Evolution of the RNA $N^6$-Methyladenosine Methylome Mediated by Genomic Duplication$^1$[OPEN]

Zhenyan Miao,$^{a,b}$ Ting Zhang,$^a$ Yuhong Qi,$^a$ Jie Song,$^a$ Zhaoxue Han,$^{a,b}$ and Chuang Ma$^{a,b,2,3}$

$^a$State Key Laboratory of Crop Stress Biology for Arid Areas, Center of Bioinformatics, College of Life Sciences, Northwest A&F University, Shaanxi, Yangling 712100, China
$^b$Key Laboratory of Biology and Genetics Improvement of Maize in Arid Area of Northwest Region, Ministry of Agriculture, Northwest A&F University, Shaanxi, Yangling 712100, China

ORCID IDs: 0000-0001-7855-8387 (Z.M.); 0000-0002-5959-1837 (J.S.); 0000-0002-6935-9535 (Z.H.); 0000-0001-9612-7898 (C.M.).

RNA $N^6$-methyladenosine (m$^6$A) modification is the most abundant form of RNA epigenetic modification in eukaryotes. Given that m$^6$A evolution is associated with the selective constraints of nucleotide sequences in mammalian genomes, we hypothesize that m$^6$A evolution can be linked, at least in part, to genomic duplication events in complex polyploid plant genomes. To test this hypothesis, we presented the maize ($Z. m y s$) m$^6$A modification landscape in a transcriptome-wide manner and identified 11,968 m$^6$A peaks carried by 5,893 and 3,811 genes from two subgenomes (maize1 and maize2, respectively). Each of these subgenomes covered over 2,200 duplicate genes. Within these duplicate genes, those carrying m$^6$A peaks exhibited significant differences in retention rate. This biased subgenome fractionation of m$^6$A-methylated genes is associated with multiple section features and is influenced by asymmetric evolutionary rates. We also characterized the coevolutionary patterns of m$^6$A-methylated genes and transposable elements, which can be mediated by whole genome duplication and tandem duplication. We revealed the evolutionary conservation and divergence of duplicated m$^6$A functional factors and the potential role of m$^6$A modification in maize responses to drought stress. This study highlights complex interplays between m$^6$A modification and gene duplication, providing a reference for understanding the mechanisms underlying m$^6$A evolution mediated by genome duplication events.

$N^6$-Methyladenosine (m$^6$A) is an internal, prevalent RNA modification and has been identified in the RNA of mammals (Adams and Cory, 1975), insects (Levis and Penman, 1978), yeast (Clancy et al., 2002), and plants, such as Arabidopsis ($A r a b i d o p s i s$ $t h i a l i a n a$; Zhong et al., 2008) and maize ($Z. m y s$; Nichols, 1979). The m$^6$A modification is formed by m$^6$A methyltransferase “writer” proteins (e.g. methyltransferase-like 3 [METTL3], METTL14, and Wilms tumor1-associating protein in mammalian cells; Bokar et al., 1997; Liu et al., 2014; Ping et al., 2014), recognized by “reader” proteins (e.g. YT512-B Homology domain proteins; Xu et al., 2014, 2015), and removed by “eraser” proteins (m$^6$A demethylases; e.g. fat mass and obesity-associated protein and alkylated DNA repair protein AlkB homolog 5 [ALKBH5]; Jia et al., 2011; Zheng et al., 2013), thus forming an epitranscriptomic system of RNA methylations directly analogous to the well-studied reversible DNA and histone modifications (Wang and He, 2014). Loss of function of core components of the m$^6$A modification system in mammals has demonstrated that m$^6$A modification affects multiple aspects of RNA metabolism, including stability (Wang et al., 2014b), translation efficiency (Shi et al., 2017), nuclear export (Roundtree et al., 2017b), and alternative splicing (Zhao et al., 2014). In Arabidopsis, the disruption of $N^6$-adenosine-methyltransferase MT-A70-like (METTL3 homolog), methyltransferase MT-A70 family protein (METTL14 homolog), and the FK506-binding protein 12 interacting protein 37 (FIP37, Wilms tumor1-associating protein homolog) leads to early embryonic lethality (Vespa et al., 2004; Zhong et al., 2008; Ruzicka et al., 2017), and the depletion of ALKBH10B (AAtg02940, ALKBH5 homolog) effects Arabidopsis floral translation (Duan et al., 2017). Additionally, two YT512-B Homology-domain proteins (EVOLUTION-ARILY CONSERVED C-TERMINAL REGION2 [ECT2] and ECT3) function as m$^6$A readers and control developmental timing and morphogenesis and trichome morphology (Arribas-Hernández et al., 2018; Scutenaire et al., 2018).
Wei et al., 2018). These pioneering biochemical and genetic researches shed light on the functional roles of RNA m6A modification, which constitutes an important regulatory mechanism in RNA biology (Roundtree et al., 2017a; Yang et al., 2018).

Recently, with the development of m6A sequencing (m6A-seq) technologies, an increasing number of m6A methylome comparison studies have begun to unravel the evolution of this important posttranscriptional modification (Dominissini et al., 2012, 2013). The evolutionary conservation of m6A modifications was detected within two yeast species (Saccharomyces mikatae and Saccharomyces Cerevisiae; Schwartz et al., 2013); across two accessions of Arabidopsis (Can-0 and Hen-16; Luo et al., 2014); and between human (Homo sapiens), chimpanzee (Pan troglodytes), and rhesus (Macaca mulatta; Ma et al., 2017). The transcriptome-wide comparison of m6A modifications from human, chimpanzee, and rhesus revealed that m6A evolution is associated with the selective constraints of DNA sequences and occurs in parallel with expression evolution of m6A-methylated genes (Ma et al., 2017). Yet virtually nothing is known about the evolution of m6A modification after genome duplications in plant evolution.

In plants, genome duplications (GDs), including whole genome duplications (WGDs), segmental duplications, tandem duplications (TDs), and translocated duplications, generate a source of specific genomic context (i.e. duplicated regions) as a dominant force of plant genome evolution (Freeling, 2009). Following GD events, duplicate genes were subjected to different levels of purifying selection, a proportion of which were lost in a process known as fractionation. There are also many duplicate genes retained in the genome as paralog pairs, in which the individual genes may be subfunctionalized (partitioning and sharing the original gene function) and/or neofunctionalized (gaining novel functions; Panchy et al., 2016). Both the bias in fractionation and the functional divergence of duplicate genes have been reported to be associated with differences in DNA methylation, rates of movement of transposable elements (TEs), gene expression, and postranscriptional regulation (Wang et al., 2014a, 2015; Panchy et al., 2016; Cheng et al., 2018). Therefore, GDs provide a source of specific genomic context that may have profound influences on transcriptional regulation and postranscriptional regulation. This raises the question of whether, and to what degree, the evolution of m6A modification is mediated by GD events in complex polyploid plant genomes.

To investigate this question, we performed deep m6A-seq on the leaf tissue of maize and explored the GD events in complex polyploid plant genomes. GD events, generating a source of speciﬁc genomic context (i.e. duplicated regions), in which the individual genes may be subfunctionalized (partitioning and sharing the original gene function) and/or neofunctionalized (gaining novel functions; Panchy et al., 2016). Both the bias in fractionation and the functional divergence of duplicate genes have been reported to be associated with differences in DNA methylation, rates of movement of transposable elements (TEs), gene expression, and postranscriptional regulation (Wang et al., 2014a, 2015; Panchy et al., 2016; Cheng et al., 2018). Therefore, GDs provide a source of specific genomic context that may have profound influences on transcriptional regulation and postranscriptional regulation. This raises the question of whether, and to what degree, the evolution of m6A modification is mediated by GD events in complex polyploid plant genomes.

To investigate this question, we performed deep m6A-seq on the leaf tissue of maize and explored the patterns of m6A evolution in maize. Maize underwent a recent WGD event after its divergence from the lineage that gave rise to sorghum (Sorghum bicolor) ~5–12 million years ago (Swigonova et al., 2004). Since that time, the two subgenomes in maize experienced a variety of changes (e.g. chromosomal rearrangements and gene conversion; Schnable et al., 2011) and were combined into a diploid genome (Wei et al., 2007; Schnable et al., 2011). Because of this unusual evolutionary history, together with the availability of high quality of the 2maize B73 reference genome (Jiao et al., 2017) and the sorghum reference genome (Paterson et al., 2009; McCormick et al., 2018), we selected maize as a model crop system to study the evolutionary implications of m6A modification in the context of GD events. Using the transcriptome-wide map of m6A generated from our study, we made an effort to address some topics of conceptual importance to understanding the evolutionary characteristics of m6A in maize. For example, do GD events contribute to the evolutionary novelty of m6A modification? How is the evolution of m6A modification associated with the expression divergence (ED) of duplicate genes? How do m6A modification and TEs experience some coevolutionary process following WGD in maize, or do they at all? Our answers to these questions are provided by examining the coordination patterns of RNA m6A modification with gene duplication, evolutionary divergence, gene expression, and TE accumulation.

RESULTS

Transcriptome-Wide Mapping of m6A in Maize

To obtain the transcriptome-wide m6A map in maize, a series of m6A-immunoprecipitation (IP) and the matched input (non-IP control) libraries was constructed and sequenced (Supplemental Table S1). This series included leaf tissue of maize B73 and Han21 seedlings under both well-watered (WW) and drought-stressed (DS) conditions, with three biological replicates each. Raw sequencing reads were processed to discard adaptor sequences and low-quality bases using the Trimmomatic v0.36 tool (Bolger et al., 2014). The resulting reads from maize B73 and Han21 samples were aligned to the maize B73 reference genome (B73_RefGen_v4) and Han21 pseudogenome, respectively, using TopHat v2.1.1 (Kim et al., 2013). To build the Han21 pseudogenome, single nucleotide polymorphisms (SNPs) between B73 and Han21 were identified by aligning Han21 RNA sequencing (RNA-seq) data to the maize B73 reference genome (B73_RefSeq_v4) using STAR (Dobin et al., 2013), followed by SNP calling using the genome analysis toolkit (GATK) UnifiedGenotyper (McKenna et al., 2010). A total of 94,761 SNPs within transcribed regions were used to replace the corresponding nucleotides in the maize B73 reference genome to generate a pseudoreference genome for Han21. Read distribution analysis showed that the reads from m6A-IP samples are highly accumulated around the stop codon and within 3′-untranslated regions (3′-UTRs) in all experimental conditions (Supplemental Fig. S1). We detected 8,224–11,134 m6A peaks in each individual biological replicate (Supplemental Fig. S2). For each experimental condition (one inbred line under one environmental condition), highly confident peaks
were identified, referring to the previous study (Yoon et al., 2017). In brief, by intersecting peak regions in a pairwise fashion among all three replicates, regions that overlap in at least two of three replicates were designated as high-confidence m6A peak regions. Strong correlations were observed for the abundance of confident peaks between biological replicates (Supplemental Fig. S3). Confident m6A peaks from different experimental conditions were further merged into a unique set of m6A peaks. As a result, a total of 11,968 unique m6A peaks with high confidence were finally detected from 11,219 maize genes (Supplemental Table S2), accounting for an average of ~1.07 m6A peaks within transcription units from each gene. The m6A peaks in maize are abundant in 3’-UTRs (74.6% of m6A peaks), near stop codons (20.8%), and coding sequences (3.2%; Fig. 1A). Enrichment analysis showed that the stop codon was the most enriched segment, representing ~8-fold enrichment, followed by the 3’-UTR (~4-fold enrichment; Fig. 1B). Similar distribution patterns of m6A peaks were also observed in the separate analysis of m6A-seq data from B73 and Han21 (Supplemental Fig. S4). At the genome level, 11,219 m6A-methylated genes (i.e. genes whose transcripts carry m6A peaks; abbreviated as m6A genes) were unevenly distributed across each chromosome (Fig. 1D).

We observed that 90.6% of 11,968 m6A peaks contain the canonical motif RRACH (where R represents A/G, A is m6A, and H represents A/C/U) in maize (Fig. 1C; Supplemental Table S3). As shown in Supplemental Table S3, this proportion is comparable to those (80.6%, 92.3% and 81.2%) estimated using m6A peaks generated from three recent Arabidopsis m6A methylome studies (Luo et al., 2014; Shen et al., 2016; Anderson et al., 2018). These 11,968 m6A peaks were further scanned for enriched motifs using the MEME suite (http://meme-suite.org/index.html; Bailey et al., 2009). As expected, the RRACH motif is significantly enriched within maize m6A peaks (Supplemental Fig. S5). We also noted an interesting motif, URUAY (where Y represents C/U; Fig. 1C), which can also be detected from m6A peaks from each replicate sample (Supplemental Fig. S6). The URUAY motif has recently been regarded as a plant-specific consensus motif recognized by m6A reader protein ECT2 (Wei et al., 2018). Indeed, as shown in Supplemental Figure S5, this URUAY motif is also enriched within m6A peaks generated from three recent Arabidopsis m6A methylome studies (Luo et al., 2014; Shen et al., 2016; Anderson et al., 2018). By using another commonly used motif enrichment analysis software, the HOMER suite (v4.10; http://homer.ucsd.edu/homer; Heinz et al., 2010), enriched URUAY motif can also be detected from maize and Arabidopsis m6A data used in our study (Supplemental Fig. S5). Dot-blot assay was also performed to validate the specificity of m6A antibody for the URUAY motif (Supplemental Fig. S7).
m6A Genes Exhibit Distinct Sequence Features from Non-m6A Genes

To identify sequence features that may associate with m6A modification, we first tested the Pearson correlation coefficient (PCC) between the frequency of m6A genes and different sequence features (gene length, exon length, exon number, GC content, intron length, and gene distance) along the maize genome in a sliding window of 100 adjacent genes (Supplemental Table S4). The statistical significance of PCC was assessed using 10,000 permutation tests; in each, 11,219 m6A genes were randomly selected from the maize B73 genome annotation, and a new PCC value was calculated for generating a background distribution. The corresponding P value was calculated as $P = (1 + N)/10,000$; here, $N$ is the number of PCCs in the background distribution, which exceeds the PCC for the original data. A $P$ value <0.0001 denotes none of the 10,000 permutation tests exceeding the PCC for the original data. We observed that the frequency of m6A genes was slightly positively correlated with exon length (PCC = 0.06; $P = 0.0955$; Supplemental Fig. S8) but significantly positively correlated with gene length (PCC = 0.33; $P$ value < 0.0001) and exon number (PCC = 0.57; $P < 0.0001$; Fig. 2, A and B; Supplemental Fig. S8). In addition, the frequency of m6A genes was significantly negatively correlated with GC content (PCC = −0.16; $P = 0.0002$) and gene distance (PCC = −0.36; $P < 0.0001$; Fig. 2, C and D; Supplemental Fig. S8).

Complementary to the correlation analysis using contiguous windows, we further performed statistical analysis using sequence features from individual genes. Maize genes were split into three groups according to the corresponding number of m6A sites from our study: low-m6A genes (10,559 genes carrying only one m6A peak), high-m6A genes (660 genes carrying at least two m6A peaks), and non-m6A genes (28,105 genes carrying no m6A peak; Supplemental Table S2). Compared with non-m6A genes, m6A genes (both low-m6A genes and high-m6A genes) had significantly more exons (Fig. 2E), lower GC content (Fig. 2F), and longer introns (Fig. 2G). The mean gene length is in the order of high-m6A genes (11,796 bp) > low-m6A genes (7,271 bp) > non-m6A genes (2,849 bp; Fig. 2H). These results may indicate that longer genes tend to have a higher probability of containing the m6A peaks.

The aforementioned statistical analysis using contiguous windows and individual genes was also performed on 10,604 and 10,085 m6A genes determined (MEDB), which is publicly available at http://bioinfo.nwafu.edu.cn/MaizeBrowse/index.html.
from m6A-seq data sets of B73 and Han21, separately. We found that there is no substantial difference between the statistical results for the two inbred lines (Supplemental Figs. S9 and S10). This result is as expected, as a substantial number of m6A genes (9,470) were overlapped in both maize inbred lines (Supplemental Fig. S11A). We observed that significant differences exist between 1,134 B73-specific and 615 Han21-specific m6A genes in the distribution of GC content and average intron length (Supplemental Fig. S11B). Next, we examined whether there are significant differences in nucleotide sequence variation between m6A genes from these two inbred lines (Han21 and B73). Comparison analysis revealed that the SNP density in m6A genes is significantly lower than that in non-m6A genes (Supplemental Fig. S12).

Overall, these results suggest that m6A modification in maize is correlated with multiple sequence features, including gene length, exon number, intron length, GC content, and SNP density.

**Biased Fractionation of m6A Genes Exists between Two Subgenomes in Maize**

The maize genome underwent its most recent WGD shortly after divergence from sorghum (Schnable et al., 2011). After the WGD event, one copy of many duplicate gene pairs in maize was lost (fractionated), leaving the other one as a singleton. Because of the biased gene fractionation, the loss of duplicate genes is uneven between two maize subgenomes (maize1 and maize2), so that the maize1 subgenome retains more genes than the maize2 subgenome (Brohammer et al., 2018). In human and yeast, the evolutionary consequences of gene duplication are associated with DNA methylation (Keller and Yi, 2011). After the WGD event, one copy of many duplicate gene pairs in maize was lost (fractionated), leaving the other one as a singleton. Because of the biased gene fractionation, the loss of duplicate genes is uneven between two maize subgenomes (maize1 and maize2), so that the maize1 subgenome retains more genes than the maize2 subgenome (Brohammer et al., 2018). In human and yeast, the evolutionary consequences of gene duplication are associated with DNA methylation (Keller and Yi, 2011). Considering that m6A modification is correlated with multiple sequence features (Fig. 2; Supplemental Fig. S11) and has a role in regulating gene expression (Yue et al., 2015; Rognant and Soller, 2017; Roundtree et al., 2017a), we raised the question of whether m6A modification might also be linked to duplicate gene retention.

We first examined the duplication status of m6A genes retained after the most recent WGD event. Genes in maize1 and maize2 were identified by performing syntenic analysis between the maize B73 reference genome and the sorghum reference genome (Brohammer et al., 2018). Among 11,219 m6A genes, 5,893 and 3,811 were annotated as maize1 and maize2 genes, respectively (Supplemental Table S2). The singleton-duplication ratio of m6A genes in maize1 (1:0.66; 3,551 vs. 2,342) is significantly higher than that of m6A genes in maize2 (1:1.43; 1,566 vs. 2,245; χ² test; P < 0.001), which is consistent with the trend in total subgenome genes (Fig. 3A). The significant difference in singleton-duplication ratio between maize1 and maize2 was also apparent when m6A genes in tandem duplication were not involved (Supplemental Fig. S13). Notably, the frequency of m6A genes in maize2 singletons is significantly higher than that in maize1 singletons (Fig. 3B). These divergences of m6A genes between two subgenomes are most likely the evolutionary consequence of biased subgenome fractionation. The biased fractionation of m6A singletons between two subgenomes may associate with multiple sequence features. As shown in Figure 3, C and D, singletons carrying m6A peaks (m6A singletons) in maize1 have significantly more exons and longer nucleotides than those in maize2 (Student’s t test; P < 0.05), but these features were not significantly different between total maize1 singletons and maize2 singletons. These suggest that the biased fractionation of m6A genes may be relative to gene length.

We then compared the evolutionary rate (ω) of m6A genes from maize1 and maize2 subgenomes. The evolutionary rates, ratio of nonsynonymous substitution (Kd)/synonymous sites (Ks), of genes in maize were estimated using interspecific comparisons with putatively orthologous sequences between maize and sorghum. The ω values of m6A genes were significantly higher than those of non-m6A genes, and Ks values of m6A genes were considerably lower than those of non-m6A genes (Student’s t test; P < 0.001; Supplemental Fig. S14). This is indicative of higher evolutionary rate and less evolutionary time of m6A genes. We also
observed that the evolutionary rate of m6A singletons in maize1 was significantly lower than that of m6A singletons in maize2, but the evolutionary rate of non-m6A singletons was not significantly different between maize1 and maize2 (Table 1). These findings indicated that m6A singletons in maize1 have experienced a higher intensity of purifying selection than those in maize2. This asymmetric purifying selection may have an influence on biased fractionation of m6A singletons. In contrast, m6A duplicates in maize1 and maize2 evolved under similar levels of purifying selection, but the evolutionary rate of non-m6A duplicates in maize1 was significantly lower than that of non-m6A duplicates in maize2 (Table 1). This indicated that m6A modification could associate with the divergence of evolutionary rate between duplicate genes in two subgenomes.

Further analysis of m6A duplicates showed that the evolutionary time (Ks) was significantly different from duplicate gene pairs with different m6A patterns: non-m6A pattern (neither of two partners carrying m6A peaks), diverged-m6A (DM) pattern (one partner had m6A peaks while the other did not), and identical-m6A (IM) pattern (both of two partners carrying m6A peaks). Duplicate genes with non-m6A pattern had the significantly highest level of Ks values, followed by duplicate genes with DM and IM patterns (Student’s t test; P < 0.001; Supplemental Fig. S15). Gene transposition could cause the separation of syntenic duplicates into two subgenomes. Protein sequence comparison between the maize and sorghum genome identified 198 and 108 pairs of transposed singletons in maize1 and maize2, respectively (Supplemental Table S2). We observed that transposed gene pairs had a significantly higher proportion of divergence of m6A status than syntenic duplicate gene pairs without transposition (χ² test; P < 0.001; Supplemental Fig. S16). These observations indicate that m6A modification divergence in young duplicate pairs was smaller than that in older duplicate pairs, and gene transposition could enhance the extent of m6A divergence between duplicate pairs.

### Coevolutionary Consequences of m6A Methyllome and TE Influence Duplicate Retention and Expression Divergence

We then explored whether there was an association between m6A modification and gene expression in maize. Similar to that for m6A-seq, a series of RNA-seq libraries was constructed using leaf tissue of maize B73 and Han21 seedlings under both WW and DS conditions, with three biological replicates each (Supplemental Table S1). For each of these RNA-seq libraries, the expression abundance of maize genes was estimated in terms of fragments per kilobase per million (FPKM). We observed that the expression abundance of m6A genes was significantly higher than that of non-m6A genes (Fig. 4A), and the singleton-duplicate ratio of m6A genes (1:0.71; 6,329 vs. 4,520) was significantly lower than that of non-m6A genes (1:0.34; 18,523 vs. 6,234; χ² test; P < 0.001), reflecting an overall higher retention rate for duplicate genes methylated by m6A (Fig. 4B). The association between m6A modification and gene expression was also revealed by the differential expression abundance between m6A duplicates and singletons in maize (Supplemental Fig. S17).

With the m6A-seq and RNA-seq data at hand, we were interested in whether m6A modification divergence associated with ED for duplicate genes. For a duplicate gene pair (G1, G2), the ED was calculated using the following formula: ED = (E1 − E2)/(E1 + E2), where E1 and E2 denote the mean expression level of G1 and G2, respectively. The ED of m6A genes with DM pattern was significantly higher than that of genes with IM pattern (Fig. 4C). For genes with DM pattern, the methylated partners exhibited a higher level of gene expression than nonmethylated partners (Fig. 4D). These findings suggested that m6A modification was more likely to occur on actively transcribed genes and could be associated with the retention rate and ED of duplicate gene pairs.

Previous studies have reported that the frequency of TEs, which is associated with local genomic stability, also affects the expression of their neighboring genes (Sahebi et al., 2018). This prompted us to investigate whether there is an evolutionary interplay between m6A and TEs. In our data, we observed that m6A genes are closer to TEs than non-m6A genes (Fig. 5A), and the frequency of TE-related genes (gene loci overlap with TE loci) carrying m6A peaks (61.1%; 6,850/11,219) was significantly higher than that of non-m6A genes (37.3%; 16,118/28,090; χ² test; P < 0.01; Fig. 5B). Moreover, by comparing the evolutionary rates of m6A genes with those of non-m6A genes, significantly higher ω values were observed from m6A genes (Fig. 5C), and ω values of TE-related m6A genes were significantly higher than those of non-TE-related m6A genes.

### Table 1. Comparison of evolutionary rates (ω) of m6A singletons and m6A duplicates, and non-m6A singletons and non-m6A duplicates in maize1 and maize2, respectively.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>m6A Singletons</th>
<th>m6A Duplicates</th>
<th>P value*</th>
<th>Non-m6A Singletons</th>
<th>Non-m6A Duplicates</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize1</td>
<td>0.2485 ± 0.1560</td>
<td>0.2256 ± 0.1617</td>
<td>&lt; 0.0001</td>
<td>0.2372 ± 0.7533</td>
<td>0.2088 ± 0.1625</td>
<td>0.0108</td>
</tr>
<tr>
<td>Maize2</td>
<td>0.2598 ± 0.1631</td>
<td>0.2283 ± 0.1568</td>
<td>&lt; 0.0001</td>
<td>0.2330 ± 0.1734</td>
<td>0.2146 ± 0.1632</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>P value*</td>
<td>0.0101</td>
<td>0.2861</td>
<td></td>
<td>–</td>
<td>0.4050</td>
<td>0.0244</td>
</tr>
</tbody>
</table>

*Statistical analysis was conducted using the Student’s t test.

Supplemental Table S1.
Evolution of the RNA m^6^A Modification

Phylogenetic analysis identified the functional counterparts of the m^6^A methyltransferases (N^6^-adenosine-methyltransferase MT-A70-like, methyltransferase MT-A70 family protein, FIP37, putative ortholog of human KIAA1429 [VIRILIZER], and the E3 ubiquitin ligase hakai [HAKAI]), demethylases (ALKBH10B), and binding proteins (ECT2, ECT3, and ECT4) in maize (Supplemental Fig. S18). According to the phylogenetic tree and genomic coordinates, five homologous pairs formed by the WGD event were identified: ZmFIP37a/ZmFIP37b, ZmHAKAla/ZmHAKAlb, ZmECTa/ZmECTb, ZmECTd/ZmECTe, and ZmALKBH10a/ZmALKBH10b (Fig. 6A). For ZmHAKAla/ZmHAKAlb and ZmECTd/ZmECTe, their corresponding partners showed very similar evolutionary rates, as reflected by \( \kappa_0 \) and \( \omega \) (Fig. 6B). In contrast, for the other three duplicate pairs, their corresponding partners exhibited notably distinct evolutionary rates, which appeared to exhibit different strengths of purifying selection. These results indicate that evolutionary conservation and divergence of m^6^A functional factors may be mediated by WGD.

We performed RNA-seq analysis to quantify the expression abundance of m^6^A functional factors in the leaf sample of two maize inbred lines’ responses to drought stress. In the drought-sensitive inbred line B73, two ALKBH10 homologs and two ECT2/ECT4 homologs were significantly upregulated by drought stress (Fig. 6C). These expression patterns were also verified by quantitative PCR (qPCR) assay (Supplemental Fig. S19). Meanwhile, the m^6^A abundance in DS samples was significantly lower than that in WW samples, indicating global m^6^A hypomethylation induced by drought stress (Supplemental Fig. S20, A and B). For each maize
line, differential methylated peaks (DMPs) were identified by comparing the abundance of m6A peaks between DS and WW conditions (Supplemental Table S6). The drought-tolerant Han21 line (2,998) has more DMPs than the drought-sensitive B73 line (386). However, the proportion of hypomethylated DMPs in Han21 (92.0%; 2,758/2,998) is comparable to that in B73 (87.8%; 339/386; Supplemental Fig. S20C). These results may reflect the phenomenon that levels of m6A modification are significantly decreased during drought stress in both inbred lines. Considering the expression patterns of m6A methyltransferase and demethylase genes, the induced expression of m6A demethylase genes most likely contributes to the decrease in m6A methylation during drought stress.

Most DMPs were located in 3′-UTR and stop codon regions in both B73 and Han21 (Supplemental Fig. S21A). A recent article reviewed that the genic locations of m6A peaks play distinct roles in mediating the functional output of m6A modification (Shen et al., 2019). Therefore, we performed gene ontology (GO) enrichment analysis of genes containing DMPs to explore functional characteristics of DMPs in the context of genic location. In B73, the GO terms “protein phosphorylation” and “regulation of apoptotic process” were specifically enriched in genes within 3′-UTR, whereas the GO terms “histone exchange” and “cytoplasmic translation” were specifically enriched in genes within DMPs near the stop codon (Supplemental Fig. S21B). In Han21, the GO terms “regulation of transcription” and “protein transport” were specifically enriched in genes with DMPs within 3′-UTR, whereas the GO terms “photoreactive repair” and “siderophore biosynthetic process” were specifically enriched in

Figure 5. Evidence for coevolution of m6A genes and TEs. A, Comparison of distance to the nearest TEs between m6A genes and non-m6A genes. B, Comparison of TE frequency between m6A genes and non-m6A genes. C, Comparison of evolutionary rates between m6A genes and non-m6A genes. D, Comparison of evolutionary rates between TE-related m6A genes and non-TE-related m6A genes. E, Comparison of frequency of m6A methylation between tandem duplicated (TD) genes and non-TD genes. F, Comparison of frequencies of DM pattern and IM pattern between TD clusters and WGD pairs. G, Comparison of distance to the nearest TEs between m6A TD genes and m6A non-TD genes. H, Comparison of frequency of TEs between m6A TD genes and m6A non-TD genes. In A, C, D, and G, box plots range from the Q1 to Q3 of the distribution and represent the IQR. A line across the box indicates the median. The whiskers are lines extending from Q1 and Q3 to end points that are defined as the most extreme data points within Q1 – 1.5 × IQR and Q3 + 1.5 × IQR, respectively. Statistical analysis was conducted using Student’s t test. *P < 0.05; **P < 0.001. In B, E, F, and H, statistical analysis was conducted using the χ² test. *P < 0.05; **P < 0.001.
genes within DMPs near DMPs (Supplemental Fig. S21B). These results revealed that genes containing DMPs in specific genic locations play roles in distinct biological processes in B73 and Han21.

The Han21 line was significantly more tolerant of drought stress than B73, which manifests as dramatic phenotypic differences observed, including increased plant height, alterations in relative water content and water loss rate, and robustness in root development (Supplemental Fig. S22). To understand the potential role of m6A hypomethylation in these phenotypic differences, we performed GO analysis of genes containing hypomethylated m6A peaks in B73 and Han21, respectively (Supplemental Fig. S23). In B73, genes involved in cuticle development, negative regulation of apoptotic signaling pathway, and response to abscisic acid are highly enriched. In Han21, genes were significantly enriched in response to abiotic stress processes, such as cellular response to oxidative stress, response to osmotic stress, acetyl-CoA metabolic process, and ethylene-mediated signaling pathway, as well as several developmental pathways, such as cell morphogenesis and development, embryo development ending in seed dormancy, Glc and starch metabolic process, and supramolecular fiber organization. These gene functions showed clear correspondence to the observed phenotypic responses of these two maize inbred lines.

Moreover, the differences in drought tolerance between B73 and Han21 can also be explored at the level of individual genes covering DMPs. VACUOLAR INVERTASE2 (VI2) is a positive regulator of root elongation in Arabidopsis (Sergeeva et al., 2006). The Han21 line showed more than a 2-fold reduction in

Figure 6. Evolutionary dynamics of genes encoding m6A functional factors. A, Orthologous relationships of genes encoding m6A functional factors between maize and sorghum. Purple boxes indicate maize genes, and yellow boxes indicate sorghum genes. Green boxes represent syntenic orthologs of m6A functional factors. B, Evolutionary rates of maize m6A functional factors compared with their syntenic homologs in sorghum. C, RNA-seq expression profiles of genes encoding m6A functional factors. Data are means ± SD (n = 3, three biological replicates). A, The significance test of differential expression was conducted using Cuffdiff software; false discovery rate (FDR)-adjusted P value ≤0.05.
methylation levels of m^6A peaks in V12 during drought stress (Supplemental Fig. S24). m^6A peaks in Actin-7 (ACT7), a member of the actin gene family involved in root growth, cell division, and root architecture in Arabidopsis (Kandasamy et al., 2009), are also less methylated in DS samples of the maize B73 line (Supplemental Fig. S24). Furthermore, increased accumulation of epicuticular waxes is known to limit water loss and increase water deficit tolerance (Aharoni et al., 2004; Zhang et al., 2005; Kosma et al., 2009). The striking disparity in relative water content and water loss rate is accompanied by marked differences in the methylation levels of m^6A peaks in genes associated with wax deposition between B73 and Han21. For example, ECERIFERUM4 and ECERIFERUM10, two genes encoding the alcohol-forming fatty acyl-CoA reductase, are involved in cuticular wax biosynthesis in Arabidopsis (Rowland et al., 2006). Methylation levels of m^6A peaks in putative maize orthologs of these two genes are significantly inhibited by drought stress in Han21 and B73, respectively (Supplemental Fig. S24). Furthermore, methylation level of m^6A peaks in wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase (WSD1), a gene encoding diacylglycerol acyltransferase that is required for stem wax ester biosynthesis in Arabidopsis (Li et al., 2008), is also significantly repressed by drought stress in Han21 (Supplemental Fig. S24). Then, we performed m^6A-IP-qPCR assay to further confirmed the m^6A methylation levels of these DMPs. As shown in Supplemental Figure S25, the results of the m^6A-IP-qPCR assay are consistent with those of m^6A-seq. Together, these observations demonstrate that the concerted hypomethylation of candidate transcripts encoding proteins may associate with wax accumulation during response to drought stress.

DISCUSSION

Over the past several years, several studies have demonstrated the important roles of epigenetic modifications in the evolutionary history of eukaryotic genomes (Zemach et al., 2010; Chang and Liao, 2012; Wang et al., 2014a; Patten, 2016). However, few transcriptome-wide studies have attempted to investigate the evolutionary patterns of RNA m^6A modification in the context of GD events. In this study, we used maize as a model system to study the evolution of m^6A methylome following GD events. Our analysis suggested that m^6A modification alteration following GD events appears to be a significant source of evolutionary novelty within plants.

The Transcriptome-wide Map of m^6A in Maize

Our study presents the transcriptome-wide map of RNA m^6A modifications in the leaf tissue of maize seedlings. This resource provides us an opportunity to globally characterize m^6A in large, complex plant genome. In our maize m^6A-seq data, the URUAY motif has a lower enrichment E value than the canonical RRACH motif. Given that the URUAY motif is identified as a plant-specific consensus motif recognized by m^6A reader protein ECT2 and the UGUA motif can be methylated by endogenous Arabidopsis m^6A writer proteins (Wei et al., 2018), these phenomena indicated that those m^6A writer proteins with methylation activity for the URUAY motif may be conserved between Arabidopsis and maize. However, our motif enrichment analysis using previously published Arabidopsis m^6A-seq data showed that the enrichment E values of the URUAY motif are generally higher than those of the RRACH motif (Supplemental Fig. S5). These differences between maize and Arabidopsis likely represent the different m^6A site biases and unique biological meanings of m^6A methylation between two species, or may result from the distinct technical biases in m^6A-seq library preparation among these studies. Further in-depth structural and functional analysis of m^6A writer and reader proteins may help to clarify these biases.

Additionally, we found that the frequency of m^6A genes is positively correlated with gene length, exon number, and intron length. We infer that the longer genes may have a higher probability of containing m^6A modification. Previous study has shown that gene length is increasing with evolutionary time (Grishkevich and Yanai, 2014), and longer genes are more likely to be retained as duplicates (McGrath et al., 2014; Guo, 2017). Further considering the higher retention rate of m^6A duplicates compared with non-m^6A genes (Fig. 4B), we suspect that gene length may be an important genetic property that linked the evolutionary relationship between m^6A modification and gene duplication in maize.

To benefit the readers in the future analysis, we developed a web browser named MEDB to host the transcriptome-wide m^6A map and to support navigation of the map and its interactive visualization, integration, comparison, and analysis. Taking advantage of the JBrowse system (Buels et al., 2016), MEDB also allows users to transfer their private m^6A methylome data to the browser as custom tracks for easy cross-study comparison. To ensure the data security, MEDB does not require the users to upload their own files to the server. Instead, it can access local files or a uniform resource locator specifying the location of a remote file. The cross-study comparison can be performed through a degree of in-browser data analysis by combining data in tracks using arithmetic and set operations, for example, finding the union, intersection, or exclusive of two tracks. Notably, combination tracks can be further used as input to other combination tracks, allowing users to build up arbitrarily complex analysis tracks.

Correlation between m^6A Methylation and Biased Subgenome Fractionation

In maize, a proportion of duplicate genes were lost because of different levels of purifying selection on
two subgenomes, a process known as biased fractionation. Genes in the overfractionated subgenome (maize1) are distinct with respect to overall fitness, in contrast to genes in the underfractionated subgenome (maize2; Schnable et al., 2011). Our analysis revealed that the singleton-duplicate ratio of m6A in maize1 was significantly higher than that in maize2 (Fig. 3A), suggesting that biased subgenome fractionation also occurred in m6A genes. Further investigation revealed that maize1 singlets exhibited significantly lower m6A frequency than maize2 singlets (Fig. 3B), and ω values were significantly lower in m6A singlets of maize1 than those of maize2 (Table 1), indicating significantly higher levels of purifying selection on the fractionated m6A genes in maize1 than those in maize2. These results are complementary evidences for the hypothesis that maize1 underwent stronger purifying selection that resulted in a higher frequency of fractionation (Schnable et al., 2012; Pophaly and Tellier, 2015).

A bias of gene expression dominance toward the less fractionated subgenome has been previously observed in maize (Schnable et al., 2011). Indeed, we showed that m6A genes had higher expression levels than non-m6A genes, and m6A modification divergence was correlated with gene ED between duplicate partners (Fig. 4; Supplemental Fig. S17). Considering that gene expression levels impose a strong constraint on gene duplications and subgenome fractionation, these observations indicate that m6A modification divergence of duplicate genes influences gene expression abundance and, ultimately, the divergence of subgenome dominance in maize.

Complex Interplays among m6A Modification, TE Accumulation, and Tandem Duplication

The disruptive effects of TEs have been extensively documented, as they can integrate into the regulatory or coding region of host genes or induce ectopic/nonallelic recombination, which is often associated with lower levels of gene stability (Jangam et al., 2017). In our study, m6A genes were found to be less distant from TEs than non-m6A genes, and the frequency of TEs in m6A genes was higher than that in non-m6A genes (Fig. 5, A and B). m6A genes involved in TD events also showed less distance from TEs than those not involved, and the frequency of TE-related genes was much higher in TD genes than in non-TD genes (Fig. 5, G and H). These findings suggest that genes flanked by TEs were more likely to be methylated by m6A, and then this coordination of m6A and TEs experienced a more relaxed selection (Fig. 5D). Together with former observations that m6A modification can enhance stability of target genes (Fig. 4), we propose preferential accumulation of TEs near m6A genes, as a coevolutionary outcome of m6A modification and TEs, involves a compromise between optimal levels of gene stability and prevention of the damage done by active TEs.

In addition, we found that gene members in TD arrays showed higher rates of divergence in m6A modification than those in syntenic duplication pairs (Fig. 5F), and m6A genes involved in TD events evolved more quickly and underwent stronger purifying selection than those not involved (Supplemental Table S5). This is hypothesized to be a consequence of the effects of gene balance, the theory being that genes encoding proteins that interact with large numbers of other proteins are more sensitive to changes in stoichiometry than those that do not (Birchler and Veitia, 2010). The stoichiometry of members of multisubunit complexes can affect the amount of functional complete product, which in turn affects patterns of gene expression and, ultimately, the phenotype and evolutionary fitness (Birchler and Veitia, 2012). The observation that TD genes have higher loss rates of m6A modification relative to non-TD genes (Fig. 5E) is a complementary evidence for the hypothesis that these genes may be undergoing subfunctionalization, representing an evolutionary outcome of m6A modification mediated by TD events.

Potential Roles of m6A Modification in Maize Response to Drought Stress

Differential gene expression has been proven to be responsible for drought responses in plants (Zhu, 2016; Miao et al., 2017). Differential levels of m6A modification under drought stress have also been observed in both B73 and Han21 (Supplemental Fig. S20), suggesting that m6A modification may be another important contributor for drought responses. Here, we discuss the evolutionary consequences of five duplicate gene pairs encoding m6A functional factors that may contribute to responses to drought stress (Fig. 6, B and C). The two members of each of two duplicate gene pairs (ZmHAKAIa/ZmHAKAlb and ZmECTd/ZmECTe) exhibited similar intensities of purifying selection, suggesting that they have been exposed to similar selective constraints. This explained why these members all showed mild expression. In contrast, the evolutionary rates of two members of ZmFIP37 pairs varied, but their expression values were smooth. It is possible that these two members have differential effects on other phenotypes or during other developmental stages, regardless of whether their specific roles in drought responses have diverged. Notably, variations in the evolutionary rates and levels of expression between members of the ZmALKBH10 pair were observed, in line with recent reports in Arabidopsis (Duan et al., 2017), and which could be the outcomes of neofunctionalization or subfunctionalization as the adaptive consequences of gene duplication. The upregulated expression of genes encoding demethylases (ZmALKBH10) appears to be the primary force driving m6A hypomethylation under drought.
stress, which indicates that m^6A hypomethylation may play a positive role in drought response.

**Future Perspectives: Exploring the Dynamics of RNA m^6A Methylation in Different Plant Species with Spatial, Temporal, and Environmental Dimensions**

In this study, we highlighted the importance of generating a transcriptome-wide m^6A map in plant species and uncovered the evolutionary patterns of m^6A modifications associated with genomic duplication using m^6A-seq data from the leaf tissue of maize seedlings. As the most abundant internal m^6A modification, m^6A has gained increasing interest in the last few years to understand its dynamic roles of posttranscriptional regulation mechanism underlying key plant developmental processes, including embryo development (Ružička et al., 2017), shoot stem cell fate (Shen et al., 2016), floral transition (Duan et al., 2017), and trichome morphogenesis (Wei et al., 2018) in Arabidopsis m^6A studies. We expect that, in the future, the m^6A modification dynamics will be explored using transcriptome-wide m^6A maps profiled from different developmental stages and tissues of maize (as well as other important plant species) under diverse environmental conditions. Comparison of m^6A modifications within and across species will be performed to further elucidate the evolutionary mode of posttranscriptional regulation in plants.

**MATERIALS AND METHODS**

**Plant Growth and Sample Preparation**

Seeds from the B73 and Han21 inbred lines of maize (*Zea mays*) were germinated, and seedlings were transferred to a growth chamber with controlled environmental conditions (28°C day/26°C night, 16-h light/8-h dark). The relative water content of soil was maintained at 80% of the soil moisture capacity. The B73 and Han21 maize seeds were harvested, immediately frozen in liquid nitrogen, and stored at −80°C for RNA isolation and sequencing.

**RNA Isolation and PolyA+ mRNA Selection**

For each leaf sample, total RNA was extracted using the RNAiso Plus (code no. 9109; TaKaRa) according to the manufacturer’s instructions. Polyadenylated (PolyA+) mRNA selection was performed using oligo(dT)25 magnetic beads (code no. 514198; New England Biolabs) following the manufacturer’s protocol.

**High-Throughput m^6A-seq and RNA-seq**

mRNA was randomly fragmented into ~200-nucleotide-long fragments by RNA Fragmentation Reagents (Ambion). Fragmented RNA was incubated for 2 h at 4°C with 0.5 mg/mL anti-m^6A polyclonal antibody (catalog no. 20203; Synaptic Systems) in IP buffer (150 mM NaCl, 0.1% [v/v] octylphenoxypolyethoxyethanol [Igepal CA-630], and 10 mM Tris-HCl [pH 7.4]) supplemented with RNase Plus RNase inhibitor (Promega). The mixture was then immunoprecipitated by incubation with protein-A beads at 4°C for an additional 2 h. After extensive washing, bound RNA was eluted from the beads with 0.5 mg/mL N^6^-methyladenosine in IP buffer and precipitated by ethanol. TrueSeq standard mRNA Sample Prep Kit (Illumina) was used to construct the libraries from immunoprecipitated RNA and input RNA according to a published protocol (Dominissini et al., 2013). Sequencing was performed on an Illumina HiSeq platform (Illumina), and 50-bp single-end reads were generated. A library for RNA-seq was generated using TrueSeq Stranded mRNA Sample Prep Kit (Illumina). The resulting libraries were sequenced on a HiSeq platform (Illumina) to produce 2×122-bp paired-end reads.

**Analysis of Sequencing Data**

Raw reads from RNA-seq and m^6A-seq were trimmed to remove adaptor sequences and low-quality bases using the Trimomatic v0.36 tool (Bolger et al., 2014). The quality of trimmed RNA-seq and m^6A-seq reads was examined using the FastQC program (https://www.bioinformatics.babraham.ac.uk/projects/fastqc).

SNPs between B73 and Han21 were identified following the GATK Best Practices workflow (Van der Auwera et al., 2013). In brief, Han21 RNA-seq reads were first mapped to the B73 reference genome using STAR with parameters set as: outFilterMultimapNmax = 1; outSAMstrandField = intronMotif; and twopassMode = Basic (Dobin et al., 2013). Then, AddOrReplaceReadGroups and MarkDuplicates functions in the Picard suite (v2.20.0, https://broadinstitute.github.io/picard/) were used to add read groups and remove duplicates, respectively. Subsequently, the SplitNCigarReads function in the GATK suite (McKenna et al., 2010) was used to split reads into exons segments and hard-clips sequences overhanging into the intronic regions. SNPs were detected using the HaploTyperCaller tool in the GATK suite with parameters set as: allowPotentiallyMisencodedQuals; don’tUseSoftClippedBases; stand_call_conf = 20; ERC = GVC; and nct = 20. The CombineCVCFs and GenotypeCVCFs functions in the GATK suite were used to merge gvcf files and to generate vcf files. The results were further filtered using the VariantFilter function in the GATK suite with recommended parameters: window = 35; cluster = 3; filterName = FS; FS > 60; filterName = QD; and QD < 2.0. Finally, Samtools v1.9 (Li et al., 2009) was used to select highly confidence SNPs with the following requirements: (1) nonreference alleles need to be consistent; (2) SNPs supported with ≥10 reads, and (3) SNPs supported with ≥2 samples.

The Han21 pseudogenome was built based on SNPs identified from our RNA-seq data. The trimmed B73 and Han21 RNA-seq reads were aligned to the maize B73 reference genome (B73_RefGen_v4) and Han21 pseudogenome, respectively, using TopHat v2.1.1 (Kim et al., 2013) with maximum intron length set to 10 kb, with default settings for other parameters. Unique mapping reads were provided as input to Cufflinks v2.2.1 (Trapnell et al., 2013) for normalization and estimation of gene expression level in terms of FPKM mapped reads (FPKM = counts of mapped fragments × 10^9/length of transcript × total count of the mapped fragments]). Differential analysis was conducted using the Cuffdiff program in Cufflinks. In this study, maize B73 reference genome sequences and annotation were downloaded from Ensemble Plants (release 41; https://plantsensembl.org; Kersey et al., 2018).

The trimmed B73 and Han21 RNA-seq reads were aligned to the maize B73 reference genome and Han21 pseudogenome, respectively, using the STAR v2.5.3a (Dobin et al., 2013) with parameters: alignIntronMin = 20; alignIntronMax = 10000; outFilterMultimapNmax = 1; and outFilterMismatchNmax = 1. Peak calling was performed using a “sliding window” method slightly modified from previous analysis (Luo et al., 2014) and implemented with the R package PEA (Zhai et al., 2018). To call m^6A peaks, the reference genome was scanned using a 25-bp sliding window. A Fisher exact test was used to identify windows enriched for m^6A by comparing normalized read counts of each window for IP and input samples. Benjamin-Hochberg was implemented to adjust the P value for FDR for multiple testing using the R function “p.adjust.” Significant windows were identified if fold change of normalized read count was more than two and FDR value was less than 0.05. Adjacent significant windows were merged together to form peak regions. Peaks with length less than 100 nt in length were excluded from the analysis. The called peaks within locally expressed genes (FPKM < 1) were discarded. For each experimental condition, peaks that overlapped in at least two of three replicates were merged as confidence m^6A peaks using slice function (lower = 2, rangesOnly = TRUE) in the IRanges package (Lawrence et al., 2013). Confident m^6A peaks from four experimental conditions (B73_WW, B73_DS, Han21_WW, and Han21_DS) were further merged into a unique set of m^6A peaks using slice function (lower = 1, rangesOnly = TRUE) in
the IRanges package. The m^A peaks with significantly differential m^A modification levels between DS and WW conditions were determined using the QNB software (Liu et al., 2017), with the criteria set as enrichment fold change $\geq 2$ or $\leq 0.5$, and FDR $\leq 0.05$. GO enrichment analysis was performed using topGO (Alexa et al., 2006).

Identification of Enriched Motifs within m^A Peaks

Two well-known motif analysis suites, MEME (Bailey et al., 2009) and HOMER (Heinz et al., 2010), were used to perform the motif enrichment analysis. The DREME (Discriminative Regular Expression Motif Elitiation) tool in the MEME suite (http://meme-suite.org/tools/dreme) was used to discover relatively short (up to $8 \times 8$ bp), ungapped motifs that are enriched within a set of target sequences (m^A peak sequences) relative to a set of control sequences (shuffled m^A peak sequences). The set of target sequences (target set) is composed of m^A peak sequences extracted from the reference genome (maize: B73, RefGen_v4; Arabidopsis [Arabidopsis thaliana]: The Arabidopsis Information Resource 10) using the fastaFromBed function in BEDTools software v2.28 (Quinlan and Hall, 2010). The set of control sequences (control set) is generated by randomly shuffling each of the m^A peak sequences while preserving the nucleotide frequencies. This shuffling process is performed by using the “fasta-shuffle-letters” utility (k = 1) provided by the MEME suite. These two sets of sequences were input to DREME for discovering motifs with the following parameters: minimum length of the motif, 5; maximum length of the motif, 7; FDR value threshold, 1E-5.

For a specified motif (e.g., RRACH or URUAY), the AME tool in the MEME suite (http://meme-suite.org/tools/ame) and findMotifs.pl script in the HOMER tool (http://homer.ucsd.edu/homer) were used to calculate the significance level of relative enrichment of this motif within target sequences relative to control sequences.

Synonymous ($K_s$) and Nonsynonymous ($K_a$) Substitutions in Homologous Maize Genes

The maize subgenomic data (genes in maize1 and maize2) in our study were obtained from Brohammer et al. (2018) and were identified via performing synteny analysis between the maize genome and the sorghum (Sorghum bicolor) genome. In brief, the SynMap pipeline (Lyons et al., 2008) was run to align the maize B73 reference genome (B73, RefGen_v4) against the sorghum reference genome (v3.1.1; http://phytozome.jgi.doe.gov) to identify maize genes in syntenic blocks relative to the ancestral state. Then, the subgenome identity of each maize chromosome was determined using a previously described method (Schnable et al., 2011). The maize tandem duplicate genes in our study were obtained from Kono et al. (2018). The coding sequences of homologous gene pairs were aligned using the MAFFT v7.271 software (Katoh and Standley, 2013). On the basis of sequence alignments, the synonymous ($K_s$) and nonsynonymous ($K_a$) substitutions and the resulting $K_a/K_s$ values were calculated using the model-averaging method in the KaKs_calculator v2.0 (Wang et al., 2010).

Identification of Gene Transposition

The protein sequences of singleton genes in maize containing syntenic orthologs in sorghum were searched against protein sequences of the singletons without syntenic orthologs in sorghum using Protein BLAST. The two sequences were used as templates to synthesize four RNA oligos that contained either m^A or A at a single internal position within the URUAY motif. Oligos were denatured at 72°C in a heat block for 3 min. Then, the samples were loaded to the Amersham Hybond-N+ membrane (RPN119B; GE Healthcare). The membrane was then ultraviolet cross-linked in an HL-2000 HybridLinker Hybridization Oven (Ultraviolet Products). After cross-linking, the membrane was washed in 10 mL of wash buffer for 5 min and then blocked in 10 mL of blocking buffer for 1 h at room temperature with gentle shaking. Subsequently, the membrane was incubated with anti-m^A antibody (diluted 1:500; catalog no. 202003; Synaptic Systems) overnight at 4°C. The membrane was then washed in 10 mL of wash buffer for 5 min three times, followed by incubation with horseradish peroxidase–conjugated mouse anti-rabbit IgG (diluted 1:10,000; catalog no. sc-2357; Santa Cruz Biotechnology) for 1 h at room temperature. The membrane was then washed in 10 mL of the wash buffer for 10 min four times. Then, the membrane was incubated with 5 mL of enhanced chemiluminescence western-blotting substrate for 5 min in darkness at room temperature and exposed with Hyperfilm enhanced chemiluminescence for a proper exposure period.

Measurement of Gene Expression Levels Using Real-Time qPCR

Total RNA was extracted following above described RNA isolation methods, and the total RNA was treated with DNAse (code no. 2270A; TaKaRa). First-strand cDNA was synthesized using PrimeScript II First Strand cDNA Synthesis Kit (code no. 6210A; TaKaRa) according to the manufacturer’s instruction. qPCR was performed in three biological replicates $\times$ three technical replicates using CFX96 Real-Time PCR Detection System (Bio-Rad) with TB Green Premix Ex Taq II (Tli RNaseH Plus, code no. RR820A; TaKaRa). The Cyclophilin (GenBank: M55021) was used as an internal control (Lin et al., 2014). The $2^{-\Delta \Delta C_t}$ method was used to calculate the gene expression levels. The primers used for real-time qPCR are listed in Supplemental Table S7.

Global m^A Quantification

Total RNA isolation and two rounds of PolyA+ mRNA selection were performed following the above described “RNA isolation and PolyA+ mRNA Selection” section methods. The change of global m^A levels in mRNA was measured by EpiQuik m^A RNA Methylation Quantification Kit (Colorimetric; catalog no. P-9005; EpiGentek) following the manufacturer’s protocol.

Measurement of m^A Methylation Levels of Specific m^A Peaks Using m^A-IP-qPCR

For validation of m^A-seq results, m^A IP was performed using WW and DS B73 and Han21 samples, respectively. RNA samples were fragmented into ~300-nucleotide-long fragments. Fragmented RNA was incubated for 2 h at 4°C with 0.5 mg/mL anti-m^A polyclonal antibody (catalog no. 202003; Synaptic Systems). After ethanol precipitation, the input RNA and immunoprecipitated RNA were subjected to reverse transcription and qPCR assays. The Cyclophilin (GenBank: M55021) was used as an internal control, since (1) Cyclophilin mRNA did not show any obvious m^A peak from m^A-seq data, (2) Cyclophilin showed relatively invariant expression levels between WW and DS samples, and (3) Cyclophilin is considered to be a housekeeping gene. Samples were performed in three biological replicates $\times$ three technical replicates. The m^A level of specific mRNA fragments was calculated by the ratio of RNA abundances, IP/input, as previously described by Shen et al. (2016). In brief, relative enrichment of each fragment was calculated by first normalizing the amount of a target cDNA fragment against that of internal control, and then
normalizing the value for the immunoprecipitated sample against that for the
input. The primers used for m6A-IP-qPCR are listed in Supplemental Table S7.

Physiological Phenotypes of Maize B73 and Han21

Seeds of the maize inbred lines, Han21 and B73, were germinated, and
seedlings were transplanted into pots filled with sand and transferred to a
growth chamber with controlled environmental conditions (16-h light/8-h dark
cycle, 28°C day/26°C night temperature). The relative water content of soil
was maintained at 80% of the soil moisture capacity for WW seedlings and at
40% of soil moisture capacity for DS seedlings.

To measure the relative water content of leaves, fresh leaves were harvested
and weighed to determine the fresh weight (FW). Leaves were then saturated in
water for 24 h at 4°C and weighed for the turgid weight. Last, leaves were dried
in an oven at 80°C for 24 h, and the dry weight (DW) was measured. The relative
water content (%) was calculated as (FW – DW)/(turgid weight – DW) × 100.

To measure the rate of water loss, fresh leaves were harvested, weighed for
the FW, and then placed in a culture dish for 24 h (22°C and 70% relative hu-
midity) to measure the dehydrated weight. Then, samples were dried in an
oven at 80°C for 24 h, and the DW was measured. The rate of water loss (%) was
calculated as (FW – dehydrated weight)/DW × 100.

Statistical Analysis

The Student’s t test was performed using the t.test function in R package.
The χ² test was performed using chisq.test function in R package.

Accession Numbers

All sequencing data have been deposited into the National Center for Bio-
technology Information’s Sequence Read Archive database under the accession
numbers SRP153627 and SRP125635.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Distribution pattern of m6A-IP reads along
transcripts.

Supplemental Figure S2. Intersection among m6A peaks identified in three
biological replicates of four experimental conditions.

Supplemental Figure S3. Correlation of m6A peak abundance among three
biological replicates in four experimental conditions.

Supplemental Figure S4. Distribution of m6A peaks in four experimental
conditions.

Supplemental Figure S5. Both RRACH and URUAY motifs are enriched
within m6A peaks in maize and Arabidopsis.

Supplemental Figure S6. The enriched RRACH and URUAY motifs identi-
ﬁed from m6A peaks in each replicated sample.

Supplemental Figure S7. Dot-blot analysis demonstrates m6A antibody
speciﬁcity for URUAY motif.

Supplemental Figure S8. Statistical signiﬁcance of correlation coefﬁcients
between the frequency of m6A genes with different gene features.

Supplemental Figure S9. Statistical signiﬁcance of correlation coefﬁcients
between the frequency of m6A genes with different gene features in B73
and Han21.

Supplemental Figure S10. Comparison of gene features (exon number, GC
content, and intron length) among different types of m6A genes in B73
and Han21, respectively.

Supplemental Figure S11. Comparison of m6A genes in B73 and Han21.

Supplemental Figure S12. Comparison of sequence variation patterns be-
 tween m6A genes and non-m6A genes.

Supplemental Figure S13. Comparison of singleton-duplication ratio of
m6A genes and total subgenome genes excluding tandem duplicates in
maize1 and maize2.

Supplemental Figure S14. Comparison of Ks (nonsynonymous substitu-
tion), Ks (synonymous substitution), and ω (evolutionary rate) values of
m6A genes and non-m6A genes.

Supplemental Figure S15. Comparison of evolutionary time (K) of three cat-
egories of duplicates (BM pattern, DM pattern, and non-m6A [NM] pattern).

Supplemental Figure S16. Pairwise comparison of frequencies of m6A di-
vergence among maize1 transposition, maize2 transposition, and dupli-
cates without transposition.

Supplemental Figure S17. Comparison of expression abundance between
duplicates and singletons of genes with m6A modification.

Supplemental Figure S18. Phylogenetic relationship of m6A functional
factors among maize, sorghum, and rice.

Supplemental Figure S19. Relative mRNA levels of m6A functional factors
in WW and DS seedling samples of B73 and Han21.

Supplemental Figure S20. Hypomethylation of m6A induced by drought
stress in B73 and Han21.

Supplemental Figure S21. Functional characteristics of DMPs in the con-
text of genic location in B73 and Han21.

Supplemental Figure S22. Phenotypic responses of B73 and Han21 under
WW and DS conditions.

Supplemental Figure S23. Genes containing drought-induced hypomethyl-
ated peaks are involved in various biological processes of plant develop-
ment and abiotic stress.

Supplemental Figure S24. Dynamic m6A peaks of five drought-responsive
genes (ZmV2, ZmACT7, ZmCRE4, ZmCRE10, and ZmWSD1) in B73 and Han21.

Supplemental Figure S25. Validation of m6A peaks in five drought-
responsive genes (ZmV2, ZmACT7, ZmCRE4, ZmCRE10, and ZmWSD1).

Supplemental Table S1. Sequenced and mapped reads in m6A-seq, input
RNA-seq, and mRNA-seq samples.

Supplemental Table S2. Characterization of maize genes regarding m6A
modification and duplicate status.

Supplemental Table S3. Frequency of m6A peaks containing RRACH mo-
tif and URUAY motif in maize and Arabidopsis.

Supplemental Table S4. Correlation of m6A gene frequency with sequence
features.

Supplemental Table S5. Comparison of evolutionary rates of m6A genes
involved and not involved in tandem duplication (TD).

Supplemental Table S6. Differentially methylated peaks in B73 and Han21
in response to drought stress.

Supplemental Table S7. Primers used in this study.

ACKNOWLEDGMENTS

We thank all our laboratory members at Northwest A&F University for their
discussion on the project.

Received March 18, 2019; accepted August 3, 2019; published August 13, 2019.

LITERATURE CITED

Adams JM, Cory S (1973) Modified nucleosides and bizarre 5’-termini in
mouse myeloma mRNA. Nature 258: 28–33

SHINE clade of AP2 domain transcription factors activates wax bio-
synthesis, alters cuticle properties, and confers drought tolerance when

Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of func-
tional groups from gene expression data bydecorrelating GO graph
structure. Bioinformatics 22: 1600–1607

Amoutzias GD, He Y, Gorden J, Mossialos D, Oliver SG, Van de Peer Y
Proc Natl Acad Sci USA 107: 2967–2971


Keller TE, Yi SY (2014) DNA methylation and evolution of duplicate genes. Proc Natl Acad Sci USA 111: 5932–5937


Kono TJY, Brohammer AB, McGaugh SE, Hirsch C (2018) Tandem duplicate genes in maize are ancient and date to two distinct periods of time. G3 8: 3049–3058


