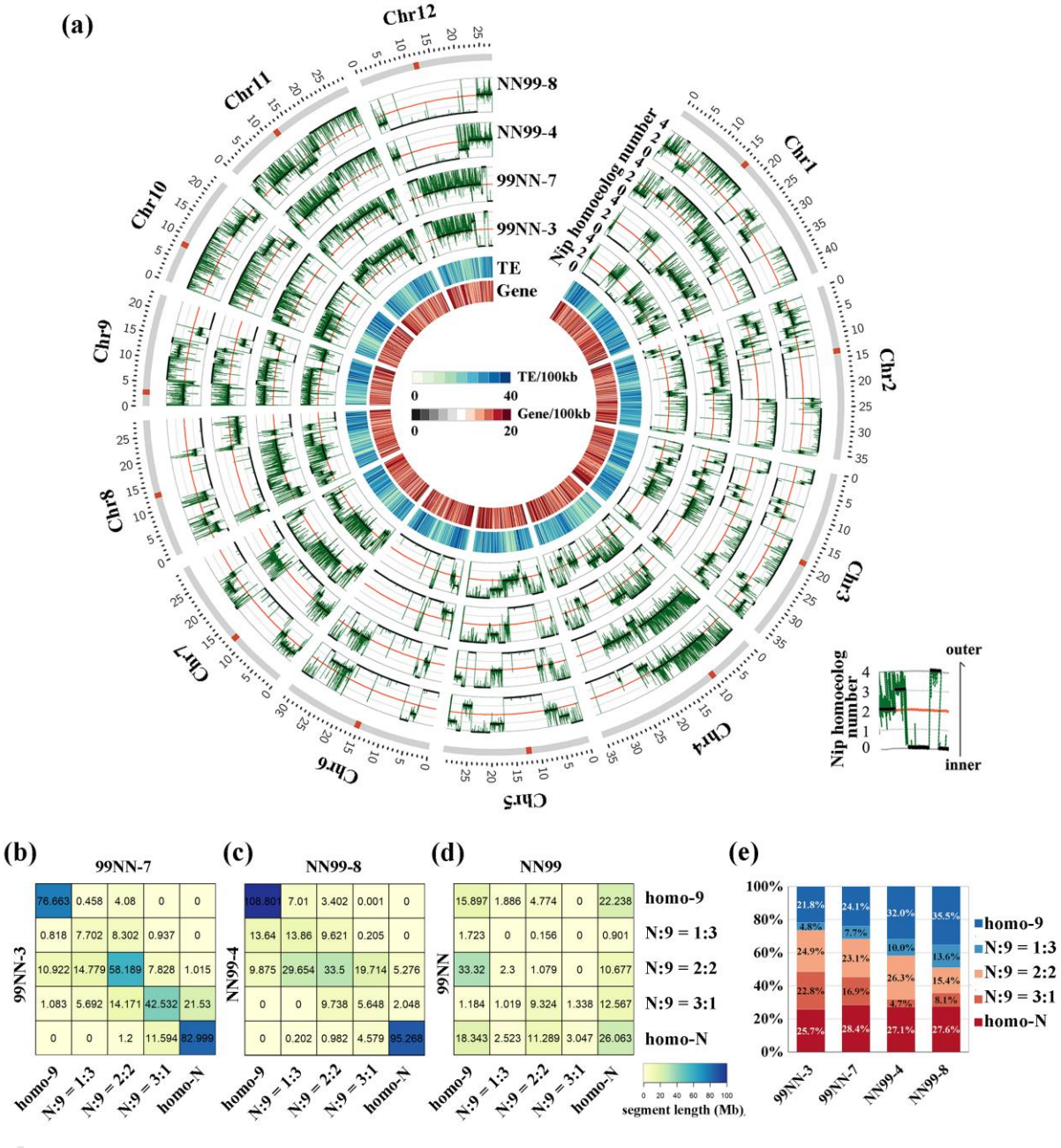
Table S6 P values from pairwise Wilcoxon rank sum tests for different genomic features with respect to CG, CHG and CHH methylation.

Table S7 Summary of DMCs in pairwise comparisons.

Table S8 Summary of DMRs in hybrids and tetraploids vs *in silico* hybrid.

Table S9 Counts and comparisons of differentially expressed genes (DEGs) located on different homoeolog composition regions to the corresponding hybrids or inbred parents.

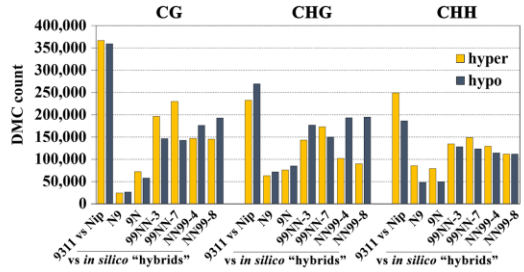
Method S1 Supplemental methods for sequencing and data analysis.



(a)

Plant	^{m}C	^{m}CG	^{m}CHG	^{m}CHH
Nipponbare	12.66%	40.56%	20.48%	3.14%
9311	13.33%	38.92%	20.74%	3.02%
N9	13.13%	38.96%	20.64%	2.94%
9N	13.77%	40.51%	21.69%	3.13%
99NN-3	15.14%	41.38%	23.30%	3.23%
99NN-7	14.29%	40.43%	22.09%	3.19%
NN99-4	13.48%	38.77%	20.93%	2.93%
NN99-8	11.72%	36.47%	17.92%	2.48%

(b)



(c)