DNA methylation signatures of duplicate gene evolution in angiosperms

Sunil K. Kenchanmane Raju[1], S. Marshall Ledford[2], Chad E. Niederhuth[1,3,*]

[1]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, U.S.A.; [2]Vassar College, Poughkeepsie, NY 12604, U.S.A. [3]AgBioResearch, Michigan State University, East Lansing, MI 48824, U.S.A. [*]Correspondence: Chad E. Niederhuth (niederhu@msu.edu)

ABSTRACT

Gene duplications have greatly shaped the gene content of plants. Multiple factors, such as the epigenome, can shape the subsequent evolution of duplicate genes and are the subject of ongoing study. We analyze genic DNA methylation patterns in 43 angiosperm species and 928 *Arabidopsis thaliana* ecotypes to finding differences in the association of whole-genome and single-gene duplicates with genic DNA methylation patterns. Whole-genome duplicates were enriched for patterns associated with higher gene expression and depleted for patterns of non-CG DNA methylation associated with gene silencing. Single-gene duplicates showed variation in DNA methylation patterns based on modes of duplication (tandem, proximal, transposed, and dispersed) and species. Age of gene duplication was a key factor in the DNA methylation of single-gene duplicates. In single-gene duplicates, non-CG DNA methylation patterns associated with silencing were younger, less conserved, and enriched for presence-absence variation. In comparison, DNA methylation patterns associated with constitutive expression were older and more highly conserved. Surprisingly, across the phylogeny, genes marked by non-CG DNA methylation were enriched for duplicate pairs with evidence of positive selection. We propose that DNA methylation has a role in maintaining gene-dosage balance and silencing by non-CG methylation and may facilitate the evolutionary fate of duplicate genes.

Keywords: Whole-genome duplication, gene duplication, epigenomics, evolution, DNA methylation, angiosperms

Introduction

Gene and genome duplication increases organismal gene content, genome size, and generates a repertoire for functional novelty (Flagel and Wendel 2009; Ohno 1970). Polyploidy or whole-

genome duplication (WGD) is much more pervasive in plants than other eukaryotic lineages (Otto and Whitton 2000). Multiple WGDs over the past 200 million years of angiosperm evolution has led to higher diversity in plant genomes evident from larger differences in genome size between closely related plant species than in other eukaryotes (Panchy, Lehti-Shiu, and Shiu 2016). Small-scale or single-gene duplications (SGDs) (Jiang, Bao, et al. 2004; W. Wang et al. 2006; Bailey et al. 2002; Zhang 2003) contribute to this gene repertoire and have been linked to environmental adaptation and domestication in plants (Michael Freeling 2009; Cuevas et al. 2016; Xiyin Wang et al. 2009; Hanada et al. 2008). Tandem duplicates (TD) are thought to occur through unequal crossing-over, creating clusters of two or more genes adjacent to each other on the same chromosome (Zhang 2003). Proximal duplicates (PD) are gene copies separated by several intervening genes and arose either through the action of local transposition or interruption of ancient tandem duplication (Zhao et al. 1998; M. Freeling et al. 2008). Transposed duplication (TRD) creates a paralogous gene copy from the ancestral location to a novel location either by retrotransposition or DNA-based duplication (Cusack and Wolfe 2007). Finally, dispersed duplication (DSD) generates copies of genes that are neither adjacent to each other nor colinear and are produced through mechanisms that are not well understood (Michael Freeling 2009; Ganko, Meyers, and Vision 2007; Qiao et al. 2019).

Following gene duplication, paralogs are retained or lost (e.g., genome fractionation), and this loss or retention is biased depending on the mode of duplication and gene function (Wendel 2015; Michael Freeling 2009). Products of WGDs are preferentially retained and tend to be enriched for genes involved in macromolecular complexes such as transcription factors, protein kinases, ribosomal proteins, and signal transduction. Multiple models have been proposed to explain the emergence and retention of gene duplicates. Prominent among them, the gene balance hypothesis posits that the stoichiometric imbalance of macromolecular complexes and regulatory networks leads to dosage-dependent phenotypic consequences (Veitia, Bottani, and Birchler 2008). However, comprehensive molecular mechanisms of duplicate gene retention and determining their functional fate remain largely unexplored (F. Cheng et al. 2018; Panchy, Lehti-Shiu, and Shiu 2016).

DNA methylation is an important chromatin modification, altering transcription, and helping to delineate heterochromatin and euchromatin (Law and Jacobsen 2010). DNA methylation plays a

critical role in diverse biological processes such as genome imprinting, X-inactivation, silencing of transposons (TEs) and repetitive regions, and the regulation of gene expression (A. Bird 2002; Edwards and Ferguson-Smith 2007; Finnegan, Peacock, and Dennis 1996). Methylation of cytosines in the CG dinucleotide context can be found throughout plants, animals, and fungi; while non-CG methylation occurring at trinucleotide CHG and CHH (where H is A, T or C) contexts is widespread in plants and known to be associated with repression of transposable elements (Gruenbaum et al. 1981; Meyer, Niedenhof, and ten Lohuis 1994; Zemach et al. 2010). In plants, maintenance of CG methylation (mCG) is carried out by METHYLTRANSFERASE 1 (MET1), a homolog of the highly conserved mammalian DNA METHYLTRANSFERASE 1 (DNMT1) (Finnegan and Dennis 1993). The establishment and maintenance of non-CG methylation involve the plant-specific CHROMOMETHYLASE (CMT) family of DNA methyltransferases and the RNA-directed DNA Methylation (RdDM) pathway (Stroud et al. 2014; Raju et al. 2019).

The functional consequence of DNA methylation is dictated by its location and sequence context and differs for TEs and genes (Law and Jacobsen 2010; Zemach et al. 2010; Sigman and Slotkin 2016). TE repression is an ancient function of DNA methylation (Goll and Bestor 2005; Henderson and Jacobsen 2007) and transcriptional silencing usually involves ubiquitous methylation in CG and non-CG contexts (Roudier, Teixeira, and Colot 2009; Zemach et al. 2010). Even though genes and TEs are methylated distinctly, DNA methylation from adjacent TEs can spread to nearby genes affecting its regulation. Such DNA methylation in gene promoters is usually associated with transcriptional repression (Suzuki and Bird 2008; Ahmed et al. 2011), however, there are notable exceptions (Gent et al. 2013; Harris et al. 2018).

Patterns of DNA methylation in coding regions of plants show strong associations with gene expression patterns. Gene body methylated (gbM) genes are characterized by enrichment of mCG in the transcribed region and depletion at the transcription start site and transcription termination sites. GbM genes comprise the bulk of methylated genes in many plant species, are often housekeeping genes with medium to high expression levels, and evolve slowly compared to unmethylated genes (UM) (Bewick et al. 2016; Takuno and Gaut 2013). UM genes have insignificant amounts of DNA methylation in coding regions, tend to be shorter in length and

number of exons (Takuno and Gaut 2012) and show more variable expression levels (Niederhuth et al. 2016). TE-like genes with persistent methylation of CHG and CHH nucleotide contexts within coding sequences, similar to transposons typically have low expression levels, show little to no conservation in DNA methylation between orthologous genes, and are typically associated with local TEs (Niederhuth et al. 2016; D. K. Seymour et al. 2014; El Baidouri et al. 2018). It is possible in some cases that these genes are mis-annotated transposons (Schnable 2019; Bennetzen et al. 2004), however, TE-like methylation has been observed in many genes of known function (Niederhuth et al. 2016; Schmitz, He, et al. 2013; Schmitz, Schultz, et al. 2013), including species-specific genes (Silveira et al. 2013).

DNA methylation may decrease genetic redundancy through silencing of duplicated gene copies (El Baidouri et al. 2018). Expression reduction models suggest that heavy DNA methylation immediately after duplication buffers the expression levels of duplicate genes (Rodin and Riggs 2003; Chang and Liao 2012). Comprehensive analysis of DNA methylation, gene expression, and chromatin accessibility in humans and mouse have shown that duplicate genes display stronger signatures of DNA methylation distinguishing younger and older duplicates, and corresponding functional divergence in chromatin accessibility and gene expression levels (Keller and Yi 2014; Chang and Liao 2012). Promoters of young duplicates were highly methylated compared to promoters of older duplicates, and gene duplicates from retrotransposition displayed more noticeable DNA methylation divergence compared to other gene pairs (Keller and Yi 2014). However, gene-body methylation did not show a relationship with evolutionary age as promoter methylation. This was in contrast to studies in rice and zebrafish that showed an association between gbM and evolutionary age (Keller and Yi 2014; Y. Wang et al. 2013; Zhong et al. 2016). Recent studies in cassava and soybean showed a strong correlation between gbM genes and expression of WGD genes (H. Wang et al. 2015; Kim et al. 2015). Moreover, gene duplicates derived from different duplication origins in rice showed different correlation directions for gbM and expression divergence of duplicates (Y. Wang et al. 2013). Although the role of DNA methylation in differential expression of gene duplicates have been explored in a few plant taxa (El Baidouri et al. 2018; C. Xu et al. 2018; H. Wang et al. 2015; Y. Wang et al. 2013; Xutong Wang et al. 2017; J. Wang, Marowsky, and Fan 2014), the complex interaction of the genome and epigenome on duplicate gene evolution remains poorly

121 understood at a broad phylogenetic scale. Here, we examine DNA methylation divergence of
122 duplicate genes across 43 diverse angiosperm species to identify general trends in the
123 relationship between DNA methylation and duplicate gene evolution.
124
125

Results:

Patterns and distribution of genic DNA methylation

Annotated gene models from 43 angiosperm species were classified based on their genic DNA methylation patterns (Niederhuth et al. 2016; Takuno and Gaut 2012) as gene body methylated (gbM), transposable element-like (TE-like), or unmethylated (UM) (Table S1). We then explored the genomic distributions of each genic DNA methylation class for each species. It has been shown that DNA methylation tends to be higher in the centromere and pericentromeric regions (Cokus et al. 2008; Lister et al. 2008) and non-CG methylation of genes is associated with local TE content and genome size (Niederhuth et al. 2016; Takuno, Ran, and Gaut 2016). Previously, El Baidouri et al. took advantage of high-resolution genetic maps in soybean to delineate pericentromeric regions to show that TE-like genes are enriched in the TE-rich pericentromeric regions (El Baidouri et al. 2018). Lacking such detail for most species, we used the number genes, TEs, and TE-base pairs in sliding windows as a proxy for regions of euchromatin and heterochromatin (Figure 1, Figure S1). Chromosomal plots of the distributions of methylated genes and TEs, helped visualize these distributions (Figure S2) It should be noted that this analysis does not capture TE content in the immediate vicinity of genes, but instead reflects large distribution patterns of both genes and TEs across chromosomes. All three classes of genic methylation classifications (gbM, TE-like, and UM genes) showed positive correlation with the distribution of genes in all species (Table-S2), except for TE-like genes in *Arabidopsis thaliana,* that showed a negative correlation with gene distribution (Pearson's r = -0.31, $p < .001$). The distribution of gbM and UM genes was negatively correlated with TE content across the majority of species (gbM: 34/43 species; UM: 30/43 species). Those species which showed a positive correlation between gbM/UM genes and TE content also showed a positive correlation between TE content and gene distribution. TE-like genes were positively correlated with TE distribution in *A. thaliana* (Figure 1A, Figure S1,S2) and most other species (28/43). Surprisingly, a number of species did not show this expected pattern (Figure 1B-F, Figure S1,S2).

We observed that many of the species deviating from this expected pattern were located in the legumes and grasses, which could suggest that there may be lineage-specific effects on the distributions. We used Pagel's lambda (λ) to test for phylogenetic signals under a 'Brownian

Motion' model of trait evolution (Münkemüller et al. 2012; Pagel 1999). The correlation between TE-like genes and TEs showed a strong phylogenetic signal (Pagel's lambda (λ) = 0.61, p = *.01*). Similarly, correlation between gbM genes and TEs also showed evidence of phylogenetic signal (Pagel's lambda (λ) = 0.82, p = *.02;* Table-S3). For gbM genes, it was previously shown that there is a lineage-specific reduction of gbM in the Brassicaceae, which could explain the observed signal for gbM. To test this, we reanalyzed the data by excluding *Eutrema salsugineum*, *Brassica rapa*, and *Brassica oleracea* from the analysis and still found significant evidence of phylogenetic signal for gbM genes and TEs (Pagel's lambda (λ) = 0.76, p = *.04)*. As genome size correlates with genic non-CG methylation (Niederhuth et al. 2016; Takuno and Gaut 2012), we analyzed genome size data from the Plant DNA C-values database (release 7.1; (Pellicer and Leitch 2020)) for all the species used in our study and did not find a phylogenetic signal for genome size (Pagel's lambda (λ) = 0.21, p = *.52*). The lack of phylogenetic signal for genome size in these species suggests that the lineage-specific changes in distribution of gbM and TE-like genes are unlikely to be derived from genome size variation.

Associations of genic DNA methylation with mode of gene duplication

To understand the relationship between genic DNA methylation and gene duplication we identified and classified duplicate genes based on their mode of duplication (Table-S4) using *DupGen_finder-unique (Qiao et al. 2019).* We then performed a two-sided Fisher's exact test to obtain an odds ratio for each mode of gene duplication with each DNA methylation classification (Figure 2, Table-S5). WGDs are depleted of TE-like genes across all species and enriched for gbM genes in most species, except *B. rapa* (odds.ratio = 0.53, $p < .001$). Several other species showed under-representation of gbM genes at a lower p-value cutoffs ($0.001 < p < 0.05$), *E. salsugineum* (odds.ratio = 0.29, p = .009), *Cucumis sativus* (odds.ratio = 0.91, p = .03) and *Citrullus lanatus* (odds.ratio = 0.88, p = .008) (Table-S5). WGDs showed enrichment for unmethylated genes in 26 out of 43 species. Of the remaining species, *Gossypium raimondii* (odds.ratio = 0.92, p = .003), *Populus trichocarpa* (odds.ratio = 0.73, $p <$ *.001*) and *Glycine max* (odds.ratio = 0.63, $p <$ *.001*) showed significant depletion of UM genes in WGDs.

For single-gene duplications (SGDs), association with genic DNA methylation differed significantly from WGD and between modes of SGD. Tandem and proximal duplicates were depleted of gbM genes and often enriched in TE-like and UM genes (Figure 2). There were

exceptions, with members of Fabaceae (except *Phaseolus vulgaris*) and Poaceae (except *Sorghum bicolor*) showing depletion of TE-like genes in tandem duplicates (Table-S5). Dispersed genes showed clear depletion of UM genes and enrichment of TE-like genes. Patterns of gbM in dispersed genes were mixed. While members of Poaceae (except *Zea mays* and *Pannicum virgatum*) showed enrichment of gbM genes, Fabaceae (except *P. vulgaris*) showed depletion of gbM genes (Table-S5). Transposed genes were depleted in UM and TE-like genes while showing enrichment for gbM genes. Exceptions are *Z. mays*, *Malus domestica*, *Fragaria x ananassa* and *G. max,* which showed depletion of gbM genes and enrichment of TE-like genes. Singletons are depleted in gbM genes and enriched in UM genes (Figure 2). These findings enable us to determine the contributions of genic DNA methylation patterns on the evolutionary fates of gene duplicates.

Stasis and switching in genic DNA methylation states

Duplicate gene pairs may have different methylation profiles. Multiple factors can influence these differences, such as methylation of parent-of-origin gene, evolutionary constraints, and chromatin environment of the new duplicate copy. We first compared duplicate pairs where each gene in the pair could be conclusively classified into one of the three genic methylation classes (Figure 3A, Table-S6). WGD pairs tend to resemble each other, with ~63% to 96% gene pairs (median - 87%) having the same methylation profiles. Similarly, tandem (~71%-94%, median – 82%), proximal (~63%-95%, median – 76%), and dispersed (~66-95%, median – 82%) duplicates also showed a higher proportion of gene pairs with similar methylation profiles. Transposed duplicates had the broadest range, from 54% gene pairs showing similar methylation profiles in *Z. mays* to 90% in *E. salsugineum* (median – 77%). This suggests that regardless of the mode of duplication, most duplicate copies retain similar DNA methylation profiles, however, switching is not infrequent. For most modes of duplication, the original and derived copies cannot be determined, so we cannot determine the directionality of switching. For transposed duplicates, however, it is possible to identify a parental and transposed copy indicating the direction of switching. A higher proportion of transposed copies were TE-like genes compared to their parental gene in all species except *C. lanatus* (Table-S7, Figure 3B). Transposed copies also showed a lower proportion of gbM (40/43 species) and UM genes (32/43 species) compared to their parent gene in most species (Table-S7, Figure 3B). While the

majority of transposed duplicates retain the same genic DNA methylation status as their parental copy, TE-like transposed copies showed a higher frequency of having switched from their original methylation status, with higher proportions of the parental copies being either gbM or UM (Figure 3C).

Genic DNA methylation patterns associate with gene age

We hypothesized that there may be a relationship between the age of gene duplication and patterns of genic DNA methylation and so used two different approaches to address this question. In the first approach, we examined the distribution of synonymous mutations (Ks) based on both the mode of gene duplication and the class of genic DNA methylation (Figure 4A,B,D,E, Figure S3). Ks distributions have been widely used to date gene duplication events, as synonymous mutations are expected to have less of an impact on gene function and so can accumulate with evolutionary age. A lower Ks, therefore suggests a more recent duplication event (Lynch and Conery 2000). We find distinct associations between WGDs and SGDs, their genic DNA methylation profiles, and the age of gene duplication (Figure 4A,B,D,E; Figure S3; Table S8). TE-like SGDs were clearly younger, while gbM SGDs were older. For WGDs, the situation is more complex. In some species, WGD gbM genes were younger, while in others, they were older. These differences likely reflect differences since the last WGD and/or a history of nested WGDs.

In the second approach, we used *MCScanX-transposed* to detect transposed gene duplication events that occurred at different periods (epochs) in the phylogenetic tree since species divergence (Table-S9) based on sequential exclusion of the closest outgroup (Y. Wang, Li, and Paterson 2013). This approach is independent of the Ks-based approach used above. We then tested transposed genes in each epoch for enrichment or depletion of genic DNA methylation (Figure 4C,F; Figure S4, Table S10). Younger transposed duplicates were depleted in gbM genes and enriched for TE-like genes, while more ancient transposed duplicates were depleted of TE-like genes and enriched for gbM genes. The trend in UM genes was unclear. *B. oleracea* and *P. virgatum* did not show a significant depletion in gbM genes in younger transposed genes, while *B. rapa* (odds.ratio = 2.45, $p < .001$), and *E. salsugineum* (odds.ratio = 13.08, $p = 0.004$) showed significant enrichment of gbM genes in younger transposed genes (Table-S10). This deviation from the depletion of gbM genes in younger transposed genes for Brassicaceae is likely due to

the known low number of gbM in these species (Bewick et al. 2016; Niederhuth et al. 2016), while we cannot explain this effect in *P. virgatum*.

We reasoned that some of the variation in the enrichment of genic DNA methylation classes amongst SGDs might be due to variation in the age of SGDs. To test this, we plotted the distribution of Ks for transposed, tandem, and proximal genes, ordered based on their median Ks (Figure S5). For transposed duplicates, species enriched for TE-like genes tended to be younger, while species depleted for TE-like genes tended to be older. For tandem and proximal genes, however, the pattern was not as clear, suggesting a much more complex scenario where other factors contribute to the enrichment or depletion of TE-like genes amongst these duplicates.

Genic DNA methylation and duplicate gene evolution

There are multiple possible evolutionary fates of duplicate genes (Panchy, Lehti-Shiu, and Shiu 2016). The ratio of nonsynonymous (Ka) to synonymous mutations (Ks) is commonly used to test for selection (Kreitman 2000). A Ka/Ks < 1 is indicative of purifying selection, Ka/Ks ~1 suggests a protein is evolving neutrally, while a Ka/Ks > 1 is suggestive of positive selection. Comparing the Ka/Ks distributions for duplicate gene pairs (Figure 5A-D, Figure S6, Table-S11) shows that the vast majority of duplicate pairs have a Ka/Ks < 1, indicating purifying selection. However, the distribution of TE-like genes, particularly for single-gene duplicates, is strongly shifted toward higher values, which suggests relaxed selective constraints on many TE-like genes. This trend was true of all TE-like gene duplicates, regardless of mode of duplication, but was particularly pronounced for transposed and dispersed SGDs (Figure S6). Interestingly, the distribution of TE-like genes in many species also shows a shift toward Ka/Ks > 1 (Figure 5, Figure-S6). Limiting our analysis to duplicate genes with a Ka/Ks > 1.1, TE-like genes were significantly overrepresented in 38/43 species(Figure 5E,F, Table-S12). Of the other five, the number of only *L. japonicus* appeared to show an opposite pattern, but this was non-significant given the low number of genes with Ka/Ks > 1.1. GbM genes, in contrast, were significantly underrepresented in 39/43 species and of the remaining four, only *E. salsugineum* showed an opposite pattern (Figure 5E,F, Table-S12). UM genes were significantly overrepresented in 18 species, and significantly underrepresented in 14 species (Figure 5E,F, Table-S12).

Larger-impact mutations besides nonsynonymous substitutions can impact genes. For instance,

differentially methylated regions in all three sequence contexts (C-DMRs) are enriched amongst structural variation in *A. thaliana* (Kawakatsu et al. 2016). Presence absence variation (PAV) is a type of structural variation where entire gene copies may be present or absent between individuals in a species. We used previously published PAV data from four species (*B. oleracea*, *Solanum lycopersicum*, *Solanum tuberosum*, and *Z. mays*) to examine whether there was a relationship between PAV and genic DNA methylation. We initially examined the relationship between genic DNA methylation for all genes (duplicate and non-duplicate) and PAVs in all four species (Figure S8; Table-S13). TE-like genes showed a significant enrichment amongst PAVs ($p < 0.001$) in all species, while gbM genes were significantly depleted ($p < 0.001$) in all but *B. oleracea*. UM genes showed conflicting patterns, being enriched in both *B. oleracea* and *S. lycopersicum* ($p < 0.001$), depleted in *Z. mays* ($p < 0.001$) and no-significant enrichment in *S. tuberosum*. These same observations held up when the analysis was limited to duplicate genes (Figure S8, Table-S13).

Population scale genic DNA methylation of duplicate genes

Within a species DNA methylation can vary across the population (Schmitz, He, et al. 2013; Kawakatsu et al. 2016; Vaughn et al. 2007; Eichten et al. 2013, 2016). We leveraged methylome data from the *A. thaliana* 1001 Epigenomes Project (Kawakatsu et al. 2016) to examine the relationship between gene duplication and DNA methylation at a population level. Genes from 928 *A. thaliana* accessions were classified as before. For each genic methylation classification, we binned the genes based on the frequency of each classification (0%, <25%, 25%-50%, 50%-75%, >75%) across all 928 accessions (see methods). Looking at the distribution of synonymous substitutions (Ks), we see clear differences in the frequency of methylation states and the age of duplication. GbM genes increase in frequency across the population with increasing age of duplication, while more recently duplicated genes are less likely to be gbM (Figure 6A). TE-like and unmethylated genes show the opposite pattern (Figure 6B,C), with increased frequency in younger duplicates

Increasing frequency of gbM across the population corresponds with an increase in Ks, while increasing frequency of both TE-like and UM corresponds to a decrease in Ks (Figure 6D-F). This suggests a relationship between the age of duplication and population-level frequencies of genic DNA methylation. We further explored how each duplication mode affects genic

304 methylation variation across diverse *A. thaliana* accessions. While dispersed duplicates with
305 gbM in <25% of accession are equally distributed among young and old genes, duplicates with
306 gbM in more accession tend to be older dispersed duplicates (Figure S9A). In contrast, we see
307 exactly the opposite trend in TE-like genes where younger dispersed genes have higher
308 proportion of TE-like methylation across the accessions. We also see a similar pattern in the
309 transposed duplicates, with older transposed duplicates with TE-like methylation showing <25%
310 accession with TE-like methylation (Figure S9B). These results suggest that DNA methylation
311 plays an important role in the evolution of duplicate gene retention through inter- and intra-
312 species variation in genic methylation patterns.
313

Discussion:

We have characterized DNA methylation patterns and modes of gene duplication across 43 Angiosperm species Our findings show that different patterns of genic methylation are associated with different modes (WGD vs SGD) and features (e.g. duplicate age) of gene duplication. The role of the epigenome, in particular, DNA methylation, in the evolution of duplicate genes has been explored in a handful of taxa (Keller and Yi 2014; Xutong Wang et al. 2017; Kim et al. 2015; C. Xu et al. 2018; Y. Wang et al. 2013; Zhong et al. 2016; Rodin and Riggs 2003; Chang and Liao 2012; El Baidouri et al. 2018). Lineage-specific differences in DNA methylation pathways, rates of gene duplication, and how the data are analyzed can all contribute to differences in the observed associations between DNA methylation and gene duplication. We focused on DNA methylation within coding regions as these have shown consistent associations with gene expression and other genic features in plants (Takuno and Gaut 2012; Niederhuth et al. 2016; D. K. Seymour et al. 2014; Takuno, Ran, and Gaut 2016).

For most species, TE-like genes are enriched in TE-rich regions, while gbM and UM genes are depleted. However, we find a number of exceptions to these trends and evidence of phylogenetic signals for the correlations of TE-like and gbM genes with TE-content. This suggests that there are lineage-specific differences in how these genes are distributed. The underlying cause(s) of these differences remains to be determined, but we suspect that it may reflect differences in chromosomal architecture and the distributions of genes and TEs. For instance, rice is known to have a high content of MITEs near genes not found in other species (Jiang, Feschotte, et al. 2004). In soybean, we found that TE-like genes were negatively correlated with TE content, which appears to contradict the enrichment of soybean TE-like genes in pericentromeres found by El Baidouri et al (2018). This discrepancy could be due to differences in classifying genes and chromosome binning between the two papers. It must be emphasized that our gene classification was very conservative, and thus, a large proportion of genes for which methylation data was ambiguous remained ‘unclassified.’

We find distinct patterns between genic DNA methylation and different modes of gene duplication. WGD genes are consistently depleted in TE-like genes and commonly enriched in gbM and unmethylated genes. A notable exception of this is *E. salsugineum* and *B. rapa*, where

it has been shown that gbM has been lost or reduced due to the evolution of the DNA methyltransferase *CHROMOMETHYLASE 3 (Bewick et al. 2016)*. The story is more complicated for SGDs. El Baidouri et al. (2018) found that in soybean, non-collinear genes, excluding tandem duplicates (e.g., transposed, dispersed), were enriched for TE-like DNA methylation and that this was due to translocation to heterochromatin rich regions. We find that when examined across species, SGDs show high variance in their enrichment or depletion of different types of genic DNA methylation. Certainly the local context of TEs is a major factor, but contrary to this, even transposed duplicates showed enrichment of gbM and depletion of TE-like methylation in some species. Variance in the correlation between TEs and genic DNA methylation patterns all suggest that there are other factors that contribute to determining DNA methylation in duplicate genes.

While we see evidence of switching of genic DNA methylation states between duplicates, the majority have the same state as its duplicate. While the high degree of correspondence between duplicates, might be most easily explained by the duplicated copy retaining the methylated state of its parent, this is not a given. Correspondence of DNA methylation states could arise by multiple mechanisms. For instance, in the *PAI* genes of Arabidopsis, an inverted tandem duplicate, *PAI4,* results in silencing in *trans* of all four members of this gene family by the RdDM pathway (Luff, Pawlowski, and Bender 1999; Melquist, Luff, and Bender 1999). So in this example, the methylation of all genes is altered but still correspond to each other. Divergence in duplicate pairs could arise immediately, when a duplicated copy is translocated to a different chromatin context (El Baidouri et al. 2018), but could also arise much later due to localized TE insertions. From our comparative analyses, we cannot with confidence determine the directionality or timing in changes of genic methylation state.

Our results show a significant relationship between the age of gene duplication and genic DNA methylation. This is especially evident for SGDs, where TE-like genes were consistently younger, and gbM genes were older. These patterns differ for WGDs, depending on the species. The consistency of the patterns observed in SGDs vs WGDs might be explained by lineage-specific differences in the number of WGDs and time since last WGD for each lineage (Garsmeur et al. 2014; Van de Peer, Mizrachi, and Marchal 2017). Within a species, there is an association between the frequency of genic DNA methylation and duplicate gene age across the population. In the case of TE-like duplicate genes with high-frequency in the population, it

seems most likely that such DNA methylation was established very early in the history of *A. thaliana*, rather than being due to multiple independent instances of silencing. This would suggest then that despite being evolutionarily younger, these genes still duplicated early enough to be present throughout the population, rather than being limited to specific lineages. In contrast, those genes with intermediate frequencies of gbM, UM, or TE-like methylation patterns indicate ongoing switching of genic DNA methylation states in Arabidopsis. Much of this variance in DNA methylation can be directly explained by local genetic variation such as novel TE insertions (Schmitz, He, et al. 2013; Kawakatsu et al. 2016; Vaughn et al. 2007; Eichten et al. 2013, 2016).

Duplicate gene sequences evolve differently based on their genic DNA methylation class. GbM duplicate genes are highly expressed in plants and enriched for housekeeping gene functions, and in comparisons of orthologous genes are more highly conserved ((Niederhuth et al. 2016; Takuno and Gaut 2012). Consistent with such a role, gbM duplicate genes, had a Ka/Ks < 1, indicative of purifying selection. In contrast, if a gene were transcriptionally silenced in all tissues, it would be expected to have little impact on phenotype and so might be expected to evolve neutrally. However, the posits that heavy DNA methylation immediately following duplication is a means to suppress duplicate gene expression, thereby shielding it from accumulating deleterious mutations and 'pseudogenization'(Chang and Liao 2012; Rodin and Riggs 2003). Similarly, Adams et al. have also proposed that silenced gene duplicates in cotton are protected from mutational loss (Adams et al. 2003). Our data shows that TE-like genes had a higher Ka/Ks than gbM or UM genes, suggestive of relaxed selection, and were enriched amongst PAVs. This would suggest that in contrast to the 'expression reduction model', most TE-like duplicates are or are on their way to becoming pseudogenes and lost. This also fits with the observed lack of conservation of TE-like methylation patterns between species(B. Seymour, Andreosso, and Seymour 2015; Niederhuth et al. 2016). However, we did find that there was an enrichment of TE-like genes with a Ka/Ks > 1, suggestive of positive selection. This would suggest that silencing by DNA methylation could facilitate functional divergence, as suggested by Rodin and Riggs (2003). SGDs, in particular, have been linked to adaptation to environmental conditions (Dassanayake et al. 2011; Hanada et al. 2008). Most TE-like genes are of unknown function or enriched in functions related to defense (e.g., chitinases, proteinases) (Niederhuth et al. 2016). Interestingly, a previous search for plant genes under positive selection primarily

identified gene families with unknown function and many of the same defense-related genes (Roth and Liberles 2006). Others have observed that genes associated with C-DMRS (differentially methylated regions in all three sequence contexts) are often specifically expressed in specific tissues, especially pollen (Schmitz, Schultz, et al. 2013). So while the evidence strongly suggests most duplicate genes with TE-like methylation are on their way to extinction, there is some evidence to support models whereby silencing facilitates functional divergence, but much more extensive study and evidence is needed to support this possibility.

DNA methylation outside of coding regions could also affect the fate of gene duplicates. In animals, promoter DNA methylation divergence increases with evolutionary age and correlates with tissue-specific expression and chromatin accessibility profiles (Keller and Yi 2014). Promoter DNA methylation of younger duplicates was also found in zebrafish and was suggested to offset detrimental mutations and pseudogenization (Chang and Liao 2012; Zhong et al. 2016). DNA methylation of upstream regions does show some associations with gene expression in some plants and is typically thought of being repressive. However, the actual effects can be quite complex, sometimes contradictory, and poorly understood. For instance, methylated CHH-islands in promoter regions are associated with increased gene expression (Li et al. 2015; Gent et al. 2013; Niederhuth et al. 2016). The exact impact of DNA methylation on transcription factor (TF) binding is often unknown; as depending on the TF it may inhibit, have no impact, or even facilitate TF binding (Medvedeva et al. 2014; O'Malley et al. 2016). Finally, *cis*-regulatory regions for most genes in most plant species are unknown or poorly defined. To fully understand how DNA methylation of regulatory regions affects duplicate gene evolution will require integration of DNA methylation, known TF binding sites, and their target genes.

We propose the following model for gene duplication and genic DNA methylation. While WGD doubles genome size and, in the case of allopolyploidy, can bring distinct genomes and their *trans* regulatory factors together, the local sequence context (*cis*-regulatory elements, neighboring TEs, etc.) for most genes does not change. For most WGD genes, it is unlikely that their genic DNA methylation states will change either. Supporting this is the high correspondence of DNA methylation between duplicates. In resynthesized allopolyploids, there has been little evidence of extensive changes to genic DNA methylation, however, generational changes are observed in flanking regions (Edger et al. 2017; K. A. Bird et al., n.d.). Features of

gbM genes, such as enriched functional categories, are consistent with features of retained WGD pairs. The enrichment of gbM and depletion of TE-like genes in WGDs, therefore, fits with expectations of the gene balance hypothesis that postulates retained WGD pairs are dosage-sensitive genes, enriched in specific functions or part of protein complexes (Veitia, Bottani, and Birchler 2008). SGDs can change the local sequence and chromatin context of a gene, by moving that gene to regions of euchromatin or heterochromatin, but also increases the dosage of that gene relative to the whole. There are, therefore, more mechanistic opportunities for silencing SGDs, and the fitness effects dosage-sensitive genes may further promote this. In most cases, these genes are destined for the dustbins of evolution. However, in rare circumstances, silencing of duplicate genes may help facilitate evolutionary novelty (Rodin and Riggs 2003). In both WGDs and SGDs, we suspect that genic DNA methylation states are typically established immediately or very soon after duplication, however, ongoing novel TE insertions or selection for functional novelty can result in DNA methylation shifts at any point in a gene's evolutionary history.

In summary, this study demonstrates that whole-genome duplication and single gene duplications show distinct patterns of genic DNA methylation suggesting the action of different chromatin-based mechanisms leading to contrasting evolutionary trajectories for duplicate gene retention and loss.

Methods:

Genomes and annotations

We used the public genomes and gene annotations of 43 species with methylome data (Table-S13) and an additional 11 species duplicate gene identification(Garcia-Mas et al. 2012; Guo et al. 2013; Ming et al. 2008; Wu et al. 2018; Dohm et al. 2014; Parkin et al. 2014; Initiative and The International Brachypodium Initiative 2010; Amborella Genome Project 2013; Lamesch et al. 2012; C.-Y. Cheng et al. 2017; Bertioli et al. 2016; Hu et al. 2011; Sato et al. 2008; Paterson et al. 2012; Schmutz et al. 2010; Edger et al. 2019, 2018; Singh et al. 2013; Bartholomé et al. 2015; Li et al. 2019; Slotte et al. 2013; D'Hont et al. 2012; R. Yang et al. 2013; Daccord et al. 2017; Bredeson et al. 2016; Hellsten et al. 2013; Tang et al. 2014; Kawahara et al. 2013; Lovell et al. 2018; Verde et al. 2017; Tuskan et al. 2006; Schmutz et al. 2014; Xue et al. 2018; McCormick et al. 2018; Bennetzen et al. 2012; Hosmani et al., n.d.; Sharma et al. 2013; Mamidi et al., n.d.; Motamayor et al. 2013; Jiao et al. 2017; Hibrand Saint-Oyant et al. 2018; VanBuren et al. 2018; Liu et al. 2014; Colle et al. 2019; VanBuren et al. 2015; Harkess et al. 2017; Hulse-Kemp et al. 2018; S. Xu et al. 2017; Bombarely et al. 2016). These were downloaded from multiple databases (Table-S13). Only the primary transcript for each gene was used. As there were differences in the availability and quality of transposon annotations, we re-identified TEs *de novo* for each species using the Extensive *de-novo* TE Annotator (EDTA) pipeline (Ou et al. 2019) provided with gene annotations and coding sequences.

DNA methylation data and analyses

We used previously published whole-genome bisulfite sequencing from forty-three Angiosperm species (Amborella Genome Project 2013; D. K. Seymour et al. 2014; Picard and Gehring 2017; Bertioli et al. 2016; Niederhuth et al. 2016; Bewick et al. 2016; Lü et al. 2018; Ong-Abdullah et al. 2015; J. Cheng et al. 2018; Kim et al. 2015; Song et al. 2017; Daccord et al. 2017; Secco et al. 2015; Dong et al. 2017; Y. Yang et al. 2019; L. Wang et al. 2018; Turco et al. 2017; Noshay et al. 2019). To the extent that was possible we chose datasets from leaf tissue, with high-sequencing coverage, a low non-conversion rate, and from the same accession as the reference genome (Table-S14). Data were mapped to their respective genomes using methylpy v1.2.9 (Schultz et al. 2015) and the non-conversion rate calculated based on either unmethylated spiked-

id phage lambda DNA or endogenous chloroplast genomes (Table-S14). Individual cytosines were called as methylated or unmethylated by methylpy. The genic DNA methylation of each gene was then classified using custom python scripts as previously done (Niederhuth et al. 2016). First a background rate of DNA methylation for coding sequences was determined by calculating the average number of methylated CG, CHG, and CHH sites for all genes in all species. A binomial test was then applied to each gene to test for enrichment of CG, CHG, or CHH methylation against the background DNA methylation rate. GbM genes had a minimum of 20 CG sites covered by three or more reads each, a significant enrichment of methylated CG sites (FDR corrected p-value < 0.05), and non-significant amounts of non-CG methylation. TE-like genes had a minimum of 20 CHG sites covered by three or more reads and significantly enriched for CHG or 20 CHH sites covered by three or more reads and significantly enriched for CHH. Unmethylated genes had a minimum of 20 cytosines covered by three reads or more each and no sites called as methylated. All other genes were classified as either 'unclassified' or if missing data, considered NA.

Genomic distribution and phylogenetic comparisons

To determine the genomic distribution of methylated genes we calculated the total number of genes, genes belonging to each of the genic DNA methylation classes, the number of TEs, and number of TE nucleotides in 100 kb sliding windows, sliding every 50 kb. Pearson correlation coefficients (*r*) were calculated using the '*rcorr*' function in the R package '*corrplot*' (Wei et al. 2017). These correlations were tested for phylogenetic signals using the function *phylosig* (method="lambda") in the R package *phytools (Revell 2012)*. Pagel's lambda measures phylogenetic signal under a Brownian motion model of trait evolution where a value of '0' indicates no phylogenetic signal and '1' as a strong phylogenetic signal (Münkemüller et al. 2012; Pagel 1999). The genome size of all species were extracted from the Plant DNA C-value database as the amount in picograms of DNA contained within a haploid nucleus of the plant species (Pellicer and Leitch 2020). These were then tested for the phylogenetic signal as above. The input phylogenetic tree and branch lengths (Dataset-S1) used in *phylosig* was created with orthofinder (Emms and Kelly 2015) with default parameters.

Gene duplication classification

For each species, DIAMOND (Buchfink, Xie, and Huson 2015) was used to perform a blastp search of all genes against itself and *A. trichopoda*. A maximum of 5 hits per gene, with a e-value < 1e-10 were kept. Duplicate genes were identified and classified by *DupGen_finder-unique* (Qiao et al. 2019). *MCScanX-transposed (Y. Wang, Li, and Paterson 2013)* was used to detect transposed duplicates occurring within different epochs since species divergence (Table-S9). Custom R scripts were used to determine the frequencies of each of the different modes of gene duplication, as well as genic methylation classification (gbM, TE-like, unmethylated, and unclassified). A two-sided Fisher's exact test was used to determine whether there was any statistically significant association between modes of duplication and genic methylation classification. A stringent p-value threshold of <.001 was set for all comparisons, however, we do report p-value significance between .001 to .05 in the heatmap. Heatmaps were plotted using the function *heatmap.2* in the R package *gplots*. The Phylogenetic tree in Figure S2 was created in R using the packages 'V.PhyloMaker' and 'phytools' (Jin and Qian, 2019).

Nucleotide evolution

Nonsynonymous substitutions (Ka), synonymous substitutions (Ks) and the ratio of Ka/Ks for each duplicate pair was determined using the calculate_Ka_Ks_pipeline.pl (Qiao et al. 2019). Briefly each pair of protein sequences are aligned using MAFFT (v7.402)(Katoh and Standley 2013) and converted into a codon alignment using PAL2NAL (Suyama, Torrents, and Bork 2006). A modified version of the Yang-Nielsen method, γ-MYN was used to calculate Ka and Ks values using the Tamura-Nei model (D. Wang et al. 2010; Qiao et al. 2019). For a subset of duplicate genes, multiple possible duplicate pairs are possible. To minimize bias in these cases, we randomly selected one set of Ka & Ks values to represent that gene. We tested the distribution of Ks and Ka/Ks for gbM, TE-like, and UM genes for divergence from the distribution of an equal number of randomly selected genes using the Kolmogorov-Smirnov test (Massey 1951), a non-parametric test. Gene pairs with a Ka/Ks > 1.1 were considered as candidates undergoing positive selection and were tested for enrichment or depletion amongst gbM, TE-like, and UM genes using a two-sided Fisher's exact test (Fisher 1935) and an FDR-corrected p-value < 0.05.

Presence absence variation

Previously published PAV variants were downloaded for *B. oleracea* (Golicz et al. 2016), *S. lycopersicum (Gao et al. 2019)*, *S. tuberosum (Hardigan et al. 2016)*, and *Z. mays (Hirsch et al. 2014)*. For both *S. tuberosum* and *Z. mays*, genes with an average read coverage of < 0.2 in at least one accession were considered to be PAVs. In all species, only genes present in the reference accession were considered for analyses, as DNA methylation data was not available for non-reference genes. PAVs were merged with genic DNA methylation data in R and tested for enrichment or depletion using a two-sided Fisher's Exact test (Fisher 1935) and an FDR-corrected p-value < 0.05..

Arabidopsis diversity

Processed WGBS data from the previously published *Arabidopsis thaliana* 1001 Epigenomes Project (Kawakatsu et al. 2016) was downloaded from the Gene Expression Omnibus (GEO Accession GSE43857). Of these, 928 accessions had been called using methylpy and were compatible with our methods. Genes were classified as done above for the 43 species using the same background methylation rates. These data were imported into R and the frequency of each genic DNA methylation class for each gene in the population binned into 0%, <25%, 25-50%, 50-75%, and >75%. Ks & Ka/Ks data were the same as those computed above.

Data availability and research reproducibility

All data used in this study are from publicly available genomes and methylomes (Table-S13). Classified gene lists, gene DNA methylation levels, Ka/Ks data, and formatted genomes and annotations suitable for reproducing our results are available at data dryad. Conda environments with exact software versions and all custom scripts are available at both https://github.com/niederhuth/DNA-methylation-signatures-of-duplicate-gene-evolution-in-angiosperms.

Acknowledgements:

We would like to thank Dr. Patrick Edger for kindly letting us use the unpublished *C. violaceae* genome in our analysis. The authors would also like to thank and dedicate this work to the healthcare workers during the COVID-19 pandemic. Dr. Niederhuth would also like to dedicate this work to his son Harrison Edgar Niederhuth who was born during the writing of this

manuscript. This work was supported by Michigan State University, by USDA National Institute of Food and Agriculture Hatch Funds (project number MICL02572), and the National Science Foundation (grant IOS-2029959). S Marshall Ledford was supported by the Plant Genomics @ MSU REU, funded by the National Science Foundation (grant DBI-1757043).

Author contributions:

S.K.K.R and C.E.N conceptualized and designed the work and analysis. S.K.K.R, C.E.N. and S.M.L performed data analysis. S.K.K.R made figures and wrote the original draft; C.E.N wrote and edited the manuscript. All authors read and approved the final manuscript.

FIGURE LEGENDS

Figure 1: Correlations of genomic features with genic DNA methylation classifications. Pearson's correlation coefficient 'r' was measured to determine the distribution of gbM, TE-like, and UM genes across the genome using TE content and TE nucleotides as a proxy for heterochromatin. Shades of blue indicate enrichment, while red indicates depletion. Correlations not significant at $p < 0.001$ are marked with 'X'. (A) *Arabidopsis thaliana*, (B) *Glycine max*, (C) *Medicago truncatula*, (D) *Panicum virgatum*, (E) *Solanum lycopersicum*, and (F) *Zea mays*.

Figure 2: Patterns of genic DNA methylation across different modes of gene duplication across 43 angiosperms. Heatmap shows enrichment/depletion of each genic DNA methylation class (gbM, TE-like, and UM) for each mode of gene duplication (WGD, tandem, proximal, dispersed, transposed). Singletons are single-copy genes (non-duplicates) present in the species and the outgroup. Unless indicated, all associations are statistically significant at a p-value < 0.001. *significant association at a lower p-value cut-off ($0.001 < p < 0.05$). 'NS' – not significant. Cyan indicates depletion, while pink denotes enrichment.

Figure 3: Stasis and switching of genic DNA methylation profiles between duplicate gene pairs. (A) The number of genes for each mode duplication with the same or different genic DNA methylation profile in *A. thaliana* and *Z. mays*. (B) Percentage of parental and transposed genes with gbM, TE-like, and UM classification for *A. thaliana* and *Z. mays* shows enrichment of TE-like methylation in the transposed copy. (C) Proportion of parental genes in each genic DNA methylation class based on the genic DNA methylation in *A. thaliana* and *Z. mays*. This indicates

596 the direction of genic DNA methylation switching.

597 **Figure 4:** Relationship between genic DNA methylation class and the age of gene duplication.
598 Density plots (A,B, D,E) show the distribution of synonymous substitutions (Ks) for each genic
599 DNA methylation class , while bar plots (C and F) show the percentage of gene copies in each
600 genic DNA methylation class for transposed genes that have duplicated during that 'epoch' since
601 divergence from the species on the x-axis. For example, in *B. distachyon* transposed genes that
602 have duplicated since *B. distachyon* diverged from *O. sativa* are shown on the x-axis under *O.*
603 *sativa*. Those shown for *Z. mays,* duplicated in the period since the common ancestor of *B.*
604 *distachyon* and *O. sativa* diverged from their common ancestor with *Z. mays*, but before the
605 divergence of *B. distachyon* and *O. sativa*. Horizontal dotted lines indicate the percentage of
606 each genic DNA methylation class in the whole genome. (A) *B. distachyon* WGD Ks
607 distributions (B) *B. distachyon* SGD Ks distributions. (C) Percentage of *B. distachyon*
608 transposed duplicates in each genic DNA methylation class at each 'epoch'. (D) *G. raimondii*
609 WGD Ks distributions (E) *G. raimondii* SGD Ks distributions (F) Percentage of *G. raimondii*
610 transposed duplicates in each genic DNA methylation class at each 'epoch'.

611 **Figure 5:** Differences in sequence evolution correspond to different genic DNA methylation
612 classes. TE-like methylation genes are under relaxed selection compared to gbM and UM genes
613 as indicated by density plots showing the ratio of non-synonymous to synonymous mutations
614 (Ka/Ks) for (A) *B. distachyon* WGDs, (B) *B. distachyon* SGDs, (**C**) *G. raimondii* WGDs and (D)
615 *G. raimondii* SGDs. However, TE-like methylated genes are also enriched in genes showing
616 evidence of positive selection. Bar graphs indicate the proportion of genes for each genic DNA
617 methylation class with a Ka/Ks > 1.1 for (E) *B. distachyon* and (F) *G. raimondii*. ***FDR-
618 corrected p-value < 0.001, **FDR-corrected p-value < 0.01, *FDR-corrected p-value < 0.05, NS
619 – not significant .

620 **Figure 6:** The frequency of genic DNA methylation across 928 *A. thaliana* ecotypes differs by
621 the age of gene duplication. Density plots showing the Ks distribution of genes at different
622 frequencies (0%, <25%, 25%-50%, 50%-75%, >75%) in the population for (A) gbM, (B) TE-
623 like, and (C) UM genes. Boxplots of Ks distributions show that Ks increases with increasing
624 frequency of (D) gbM in the population, but decreases with the increasing frequency of (E) TE-
625 like and (F) UM genes.

SUPPLEMENTARY TABLES

Table S1: Classification of genic DNA methylation of all genes in each species. Genes were classified as gene body methylated (gbM), transposable element-like (TE-like), unmethylated (UM), or remained unclassified.'NA' represents genes with missing methylation data.

Table S2: Correlations between genic DNA methylation class (gbM, TE-like, UM) and genomic features (number of genes, TEs, and TE nucleotides) in 100kb sliding windows with a 50kb step size. Positive correlations are marked in blue, negative correlations in red. P-value significance is indicated by shades of blue, dark blue for p-value < .001 and light blue for p-value < .05.

Table S3: Pagel's lambda test for phylogenetic signal of trait correlations in Table S2. A lambda value of '0' indicates no phylogenetic signal, while '1' indicates a strong phylogenetic signal. Correlations in blue show a statistically significant phylogenetic signal.

Table S4: Number of genes derived from different modes of duplications in each species.

Table S5: Fisher's Exact test results for enrichment and depletion of genic DNA methylation classifications across different modes of duplication. Odds ratios of enriched associations are colored green, depleted associations are in orange. P-value < .001 are in dark blue, while p-value < .05 are in light blue.

Table S6: Number of duplicate pairs with different or the same genic DNA methylation status for each mode of duplication for each species.

Table S7: Proportions of each genic DNA methylation class genes for parental genes and transposed copies for each species.

Table S8: Kolmogorov-Smirnov test results for differences in the distribution of synonymous substitution rates (Ks) for gbM, TE-like and UM genes compared to a random distribution. Blue indicates distribution is significantly different at a FDR adjusted p-value < 0.05.

Table S9: Outgroup species used for each epoch as part of MCscanX-transposed.

Table S10: Fisher's Exact test results for enrichment and depletion of genic DNA methylation classifications across different epochs of transposed duplicates for all species. Odds ratios of

enriched associations are colored green, depleted associations are in orange. Blue indicates distribution is significantly different at a FDR adjusted p-value < 0.05.

Table S11: Kolmogorov-Smirnov test results for differences in the distribution of non-synonymous substitution to synonymous substitution ratios (Ka/Ks) for gbM, TE-like and UM genes compared to a random distribution. Blue indicates distribution is significantly different at a FDR adjusted p-value < 0.05.

Table S12: Fisher's Exact Test results for enrichment and depletion of gbM, TE-like, and unmethylated genes with Ka/Ks ratio > 1.1. Odds ratios of enriched associations are colored green, depleted associations are in orange. Blue indicates distribution is significantly different at a FDR adjusted p-value < 0.05.

Table S13: Fisher's Exact Test results for enrichment and depletion of known presence-absence variants for gbM, TE-like, and unmethylated genes. Odds ratios of enriched associations are colored green, depleted associations are in orange. Blue indicates distribution is significantly different at a FDR adjusted p-value < 0.05.

Table S14: Data source and mapping statistics for all methylome data used in the study.

SUPPLEMENTARY FIGURES

Figure S1: Correlations between genes, transposons (TEs), and different classes of methylated genes. Increaseing blue indicates a positive correlation, increasing red indicates a negative correlation. Boxes marked with an X are statistically non-significant.

Figure S2: Distribution of genic methylation classified genes and genomic features across the largest chromosomes in representative species.

Figure S3: Distribution of genic methylation classified genes based on synonymous substitution rates (Ks) across different modes of gene duplication. WGD = whole-genome duplication. SGD = single-gene duplication and combines data from tandem, proximal, transposed, and dispersed modes of duplication.

Figure S4: The percentage of gene copies in each genic DNA methylation class for transposed

genes that have duplicated during that 'epoch' since divergence from the species on the x-axis. For example, in *B. distachyon* transposed genes that have duplicated since *B. distachyon* diverged from *O. sativa* are shown on the x-axis under *O. sativa*. Those shown for *Z. mays* duplicated in the period since the common ancestor of *B. distachyon* and *O. sativa* diverged from their common ancestor with *Z. mays*, but before the divergence of *B. distachyon* and *O. sativa*. Horizontal dotted lines indicate the percentage of each genic DNA methylation class in the whole genome.

Figure S5: Synonymous substitution rate (Ks) plots of TE-like genes in 43 species. Species are ordered by median Ks of TE-like genes for transposed (A), tandem (B), and proximal duplicates (C). Magenta boxplots = species enriched for TE-like genes, blue = depleted, white = NS. Significance at p-value < 0.05

Figure S6: Distribution of genic methylation classified genes based on the ratio of non-synonymous to synonymous substitutions (Ka/Ks) across different modes of gene duplication. WGD = whole-genome duplication. SGD = single-gene duplication and combines data from tandem, proximal, transposed, and dispersed modes of duplication.

Figure S7: Percentage of genes for each genic DNA Methylation class with a Ka/Ks > 1.1. *FDR adjusted p-value < 0.05, **FDR adjusted p-value < 0.01, ***FDR adjusted p-value < 0.001, NS = Not Significant.

Figure S8: Percentage of Total (all genes), gbM, TE-like, and unmethylated genes with known presence absences variation. A two-sided Fisher's Exact test was used to test for depletion or enrichment of PAVs amongst each category of genic DNA methylation. *FDR corrected p-value < 0.05, **FDR corrected p-value < 0.01, ***FDR corrected p-value < 0.001, NS – Not significantly different.

Figure S9: Genic methylation variation across different modes of gene duplication and its relationship with age (Ks).

Dataset S1: Rooted species tree with branch lengths.

705

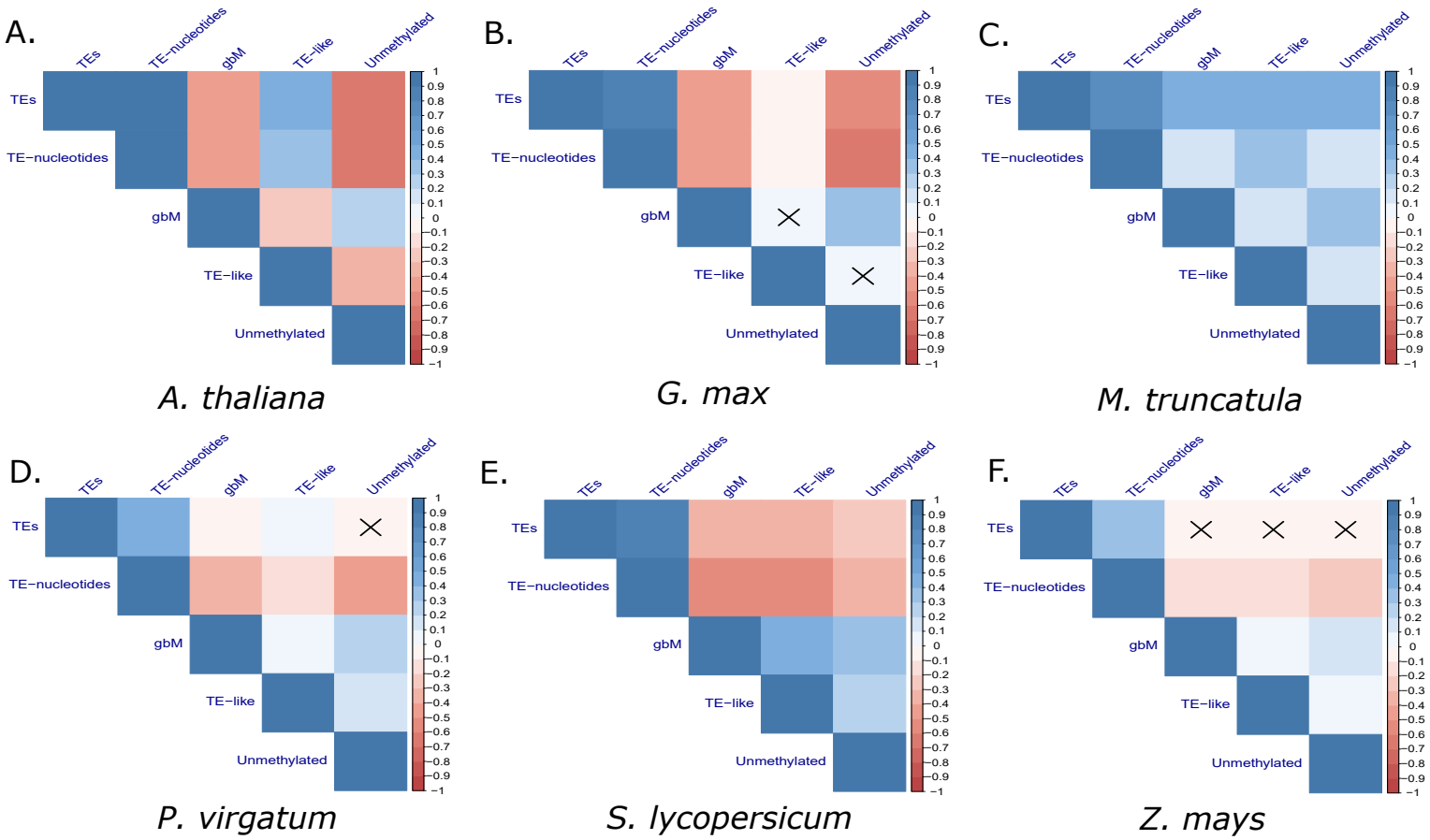

Figure 1: Correlations of genomic features with genic DNA methylation classifications. Pearson's correlation coefficient 'r' was measured to determine the distribution of gbM, TE-like, and UM genes across the genome using TE content and TE nucleotides as a proxy for heterochromatin. Shades of blue indicate enrichment, while red indicates depletion. Correlations not significant at $p < 0.001$ are marked with 'X'. (A) Arabidopsis thaliana, (B) Glycine max, (C) Medicago truncatula, (D) Panicum virgatum, (E) Solanum lycopersicum, and (F) Zea mays.

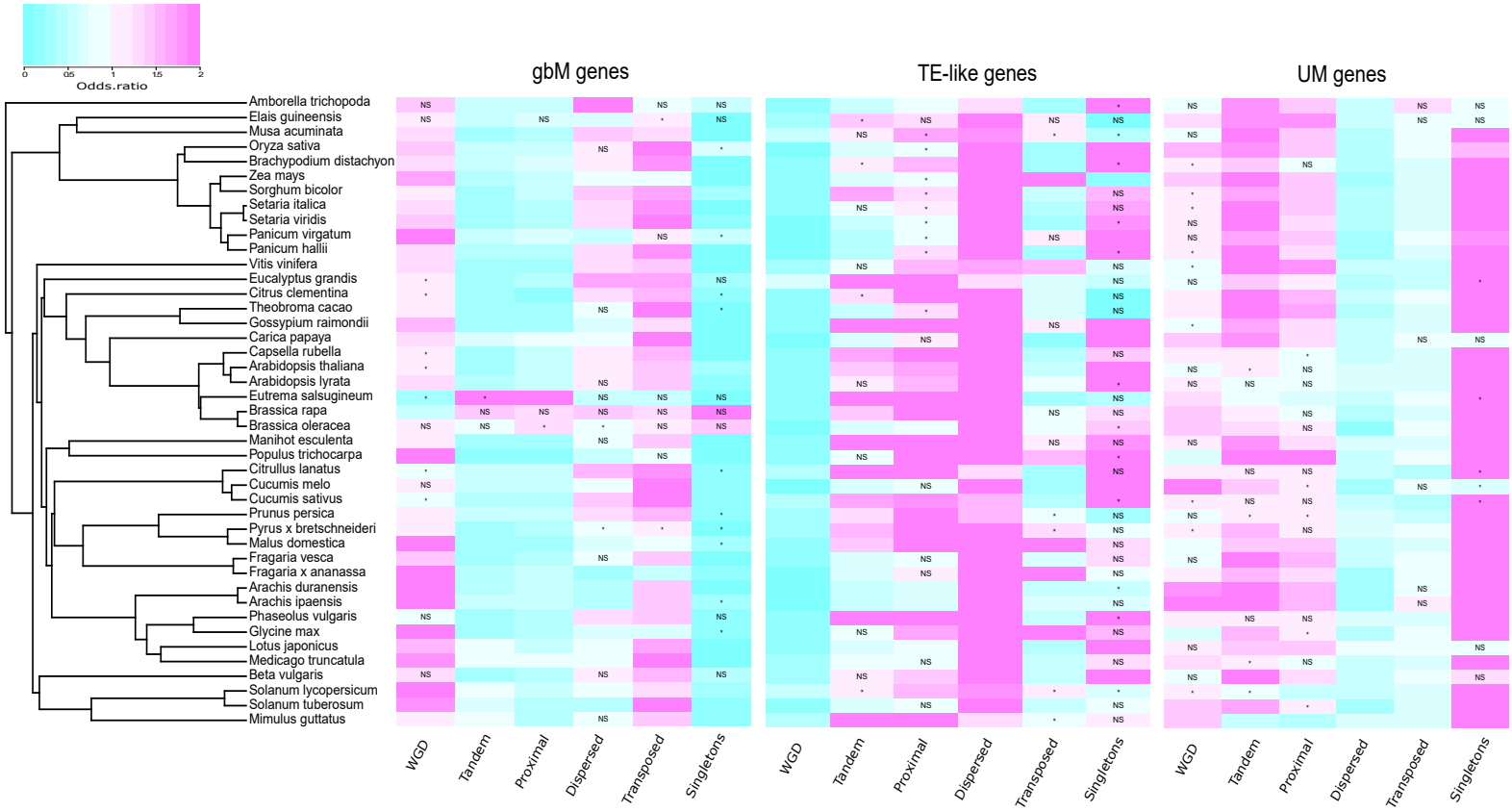

Figure 2: Patterns of genic DNA methylation across different modes of gene duplication across 43 angiosperms. Heatmap shows enrichment/depletion of each genic DNA methylation class (gbM, TE-like, and UM) for each mode of gene duplication (WGD, tandem, proximal, dispersed, transposed). Singletons are single-copy genes (non-duplicates) present in the species and the outgroup. Unless indicated, all associations are statistically significant at a p-value < 0.001. *significant association at a lower p-value cut-off (0.001 < p < 0.05). 'NS' – not significant. Cyan indicates depletion, while pink denotes enrichment.

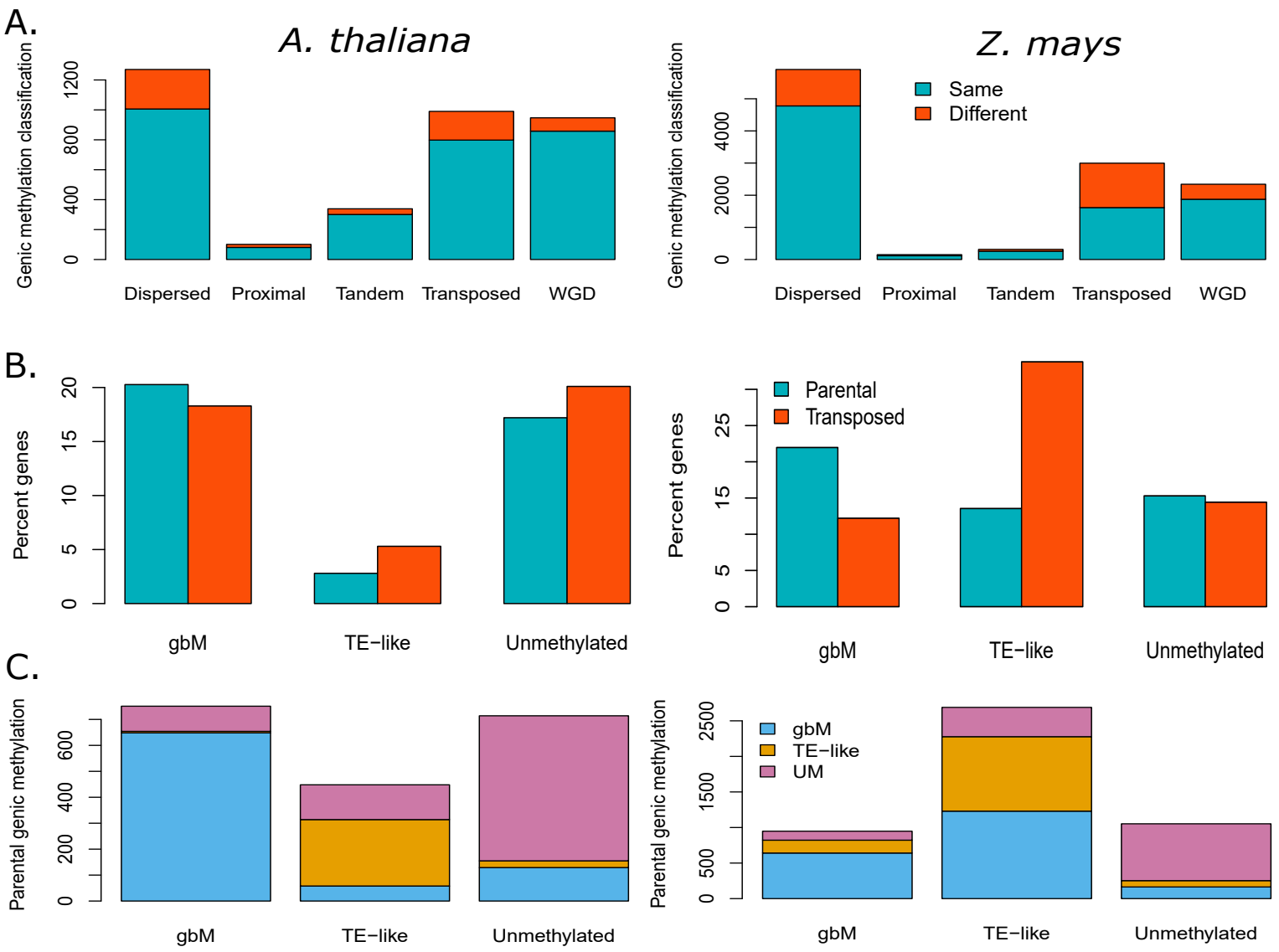

Figure 3: Stasis and switching of genic DNA methylation profiles between duplicate gene pairs. (A) The number of genes for each mode duplication with the same or different genic DNA methylation profile in A. thaliana and Z. mays. (B) Percentage of parental and transposed genes with gbM, TE-like, and UM classification for A. thaliana and Z. mays shows enrichment of TE-like methylation in the transposed copy. (C) Proportion of parental genes in each genic DNA methylation class based on the genic DNA methylation in A. thaliana and Z. mays. This indicates the direction of genic DNA methylation switching.

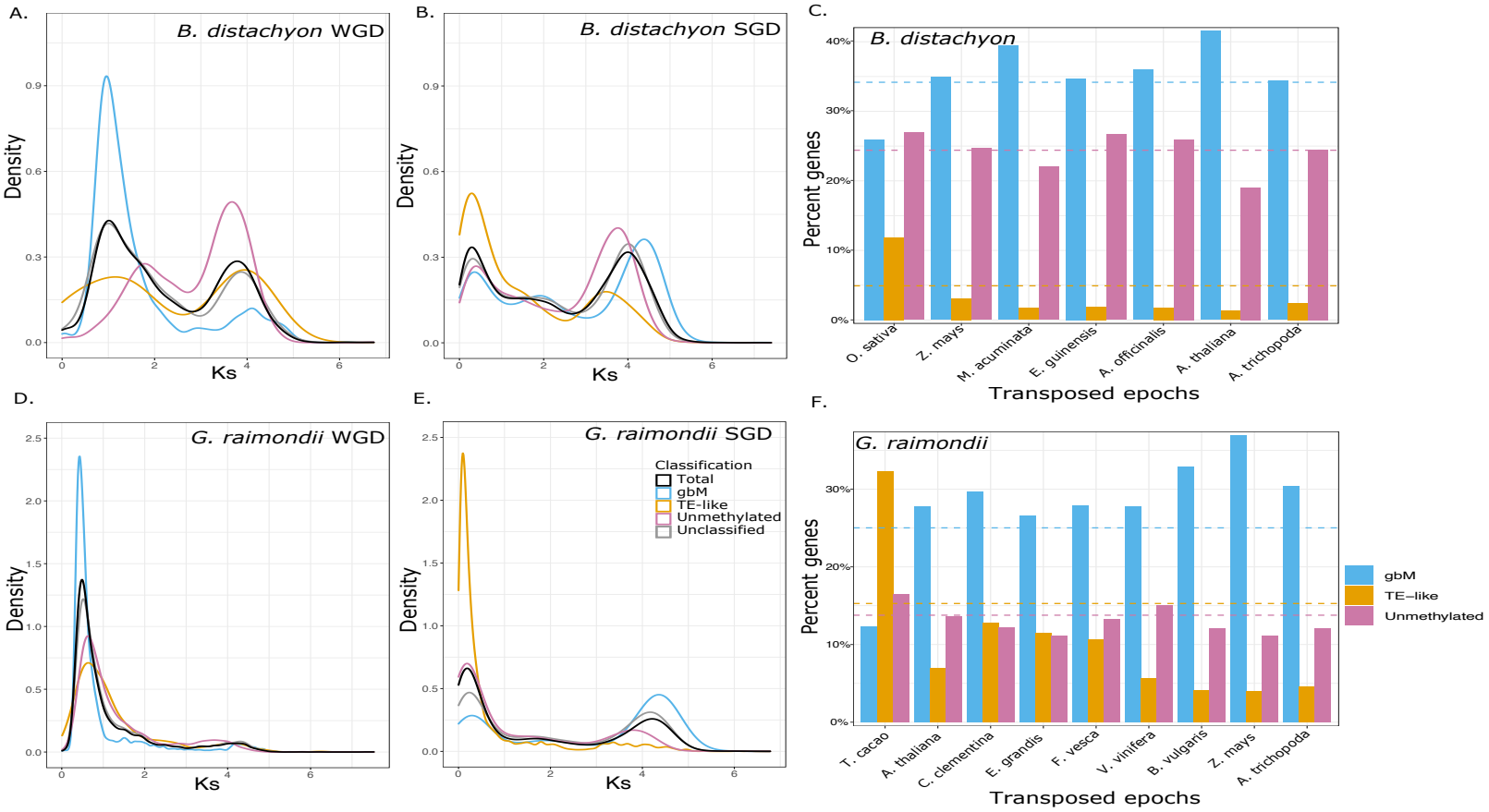

Figure 4: Relationship between genic DNA methylation class and the age of gene duplication. Density plots (A,B, D,E) show the distribution of synonymous substitutions (Ks) for each genic DNA methylation class , while bar plots (C and F) show the percentage of gene copies in each genic DNA methylation class for transposed genes that have duplicated during that 'epoch' since divergence from the species on the x-axis. For example, in B. distachyon transposed genes that have duplicated since B. distachyon diverged from O. sativa are shown on the x-axis under O. sativa. Those shown for Z. mays duplicated in the period since the common ancestor of B. distachyon and O. sativa diverged from their common ancestor with Z. mays, but before the divergence of B. distachyon and O. sativa. Horizontal dotted lines indicate the percentage of each genic DNA methylation class in the whole genome. (A) B. distachyon WGD Ks distributions (B) B. distachyon SGD Ks distributions. (C) Percentage of B. distachyon transposed duplicates in each genic DNA methylation class at each 'epoch'. (D) G. raimondii WGD Ks distributions (E) G. raimondii SGD Ks distributions (F) Percentage of G. raimondii transposed duplicates in each genic DNA methylation class at each 'epoch'.

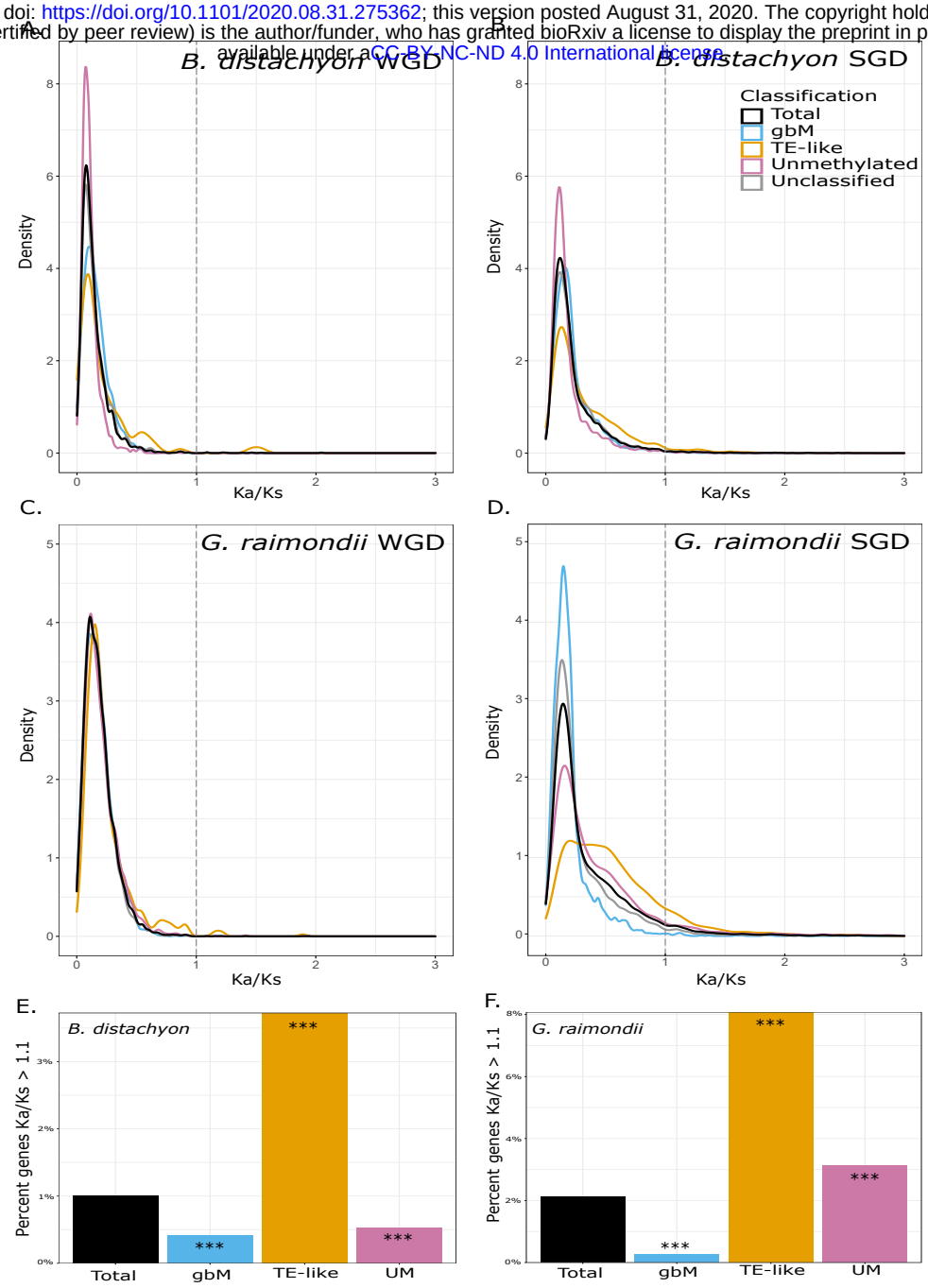

Figure 5: Differences in sequence evolution correspond to different genic DNA methylation classes. TE-like methylation genes are under relaxed selection compared to gbM and UM genes as indicated by density plots showing the ratio of non-synonymous to synonymous mutations (Ka/Ks) for (A) B. distachyon WGDs, (B) B. distachyon SGDs, (C) G. raimondii WGDs and (D) G. raimondii SGDs. However, TE-like methylated genes are also enriched in genes showing evidence of positive selection. Bar graphs indicate the proportion of genes for each genic DNA methylation class with a Ka/Ks > 1.1 for (E) B. distachyon and (F) G. raimondii. ***FDR-corrected p-value < 0.001, **FDR-corrected p-value < 0.01, *FDR-corrected p-value < 0.05, NS – not significant .

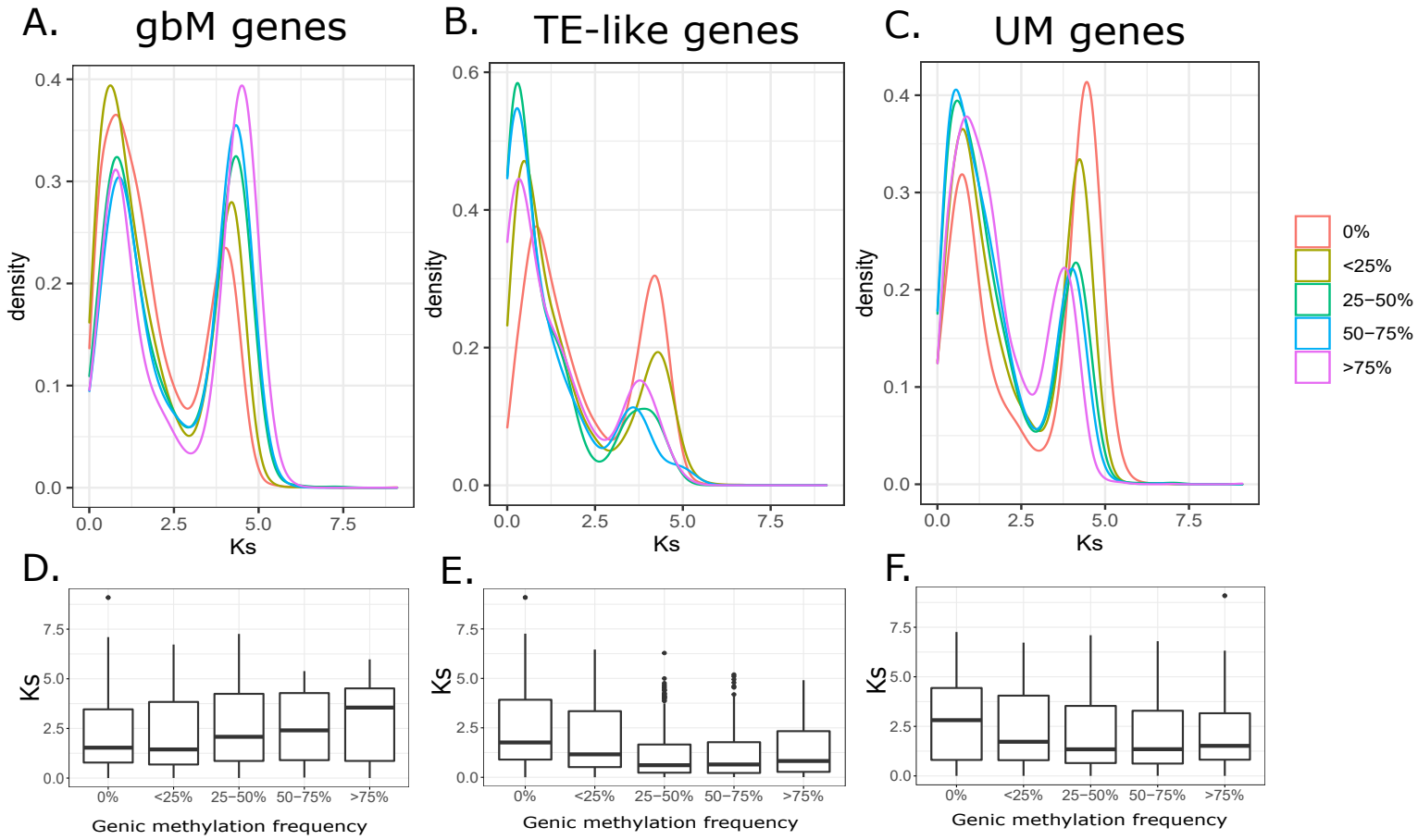

Figure 6: The frequency of genic DNA methylation across 928 A. thaliana ecotypes differs by the age of gene duplication. Density plots showing the Ks distribution of genes at different frequencies (0%, <25%, 25%-50%, 50%-75%, >75%) in the population for (A) gbM, (B) TE-like, and (C) UM genes. Boxplots of Ks distributions show that Ks increases with increasing frequency of (D) gbM in the population, but decreases with the increasing frequency of (E) TE-like and (F) UM genes.

Parsed Citations

Adams, K. L., R. Cronn, R. Percifield, and J. F. Wendel. 2003. "Genes Duplicated by Polyploidy Show Unequal Contributions to the Transcriptome and Organ-Specific Reciprocal Silencing." Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.0630618100.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ahmed, Ikhlak, Alexis Sarazin, Chris Bowler, Vincent Colot, and Hadi Quesneville. 2011. "Genome-Wide Evidence for Local DNA Methylation Spreading from Small RNA-Targeted Sequences in Arabidopsis." Nucleic Acids Research 39 (16): 6919–31.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Amborella Genome Project. 2013. "The Amborella Genome and the Evolution of Flowering Plants." Science 342 (6165): 1241089.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bailey, Jeffrey A., Zhiping Gu, Royden A. Clark, Knut Reinert, Rhea V. Samonte, Stuart Schwartz, Mark D. Adams, Eugene W. Myers, Peter W. Li, and Evan E. Eichler. 2002. "Recent Segmental Duplications in the Human Genome." Science 297 (5583): 1003–7.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bartholomé, Jérôme, Eric Mandrou, André Mabiala, Jerry Jenkins, Ibouniyamine Nabihoudine, Christophe Klopp, Jeremy Schmutz, Christophe Plomion, and Jean-Marc Gion. 2015. "High-Resolution Genetic Maps of Eucalyptus Improve Eucalyptus Grandis Genome Assembly." The New Phytologist 206 (4): 1283–96.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bennetzen, Jeffrey L., Craig Coleman, Renyi Liu, Jianxin Ma, and Wusirika Ramakrishna. 2004. "Consistent over-Estimation of Gene Number in Complex Plant Genomes." Current Opinion in Plant Biology 7 (6): 732–36.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bennetzen, Jeffrey L., Jeremy Schmutz, Hao Wang, Ryan Percifield, Jennifer Hawkins, Ana C. Pontaroli, Matt Estep, et al. 2012. "Reference Genome Sequence of the Model Plant Setaria." Nature Biotechnology 30 (6): 555–61.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bertioli, David John, Steven B. Cannon, Lutz Froenicke, Guodong Huang, Andrew D. Farmer, Ethalinda K. S. Cannon, Xin Liu, et al. 2016. "The Genome Sequences of Arachis Duranensis and Arachis Ipaensis, the Diploid Ancestors of Cultivated Peanut." Nature Genetics 48 (4): 438–46.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bewick, Adam J., Lexiang Ji, Chad E. Niederhuth, Eva-Maria Willing, Brigitte T. Hofmeister, Xiuling Shi, Li Wang, et al. 2016. "On the Origin and Evolutionary Consequences of Gene Body DNA Methylation." Proceedings of the National Academy of Sciences of the United States of America 113 (32): 9111–16.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bird, A. 2002. "DNA Methylation Patterns and Epigenetic Memory." Genes & Development. https://doi.org/10.1101/gad.947102.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bird, Kevin A., Chad Niederhuth, Shujun Ou, Malia Gehan, J. Chris Pires, Zhiyong Xiong, Robert VanBuren, and Patrick P. Edger. n.d. "Replaying the Evolutionary Tape to Investigate Subgenome Dominance in Allopolyploid Brassica Napus." https://doi.org/10.1101/814491.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bombarely, Aureliano, Michel Moser, Avichai Amrad, Manzhu Bao, Laure Bapaume, Cornelius S. Barry, Mattijs Bliek, et al. 2016. "Insight into the Evolution of the Solanaceae from the Parental Genomes of Petunia Hybrida." Nature Plants 2 (6): 16074.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bredeson, Jessen V., Jessica B. Lyons, Simon E. Prochnik, G. Albert Wu, Cindy M. Ha, Eric Edsinger-Gonzales, Jane Grimwood, et al. 2016. "Sequencing Wild and Cultivated Cassava and Related Species Reveals Extensive Interspecific Hybridization and Genetic Diversity." Nature Biotechnology 34 (5): 562–70.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Buchfink, Benjamin, Chao Xie, and Daniel H. Huson. 2015. "Fast and Sensitive Protein Alignment Using DIAMOND." Nature Methods. https://doi.org/10.1038/nmeth.3176.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chang, Andrew Ying-Fei, and Ben-Yang Liao. 2012. "DNA Methylation Rebalances Gene Dosage after Mammalian Gene Duplications." Molecular Biology and Evolution 29 (1): 133–44.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cheng, Chia-Yi, Vivek Krishnakumar, Agnes P. Chan, Françoise Thibaud-Nissen, Seth Schobel, and Christopher D. Town. 2017. "Araport11: A Complete Reannotation of the Arabidopsis Thaliana Reference Genome." The Plant Journal: For Cell and Molecular Biology 89 (4): 789–804.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cheng, Feng, Jian Wu, Xu Cai, Jianli Liang, Michael Freeling, and Xiaowu Wang. 2018. "Gene Retention, Fractionation and Subgenome Differences in Polyploid Plants." Nature Plants 4 (5): 258–68.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cheng, Jingfei, Qingfeng Niu, Bo Zhang, Kunsong Chen, Ruihua Yang, Jian-Kang Zhu, Yijing Zhang, and Zhaobo Lang. 2018. "Downregulation of RdDM during Strawberry Fruit Ripening." Genome Biology 19 (1): 212.

Cokus, Shawn J., Suhua Feng, Xiaoyu Zhang, Zugen Chen, Barry Merriman, Christian D. Haudenschild, Sriharsa Pradhan, Stanley F. Nelson, Matteo Pellegrini, and Steven E. Jacobsen. 2008. "Shotgun Bisulphite Sequencing of the Arabidopsis Genome Reveals DNA Methylation Patterning." Nature 452 (7184): 215–19.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Colle, Marivi, Courtney P. Leisner, Ching Man Wai, Shujun Ou, Kevin A. Bird, Jie Wang, Jennifer H. Wisecaver, et al. 2019. "Haplotype-Phased Genome and Evolution of Phytonutrient Pathways of Tetraploid Blueberry." GigaScience 8 (3). https://doi.org/10.1093/gigascience/giz012.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cuevas, Hugo E., Chengbo Zhou, Haibao Tang, Prashant P. Khadke, Sayan Das, Yann-Rong Lin, Zhengxiang Ge, et al. 2016. "The Evolution of Photoperiod-Insensitive Flowering in Sorghum, A Genomic Model for Panicoid Grasses." Molecular Biology and Evolution 33 (9): 2417–28.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cusack, Brian P., and Kenneth H. Wolfe. 2007. "Not Born Equal: Increased Rate Asymmetry in Relocated and Retrotransposed Rodent Gene Duplicates." Molecular Biology and Evolution 24 (3): 679–86.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Daccord, Nicolas, Jean-Marc Celton, Gareth Linsmith, Claude Becker, Nathalie Choisne, Elio Schijlen, Henri van de Geest, et al. 2017. "High-Quality de Novo Assembly of the Apple Genome and Methylome Dynamics of Early Fruit Development." Nature Genetics 49 (7): 1099–1106.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Dassanayake, Maheshi, Dong-Ha Oh, Jeffrey S. Haas, Alvaro Hernandez, Hyewon Hong, Shahjahan Ali, Dae-Jin Yun, et al. 2011. "The Genome of the Extremophile Crucifer Thellungiella Parvula." Nature Genetics 43 (9): 913–18.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

D'Hont, Angélique, France Denoeud, Jean-Marc Aury, Franc-Christophe Baurens, Françoise Carreel, Olivier Garsmeur, Benjamin Noel, et al. 2012. "The Banana (Musa Acuminata) Genome and the Evolution of Monocotyledonous Plants." Nature 488 (7410): 213–17.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Dohm, Juliane C., André E. Minoche, Daniela Holtgräwe, Salvador Capella-Gutiérrez, Falk Zakrzewski, Hakim Tafer, Oliver Rupp, et al. 2014. "The Genome of the Recently Domesticated Crop Plant Sugar Beet (Beta Vulgaris)." Nature 505 (7484): 546–49.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Dong, Pengfei, Xiaoyu Tu, Po-Yu Chu, Peitao Lü, Ning Zhu, Donald Grierson, Baijuan Du, Pinghua Li, and Silin Zhong. 2017. "3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments." Molecular Plant. https://doi.org/10.1016/j.molp.2017.11.005.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Edger, Patrick P., Thomas J. Poorten, Robert VanBuren, Michael A. Hardigan, Marivi Colle, Michael R. McKain, Ronald D. Smith, et al. 2019. "Origin and Evolution of the Octoploid Strawberry Genome." Nature Genetics 51 (3): 541–47.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Edger, Patrick P., Ronald Smith, Michael R. McKain, Arielle M. Cooley, Mario Vallejo-Marin, Yaowu Yuan, Adam J. Bewick, et al. 2017. "Subgenome Dominance in an Interspecific Hybrid, Synthetic Allopolyploid, and a 140-Year-Old Naturally Established Neo-Allopolyploid Monkeyflower." The Plant Cell. https://doi.org/10.1105/tpc.17.00010.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Edger, Patrick P., Robert VanBuren, Marivi Colle, Thomas J. Poorten, Ching Man Wai, Chad E. Niederhuth, Elizabeth I. Alger, et al. 2018. "Single-Molecule Sequencing and Optical Mapping Yields an Improved Genome of Woodland Strawberry (Fragaria Vesca) with Chromosome-Scale Contiguity." GigaScience 7 (2): 1–7.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Edwards, Carol A., and Anne C. Ferguson-Smith. 2007. "Mechanisms Regulating Imprinted Genes in Clusters." Current Opinion in Cell Biology 19 (3): 281–89.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Eichten, Steven R., Roman Briskine, Jawon Song, Qing Li, Ruth Swanson-Wagner, Peter J. Hermanson, Amanda J. Waters, et al. 2013. "Epigenetic and Genetic Influences on DNA Methylation Variation in Maize Populations." The Plant Cell 25 (8): 2783–97.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Eichten, Steven R., Tim Stuart, Akanksha Srivastava, Ryan Lister, and Justin O. Borevitz. 2016. "DNA Methylation Profiles of Diverse Brachypodium Distachyon Align with Underlying Genetic Diversity." Genome Research 26 (11): 1520–31.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

El Baidouri, Moaine, Kyung Do Kim, Brian Abernathy, Ying-Hui Li, Li-Juan Qiu, and Scott A. Jackson. 2018. "Genic C-Methylation in Soybean Is Associated with Gene Paralogs Relocated to Transposable Element-Rich Pericentromeres." Molecular Plant 11 (3): 485–95.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Emms, David M., and Steven Kelly. 2015. "OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy." Genome Biology 16 (August): 157.

Finnegan, E. J., and E. S. Dennis. 1993. "Isolation and Identification by Sequence Homology of a Putative Cytosine Methyltransferase from Arabidopsis Thaliana." Nucleic Acids Research 21 (10): 2383–88.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Finnegan, E. J., W. J. Peacock, and E. S. Dennis. 1996. "Reduced DNA Methylation in Arabidopsis Thaliana Results in Abnormal Plant Development." Proceedings of the National Academy of Sciences of the United States of America 93 (16): 8449–54.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Fisher, R. A. 1935. "The Logic of Inductive Inference." Journal of the Royal Statistical Society. https://doi.org/10.2307/2342435.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Flagel, Lex E., and Jonathan F. Wendel. 2009. "Gene Duplication and Evolutionary Novelty in Plants." New Phytologist. https://doi.org/10.1111/j.1469-8137.2009.02923.x.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Freeling, Michael. 2009. "Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition." Annual Review of Plant Biology. https://doi.org/10.1146/annurev.arplant.043008.092122.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Freeling, M., E. Lyons, B. Pedersen, M. Alam, R. Ming, and D. Lisch. 2008. "Many or Most Genes in Arabidopsis Transposed after the Origin of the Order Brassicales." Genome Research. https://doi.org/10.1101/gr.081026.108.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ganko, Eric W., Blake C. Meyers, and Todd J. Vision. 2007. "Divergence in Expression between Duplicated Genes in Arabidopsis." Molecular Biology and Evolution 24 (10): 2298–2309.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Gao, Lei, Itay Gonda, Honghe Sun, Qiyue Ma, Kan Bao, Denise M. Tieman, Elizabeth A. Burzynski-Chang, et al. 2019. "The Tomato Pan-Genome Uncovers New Genes and a Rare Allele Regulating Fruit Flavor." Nature Genetics 51 (6): 1044–51.

Pubmed: Author and Title

Google Scholar: Author Only Title Only Author and Title

Garcia-Mas, Jordi, Andrej Benjak, Walter Sanseverino, Michael Bourgeois, Gisela Mir, Víctor M. González, Elizabeth Hénaff, et al. 2012. "The Genome of Melon (Cucumis Melo L.)." Proceedings of the National Academy of Sciences of the United States of America 109 (29): 11872–77.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Garsmeur, Olivier, James C. Schnable, Ana Almeida, Cyril Jourda, Angélique D'Hont, and Michael Freeling. 2014. "Two Evolutionarily Distinct Classes of Paleopolyploidy." Molecular Biology and Evolution 31 (2): 448–54.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Gent, Jonathan I., Nathanael A. Ellis, Lin Guo, Alex E. Harkess, Yingyin Yao, Xiaoyu Zhang, and R. Kelly Dawe. 2013. "CHH Islands: De Novo DNA Methylation in near-Gene Chromatin Regulation in Maize." Genome Research 23 (4): 628–37.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Golicz, Agnieszka A., Philipp E. Bayer, Guy C. Barker, Patrick P. Edger, Hyeran Kim, Paula A. Martinez, Chon Kit Kenneth Chan, et al. 2016. "The Pangenome of an Agronomically Important Crop Plant Brassica Oleracea." Nature Communications 7 (November): 13390.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Goll, Mary Grace, and Timothy H. Bestor. 2005. "EUKARYOTIC CYTOSINE METHYLTRANSFERASES." Annual Review of Biochemistry. https://doi.org/10.1146/annurev.biochem.74.010904.153721.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Gruenbaum, Yosef, Tally Naveh-Many, Howard Cedar, and Aharon Razin. 1981. "Sequence Specificity of Methylation in Higher Plant DNA." Nature. https://doi.org/10.1038/292860a0.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Guo, Shaogui, Jianguo Zhang, Honghe Sun, Jerome Salse, William J. Lucas, Haiying Zhang, Yi Zheng, et al. 2013. "The Draft Genome of Watermelon (Citrullus Lanatus) and Resequencing of 20 Diverse Accessions." Nature Genetics 45 (1): 51–58.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hanada, Kousuke, Cheng Zou, Melissa D. Lehti-Shiu, Kazuo Shinozaki, and Shin-Han Shiu. 2008. "Importance of Lineage-Specific Expansion of Plant Tandem Duplicates in the Adaptive Response to Environmental Stimuli." Plant Physiology. https://doi.org/10.1104/pp.108.122457.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hardigan, Michael A., Emily Crisovan, John P. Hamilton, Jeongwoon Kim, Parker Laimbeer, Courtney P. Leisner, Norma C. Manrique-Carpintero, et al. 2016. "Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated Solanum Tuberosum." The Plant Cell 28 (2): 388–405.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Harkess, Alex, Jinsong Zhou, Chunyan Xu, John E. Bowers, Ron Van der Hulst, Saravanaraj Ayyampalayam, Francesco Mercati, et al. 2017. "The Asparagus Genome Sheds Light on the Origin and Evolution of a Young Y Chromosome." Nature Communications. https://doi.org/10.1038/s41467-017-01064-8.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Harris, C. Jake, Marion Scheibe, Somsakul Pop Wongpalee, Wanlu Liu, Evan M. Cornett, Robert M. Vaughan, Xueqin Li, et al. 2018. "A DNA Methylation Reader Complex That Enhances Gene Transcription." Science 362 (6419): 1182–86.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hellsten, Uffe, Kevin M. Wright, Jerry Jenkins, Shengqiang Shu, Yaowu Yuan, Susan R. Wessler, Jeremy Schmutz, John H. Willis, and Daniel S. Rokhsar. 2013. "Fine-Scale Variation in Meiotic Recombination in Mimulus Inferred from Population Shotgun Sequencing." Proceedings of the National Academy of Sciences of the United States of America 110 (48): 19478–82.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Henderson, Ian R., and Steven E. Jacobsen. 2007. "Epigenetic Inheritance in Plants." Nature 447 (7143): 418–24.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hibrand Saint-Oyant, L., T. Ruttink, L. Hamama, I. Kirov, D. Lakhwani, N. N. Zhou, P. M. Bourke, et al. 2018. "A High-Quality Genome Sequence of Rosa Chinensis to Elucidate Ornamental Traits." Nature Plants 4 (7): 473–84.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hirsch, Candice N., Jillian M. Foerster, James M. Johnson, Rajandeep S. Sekhon, German Muttoni, Brieanne Vaillancourt, Francisco Peñagaricano, et al. 2014. "Insights into the Maize Pan-Genome and Pan-Transcriptome." The Plant Cell 26 (1): 121–35.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hosmani, Prashant S., Mirella Flores-Gonzalez, Henri van de Geest, Florian Maumus, Linda V. Bakker, Elio Schijlen, Jan van Haarst, et al. n.d. "An Improved de Novo Assembly and Annotation of the Tomato Reference Genome Using Single-Molecule Sequencing, Hi-C Proximity Ligation and Optical Maps." https://doi.org/10.1101/767764.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hulse-Kemp, Amanda M., Shamoni Maheshwari, Kevin Stoffel, Theresa A. Hill, David Jaffe, Stephen R. Williams, Neil Weisenfeld, et al. 2018. "Reference Quality Assembly of the 3.5-Gb Genome of Capsicum Annuum from a Single Linked-Read Library." Horticulture Research. https://doi.org/10.1038/s41438-017-0011-0.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hu, Tina T., Pedro Pattyn, Erica G. Bakker, Jun Cao, Jan-Fang Cheng, Richard M. Clark, Noah Fahlgren, et al. 2011. "The Arabidopsis Lyrata Genome Sequence and the Basis of Rapid Genome Size Change." Nature Genetics 43 (5): 476–81.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Initiative, The International Brachypodium, and The International Brachypodium Initiative. 2010. "Genome Sequencing and Analysis of the Model Grass Brachypodium Distachyon." Nature. https://doi.org/10.1038/nature08747.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Jiang, Ning, Zhirong Bao, Xiaoyu Zhang, Sean R. Eddy, and Susan R. Wessler. 2004. "Pack-MULE Transposable Elements Mediate Gene Evolution in Plants." Nature 431 (7008): 569–73.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Jiang, Ning, Cédric Feschotte, Xiaoyu Zhang, and Susan R. Wessler. 2004. "Using Rice to Understand the Origin and Amplification of Miniature Inverted Repeat Transposable Elements (MITEs)." Current Opinion in Plant Biology 7 (2): 115–19.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Jiao, Yinping, Paul Peluso, Jinghua Shi, Tiffany Liang, Michelle C. Stitzer, Bo Wang, Michael S. Campbell, et al. 2017. "Improved Maize Reference Genome with Single-Molecule Technologies." Nature 546 (7659): 524–27.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Katoh, Kazutaka, and Daron M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." Molecular Biology and Evolution 30 (4): 772–80.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kawahara, Yoshihiro, Melissa de la Bastide, John P. Hamilton, Hiroyuki Kanamori, W. Richard McCombie, Shu Ouyang, David C. Schwartz, et al. 2013. "Improvement of the Oryza Sativa Nipponbare Reference Genome Using next Generation Sequence and Optical Map Data." Rice 6 (1): 4.

Kawakatsu, Taiji, Shao-Shan Carol Huang, Florian Jupe, Eriko Sasaki, Robert J. Schmitz, Mark A. Urich, Rosa Castanon, et al. 2016. "Epigenomic Diversity in a Global Collection of Arabidopsis Thaliana Accessions." Cell 166 (2): 492–505.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Keller, T. E., and S. V. Yi. 2014. "DNA Methylation and Evolution of Duplicate Genes." Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.1321420111.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kim, Kyung Do, Moaine El Baidouri, Brian Abernathy, Aiko Iwata-Otsubo, Carolina Chavarro, Michael Gonzales, Marc Libault, Jane Grimwood, and Scott A. Jackson. 2015. "A Comparative Epigenomic Analysis of Polyploidy-Derived Genes in Soybean and Common Bean." Plant Physiology 168 (4): 1433–47.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kreitman, Martin. 2000. "METHODS TODETECTSELECTION INPOPULATIONS WITHAPPLICATIONS TO THEHUMAN." Annual Review of Genomics and Human Genetics. https://doi.org/10.1146/annurev.genom.1.1.539.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lamesch, Philippe, Tanya Z. Berardini, Donghui Li, David Swarbreck, Christopher Wilks, Rajkumar Sasidharan, Robert Muller, et al.

2012. "The Arabidopsis Information Resource (TAIR): Improved Gene Annotation and New Tools." Nucleic Acids Research 40 (Database issue): D1202–10.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Law, Julie A., and Steven E. Jacobsen. 2010. "Establishing, Maintaining and Modifying DNA Methylation Patterns in Plants and Animals." Nature Reviews. Genetics 11 (3): 204–20.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Li, Qing, Jonathan I. Gent, Greg Zynda, Jawon Song, Irina Makarevitch, Cory D. Hirsch, Candice N. Hirsch, et al. 2015. "RNA-Directed DNA Methylation Enforces Boundaries between Heterochromatin and Euchromatin in the Maize Genome." Proceedings of the National Academy of Sciences of the United States of America 112 (47): 14728–33.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Li, Qing, Hongbo Li, Wu Huang, Yuanchao Xu, Qian Zhou, Shenhao Wang, Jue Ruan, Sanwen Huang, and Zhonghua Zhang. 2019. "A Chromosome-Scale Genome Assembly of Cucumber (Cucumis Sativus L.)." GigaScience 8 (6). https://doi.org/10.1093/gigascience/giz072.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lister, Ryan, Ronan C. O'Malley, Julian Tonti-Filippini, Brian D. Gregory, Charles C. Berry, A. Harvey Millar, and Joseph R. Ecker. 2008. "Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis." Cell 133 (3): 523–36.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Liu, Meng-Jun, Jin Zhao, Qing-Le Cai, Guo-Cheng Liu, Jiu-Rui Wang, Zhi-Hui Zhao, Ping Liu, et al. 2014. "The Complex Jujube Genome Provides Insights into Fruit Tree Biology." Nature Communications 5 (October): 5315.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lovell, John T., Jerry Jenkins, David B. Lowry, Sujan Mamidi, Avinash Sreedasyam, Xiaoyu Weng, Kerrie Barry, et al. 2018. "The Genomic Landscape of Molecular Responses to Natural Drought Stress in Panicum Hallii." Nature Communications 9 (1): 5213.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Luff, B., L. Pawlowski, and J. Bender. 1999. "An Inverted Repeat Triggers Cytosine Methylation of Identical Sequences in Arabidopsis." Molecular Cell 3 (4): 505–11.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lü, Peitao, Sheng Yu, Ning Zhu, Yun-Ru Chen, Biyan Zhou, Yu Pan, David Tzeng, et al. 2018. "Genome Encode Analyses Reveal the Basis of Convergent Evolution of Fleshy Fruit Ripening." Nature Plants 4 (10): 784–91.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lynch, M., and J. S. Conery. 2000. "The Evolutionary Fate and Consequences of Duplicate Genes." Science.

Mamidi, Sujan, Adam Healey, Pu Huang, Jane Grimwood, Jerry Jenkins, Kerrie Barry, Avinash Sreedasyam, et al. n.d. "The Setaria Viridis Genome and Diversity Panel Enables Discovery of a Novel Domestication Gene." https://doi.org/10.1101/744557.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Massey, Frank J. 1951. "The Kolmogorov-Smirnov Test for Goodness of Fit." Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1951.10500769.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

McCormick, Ryan F., Sandra K. Truong, Avinash Sreedasyam, Jerry Jenkins, Shengqiang Shu, David Sims, Megan Kennedy, et al. 2018. "The Sorghum Bicolor Reference Genome: Improved Assembly, Gene Annotations, a Transcriptome Atlas, and Signatures of Genome Organization." The Plant Journal: For Cell and Molecular Biology 93 (2): 338–54.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Medvedeva, Yulia A., Abdullah M. Khamis, Ivan V. Kulakovskiy, Wail Ba-Alawi, Md Shariful I. Bhuyan, Hideya Kawaji, Timo Lassmann, et al. 2014. "Effects of Cytosine Methylation on Transcription Factor Binding Sites." BMC Genomics. https://doi.org/10.1186/1471-2164-15-119.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Melquist, S., B. Luff, and J. Bender. 1999. "Arabidopsis PAI Gene Arrangements, Cytosine Methylation and Expression." Genetics 153 (1): 401–13.
Pubmed: Author and Title

Google Scholar: Author Only Title Only Author and Title

Meyer, P., I. Niedenhof, and M. ten Lohuis. 1994. "Evidence for Cytosine Methylation of Non-Symmetrical Sequences in Transgenic Petunia Hybrida." The EMBO Journal 13 (9): 2084–88.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ming, Ray, Shaobin Hou, Yun Feng, Qingyi Yu, Alexandre Dionne-Laporte, Jimmy H. Saw, Pavel Senin, et al. 2008. "The Draft Genome of the Transgenic Tropical Fruit Tree Papaya (Carica Papaya Linnaeus)." Nature 452 (7190): 991–96.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Motamayor, Juan C., Keithanne Mockaitis, Jeremy Schmutz, Niina Haiminen, Donald Livingstone 3rd, Omar Cornejo, Seth D. Findley, et al. 2013. "The Genome Sequence of the Most Widely Cultivated Cacao Type and Its Use to Identify Candidate Genes Regulating Pod Color." Genome Biology 14 (6): r53.

Münkemüller, Tamara, Sébastien Lavergne, Bruno Bzeznik, Stéphane Dray, Thibaut Jombart, Katja Schiffers, and Wilfried Thuiller. 2012. "How to Measure and Test Phylogenetic Signal." Methods in Ecology and Evolution. https://doi.org/10.1111/j.2041-210x.2012.00196.x.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Niederhuth, Chad E., Adam J. Bewick, Lexiang Ji, Magdy S. Alabady, Kyung Do Kim, Qing Li, Nicholas A. Rohr, et al. 2016. "Widespread Natural Variation of DNA Methylation within Angiosperms." Genome Biology 17 (1): 194.

Noshay, Jaclyn M., Sarah N. Anderson, Peng Zhou, Lexiang Ji, William Ricci, Zefu Lu, Michelle C. Stitzer, et al. 2019. "Monitoring the Interplay between Transposable Element Families and DNA Methylation in Maize." PLoS Genetics 15 (9): e1008291.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ohno, Susumu. 1970. "Evolution by Gene Duplication." https://doi.org/10.1007/978-3-642-86659-3.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

O'Malley, Ronan C., Shao-Shan Carol Huang, Liang Song, Mathew G. Lewsey, Anna Bartlett, Joseph R. Nery, Mary Galli, Andrea Gallavotti, and Joseph R. Ecker. 2016. "Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape." Cell 166 (6): 1598.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ong-Abdullah, Meilina, Jared M. Ordway, Nan Jiang, Siew-Eng Ooi, Sau-Yee Kok, Norashikin Sarpan, Nuraziyan Azimi, et al. 2015. "Loss of Karma Transposon Methylation Underlies the Mantled Somaclonal Variant of Oil Palm." Nature 525 (7570): 533–37.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Otto, Sarah P., and Jeannette Whitton. 2000. "Polyploid Incidence and Evolution." Annual Review of Genetics. https://doi.org/10.1146/annurev.genet.34.1.401.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ou, Shujun, Weija Su, Yi Liao, Kapeel Chougule, Jireh R. A. Agda, Adam J. Hellinga, Carlos Santiago Blanco Lugo, et al. 2019. "Benchmarking Transposable Element Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline." Genome Biology 20 (1): 275.

Pagel, Mark. 1999. "Inferring the Historical Patterns of Biological Evolution." Nature. https://doi.org/10.1038/44766.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Panchy, Nicholas, Melissa D. Lehti-Shiu, and Shin-Han Shiu. 2016. "Evolution of Gene Duplication in Plants." Plant Physiology. https://doi.org/10.1104/pp.16.00523.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Parkin, Isobel A. P., Chushin Koh, Haibao Tang, Stephen J. Robinson, Sateesh Kagale, Wayne E. Clarke, Chris D. Town, et al. 2014. "Transcriptome and Methylome Profiling Reveals Relics of Genome Dominance in the Mesopolyploid Brassica Oleracea." Genome Biology 15 (6): R77.

Paterson, Andrew H., Jonathan F. Wendel, Heidrun Gundlach, Hui Guo, Jerry Jenkins, Dianchuan Jin, Danny Llewellyn, et al. 2012. "Repeated Polyploidization of Gossypium Genomes and the Evolution of Spinnable Cotton Fibres." Nature 492 (7429): 423–27.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pellicer, Jaume, and Ilia J. Leitch. 2020. "The Plant DNA C-values Database (release 7.1): An Updated Online Repository of Plant Genome Size Data for Comparative Studies." New Phytologist. https://doi.org/10.1111/nph.16261.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Picard, Colette L., and Mary Gehring. 2017. "Proximal Methylation Features Associated with Nonrandom Changes in Gene Body Methylation." Genome Biology 18 (1): 73.

Qiao, Xin, Qionghou Li, Hao Yin, Kaijie Qi, Leiting Li, Runze Wang, Shaoling Zhang, and Andrew H. Paterson. 2019. "Gene Duplication and Evolution in Recurring Polyploidization-Diploidization Cycles in Plants." Genome Biology 20 (1): 38.

Raju, Sunil K. Kenchanmane, Sunil K. Kenchanmane Raju, Eleanore Jeanne Ritter, and Chad E. Niederhuth. 2019. "Establishment, Maintenance, and Biological Roles of Non-CG Methylation in Plants." Essays in Biochemistry. https://doi.org/10.1042/ebc20190032.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Revell, Liam J. 2012. "Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things)." Methods in Ecology and Evolution. https://doi.org/10.1111/j.2041-210x.2011.00169.x.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Rodin, Sergei N., and Arthur D. Riggs. 2003. "Epigenetic Silencing May Aid Evolution by Gene Duplication." Journal of Molecular Evolution 56 (6): 718–29.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Roth, Christian, and David A. Liberles. 2006. "A Systematic Search for Positive Selection in Higher Plants (Embryophytes)." BMC Plant Biology 6 (June): 12.

Roudier, François, Felipe Karam Teixeira, and Vincent Colot. 2009. "Chromatin Indexing in Arabidopsis: An Epigenomic Tale of Tails and More." Trends in Genetics: TIG 25 (11): 511–17.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Sato, Shusei, Yasukazu Nakamura, Takakazu Kaneko, Erika Asamizu, Tomohiko Kato, Mitsuteru Nakao, Shigemi Sasamoto, et al. 2008. "Genome Structure of the Legume, Lotus Japonicus." DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes 15 (4): 227–39.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schmitz, Robert J., Yupeng He, Oswaldo Valdés-López, Saad M. Khan, Trupti Joshi, Mark A. Urich, Joseph R. Nery, et al. 2013. "Epigenome-Wide Inheritance of Cytosine Methylation Variants in a Recombinant Inbred Population." Genome Research 23 (10): 1663–74.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schmitz, Robert J., Matthew D. Schultz, Mark A. Urich, Joseph R. Nery, Mattia Pelizzola, Ondrej Libiger, Andrew Alix, et al. 2013. "Patterns of Population Epigenomic Diversity." Nature 495 (7440): 193–98.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schmutz, Jeremy, Steven B. Cannon, Jessica Schlueter, Jianxin Ma, Therese Mitros, William Nelson, David L. Hyten, et al. 2010. "Genome Sequence of the Palaeopolyploid Soybean." Nature 463 (7278): 178–83.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schmutz, Jeremy, Phillip E. McClean, Sujan Mamidi, G. Albert Wu, Steven B. Cannon, Jane Grimwood, Jerry Jenkins, et al. 2014. "A Reference Genome for Common Bean and Genome-Wide Analysis of Dual Domestications." Nature Genetics 46 (7): 707–13.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schnable, James C. 2019. "Genes and Gene Models, an Important Distinction." The New Phytologist, June. https://doi.org/10.1111/nph.16011.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schultz, Matthew D., Yupeng He, John W. Whitaker, Manoj Hariharan, Eran A. Mukamel, Danny Leung, Nisha Rajagopal, et al. 2015. "Human Body Epigenome Maps Reveal Noncanonical DNA Methylation Variation." Nature 523 (7559): 212–16.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Secco, David, Chuang Wang, Huixia Shou, Matthew D. Schultz, Serge Chiarenza, Laurent Nussaume, Joseph R. Ecker, James Whelan, and Ryan Lister. 2015. "Stress Induced Gene Expression Drives Transient DNA Methylation Changes at Adjacent Repetitive Elements." eLife 4 (July). https://doi.org/10.7554/eLife.09343.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Seymour, Benjamin, Athena Andreosso, and Jamie Seymour. 2015. "Cardiovascular Toxicity from Marine Envenomation." Heart and Toxins. https://doi.org/10.1016/b978-0-12-416595-3.00007-4.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Seymour, Danelle K., Daniel Koenig, Jörg Hagmann, Claude Becker, and Detlef Weigel. 2014. "Evolution of DNA Methylation Patterns in the Brassicaceae Is Driven by Differences in Genome Organization." PLoS Genetics. https://doi.org/10.1371/journal.pgen.1004785.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Sharma, Sanjeev Kumar, Daniel Bolser, Jan de Boer, Mads Sønderkær, Walter Amoros, Martin Federico Carboni, Juan Martín D'Ambrosio, et al. 2013. "Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps." G3 3 (11): 2031–47.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Sigman, Meredith J., and R. Keith Slotkin. 2016. "The First Rule of Plant Transposable Element Silencing: Location, Location, Location." The Plant Cell 28 (2): 304–13.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Silveira, Amanda Bortolini, Charlotte Trontin, Sandra Cortijo, Joan Barau, Luiz Eduardo Vieira Del Bem, Olivier Loudet, Vincent Colot, and Michel Vincentz. 2013. "Extensive Natural Epigenetic Variation at a De Novo Originated Gene." PLoS Genetics. https://doi.org/10.1371/journal.pgen.1003437.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Singh, Rajinder, Meilina Ong-Abdullah, Eng-Ti Leslie Low, Mohamad Arif Abdul Manaf, Rozana Rosli, Rajanaidu Nookiah, Leslie Cheng-Li Ooi, et al. 2013. "Oil Palm Genome Sequence Reveals Divergence of Interfertile Species in Old and New Worlds." Nature 500 (7462): 335–39.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Slotte, Tanja, Khaled M. Hazzouri, J. Arvid Ågren, Daniel Koenig, Florian Maumus, Ya-Long Guo, Kim Steige, et al. 2013. "The Capsella Rubella Genome and the Genomic Consequences of Rapid Mating System Evolution." Nature Genetics 45 (7): 831–35.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Song, Qingxin, Tianzhen Zhang, David M. Stelly, and Z. Jeffrey Chen. 2017. "Epigenomic and Functional Analyses Reveal Roles of Epialleles in the Loss of Photoperiod Sensitivity during Domestication of Allotetraploid Cottons." Genome Biology 18 (1): 99.

Stroud, Hume, Truman Do, Jiamu Du, Xuehua Zhong, Suhua Feng, Lianna Johnson, Dinshaw J. Patel, and Steven E. Jacobsen. 2014. "Non-CG Methylation Patterns Shape the Epigenetic Landscape in Arabidopsis." Nature Structural & Molecular Biology 21 (1): 64–72.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Suyama, Mikita, David Torrents, and Peer Bork. 2006. "PAL2NAL: Robust Conversion of Protein Sequence Alignments into the Corresponding Codon Alignments." Nucleic Acids Research 34 (Web Server issue): W609–12.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Suzuki, Miho M., and Adrian Bird. 2008. "DNA Methylation Landscapes: Provocative Insights from Epigenomics." Nature Reviews Genetics. https://doi.org/10.1038/nrg2341.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Takuno, Shohei, and Brandon S. Gaut. 2012. "Body-Methylated Genes in Arabidopsis Thaliana Are Functionally Important and Evolve Slowly." Molecular Biology and Evolution 29 (1): 219–27.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

---. 2013. "Gene Body Methylation Is Conserved between Plant Orthologs and Is of Evolutionary Consequence." Proceedings of the National Academy of Sciences of the United States of America 110 (5): 1797–1802.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Takuno, Shohei, Jin-Hua Ran, and Brandon S. Gaut. 2016. "Evolutionary Patterns of Genic DNA Methylation Vary across Land Plants." Nature Plants 2 (January): 15222.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Tang, Haibao, Vivek Krishnakumar, Shelby Bidwell, Benjamin Rosen, Agnes Chan, Shiguo Zhou, Laurent Gentzbittel, et al. 2014. "An Improved Genome Release (version Mt4.0) for the Model Legume Medicago Truncatula." BMC Genomics 15 (April): 312.

Turco, Gina M., Kaisa Kajala, Govindarajan Kunde-Ramamoorthy, Chew-Yee Ngan, Andrew Olson, Shweta Deshphande, Denis Tolkunov, et al. 2017. "DNA Methylation and Gene Expression Regulation Associated with Vascularization in Sorghum Bicolor." The

New Phytologist 214 (3): 1213–29.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Tuskan, G. A., S. Difazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, et al. 2006. "The Genome of Black Cottonwood, Populus Trichocarpa (Torr. & Gray)." Science 313 (5793): 1596–1604.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

VanBuren, Robert, Doug Bryant, Patrick P. Edger, Haibao Tang, Diane Burgess, Dinakar Challabathula, Kristi Spittle, et al. 2015. "Single-Molecule Sequencing of the Desiccation-Tolerant Grass Oropetium Thomaeum." Nature 527 (7579): 508–11.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

VanBuren, Robert, Ching Man Wai, Marivi Colle, Jie Wang, Shawn Sullivan, Jill M. Bushakra, Ivan Liachko, et al. 2018. "A near Complete, Chromosome-Scale Assembly of the Black Raspberry (Rubus Occidentalis) Genome." GigaScience 7 (8). https://doi.org/10.1093/gigascience/giy094.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Van de Peer, Yves, Eshchar Mizrachi, and Kathleen Marchal. 2017. "The Evolutionary Significance of Polyploidy." Nature Reviews. Genetics 18 (7): 411–24.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Vaughn, Matthew W., Miloš Tanurdžić, Zachary Lippman, Hongmei Jiang, Robert Carrasquillo, Pablo D. Rabinowicz, Neilay Dedhia, et al. 2007. "Epigenetic Natural Variation in Arabidopsis Thaliana." PLoS Biology. https://doi.org/10.1371/journal.pbio.0050174.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Veitia, Reiner A., Samuel Bottani, and James A. Birchler. 2008. "Cellular Reactions to Gene Dosage Imbalance: Genomic, Transcriptomic and Proteomic Effects." Trends in Genetics: TIG 24 (8): 390–97.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Verde, Ignazio, Jerry Jenkins, Luca Dondini, Sabrina Micali, Giulia Pagliarani, Elisa Vendramin, Roberta Paris, et al. 2017. "The Peach v2.0 Release: High-Resolution Linkage Mapping and Deep Resequencing Improve Chromosome-Scale Assembly and Contiguity." BMC Genomics 18 (1): 225.

Wang, Dapeng, Yubin Zhang, Zhang Zhang, Jiang Zhu, and Jun Yu. 2010. "KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies." Genomics, Proteomics & Bioinformatics 8 (1): 77–80.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wang, Haifeng, Getu Beyene, Jixian Zhai, Suhua Feng, Noah Fahlgren, Nigel J. Taylor, Rebecca Bart, James C. Carrington, Steven E. Jacobsen, and Israel Ausin. 2015. "CG Gene Body DNA Methylation Changes and Evolution of Duplicated Genes in Cassava." Proceedings of the National Academy of Sciences of the United States of America 112 (44): 13729–34.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wang, Jun, Nicholas C. Marowsky, and Chuanzhu Fan. 2014. "Divergence of Gene Body DNA Methylation and Evolution of Plant Duplicate Genes." PloS One 9 (10): e110357.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wang, Lin, Jiahui Xie, Jiantuan Hu, Binyuan Lan, Chenjiang You, Fenglan Li, Zhengjia Wang, and Haifeng Wang. 2018. "Comparative Epigenomics Reveals Evolution of Duplicated Genes in Potato and Tomato." The Plant Journal: For Cell and Molecular Biology 93 (3): 460–71.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wang, Wen, Hongkun Zheng, Chuanzhu Fan, Jun Li, Junjie Shi, Zhengqiu Cai, Guojie Zhang, et al. 2006. "High Rate of Chimeric Gene Origination by Retroposition in Plant Genomes." The Plant Cell 18 (8): 1791–1802.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wang, Xiyin, Udo Gowik, Haibao Tang, John E. Bowers, Peter Westhoff, and Andrew H. Paterson. 2009. "Comparative Genomic Analysis of C4 Photosynthetic Pathway Evolution in Grasses." Genome Biology 10 (6): R68.

Wang, Xutong, Zhibin Zhang, Tiansi Fu, Lanjuan Hu, Chunming Xu, Lei Gong, Jonathan F. Wendel, and Bao Liu. 2017. "Gene-Body CG Methylation and Divergent Expression of Duplicate Genes in Rice." Scientific Reports. https://doi.org/10.1038/s41598-017-02860-4.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wang, Yupeng, Jingping Li, and Andrew H. Paterson. 2013. "MCScanX-Transposed: Detecting Transposed Gene Duplications Based on Multiple Colinearity Scans." Bioinformatics 29 (11): 1458–60.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wang, Yupeng, Xiyin Wang, Tae-Ho Lee, Shahid Mansoor, and Andrew H. Paterson. 2013. "Gene Body Methylation Shows Distinct Patterns Associated with Different Gene Origins and Duplication Modes and Has a Heterogeneous Relationship with Gene Expression inOryza Sativa(rice)." New Phytologist. https://doi.org/10.1111/nph.12137.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wei, Taiyun and Viliam Simko (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wendel, Jonathan F. 2015. "The Wondrous Cycles of Polyploidy in Plants." American Journal of Botany. https://doi.org/10.3732/ajb.1500320.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wu, Guohong Albert, Javier Terol, Victoria Ibanez, Antonio López-García, Estela Pérez-Román, Carles Borredá, Concha Domingo, et al. 2018. "Genomics of the Origin and Evolution of Citrus." Nature 554 (7692): 311–16.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Xu, Chunming, Brian D. Nadon, Kyung Do Kim, and Scott A. Jackson. 2018. "Genetic and Epigenetic Divergence of Duplicate Genes in Two Legume Species." Plant, Cell & Environment 41 (9): 2033–44.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Xue, Huabai, Suke Wang, Jia-Long Yao, Cecilia H. Deng, Long Wang, Yanli Su, Huirong Zhang, et al. 2018. "Chromosome Level High-Density Integrated Genetic Maps Improve the Pyrus Bretschneideri 'DangshanSuli' v1.0 Genome." BMC Genomics. https://doi.org/10.1186/s12864-018-5224-6.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Xu, Shuqing, Thomas Brockmöller, Aura Navarro-Quezada, Heiner Kuhl, Klaus Gase, Zhihao Ling, Wenwu Zhou, et al. 2017. "Wild Tobacco Genomes Reveal the Evolution of Nicotine Biosynthesis." Proceedings of the National Academy of Sciences of the United States of America 114 (23): 6133–38.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Yang, Ruolin, David E. Jarvis, Hao Chen, Mark A. Beilstein, Jane Grimwood, Jerry Jenkins, Shengqiang Shu, et al. 2013. "The Reference Genome of the Halophytic Plant Eutrema Salsugineum." Frontiers in Plant Science 4 (March): 46.

Yang, Yu, Kai Tang, Tatsiana U. Datsenka, Wenshan Liu, Suhui Lv, Zhaobo Lang, Xingang Wang, et al. 2019. "Critical Function of DNA Methyltransferase 1 in Tomato Development and Regulation of the DNA Methylome and Transcriptome." Journal of Integrative Plant Biology 61 (12): 1224–42.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zemach, Assaf, Ivy E. McDaniel, Pedro Silva, and Daniel Zilberman. 2010. "Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation." Science 328 (5980): 916–19.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zhang, Jianzhi. 2003. "Evolution by Gene Duplication: An Update." Trends in Ecology & Evolution. https://doi.org/10.1016/s0169-5347(03)00033-8.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zhao, X. P., Y. Si, R. E. Hanson, C. F. Crane, H. J. Price, D. M. Stelly, J. F. Wendel, and A. H. Paterson. 1998. "Dispersed Repetitive DNA Has Spread to New Genomes since Polyploid Formation in Cotton." Genome Research 8 (5): 479–92.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zhong, Zaixuan, Kang Du, Qian Yu, Yong E. Zhang, and Shunping He. 2016. "Divergent DNA Methylation Provides Insights into the Evolution of Duplicate Genes in Zebrafish." G3 6 (11): 3581–91.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title