The Plant Cell, Vol. 21: 1053–1069, April 2009, www.plantcell.org © 2009 American Society of Plant Biologists

RESEARCH ARTICLES

Genome-Wide and Organ-Specific Landscapes of Epigenetic Modifications and Their Relationships to mRNA and Small RNA Transcriptomes in Maize[W]

Xiangfeng Wang,[a,b,c,d,1] Axel A. Elling,[c,1] Xueyong Li,[b,c,1] Ning Li,[e,1] Zhiyu Peng,[a,e] Guangming He,[b] Hui Sun,[c] Yijun Qi,[b] X. Shirley Liu,[d] and Xing Wang Deng[a,b,c,2]

[a] Peking-Yale Joint Center of Plant Molecular Genetics and Agrobiotechnology, College of Life Sciences, Peking University, Beijing 100871, China

[b] National Institute of Biological Sciences, Zhongguancun Life Science Park, Beijing 102206, China

[c] Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520

[d] Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts 02115

[e] Beijing Genomics Institute at Shenzhen, Shenzhen 518083, China

Maize (*Zea mays*) has an exceptionally complex genome with a rich history in both epigenetics and evolution. We report genomic landscapes of representative epigenetic modifications and their relationships to mRNA and small RNA (smRNA) transcriptomes in maize shoots and roots. The epigenetic patterns differed dramatically between genes and transposable elements, and two repressive marks (H3K27me3 and DNA methylation) were usually mutually exclusive. We found an organ-specific distribution of canonical microRNAs (miRNAs) and endogenous small interfering RNAs (siRNAs), indicative of their tissue-specific biogenesis. Furthermore, we observed that a decreasing level of *mop1* led to a concomitant decrease of 24-nucleotide siRNAs relative to 21-nucleotide miRNAs in a tissue-specific manner. A group of 22-nucleotide siRNAs may originate from long-hairpin double-stranded RNAs and preferentially target gene-coding regions. Additionally, a class of miRNA-like smRNAs, whose putative precursors can form short hairpins, potentially targets genes in trans. In summary, our data provide a critical analysis of the maize epigenome and its relationships to mRNA and smRNA transcriptomes.

INTRODUCTION

Histones are decorated by numerous epigenetic modifications, particularly at their N-terminal ends (Fuchs et al., 2006; Kouzarides, 2007). It has been proposed that combinations of different histone modifications form a histone code (Jenuwein and Allis, 2001), which extends the genetic code embedded in the DNA nucleotide sequence. Numerous studies have demonstrated that histone modifications influence gene expression genome-wide. Whereas histone acetylation generally is associated with gene activation (e.g., Wang et al., 2008), histone methylation can lead to either gene repression or activation depending on the modification site (Shi and Dawe, 2006; Barski et al., 2007; Mikkelsen et al., 2007; Zhang et al., 2007).

DNA methylation adds another layer of heritable epigenetic changes. In higher plants, methylation of cytosines is present in CG, CHG (where H is A, C, or T), and asymmetric CHH sequence contexts (Henderson and Jacobsen, 2007). Recent studies have shown that cytosines are methylated not only in plant repetitive sequences and transposable elements (TEs) but also in promoters and gene bodies and that DNA methylation is highly correlated with transcription (Rabinowicz et al., 2005; Zhang et al., 2006; Vaughn et al., 2007; Zilberman et al., 2007; Cokus et al., 2008; Li et al., 2008c; Lister et al., 2008). Epigenetic changes, such as DNA methylation and histone modifications, do not act in isolation but rather in concert with each other, allowing for complex interdependencies. For example, in *Arabidopsis thaliana*, CHG DNA methylation is associated with dimethylation of histone H3K9 (Bernatavichute et al., 2008), and CG DNA methylation is necessary for transgenerational epigenetic stability, including H3K9 methylation (Mathieu et al., 2007). Moreover, histone deacetylase HDA6 and histone methyltransferase KRYPTONITE are known to control DNA methylation (Aufsatz et al., 2002; Jackson et al., 2002). Other histone methylations and acetylations have been shown to be excluded by chromatin structure remodeling induced by DNA methylation (Lorincz et al., 2004; Okitsu and Hsieh, 2007). A complex interplay between DNA methylation, histone modifications, and gene expression has been reported in rice (*Oryza sativa*; Li et al., 2008c).

[1] These authors contributed equally to this work.

[2] Address correspondence to xingwang.deng@yale.edu.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Xing Wang Deng (xingwang.deng@yale.edu).

[W]Online version contains Web-only data.

www.plantcell.org/cgi/doi/10.1105/tpc.109.065714

In addition, recent studies have shown that small RNAs (smRNAs) are associated with DNA methylation (Lister et al., 2008) and that small interfering RNAs (siRNAs) target epigenetic changes to specific regions of the genome (Martienssen et al., 2005). In *Arabidopsis*, siRNAs are highly correlated with repetitive regions (Kasschau et al., 2007). Epigenetic modifications achieve an additional layer of complexity through the involvement of TEs, whose DNA is generally highly methylated and can attract the RNA silencing machinery and interact with histone modifications (Lippman et al., 2003, 2004). Epigenetic changes of TEs are not restricted to the TEs themselves, but in turn also regulate neighboring genes, which gives TEs a key role in the genome-wide distribution of epigenetic marks and smRNAs (Slotkin and Martienssen, 2007; Weil and Martienssen, 2008). This aspect is of particular importance in maize (*Zea mays*), since >60% of its genome consists of TEs (Meyers et al., 2001; Haberer et al., 2005; Messing and Dooner, 2006). Moreover, although genes are estimated to make up 8 to 20% of the maize genome, we now know that they are organized in islands surrounded by TEs (Chandler and Brendel, 2002; Messing et al., 2004; Rabinowicz and Bennetzen, 2006). In early 2008, a first draft of the sequence of the maize inbred line B73 genome was released, the largest and most complex plant genome ever sequenced. Sequencing projects for Mo17, another well-studied inbred line, and a popcorn strain are also scheduled to be completed shortly (Pennisi, 2008). However, presently, the maize genome is only sparsely annotated and assembled, which hampers its full exploitation.

Here, we describe an integrated genome-wide analysis of DNA methylation, histone modifications, smRNAs, and mRNA transcriptional activity, using maize as a model. We surveyed the epigenomes of the maize inbred line B73 in shoot and root tissue by Illumina/Solexa 1G parallel sequencing after digesting genomic DNA with a methylation-sensitive restriction enzyme and after conducting chromatin immunoprecipitation (ChIP) using antibodies that target specific histone modifications (H3K4me3, H3K9ac, H3K27me3, and H3K36me3). Additionally, we profiled RNA pools (microRNA [miRNA], siRNA, and mRNA) using the same sequencing strategy. This study provides a comprehensive and integrated organ-specific analysis of diverse epigenetic marks, smRNAs, and transcriptional activity and also gives new insight into the organization of the maize genome, which will aid in its continued assembly and annotation.

RESULTS

Direct Sequence Profiling of Maize Transcripts, Epigenetically Modified Genomic Regions, and smRNAs

To survey the mRNA transcriptome, epigenetic landscapes, and smRNAs in a maize inbred line, we isolated total RNA and genomic DNA from shoots and roots of 14-d-old B73 seedlings. We extracted mRNA from total RNA using Dynabeads and enriched for smRNA by running total RNA on a PAGE gel for gel purification of RNAs in the 19- to 24-nucleotide size range, respectively. Methylated regions of the genome were enriched by digesting genomic DNA with the methylation-sensitive restriction enzyme McrBC. Genomic regions populated by epigenetically modified histone H3 proteins were enriched by a ChIP approach using antibodies targeting H3K4me3, H3K9ac, H3K27me3, or H3K36me3, respectively (see Methods). We used the resulting fractions to build libraries for Illumina/Solexa 1G high-throughput parallel sequencing, which generated 8.4 to 35.9 million reads for the individual libraries (Figure 1A; see Supplemental Figure 1A and Supplemental Table 1 online).

Previous studies estimated that repetitive elements make up 80% or more of the maize genome (Chandler and Brendel, 2002; Messing et al., 2004; Rabinowicz and Bennetzen, 2006). This poses a major challenge to map Illumina/Solexa 1G sequencing reads to the maize genome accurately, since each read is usually 36 nucleotides or less in length. We used MAQ software (Li et al., 2008b; see also Supplemental Methods online) to map our 196 million sequencing reads to the currently available 2.4 Gb of B73 genome sequence represented by 16,205 BACs at http://www.maizesequence.org (dated June 4, 2008). The MAQ algorithm uses quality (MQ) scores to evaluate the reliability of a read based on both the uniqueness of the mapping position and the probability of sequencing errors. This allowed us to exploit sequencing data even for repetitive regions. A statistical model for calculating MQ scores and a detailed mapping procedure are described in Supplemental Methods online. Using MAQ, we mapped the proportion of reads corresponding to unique positions in the B73 genome as follows for shoot (root) libraries: H3K4me3, 31% (25%); H3K27me3, 14% (12%); H3K9ac, 30% (19%); H3K36me3, 34% (25%); DNA methylation, 8% (8%); smRNAs, 21% (23%); and mRNA, 44% (42%) (Figure 1B; see Supplemental Figure 1B and Supplemental Table 1 online). Using our criteria, we could map ~85% of all mRNA reads to unique or non-unique positions. This indicates that even though the sequencing project is still ongoing, the currently available B73 genomic sequence is nearly complete. It also indicates that Illumina/Solexa 1G sequencing is a feasible alternative to previous large-scale transcriptome studies in maize (Ma et al., 2006; Fernandes et al., 2008).

Most reads that could not be correctly mapped to unique locations matched repetitive sequences, which are widespread in the maize genome. To classify recognizable repeat types, we used RepeatMasker software (http://www.repeatmasker.org) and found that 504 Mb of the B73 genome sequence were made up of long terminal repeat (LTR) retrotransposons of the *Copia* class, while 818 Mb were made up of LTR retrotransposons of the *Gypsy* class. Similarly, we found that 14 Mb of the genome sequence were occupied by DNA transposons and 22 Mb by other repeats (Figure 1C). For example, BAC AC199189.3 shows that maize genes are surrounded by a vast number of TEs, which is a key characteristic of the maize genome. As indicated for this representative BAC, we found that, in general, TE-rich regions were less commonly modified by H3K4me3, H3K9ac, H3K27me3, and H3K36me3 relative to non-TE regions and TE-free intergenic regions between non-TE genes (Figure 1D).

To visualize the epigenetic profiles of TEs and non-TE genes in more detail, we developed a pipeline to display a continuous 20-Mb stretch of the B73 genome (see Supplemental Figure 2 online). As illustrated for a representative section of this 20-Mb region, mRNA signals showed a strong correlation with predicted

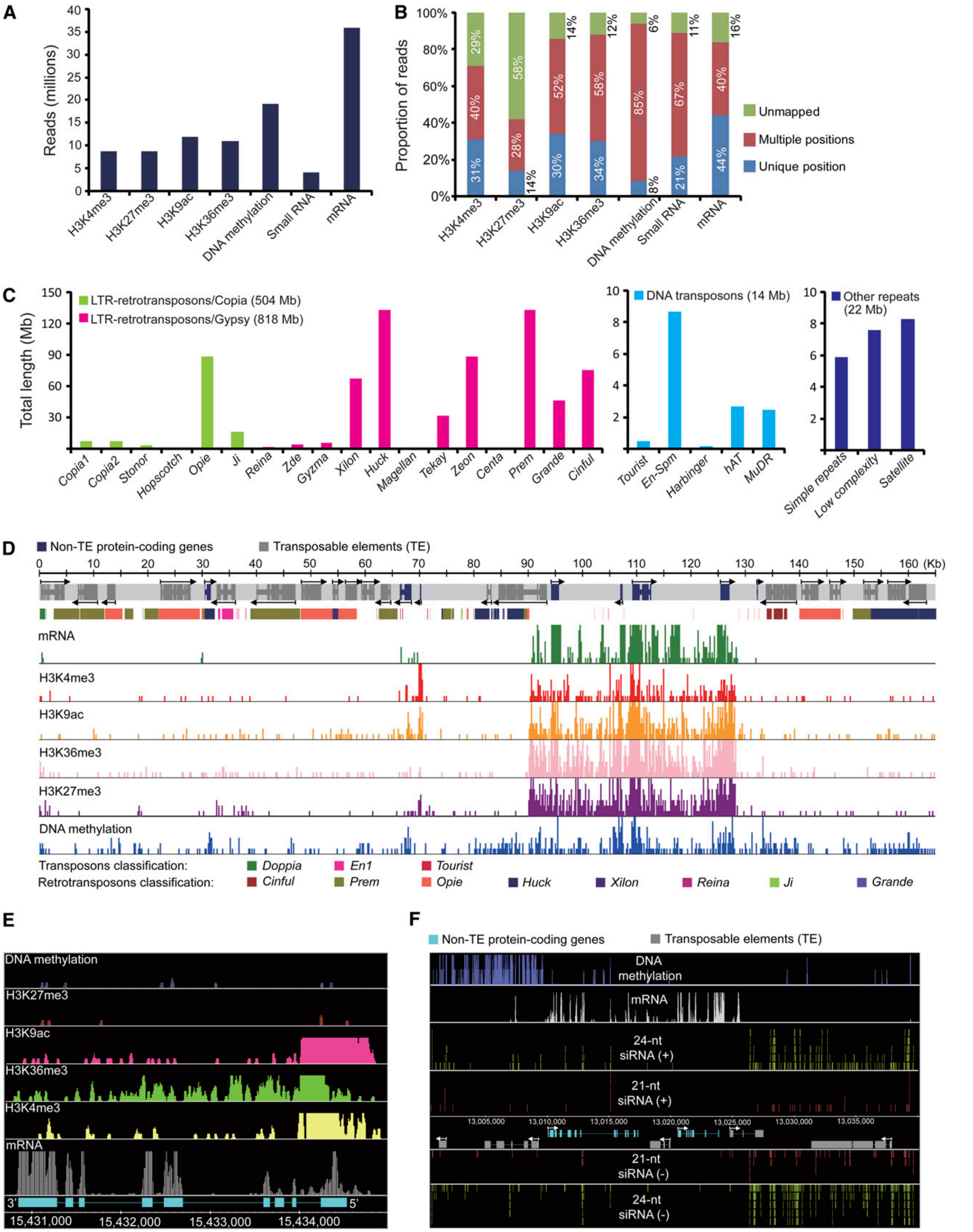


Figure 1. Sequencing, Mapping, and Visualization of the Maize Transcriptome, Epigenome, and smRNAome.

gene structures. Sequencing reads for the studied activating epigenetic marks (H3K4me3, H3K9ac, and H3K36me3) were generally present at high levels at transcribed genes in this region (Figure 1E). As shown for a larger region of these 20 Mb, TEs were generally heavily DNA methylated and lacked transcriptional activity, while non-TE genes were transcriptionally active and lacked significant DNA methylation (Figure 1F). Interestingly, we detected many smRNA reads at TEs whose DNA was not methylated. Conversely, we found that TEs whose DNA was highly methylated were relatively devoid of smRNAs (Figure 1F).

An Initial Estimate of Transcriptional Activity in the Maize Genome Using mRNA-seq

We used two gene sets for analyzing the transcriptional activity of the maize genome: a set of 11,742 full-length cDNAs (flcDNAs) obtained from http://maizecdna.org and prediction results from the FgeneSH gene finding software for 16,205 BACs obtained from http://maizesequence.org. Compiling these flcDNAs resulted in 9451 nonredundant sequences mapped to the maize genome, including 7141 flcDNAs with only one best location and 2310 with multiple best locations (see Supplemental Figure 3A online).

To estimate the transcriptional activity of the maize genome using mRNA-seq data, we developed a pipeline for de novo scanning of transcribed exons by merging overlapping Illumina/Solexa reads into contiguous regions (see Supplemental Figure 2 online). For this part of our analysis, we combined 16 lanes of mRNA reads (71 million) from both shoot and root libraries to achieve a maximum coverage. We then scanned for putative exons using MQ scores larger or equal to 0, 13, 20, and 30 (Figures 2B and 2C). We identified up to 1,122,064 putative exons representing 87,606,799 transcribed bases using our de novo scanning approach. To evaluate the coverage of mRNA-seq, we matched the de novo detected exons with flcDNAs representing bona fide genes. At MQ 0, the detected exons covered 99% at gene level, 95% at exon level, and 87% at base level, while at MQ 13 only 79, 65, and 56% were covered, respectively (Figures 2D to 2G).

We next matched the de novo detected exons as derived from our mRNA-seq data of shoot and root libraries with FgeneSH predicted genes. This resulted in the identification of nearly 45,000 validated protein-coding genes (Figure 2H; see Supplemental Table 2 online). Because the maize genome is not completely sequenced and because the available sequence data is marginally annotated, we were unable to estimate all transcribed regions. However, our pilot survey of transcriptional activity in maize suggests that even though the maize genome is about six times larger than the rice genome (Goff et al., 2002; Yu et al., 2002), the number of genes is likely to be similar. To complement these data, a series of protein-level comparative analyses, including functional comparisons based on pathway enrichment and Gene Ontology (The Gene Ontology Consortium, 2000) for maize, rice, and *Arabidopsis*, were performed (see Supplemental Figures 4 to 7 and Supplemental Data Sets 1 and 2 online). This analysis assigned the products of ~20,000 genes to known Gene Ontology pathways.

Epigenetic Marks Differ in Their Absolute and Relative Distributions on a Whole-Genome and Gene Level

To analyze the extent of epigenetic modifications on a whole-genome level, we determined how many regions were covered by DNA methylation, H3K4me3, H3K9ac, H3K27me3, or H3K36me3 (Figure 3A) using MACS software (Zhang et al., 2008; see Supplemental Methods online). We found that DNA methylation was the most prevalent modification in both shoots and roots, covering ~60,000 regions in shoots and 40,000 regions in roots, respectively. Two of the studied activating histone modifications, H3K4me3 and H3K9ac, were also found at high frequencies. Interestingly, the number of regions modified by H3K9ac or H3K27me3 was almost twice as high in shoots compared with roots, which might indicate genome-wide tissue-specific epigenetic alterations. The length and frequency of modified regions varied dramatically. DNA methylation was found at more regions than any other modification, but the average length of the affected genomic regions was only ~200 bp, by far the shortest of all modifications studied. Conversely, H3K36me3 was present at relatively few regions, but their average length was almost 1600 bp; significantly longer than any other modification (Figure 3B). Similar conclusions can be drawn when the total lengths of modified regions are considered rather than the average lengths or number of regions (Figure 3C).

To study the level of epigenetic modifications in different regions of genes and TEs, we aligned all flcDNAs at their transcript start site (TSS) and all predicted non-TE genes and TEs at their start codon (ATG). We defined the region of a gene or TE as its body (annotated transcribed region) plus 2 kb upstream. We observed no significant differences in the distributions of the epigenetic marks on aligned genes when we compared flcDNAs and predicted non-TE genes (Figures 3D and 3E; see Supplemental Figures 8A and 8B online). H3K4me3 and H3K9ac formed a strong peak at or near the TSS or ATG, respectively, and were present at relatively low levels in the gene body. By contrast, H3K36me3 was found throughout the gene body in shoots, but formed a more distinct peak at the TSS or ATG in roots (Figures 3D and 3E; see Supplemental Figures 8A and 8B online). As expected, DNA methylation was present at very low levels in

Figure 1. (continued).

(A) Counts of quality reads from Illumina/Solexa 1G sequencing.

(B) Proportions of unmapped and mapped reads with unique and multiple locations.

(C) Distribution of classified repetitive sequences in maize 2.4-Gb BAC sequences.

(D) A representative BAC (AC199189.3) showing predicted gene models with mRNA and epigenetic landscapes in shoots.

(E) Distribution of epigenetic patterns on an actively transcribed gene in shoots.

(F) The 21- and 24-nucleotide siRNAs are enriched in methylation-depleted regions in shoots.

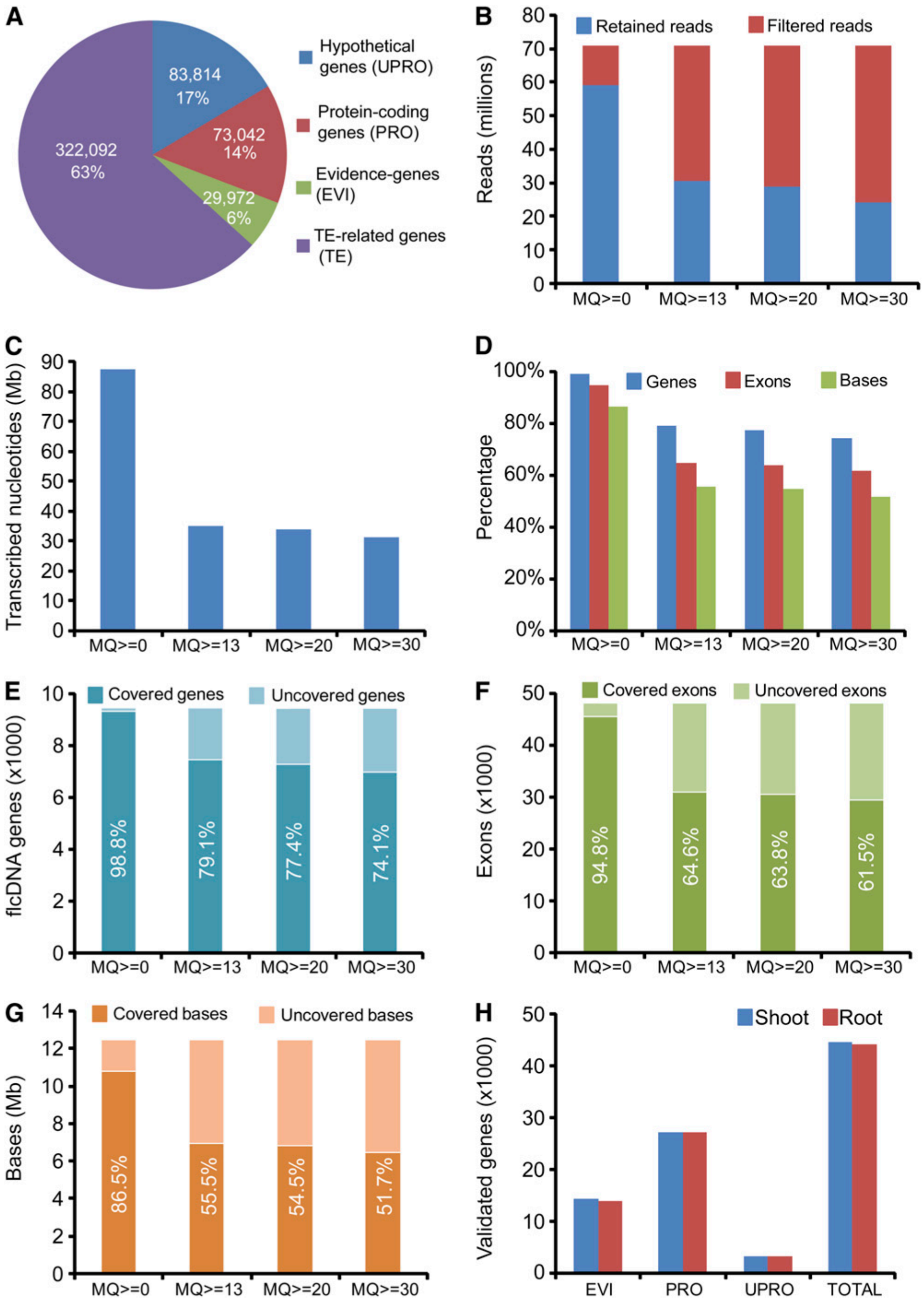


Figure 2. Validation of flcDNAs and FgeneSH-Predicted Genes.

(A) FgeneSH-predicted maize genes in different groups.

(B) Numbers of retained and filtered reads in 16 merged lanes of mRNA-seq reads using different mapping quality (MQ) scores.

(C) Total lengths of transcribed nucleotides by adding up de novo exons using different MQ scores.

(D) to **(G)** Percentages and numbers of validated flcDNAs at gene, exon, and base level.

(H) Numbers of validated non-TE genes in different groups.

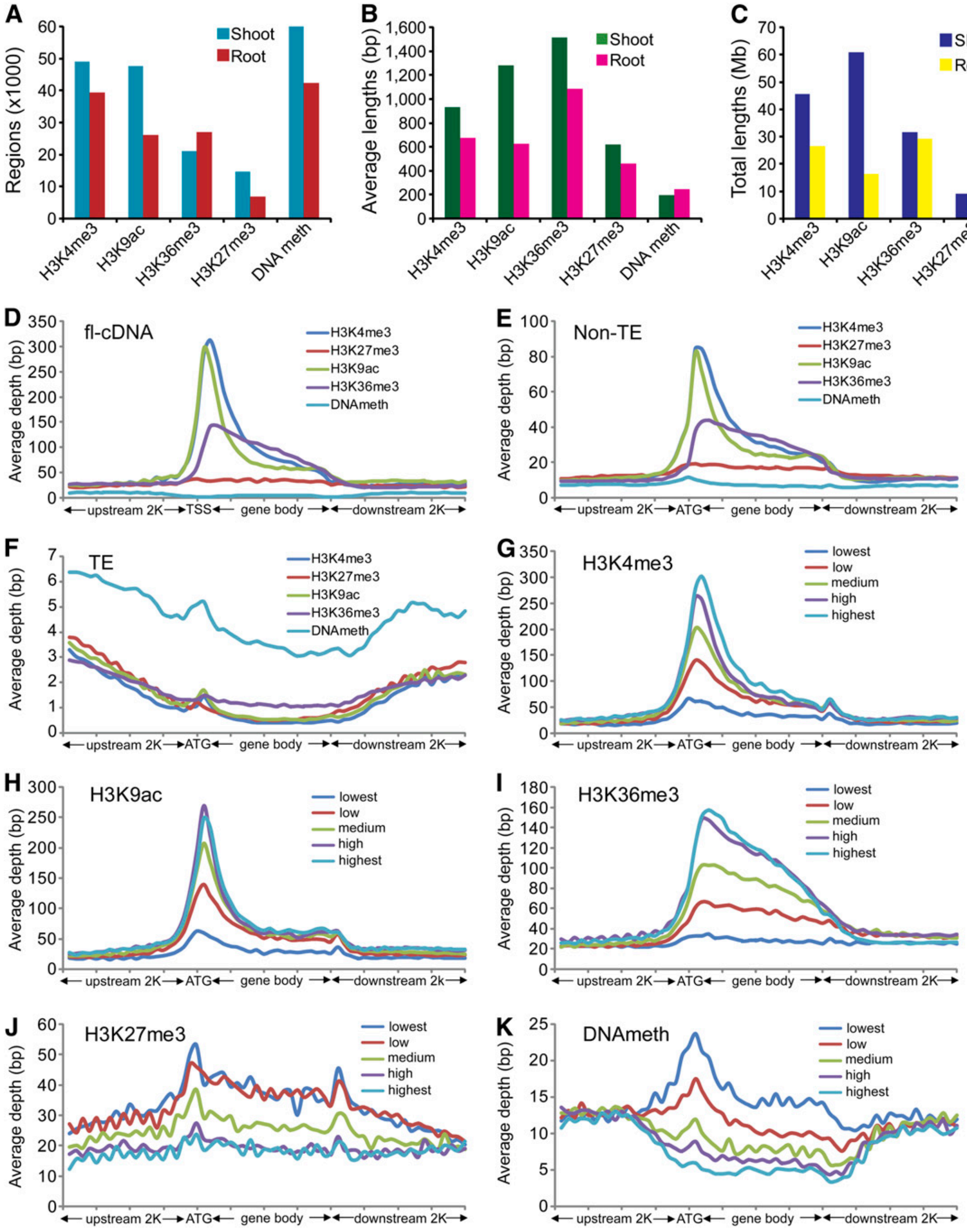


Figure 3. Genome-Wide and Genic Distribution Patterns of Epigenetic Modifications.

(A) to **(C)** Numbers, average lengths, and total lengths of epigenetically modified regions detected by MACS software.

(D) to **(F)** Distribution of H3K4me3, H3K27me3, H3K9ac, H3K36me3, and DNA methylation levels within flcDNAs, predicted TE-related, and non-TE genes aligned from TSSs and ATG, respectively. The *y* axis shows the average depth, which is the frequency of piled-up reads at each base divided by the bin size. The *x* axis represents the aligned genes that were equally binned into 40 portions, including 2K up- and downstream regions.

(G) to **(K)** Distribution of H3K4me3, H3K27me3, H3K9ac, H3K36me3, and DNA methylation within five groups of genes with different expression levels summarized from validated non-TE genes.

genes but was the most prevalent modification in TEs (Figure 3F; see Supplemental Figure 8C online).

To determine the effect of individual epigenetic modifications on transcriptional activity, we sorted all protein-coding genes (~45,000) as identified above based on their expression levels derived from mRNA-seq reads using percentile grouping. The top ~9000 most highly expressed genes were labeled "highest," the next ~9000 genes "high," the next ~9000 genes "medium," etc., such that five groups of equal size were obtained, for each of which we analyzed the distribution of each epigenetic modification of interest. We found that in both shoots and roots, the genes with the highest expression levels showed the highest amounts of H3K4me3, H3K9ac, or H3K36me3 (Figures 3G to 3I; see Supplemental Figures 8D to 8F online). By contrast, genes with the lowest expression levels had the highest amounts of H3K27me3 or DNA methylation in both tissue types (Figures 3J and 3K; see Supplemental Figures 8G and 8H online). Whereas H3K27me3 was present throughout the gene body, DNA methylation peaked at the ATG for genes in the lowest expression group. In addition, we determined the average levels of all four histone modifications of interest relative to the expression levels of genes (see Supplemental Figure 9 online). We found that in shoots and roots, genes with the highest expression levels tended to have the most H3K4me3, H3K9ac, or H3K36me3. By contrast, genes with the lowest expression levels tended to have the most H3K27me3, albeit at markedly lower levels relative to activating histone marks in highly expressed genes.

Epigenetic Modifications Show Differential Targeting of Genes and TEs and Display Combinatorial Effects in Maize Shoots and Roots

To analyze whether epigenetic modifications target genes and TEs differentially, we determined how many flcDNAs, predicted non-TE genes, and TEs show specific epigenetic modifications. We found that for both shoots and roots, genes (represented by either a flcDNA or as predicted non-TE gene) were less commonly affected by H3K27me3 or DNA methylation than by H3K4me3, H3K9ac, or H3K36me3 (Figures 4A and 4B). By contrast, TEs were epigenetically modified by DNA methylation up to 8 times more often than by modification of histone H3 (Figure 4C).

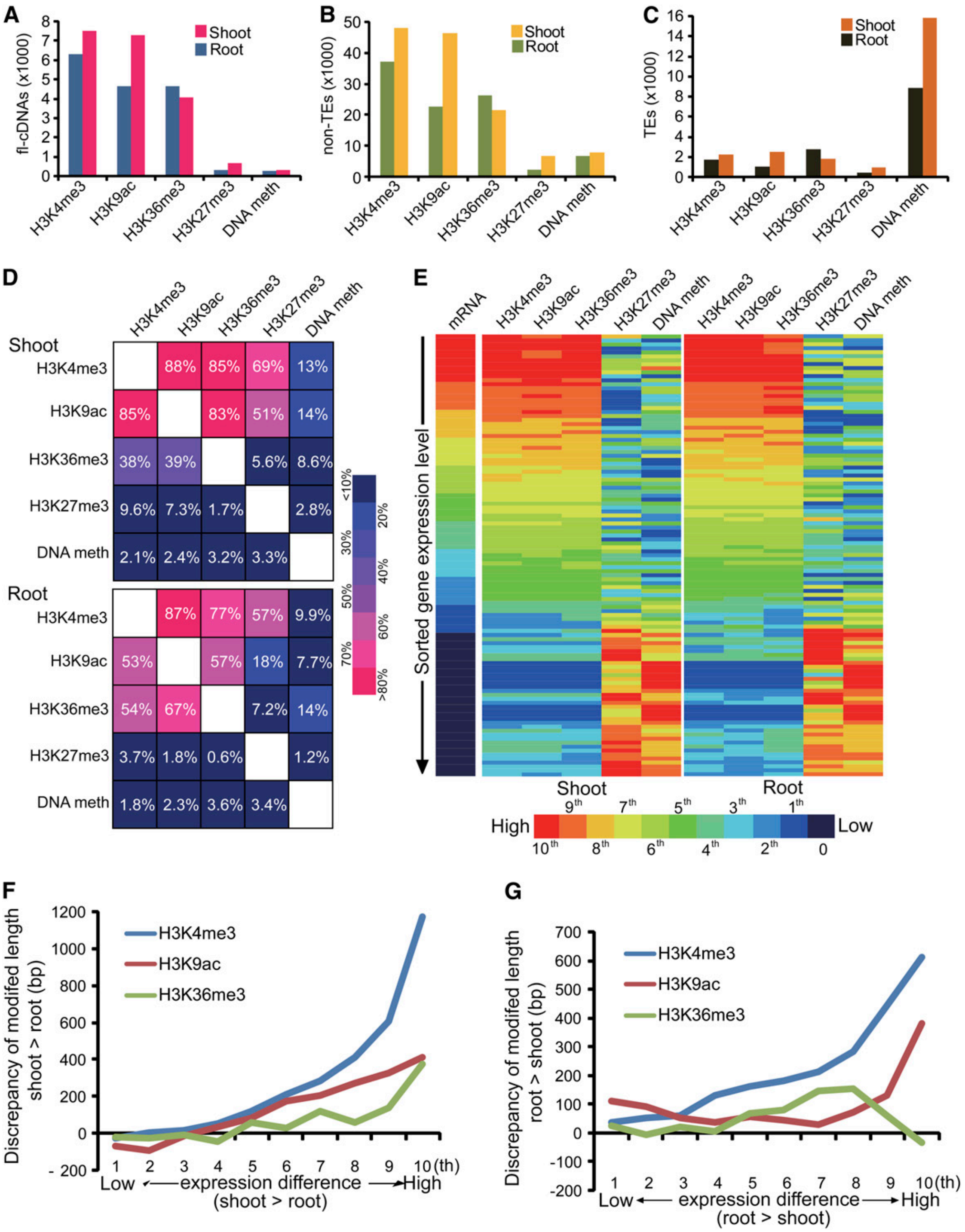
Furthermore, we analyzed whether different epigenetic marks showed distinct combinatorial effects. We found that in both shoots and roots, a significant and similar proportion of regions that were modified by one of the activating marks studied (H3K4me3, H3K9ac, and H3K36me3) were also modified by a second activating epigenetic mark (Figure 4D). While most pairwise combinations of activating epigenetic marks did not differ drastically between shoots and roots, 51% of all shoot-derived regions that were modified by H3K27me3 were comodified by H3K9ac, while only 18% of root-derived regions showed the same comodification pattern.

Additionally, we analyzed the influence of various combinations of epigenetic marks on the mRNA level of such modified genes. We observed that while all three activating epigenetic modifications under study were cooperatively present in genes with high mRNA levels and lacking in genes with low mRNA levels, the two repressive marks showed a mutually exclusive pattern (Figure 4E). In both shoots and roots, genes with low mRNA levels were marked with either H3K27me3 or methylated DNA, but genes marked with one of these modifications had low levels of the other, indicating a mutually exclusive effect between these two modifications. The mutually exclusive effect of those two repressive marks could also be observed for genes with high mRNA levels.

We observed that H3K9ac was more enriched in shoots than in roots (see Supplemental Figure 10 online). To analyze tissue-specific epigenetic effects in more detail, we grouped all non-TE genes into 10 percentiles based on their mRNA levels and plotted them against differences in the respective epigenetic modifications in shoots and roots (Figures 4F and 4G; see Supplemental Figure 11 online). We observed that H3K4me3, H3K9ac, and H3K36me3 were all correlated with tissue-specific gene expression, albeit to different degrees. Whereas a very distinct trend could be determined for H3K4me3, which was positively correlated with expression levels, H3K36me3 was less correlated with differential gene expression between shoots and roots. The different degrees of correlation with gene expression between these two activating histone modifications are unclear at this point. Interestingly, H3K36me3 continued to increase in the highest expression percentiles for genes that were more strongly expressed in shoots than in roots, but in contrast, it dramatically dropped in the highest gene expression percentiles for genes that were more strongly expressed in roots than in shoots (Figures 4F and 4G). Neither H3K27me3 nor DNA methylation displayed a clear trend like the activating epigenetic marks of interest, which indicates that in our study, neither H3K27me3 nor DNA methylation had a clear effect on differential expression of genes in maize shoots and roots at the genome scale (see Supplemental Figure 11 online).

Changes in smRNA Populations Follow *mop1* Gene Expression

To profile smRNA populations in maize seedling shoot and root tissue, we generated smRNA libraries for Illumina/Solexa 1G sequencing. After removing reads that likely originated from rRNA or tRNA contamination, we obtained 4,406,055 adaptor-trimmed sequences representing 1,639,984 unique smRNAs from shoots and 3,960,345 sequences representing 709,440 unique smRNAs from roots, respectively (see Supplemental Figure 12 online). We noted a tissue-specific smRNA size distribution: 24-nucleotide smRNAs were the predominant size class in shoots, whereas the predominant smRNAs in roots were 21 nucleotides (Figure 5A). This observation indicates that in maize, miRNAs, most of which are 20 to 22 nucleotides in length, are relatively enriched in roots, while siRNAs, which are mostly 24 nucleotides long, are relatively more prevalent in shoots. Interestingly, we did not observe a dramatic enrichment of 24-nucleotide siRNAs, as recently reported for maize flower organs (Nobuta et al., 2008) and for *Arabidopsis* immature floral tissue (Lister et al., 2008). It has been previously described (Henderson and Jacobsen, 2007) that in *Arabidopsis* the endogenous siRNA biogenesis pathway requires RNA-dependent RNA polymerase-2 (RDR2). In maize, MOP1 is homologous to RDR2, and it has been shown that a loss of function of RDR2 and MOP1 caused dramatic reduction of 24-nucleotide siRNAs in *Arabidopsis* and



Shoot	H3K4me3	H3K9ac	H3K36me3	H3K27me3	DNA meth
H3K4me3		88%	85%	69%	13%
H3K9ac	85%		83%	51%	14%
H3K36me3	38%	39%		5.6%	8.6%
H3K27me3	9.6%	7.3%	1.7%		2.8%
DNA meth	2.1%	2.4%	3.2%	3.3%	

Root	H3K4me3	H3K9ac	H3K36me3	H3K27me3	DNA meth
H3K4me3		87%	77%	57%	9.9%
H3K9ac	53%		57%	18%	7.7%
H3K36me3	54%	67%		7.2%	14%
H3K27me3	3.7%	1.8%	0.6%		1.2%
DNA meth	1.8%	2.3%	3.6%	3.4%	

Figure 4. Combinatory Modifications and Correlation with Gene Expression.

(A) to **(C)** Numbers of modified flcDNAs, non-TEs, and TEs by H3K4me3, H3K9ac, H3K36me3, H3K27me3, and DNA methylation in shoot and root.

(D) Frequencies of concurrent modifications on genes. Above the diagonal, numbers indicate the percentage of genes modified by X also have modification Y, while below the diagonal, percentages indicate how many genes were modified by Y and also modified by X.

(E) Heat maps of epigenetic modification levels on ~60,000 genes sorted by their expression measured by mRNA-seq. Gene expression levels and modifications levels were transformed to 100 percentiles, and each bar represents the averaged level of ~600 genes within each percentile.

(F) and **(G)** Correlation of differential modifications and differential gene expression in shoot and root. The *y* axis shows differences in the modification level of shoot higher than root and vice versa. The *x* axis shows the difference in the expression level of shoot higher than root and vice versa.

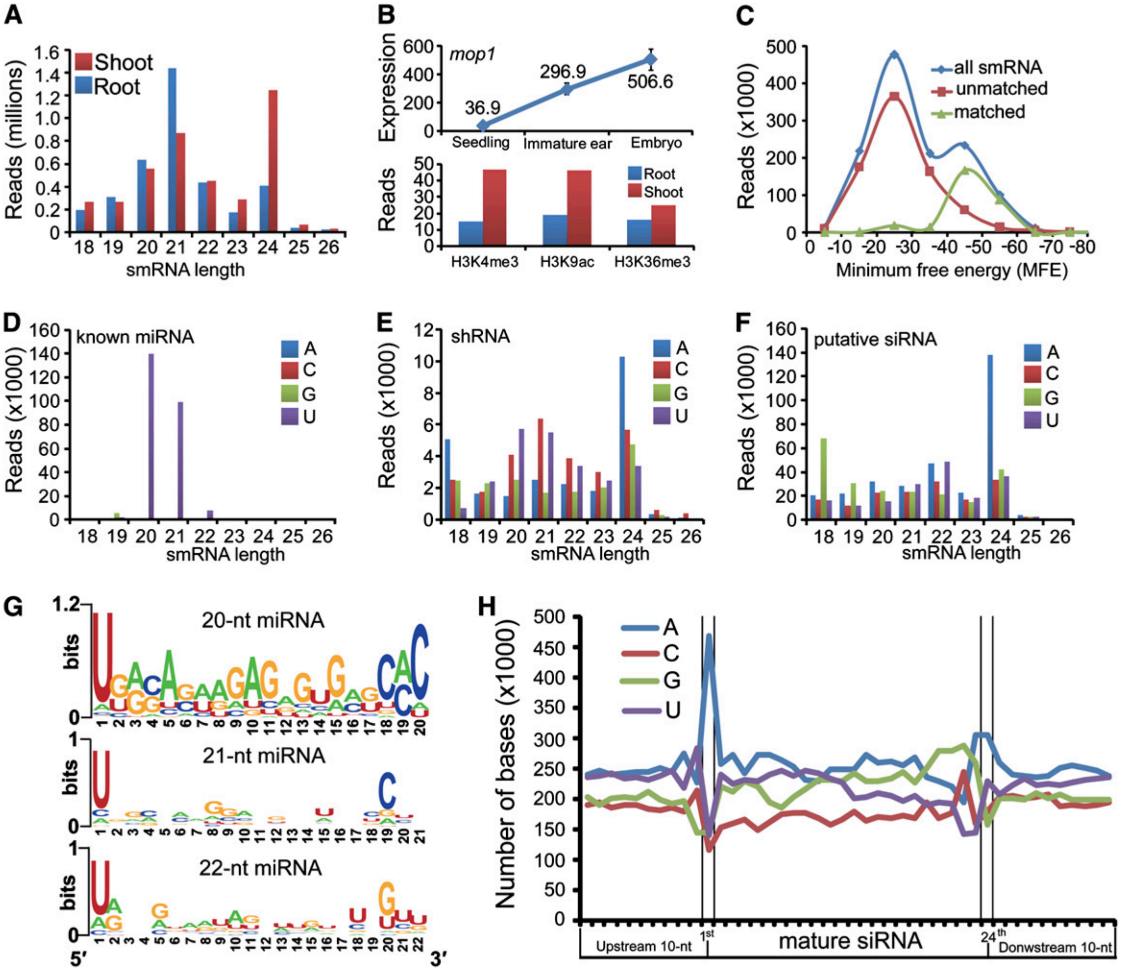


Figure 5. In Silico Classification Indicates Dynamic smRNA Populations in Maize Shoots and Roots.

(A) smRNA length distributions in shoots and roots.

(B) Tissue-specific expression and epigenetic modification of maize *mop1* gene.

(C) Distribution of smRNAs and matched and unmatched known miRNAs in miRBase within different MFE bins.

(D) to **(F)** Length distributions of known miRNA, shRNAs, and putative siRNAs with different 5′ terminal nucleotides.

(G) Sequence motifs of 20-, 21-, and 22-nucleotide miRNAs analyzed by WebLogo (Crooks et al., 2004).

(H) Nucleotide composition of mature 24-nucleotide putative siRNAs.

maize, respectively (Nobuta et al., 2008; Woodhouse et al., 2006). To determine whether differences in *mop1* expression levels could explain the different compositions of smRNA populations in maize seedling and floral tissue, we examined the *mop1* expression level across different organs using published microarray data (Stupar and Springer, 2006) and our mRNA-seq reads for shoots and roots. We found that *mop1* expression in seedlings was significantly lower than in immature ears and embryos (Figure 5B), confirming previous findings for maize (Woodhouse et al., 2006) and for *RDR* homologs in rice (Kapoor et al., 2008). In fact, when examining the mRNA-seq data from our study, we found that only ~40 reads, mostly from shoots, mapped to the *mop1* gene, which indicates a very low expression level in seedling tissues. Moreover, we found that the three activating epigenetic marks H3K4me3, H3K9ac, and H3K36me3 were slightly more abundant for *mop1* in shoots compared with roots (Figure 5B). In summary, these findings suggest that decreasing *mop1* expression leads to a concomitant decrease of 24-nucleotide siRNAs relative to 21-nucleotide miRNAs in a tissue-specific manner progressing from floral organs, to shoots, to roots.

Classification of smRNAs Based on Secondary Structure Predictions of Precursors

The smRNA population within a cell is composed of miRNA and natural antisense transcript-derived miRNA (Lu et al., 2008) as well as several classes of endogenous siRNAs, including repeat-associated RNA, natural antisense transcript-derived siRNA, and *trans*-acting siRNA (Bonnet et al., 2006; Ramachandran and

Chen, 2008). To separate miRNAs from siRNAs, we aligned all smRNA reads with known miRNA sequences from miRBase (Griffiths-Jones et al., 2006). Since sequence similarity alone does not necessarily guarantee that the smRNA in question is a miRNA, we next determined whether the respective smRNA precursor sequences were able to form a stem-loop structure indicative of miRNAs, which are derived from short hairpin structures, whereas siRNAs generally form from long double-stranded RNA (dsRNA) molecules. To determine putative precursor sequences, we adopted a more stringent mapping method using SOAP software (Li et al., 2008a) to retrieve all perfectly mapped locations for each smRNA and then extended 20 nucleotides at the 5′-end and 70 nucleotides at the 3′-end (see Supplemental Methods online). Using this approach, we obtained 37,763,920 and 18,734,677 putative precursor sequences from 2,890,098 smRNAs in shoots and 1,650,153 smRNAs in roots, respectively. We employed RNAfold (Hofacker et al., 1994) to calculate a minimum free energy (MFE) for each putative precursor. The lower the MFE, the higher the possibility that a precursor can form a stem-loop structure (Hofacker et al., 1994). To determine the minimum threshold, we compared the MFE for the smRNAs that matched known miRNAs in miRBase and those with unmatched sequences (Figure 5C; see Supplemental Figure 14A and Supplemental Tables 3 and 4 online). For the overall set of smRNAs, we observed two distinct peaks at −25 and −45, indicating a mixture of miRNAs and siRNAs, while for the matched and the unmatched smRNAs, single peaks were found to center at −45 and −25, respectively. Therefore, we set the MFE minimum threshold at −40 to determine the ability of a smRNA's precursor to form a hairpin structure.

Based on these criteria, we categorized all smRNAs into three groups (see Supplemental Figure 13 online). Group I, "known miRNAs" with matches in miRBase and MFE < −40, consisted of 526,961 reads representing 155 unique sequences from shoots and 252,505 reads representing 126 unique sequences from roots. Group II, "small hairpin RNAs (shRNAs)" without matches in miRBase but MFE < −40, consisted of 120,227 reads representing 10,314 unique sequences from shoots and 131,553 reads representing 31,856 unique sequences from roots. This group might include unidentified miRNAs and other smRNA species. Group III consisted of all remaining smRNAs whose precursors could not form hairpins. We classified all smRNAs in group III as "putative siRNAs," consisting of 1,768,555 reads representing 984,890 unique sequences from shoots and 800,094 reads representing 379,199 unique sequences from roots, respectively. Interestingly, these three groups of smRNAs had distinctly different average frequencies with ∼3400 copies for known miRNA, ∼120 copies for shRNAs, and ∼1.8 copies for putative siRNAs.

Three Groups of smRNAs Exhibited Distinct Signatures of 5′ Terminal Nucleotide Identities and Overall Nucleotide Compositions

It has been shown that in *Arabidopsis*, the 5′ terminal nucleotide is a key characteristic to direct distinct smRNA classes to different Argonaute (AGO) complexes (Mi et al., 2008). Therefore, we examined the size distributions of smRNAs in these three groups based on their 5′ terminal nucleotides. We found that virtually all known miRNAs (Group I) had a 5′ U, the signature of canonical miRNAs (Figure 5D; see Supplemental Figure 14B online), while most 24-nucleotide putative siRNAs (Group III) had a 5′ A, a signature feature of canonical siRNAs (Figure 5F; see Supplemental Figure 14D online).

Unexpectedly, smRNAs in Group II demonstrated a more complex distribution (Figure 5E). Within this group, a large number of 20-, 21-, and 22-nucleotide smRNAs had a 5′ terminal U, indicative of canonical miRNAs. However, an equally large number of smRNAs in these size classes also had a 5′ terminal C, which might represent either a novel group of miRNAs or unknown small hairpin siRNAs. Furthermore, this group of smRNAs also contained a large number of 24-nucleotide siRNAs with a 5′ A, suggesting that certain siRNA species need a hairpin precursor state for processing through DICER. The complex composition of this group of smRNAs, which most likely includes miRNAs and siRNAs as well as potentially other unknown smRNA species, led us to classify these smRNAs collectively as shRNAs. In shoots, 20- to 22-nucleotide smRNAs with a 5′ terminal C were not detected, indicating that 5′ C shRNAs might potentially represent a group of uncharacterized tissue-specific smRNAs (see Supplemental Figure 14C online).

To further characterize the sequence patterns of these three groups of smRNAs and to explore smRNAs in irregular lengths other than 21 and 24 nucleotides, we calculated the frequencies for each nucleotide within the mature smRNA and extended the mature RNA by 10 nucleotides at both ends. For the known miRNAs in lengths of 20, 21, and 22 nucleotides, sequence motifs were analyzed by WebLogo (Crooks et al., 2004). The sequence motifs reflected the most enriched miRNA families (Figure 5G; see Supplemental Figures 15A and 15B online). Overall, we observed a high frequency of upstream As and Us for half of the putative siRNA group and a sharp peak for 5′ terminal A (Figure 5H). This result is congruent with sequence patterns found in *Arabidopsis* (Lister et al., 2008). However, the relative enrichment of 3′ Gs seems to be a unique feature of maize when compared with *Arabidopsis*. For putative 20- to 26-nucleotide siRNAs (excluding the 24-nucleotide class), we observed a relatively high frequency of As up to two nucleotides upstream of the 5′ terminus as well as for the 3′ terminal nucleotide (see Supplemental Figure 16 online). This result indicates that the siRNA of other lengths could be variations of canonical siRNAs. Overall, the nucleotide composition of the shRNA group showed the highest amount of GC from −10 nucleotides to +10 nucleotides in the mature smRNAs, indicating distinct differences in the nature of shRNA compared with miRNAs and siRNAs (see Supplemental Figure 17 online).

22-Nucleotide siRNAs Are Differentially Enriched in Long Hairpin dsRNAs

In both shoots and roots, we found that siRNA populations were enriched primarily in 24-nucleotide and secondarily in 22-nucleotide species (see Supplemental Figures 13E and 13F online). A recent study showed that 22-nucleotide siRNAs were specifically enriched in maize compared with other plants, which led to the hypothesis that this size class might potentially represent a new species of smRNA in addition to the canonical 21- and 24-nucleotide siRNA (Nobuta et al., 2008). It is possible that a yet to be identified siRNA biogenesis pathway exists in maize (Nobuta et al., 2008). Two other recent reports summarizing work in mouse delivered evidence that siRNAs found in naturally formed endogenous long hairpin dsRNA molecules are responsible for generating a certain class of smRNAs (Tam et al., 2008; Watanabe et al., 2008). It has also been shown in maize that smRNAs produced from a hairpin version of *MuDR*, Muk, are not lost in a *mop1* mutant background (Woodhouse et al., 2006). Taken together, these findings led us to explore whether long hairpin dsRNAs are the sources of 22-nucleotide siRNAs in maize because a naturally formed RNA duplex could be independent of *mop1*, whose expression we found to be very low in seedling tissues.

We performed de novo scanning of 2.4-Gb maize BACs using the *einverted* program (see Supplemental Methods online) and identified 1086 long hairpin dsRNAs with a stem length of at least 1000 nucleotides and at least 90% base pair complementation within the stem sequence. By mapping the putative siRNAs onto long hairpin dsRNAs, we indeed observed a higher relative enrichment of 22-nucleotide compared with 24-nucleotide siRNAs in both shoots and roots (Figures 6A to 6D), which differed from the siRNAs mapped onto LTR-TEs (Figures 6E to 6G). A detailed comparison of siRNAs derived from long hairpin dsRNAs or LTR-TEs revealed more unique features of this novel siRNA species. First, we found that these siRNAs had a higher copy number (305,288 reads representing 58,210 unique sequences from shoots and 238,313 reads representing 30,138 unique sequences from roots). Second, we identified shorter siRNAs (18 to 22 nucleotides), which were replicated in even higher frequencies (e.g., 30 times for 20-nucleotide siRNAs in roots). Third, 19-, 20-, 21-, and 24-nucleotide siRNAs bore a signature 5′ terminal A, whereas 22-nucleotide siRNAs had approximately equal amounts of 5′ A and 5′ U. In summary, our observations indicate that siRNAs derived from long hairpin dsRNA might be a miRNA-like species, even though they bear canonical siRNA features.

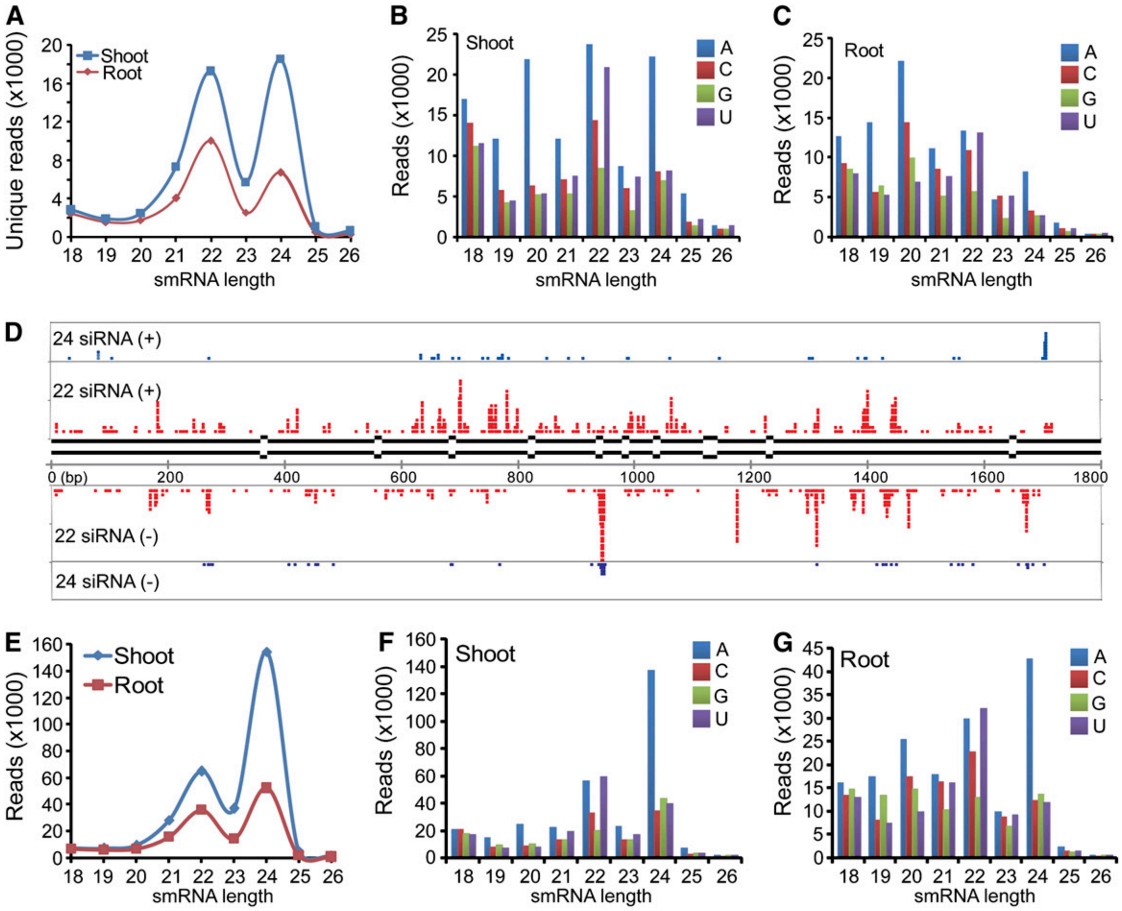


Figure 6. 22-Nucleotide siRNAs Are Differentially Enriched in Long Hairpin dsRNAs Rather Than in LTR-TEs.

(A) to **(C)** Length distributions of putative siRNAs mapped on long hairpin dsRNAs. **(A)** Count of unique sequences; **(B)** and **(C)** total reads.

(D) An example of a long hairpin dsRNAs generating more 22-nucleotide siRNAs than 24-nucleotide siRNAs. The loop region of ~500 bp is not shown, and paired regions in stem are 99% in identity. Bubbles indicate unmatched nucleotides.

(E) to **(G)** Length distributions of putative siRNAs mapped on full-length LTR-retrotransposons. **(E)** Count of unique sequences; **(F)** and **(G)** total reads.

smRNAs Target Distinct Regions in Genes and Full-Length LTR-RetroTEs

Traditional annotation of TEs is based on open reading frame predictions followed by comparison with known repeat types in public databases. However, TEs predicted following this strategy cannot represent a complete unit, especially in the case of LTR-retrotransposons, which have a complicated architecture. Therefore, we used a program called LTR-finder (Zhao and Wang, 2007) and identified 75,015 full-length LTR retrotransposons de novo, representing 880 Mb of DNA sequence (see Supplemental Methods online). By mapping putative siRNAs to LTR-TEs, we found 753,512 siRNA reads representing 314,044 unique sequences from shoots and 455,881 reads representing 138,853 unique sequences from roots, respectively. When we analyzed the distribution of the 5′ terminal nucleotides for siRNAs matching LTR-TEs, we found that in both shoots and roots, most 24-nucleotide siRNAs had the characteristic 5′ terminal A, but that 22-nucleotide siRNAs started with an A or U in about equal proportions (Figures 6E to 6G). This result might indicate different mechanisms in 22- and 24-nucleotide siRNA biogenesis as well as tissue-specific siRNA populations.

siRNAs have two main known functions. The majority of repeat-associated 24-nucleotide siRNAs contribute to the formation of DNA methylation, while a small portion of siRNAs including 21- and 24-nucleotide classes contribute to the RNA interference machinery targeting genes and TEs either in *trans*-acting or natural-antisense-transcript mode (Bonnet et al., 2006). Therefore, we analyzed the distributions of the respective siRNA classes surrounding and within genes and LTR-TEs. Since most siRNAs are associated with repetitive sequences, keeping a randomly selected subset of all siRNAs would lead to a significant bias. For this reason, we adopted a method based on a best possible compromise using the following formula: coverage of one siRNA divided by all the locations this siRNA could be mapped to in the genome (see Supplemental Methods online). As a basic classification, we assumed here that if a smRNA is mapped to the sense strand of a genomic locus, this smRNA might originate from this site, while a smRNA mapped on the antisense strand of a locus might indicate that this smRNA targets this site. However, this approximation does not take other, more complicated scenarios into account (e.g., origination of siRNAs from antisense mRNAs and base-pairing of siRNAs to genomic DNA in addition to sense mRNAs). To determine whether different size classes of siRNAs and shRNAs target different regions in genes or LTR-TEs, we examined the distribution of the respective smRNAs over gene regions and LTR-TEs in shoots and roots (Figures 7A to 7H; see Supplemental Figures 18A to 18H online).

Interestingly, each class of siRNAs exhibited a distinct pattern on genes and LTR-TEs. For the 24-nucleotide siRNAs on flcDNA genes, we observed a distinct bias toward the sense and antisense strand in specific regions. A sharp 5′ peak indicated that a group of 24-nucleotide siRNAs originated from the immediate upstream 100- to 200-bp region on the sense strand of genes, while another equal amount of 24-nucleotide siRNAs targeted the immediate downstream 100 to 200 bp, which would still be within the 3′ untranslated regions. This group of 24-nucleotide siRNAs potentially represents natural antisense transcript-derived siRNAs. For the 24-nucleotide siRNAs on LTR-TEs, we found that the overall distribution on the sense strand was mirrored on the antisense strand and that the transcribed regions had a higher proportion of 24-nucleotide siRNAs than the 5′ and 3′ LTR regions (Figures 7A and 7B; see Supplemental Figures 18A and 18B online).

The 21-nucleotide siRNAs exhibited a similar pattern compared with 24-nucleotide siRNAs in genes, but differed in their origin regions. On the LTR-TEs, the distribution of 21-nucleotide siRNAs on the sense and antisense strands was dissimilar. We found more 21-nucleotide siRNAs on the sense strand at the 5′ end, indicating more origin sites, while we observed more 21-nucleotide siRNAs on the antisense strand at the 3′ end, indicating more targeting sites in this region (Figures 7C and 7D; see Supplemental Figures 18C and 18D online).

Similarly, 22-nucleotide siRNAs exhibited a strand-specific distribution in the transcribed region of genes (Figure 7E; see Supplemental Figure 18E online). However, the origins of 22-nucleotide siRNAs were biased toward the 3′ end, while the targeting sites were biased toward the 5′ end within the transcribed regions. The 22-nucleotide siRNAs showed a similar pattern on LTR-TEs (Figure 7F; see Supplemental Figure 18F online). This pattern indicates 22-nucleotide siRNAs might fulfill their silencing function in a different way compared with 21- and 24-nucleotide siRNAs.

Interestingly, shRNAs were extremely strand-specific in both flcDNAs and LTR-TEs (Figures 7G and 7H; see Supplemental Figures 18G and 18H online). Virtually all shRNAs mapped to the antisense strand in both 5′ and 3′ regions of flcDNAs, indicating that shRNAs could function in a *trans*-acting fashion. In LTR-TEs, almost all shRNAs mapped to the sense strand in the 3′ coding region.

Overall, our findings indicate that different siRNA classes target different regions in genes and LTR-TEs and target different transposon classes (see Supplemental Figure 19 online), pointing at specialized regulatory roles during epigenetic regulation of these siRNAs (Figures 7I to 7K).

DISCUSSION

Using maize, we have generated an integrated genome-wide and organ-specific survey of epigenetic marks together with transcriptional outputs. Our results show that Illumina/Solexa 1G sequencing and read mapping are feasible with high accuracy even in large and repeat-rich plant genomes, opening the door to exploring similarly complex genomes in the future.

Epigenetic changes, including histone modifications and DNA methylation, have a profound impact on gene regulation. We observed that H3K4me3, H3K9ac, and H3K36me3 were associated with transcriptionally active genes, while H3K27me3 and DNA methylation were predominantly found in transcriptionally inactive genes and repetitive elements, supporting the findings of previous studies in other organisms (e.g., Martens et al., 2005; Zhang et al., 2006; Barski et al., 2007; Zilberman et al., 2007; Li et al., 2008c; Wang et al., 2008). Interestingly, we found that genic DNA methylation patterns in maize are very similar to rice, but very different from *Arabidopsis*. While in maize and rice, genic

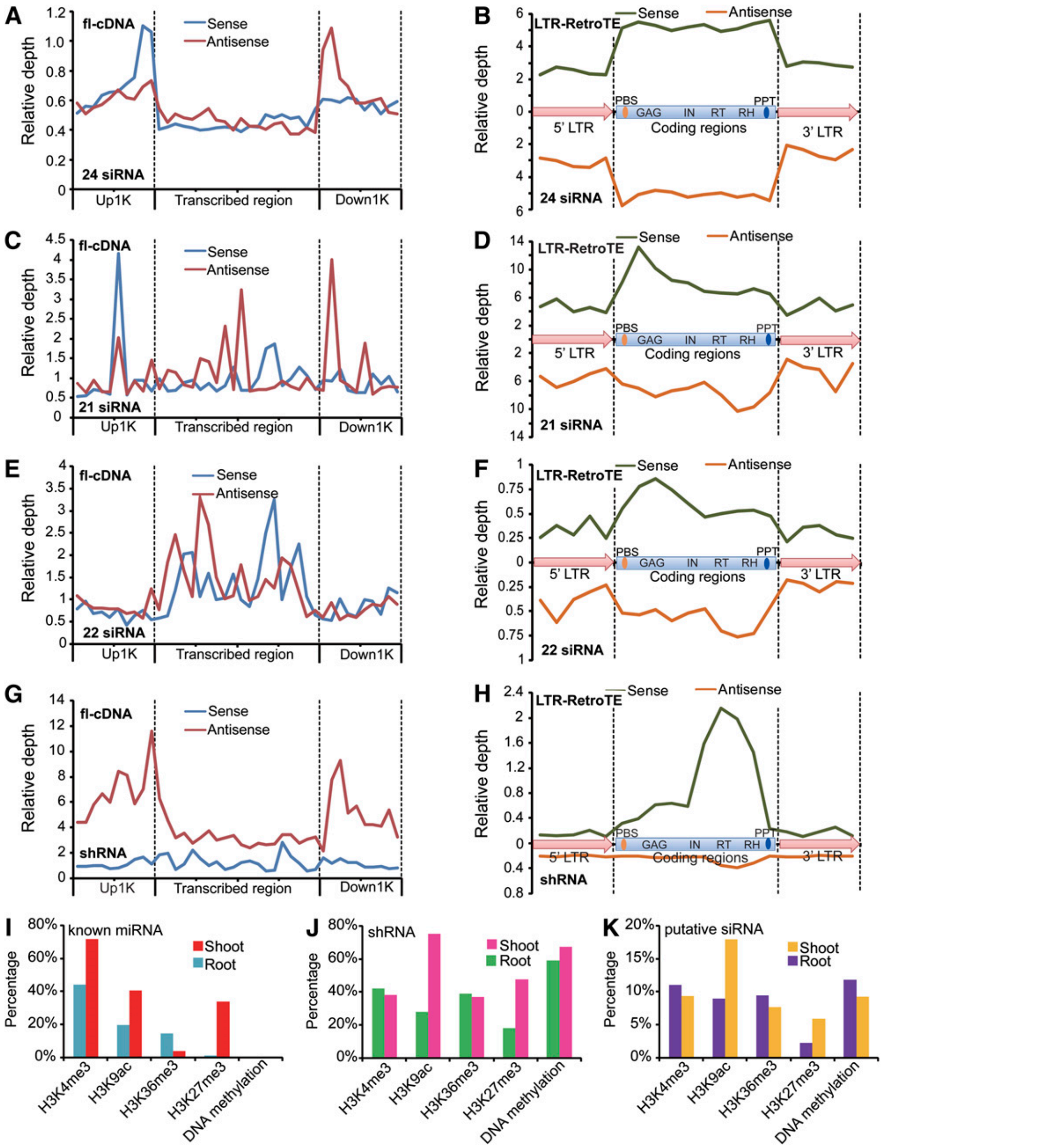


Figure 7. Origin and Target Sites on Genes and LTR-TEs for Different Classes of Putative siRNAs.

(A), **(C)**, **(E)**, and **(G)** The 24-, 21-, and 22-nucleotide siRNAs and shRNAs on flcDNA genes show significant strand bias on different positions in originating and targeting strands.

(B), **(D)**, **(F)**, and **(H)** The 24-, 21-, and 22-nucleotide siRNAs and shRNAs on LTR-TEs. Calculation of relative depth and de novo identification of LTR-TEs is described in the supplemental data online.

(I) to **(K)** Percentages of unique smRNA loci situated in epigenetic regions of H3K4me3, H3K9ac, H3K36me3, H3K27me3, and DNA methylation.

DNA methylation peaks around the ATG (Figure 3K; Li et al., 2008c), it is most prevalent in the transcribed region in *Arabidopsis* genes (Zhang et al., 2006; Zilberman et al., 2007). Moreover, we found that the differential accumulation of distinct epigenetic marks in genes and repetitive elements was reflected in the proportion of reads mapped to unique or nonunique positions in the genome. As expected for strongly repeat-associated modifications, we only identified a small number of unique genome positions for H3K27me3 and DNA methylation (Figure 1B; see Supplemental Figure 1B online). Interestingly, we found that while multiple activating epigenetic marks tended to occur together, the two repressive marks under study, H3K27me3 and DNA methylation, were more likely to exclude each other at the same loci (Figures 4D and 4E). This supports similar findings for genome-wide studies in *Arabidopsis* (Mathieu et al., 2005; Zhang et al., 2007) and for a locus-specific analysis in mouse (Lindroth et al., 2008). For example, in *Arabidopsis*, <10% of H3K27me3-covered regions overlapped with DNA methylation (Zhang et al., 2007). Even though the reason behind this antagonism is unclear, it suggests a very different mode of action compared with activating epigenetic marks, which generally do not seem to be mutually exclusive. H3K27me3 is regarded as a mark of transcriptional quiescence, but a recent study (Riclet et al., 2009) showed that in mouse, upon loss of heterochromatin protein 1 on the *mesoderm-specific transcript* promoter, H3K27me3 associates with gene activation and correlates with DNA hypomethylation. In animals, H3K27me3 regions typically form large domains (>5 kb) and include multiple genes (Bernstein et al., 2006; Schwartz et al., 2006). In plants, it covers much shorter regions (typically <1 kb), and it tends to be restricted to the coding region of single genes (Figures 3B and 3J; Zhang et al., 2007). Taken together, these results suggest that H3K27me3 might be based on different spreading and maintenance mechanisms and that it might also have different functions in gene activation and gene repression in plants and animals.

smRNAs have been increasingly recognized as key regulators of gene activity that can have major effects. For example, a recent study has shown that miRNAs were involved in the domestication of maize (Chuck et al., 2007). Whereas most endogenous plant siRNAs are 21 to 24 nucleotides long (Ramachandran and Chen, 2008), maize possesses an additional class of 22-nucleotide siRNAs. Interestingly, other monocots, such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), or rice, lack 22-nucleotide siRNAs (Nobuta et al., 2008). To elucidate the biogenesis of this 22-nucleotide class, we examined potential sources of these siRNAs and found that they might be generated from long dsRNAs. We hypothesize that the respective long dsRNAs might be encoded by pseudogenes similar to those found in mouse, where duplexes formed by their sense and antisense transcripts have been shown to produce siRNAs without requiring RdRP activities (Tam et al., 2008; Watanabe et al., 2008). Canonical 24-nucleotide siRNAs bear a 5′ terminal A, which is recognized by AGO2 and AGO4 (Mi et al., 2008). Interestingly, we found that 22-nucleotide siRNAs matching long dsRNAs bear all four nucleotides at their 5′ end, which indicates the involvement of other AGO proteins or potentially non-AGO proteins during 22-nucleotide siRNA-mediated silencing processes. We observed marked differences in the distributions of siRNAs derived from long hairpin dsRNAs compared with those derived from LTR-TEs. For example, long hairpin dsRNA-derived siRNAs were relatively enriched for small sizes (18 to 22 nucleotides) and had a high copy frequency (Figures 6B and 6C), while for LTR-TE–derived siRNAs, the copy frequency was relatively low. These differences might indicate two distinct siRNA biogenesis pathways in maize, in which RdRP is necessary to generate siRNAs from LTR-TEs but not from long hairpin dsRNAs. We found that the expression level of one RdRP gene, *mop1*, correlated with a decrease of 24-nucleotide siRNAs relative to 21-nucleotide miRNAs in a tissue-specific manner progressing from floral organs to shoots and roots. Intriguingly, *mop1* also seems to be involved in a tissue-specific regulation of paramutation and silencing at the *p1* locus in maize (Sidorenko and Chandler, 2008), which opens the possibility that siRNAs might be involved in tissue-specific and targeted paramutation.

Maize was one of the first model organisms for biological research and has a rich history in the study of epigenetics, plant domestication, and evolution. With the recent release of its first draft genomic sequence, it is once again taking center stage in both plant biology and crop improvement. We hope that the epigenetic and transcriptomic survey we have described here will aid in further annotating and understanding the maize genome. It will also be useful for exploring epigenetic principles and even more complex smRNA biology, as well as the interplay between epigenomes and transcriptomes. In summary, we hereby have delivered a critical analysis of the overall landscapes of epigenetic histone marks and DNA methylation, together with mRNA and smRNA transcriptomes in maize.

METHODS

Plant Growth Conditions

Maize (*Zea mays*) inbred line B73 was obtained from the USDA–Agricultural Research Service North Central Regional Plant Introduction Station in Ames, IA. Seeds were planted in individual pots containing a mixture of three parts soil (Premier Pro-Mix Bx Professional; Premier Horticulture) and two parts vermiculite (D3 Fine Graded Horticultural Vermiculite; Whittemore). Plants were grown under controlled environmental conditions (15 h light/25°C, 9 h dark/20°C) in a growth chamber, and the soil mixture was kept moist by watering the pots with 0.7 mM $Ca(NO_3)_2$. Seedlings were harvested after 14 d, separated into shoots and roots, frozen in liquid nitrogen, and stored at –80°C or processed directly after harvesting for ChIP.

Sample Preparation and Solexa Library Construction

Maize tissue from 10 different seedlings was ground in liquid nitrogen, and genomic DNA was extracted from 1 g pooled tissue using a Qiagen DNeasy plant maxi kit. To enrich for methylated genomic DNA, 20 μg genomic DNA were digested with 200 units McrBC (New England Biolabs) overnight, and fragments 500 nucleotides and smaller were gel purified and used for library construction following the manufacturer's instructions, but adding a final gel purification step. To enrich for histone-modified regions, ChIP was conducted using 5 g fresh maize tissue from 10 seedlings following a previously described procedure (Lee et al., 2007). The following antibodies were used: H3K9ac (Upstate; 07-352), H3K27me3 (Upstate; 07-449), H3K4me3 (Abcam; ab8580), and H3K36me3 (Abcam; ab9050). For each 1-mL ChIP reaction, 5 μL antibody were added. The ChIPed DNA from three reactions was pooled to construct Solexa libraries essentially following the manufacturer's standard protocol but running 18 PCR cycles before gel purification of the samples. Total RNA was isolated using TRIzol reagent (Invitrogen) following the manufacturer's instructions. mRNA was extracted from total RNA using Dynabeads Oligo(dT) (Invitrogen Dynal) following the manufacturer's directions. After elution from the beads, first- and second-strand cDNA was generated using SuperscriptII reverse transcriptase (Invitrogen), and the standard Solexa protocol was followed thereafter to create mRNA libraries. smRNA was extracted by running total RNA on a 15% PAGE gel and cutting out bands in the ∼19- to 24-nucleotide size range. Libraries for smRNAs were constructed following previously published procedures (Mi et al., 2008; see Supplemental Methods online for details). All samples were prepared for sequencing following the manufacturer's standard protocol.

Sequence Data

The data for this article have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE15286. All data also can be freely accessed at http://plantgenomics.biology.yale.edu.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Sequencing and Mapping of mRNA Transcripts, H3K4me3, H3K9ac, H3K36me3, H3K27me3, DNA Methylation, and smRNAs in Maize Roots.

Supplemental Figure 2. Detection of Individual Transcribed Exons by de Novo Scanning of mRNA-seq Reads across the Maize Genome Sequence.

Supplemental Figure 3. Mapping of 11,742 Maize flcDNAs to the Maize Genome Sequence.

Supplemental Figure 4. Comparison of Gene Homology in Maize, Rice, and *Arabidopsis*.

Supplemental Figure 5. Pathway Annotation of Maize Genes.

Supplemental Figure 6. Comparison of Pathway Enrichment in Maize versus Rice and Maize versus *Arabidopsis*.

Supplemental Figure 7. Comparison of Gene Ontology Enrichment between Maize/Rice and Maize/*Arabidopsis*.

Supplemental Figure 8. Distribution of Epigenetic Patterns within Maize Genes in Roots.

Supplemental Figure 9. Effect of Modification Levels on Gene Expression.

Supplemental Figure 10. A Differentially Expressed Gene Shows a Different Epigenetic Pattern.

Supplemental Figure 11. Correlation of Differential Modifications of H3K27me3 and DNA Methylation with Differential Gene Expression in Shoots and Roots.

Supplemental Figure 12. Length Distributions of Removed smRNA Reads Matched with tRNA and rRNA Sequences.

Supplemental Figure 13. Length Distribution of Three Groups of smRNAs.

Supplemental Figure 14. Classification of smRNA Population in Shoots.

Supplemental Figure 15. Nucleotide Composition of Known miRNAs.

Supplemental Figure 16. Nucleotide Composition of Putative siRNAs.

Supplemental Figure 17. Nucleotide Composition of shRNAs.

Supplemental Figure 18. Origin and Target Sites on Genes and LTR-TEs for Different Classes of Putative siRNAs in Roots.

Supplemental Figure 19. Percentages of Different Types of Repeats Generating smRNAs in Different Lengths.

Supplemental Figure 20. Length Distribution of Unmapped smRNAs Classified by 5′ Terminal Nucleotides.

Supplemental Table 1. Solexa Sequencing and Mapping Statistics.

Supplemental Table 2. Validation Statistics of FgeneSH Predicted Genes.

Supplemental Table 3. Number of smRNAs Hitting Known miRNAs in Different Minimum Free Energy Ranges.

Supplemental Table 4. Solexa Sequencing Reads for miRNAs.

Supplemental Data Set 1. Comparisons of Maize and Rice Pathways.

Supplemental Data Set 2. Comparisons of Maize and *Arabidopsis* Pathways.

Supplemental Methods.

ACKNOWLEDGMENTS

The Deng laboratory at Yale University was supported by grants from the National Institutes of Health (GM047850), the National Science Foundation Plant Genome Program (DBI0421675), and the National Science Foundation 2010 program. Studies conducted at the National Institute of Biological Sciences were supported by special funds from the Ministry of Science and Technology of China and Beijing Commission of Science and Technology. We thank the Maize Genome Sequencing Consortium (http://www.maizesequence.org) for communicating maize sequence information before publication and Will Terzaghi and Tim Nelson for critical reading of the manuscript. A.A.E. was supported by a Yale University Brown-Coxe Postdoctoral Fellowship.

Received January 17, 2009; revised March 4, 2009; accepted April 1, 2009; published April 17, 2009.

REFERENCES

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J.M. (2002). HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. EMBO J. **21:** 6832–6841.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell **129:** 823–837.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana.* PLoS One **3:** e3156.

Bernstein, B.E., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell **125:** 315–326.

Bonnet, E., Van de Peer, Y., and Rouzé, P. (2006). The small RNA world of plants. New Phytol. **171:** 451–468.

Chandler, V.L., and Brendel, V. (2002). The maize genome sequencing project. Plant Physiol. **130:** 1594–1597.

Chuck, G., Cigan, A.M., Saeteurn, K., and Hake, S. (2007). The heterochronic maize mutant *Corngrass1* results from overexpression of a tandem microRNA. Nat. Genet. **39:** 544–549.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature **452:** 215–219.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Res. **14:** 1188–1190.

Fernandes, J., Morrow, D.J., Casati, P., and Walbot, V. (2008). Distinctive transcriptome responses to adverse environmental conditions in *Zea mays* L. Plant Biotechnol. J. **6:** 782–798.

Fuchs, J., Demidov, D., Houben, A., and Schubert, I. (2006). Chromosomal histone modification patterns - from conservation to diversity. Trends Plant Sci. **11:** 199–208.

Goff, S.A., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296:** 92–100.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. **34:** D140–D144.

Haberer, G., al. (2005). Structure and architecture of the maize genome. Plant Physiol. **139:** 1612–1624.

Henderson, I.R., and Jacobsen, S.E. (2007). Epigenetic inheritance in plants. Nature **447:** 418–424.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S., Tacker, M., and Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. Monatshefte f. Chemie **125:** 167–188.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature **416:** 556–560.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science **293:** 1074–1080.

Kapoor, M., Arora, R., Lama, T., Nijhawan, A., Khurana, J.P., Tyagi, A.K., and Kapoor, S. (2008). Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analysis during reproductive development and stress in rice. BMC Genomics **9:** 451.

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of *Arabidopsis* siRNAs. PLoS Biol. **5:** e57.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell **128:** 693–705.

Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I., and Deng, X.W. (2007). Analysis of transcription factor HY5 binding sites revealed its hierarchical role in light regulation of development. Plant Cell **19:** 731–749.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008a). SOAP: Short oligonucleotide alignment program. Bioinformatics **24:** 713–714.

Li, H., Ruan, J., and Durbin, R. (2008b). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18:** 1851–1858.

Li, X., et al. (2008c). High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. Plant Cell **20:** 259–276.

Lindroth, A.M., Park, Y.J., McLean, C.M., Dokshin, G.A., Persson, J. M., Herman, H., Pasini, D., Miró, X., Donohoe, M.E., Lee, J.T., Helin, K., and Soloway, P.D. (2008). Antagonism between DNA methylation and H3K27 methylation at the imprinted *Rasgrf1* locus. PLoS Genet. **4:** e1000145.

Lippman, Z., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature **430:** 471–476.

Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol. **1:** e67.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C. C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis.* Cell **133:** 523–536.

Lorincz, M.C., Dickerson, D.R., Schmitt, M., and Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat. Struct. Mol. Biol. **11:** 1068–1075.

Lu, C., et al. (2008). Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc. Natl. Acad. Sci. USA **105:** 4951–4956.

Ma, J., Morrow, D.J., Fernandes, J., and Walbot, V. (2006). Comparative profiling of the sense and antisense transcriptome of maize lines. Genome Biol. **7:** R22.

Martens, J.H.A., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. **24:** 800–812.

Martienssen, R.A., Doerge, R.W., and Colot, V. (2005). Epigenomic mapping in Arabidopsis using tiling microarrays. Chromosome Res. **13:** 299–308.

Mathieu, O., Probst, A.V., and Paszkowski, J. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis.* EMBO J. **24:** 2783–2791.

Mathieu, O., Reinders, J., Čaikovsky, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell **130:** 851–862.

Messing, J., Bharti, A.K., Karlowski, W.M., Gundlach, H., Kim, H.R., Yu, Y., Wei, F., Fuks, G., Soderlund, C.A., Mayer, K.F., and Wing, R.A. (2004). Sequence composition and genome organization of maize. Proc. Natl. Acad. Sci. USA **101:** 14349–14354.

Messing, J., and Dooner, H.K. (2006). Organization and variability of the maize genome. Curr. Opin. Plant Biol. **9:** 157–163.

Meyers, B.C., Tingey, S.V., and Morgante, M. (2001). Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res. **11:** 1660–1676.

Mi, S., et al. (2008). Sorting of small RNAs into Arabidopsis Argonaute complexes is directed by the 5′ terminal nucleotide. Cell **133:** 116–127.

Mikkelsen, T.S., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature **448:** 553–560.

Nobuta, K., et al. (2008). Distinct size distribution of endogenous siRNAs in maize: Evidence from deep sequencing in the *mop1-1* mutant. Proc. Natl. Acad. Sci. USA **105:** 14958–14963.

Okitsu, C.Y., and Hsieh, C.L. (2007). DNA methylation dictates histone H3K4 methylation. Mol. Cell. Biol. **27:** 2746–2757.

Pennisi, E. (2008). Corn genomics pops wide open. Science **319:** 1333.

Rabinowicz, P.D., and Bennetzen, J.L. (2006). The maize genome as a model for efficient sequence analysis of large plant genomes. Curr. Opin. Plant Biol. **9:** 146–156.

Rabinowicz, P.D., Citek, R., Budiman, M.A., Nunberg, A., Bedell, J.A., Lakey, N., O'Shaughnessy, A.L., Nascimento, L.U., McCombie, W.R., and Martienssen, R.A. (2005). Differential methylation of genes and repeats in land plants. Genome Res. **15:** 1431–1440.

Ramachandran, V., and Chen, X. (2008). Small RNA metabolism in *Arabidopsis.* Trends Plant Sci. **13:** 368–374.

Riclet, R., Chendeb, M., Vonesch, J.-L., Koczan, D., Thiesen, H.-J., Losson, R., and Cammas, F. (2009). Disruption of the interaction between transcriptional intermediary factor 1β and heterochromatin protein 1 leads to a switch from DNA hyper- to hypomethylation and H3K9 to H3K27 trimethylation on the *MEST* promoter correlating with gene reactivation. Mol. Biol. Cell **20:** 296–305.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster.* Nat. Genet. **38:** 700–705.

Shi, J., and Dawe, R.K. (2006). Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. Genetics **172:** 1571–1583.

Sidorenko, L., and Chandler, V. (2008). RNA-dependent RNA polymerase is required for enhancer-mediated transcriptional silencing associated with paramutation at the maize p1 gene. Genetics **180:** 1983–1993.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. **8:** 272–285.

Stupar, R.M., and Springer, N. (2006). *Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F_1 hybrid. Genetics **173:** 2199–2210.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz,

R.M., and Hannon, G.J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature **453:** 534–539.

The Gene Ontology Consortium (2000). Gene Ontology: Tool for the unification of biology. Nat. Genet. **25:** 25–29.

Vaughn, M.W., et al. (2007). Epigenetic natural variation in *Arabidopsis thaliana.* PLoS Biol. **5:** e174.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. **40:** 897–903.

Watanabe, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature **453:** 539–544.

Weil, C., and Martienssen, R. (2008). Epigenetic interactions between transposons and genes: Lessons from plants. Curr. Opin. Genet. Dev. **18:** 188–192.

Woodhouse, M.R., Freeling, M., and Lisch, D. (2006). Initiation, establishment, and maintenance of heritable *MuDR* transposon silencing in maize are mediated by distinct factors. PLoS Biol. **4(10):** e339 (online). doi:10.1371/journal.pbio.0040339.

Yu, J., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296:** 79–92.

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

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, J.R., and Ecker, J.R. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis.* Cell **126:** 1189–1201.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. **9:** R137.

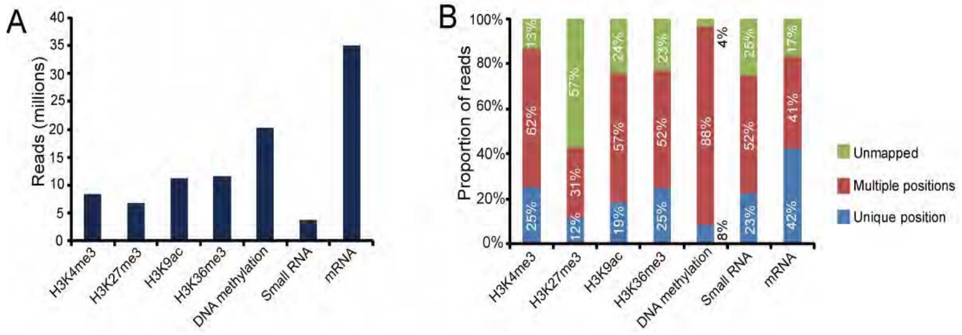
Zhao, X., and Wang, H. (2007). LTR-FINDER: An efficient tool for the prediction of full-length LTR retrotransposons (2007). Nucleic Acids Res. **35:** W265–W268.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. **39:** 61–69.

Genome-Wide and Organ-Specific Landscapes of Epigenetic Modifications and Their Relationships to mRNA and Small RNA Transcriptomes in Maize

Xiangfeng Wang, Axel A. Elling, Xueyong Li, Ning Li, Zhiyu Peng, Guangming He, Hui Sun, Yijun Qi, X. Shirley Liu and Xing Wang Deng

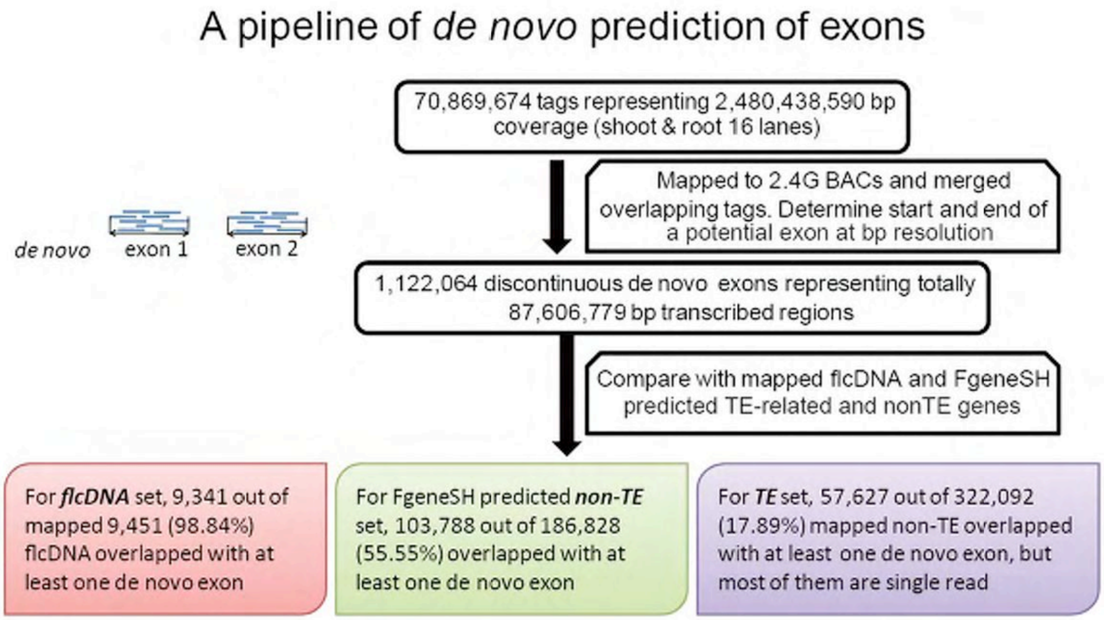
PLANT CELL 2009;21;1053-1069; originally published online Apr 17, 2009;
DOI: 10.1105/tpc.109.065714



Supplemental Figure 1. Sequencing and mapping of mRNA transcripts, H3K4me3, H3K9ac, H3K36me3, H3K27me3, DNA methylation and smRNAs in maize roots.

(A) Counts of raw reads from Illumina/Solexa 1G Sequencing.

(B) Proportions of unmapped and mapped reads with unique and multiple locations.



A pipeline of *de novo* prediction of exons

70,869,674 tags representing 2,480,438,590 bp coverage (shoot & root 16 lanes)

Mapped to 2.4G BACs and merged overlapping tags. Determine start and end of a potential exon at bp resolution

1,122,064 discontinuous de novo exons representing totally 87,606,779 bp transcribed regions

Compare with mapped flcDNA and FgeneSH predicted TE-related and nonTE genes

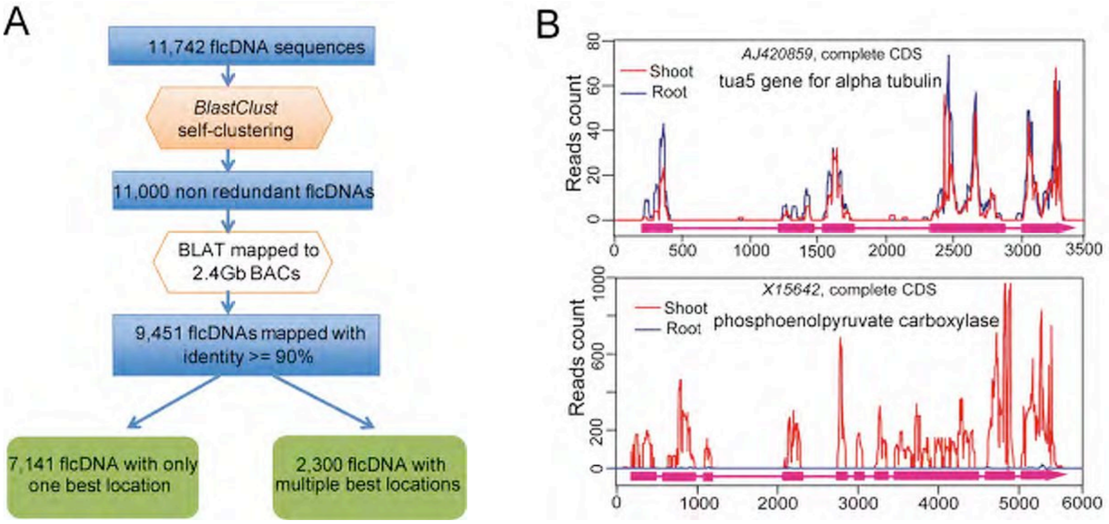
For ***flcDNA*** set, 9,341 out of mapped 9,451 (98.84%) flcDNA overlapped with at least one de novo exon

For FgeneSH predicted ***non-TE*** set, 103,788 out of 186,828 (55.55%) overlapped with at least one de novo exon

For ***TE*** set, 57,627 out of 322,092 (17.89%) mapped non-TE overlapped with at least one de novo exon, but most of them are single read

Supplemental Figure 2. Detection of individual transcribed exons by *de novo* scanning of mRNA-seq reads across the maize genome sequence.

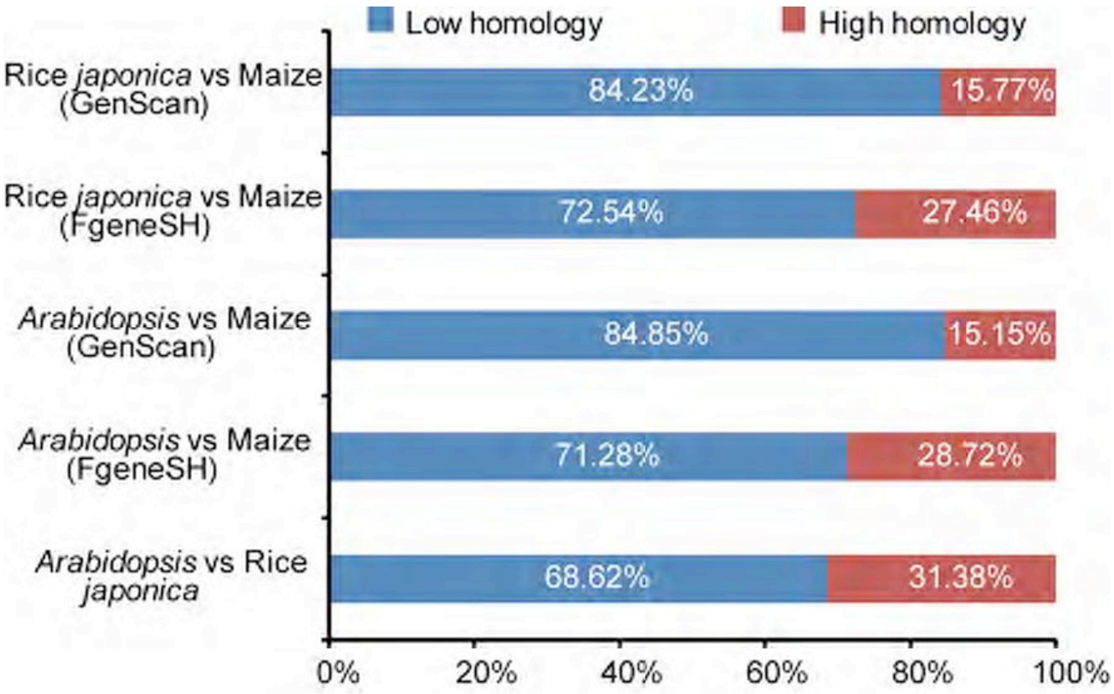
A *de novo* exon is represented by a cluster of piled-up mRNA-seq reads. We performed a *de novo* scanning of the mRNA-seq reads across the genome to identify the bp-resolution margins for each exon. By adding up all *de nov*o exons, we found that the total maize transcription activity mapped to 87.6 Mb. We then aligned the *de novo* exons with flcDNAs, TEs and non-TE genes to validate the gene prediction at gene, exon, and base-level.



Supplemental Figure 3. Mapping of 11,742 maize flcDNAs to the maize genome sequence.

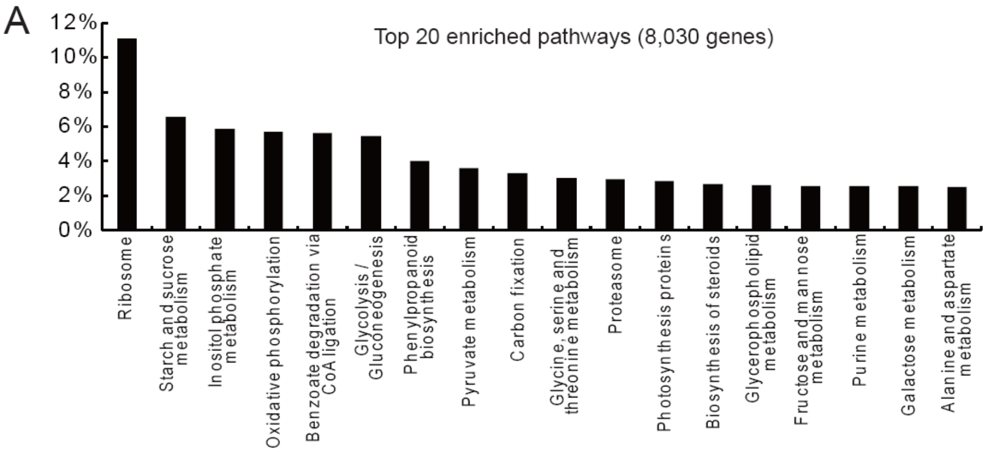
(A) Procedure for mapping flcDNAs onto 2.4 Gb of maize BACs.

(B) Two examples of flcDNA genes matched with mRNA-seq reads. mRNA-seq reads are highly consistent with gene structure, indicating mRNA-seq data can be used for experimentally validating the genome annotation. AJ420859 displayed equal expression levels in shoots and roots, while the second gene, X15642, showed differential expression.



Supplemental Figure 4. Comparison of gene homology in maize, rice and *Arabidopsis*.

Rice and *Arabidopsis* protein sequences were downloaded from TIGR (version 5) and TAIR (version 8), respectively. We compared the translated genes predicted by GenScan and FgeneSH the protein sequences from these two plant species. Since there are nearly 200,000 non-TE genes predicted in maize, and only ~30,000 and ~50,000 genes in *Arabidopsis* and rice, respectively, the percentages in the bars only show how many homologous genes are found in *Arabidopsis* or rice.



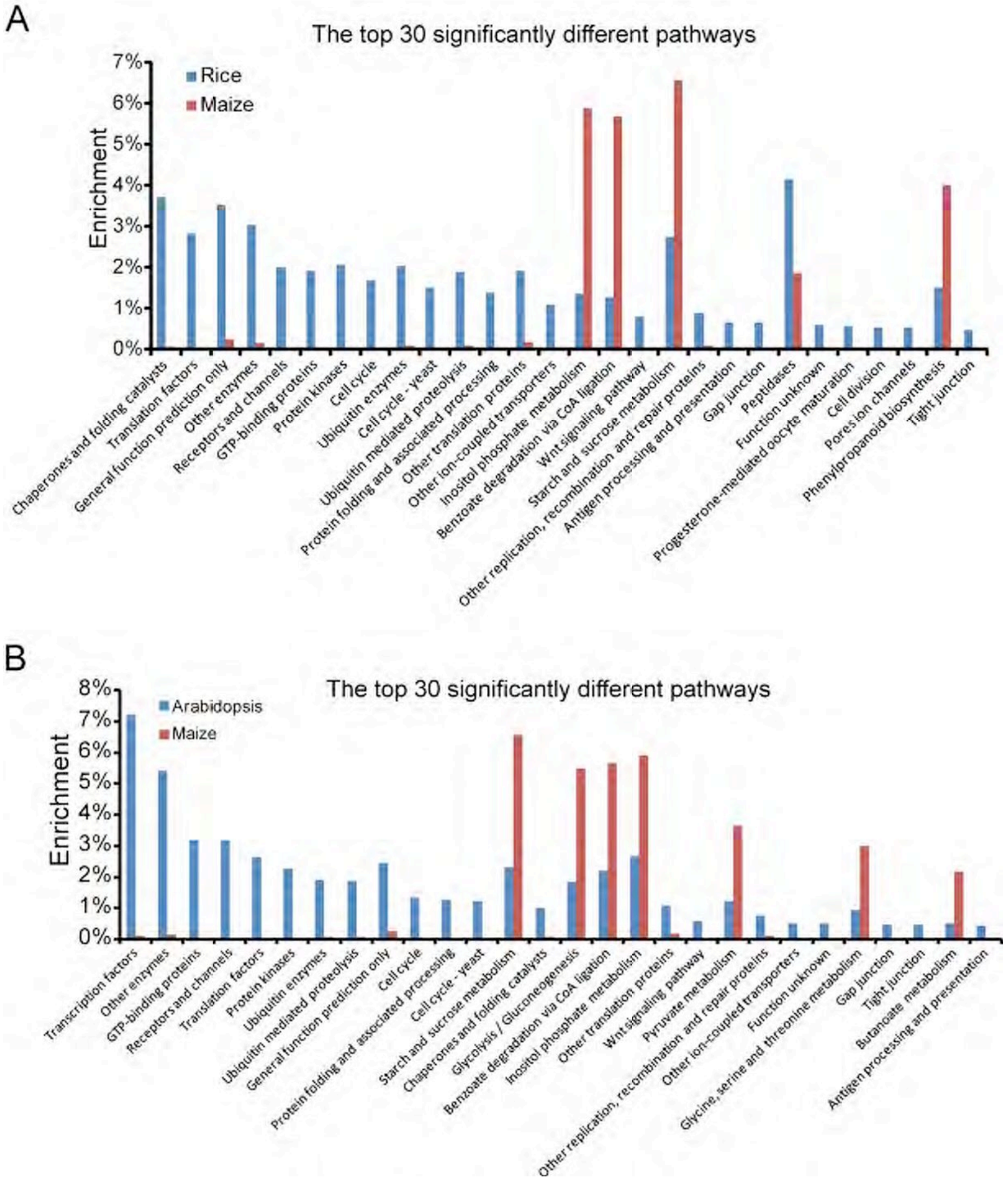
B

The remaining pathways (13,523 genes)

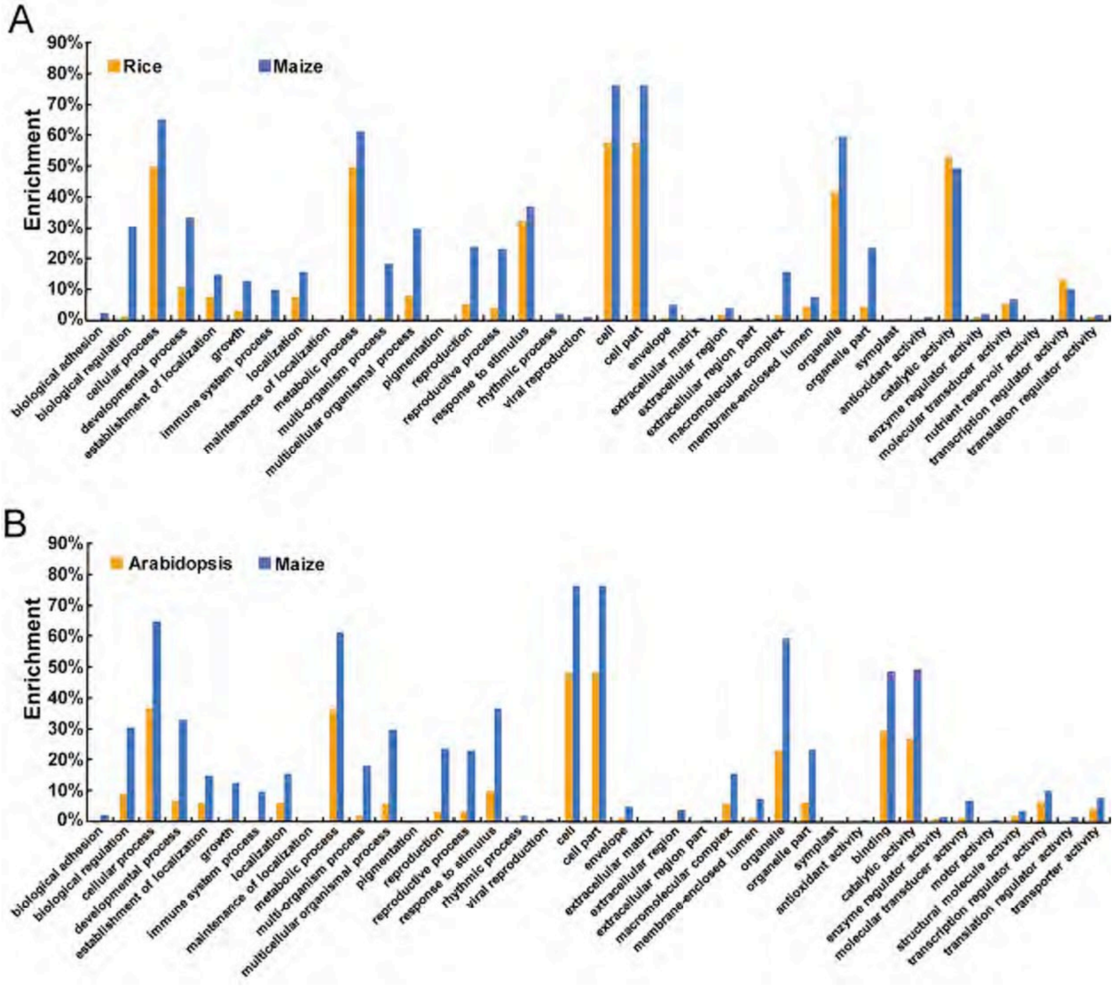
Pathways	#	%	Pathways	#	%	Pathways	#	%
Aminoacyl-tRNA biosynthesis	266	2.51%	beta-Alanine metabolism	111	1.05%	Synthesis and degradation of ketone bodies	39	0.37%
Nitrogen metabolism	255	2.41%	Methionine metabolism	110	1.04%	Type II diabetes mellitus	38	0.36%
Citrate cycle (TCA cycle)	248	2.34%	Ether lipid metabolism	110	1.04%	Glycosaminoglycan degradation	36	0.34%
Photosynthesis	230	2.17%	Ascorbate and aldarate metabolism	110	1.04%	N-Glycan degradation	35	0.33%
Butanoate metabolism	228	2.15%	Metabolism of xenobiotics by cytochrome P450	109	1.03%	Inositol metabolism	34	0.32%
Glutamate metabolism	217	2.05%	Two-component system	108	1.02%	Glycosphingolipid biosynthesis - ganglioseries	34	0.32%
Tyrosine metabolism	216	2.04%	Lysine biosynthesis	107	1.01%	Glycosphingolipid biosynthesis - neo-lactoseries	31	0.29%
Phosphatidylinositol signaling system	213	2.01%	DNA polymerase	107	1.01%	Bisphenol A degradation	29	0.27%
Pentose phosphate pathway	212	2.00%	Regulation of actin cytoskeleton	86	0.81%	Novobiocin biosynthesis	28	0.26%
Valine, leucine and isoleucine degradation	212	2.00%	Terpenoid biosynthesis	83	0.78%	Glycosphingolipid biosynthesis - lactoseries	28	0.26%
Phenylalanine, tyrosine and tryptophan biosynthesis	209	1.97%	Cysteine metabolism	82	0.77%	Taurine and hypotaurine metabolism	27	0.25%
Cyanoamino acid metabolism	209	1.97%	Folate biosynthesis	82	0.77%	1,2-Dichloroethane degradation	26	0.25%
RNA polymerase	206	1.95%	PPAR signaling pathway	80	0.76%	General function prediction only	26	0.25%
Limonene and pinene degradation	199	1.88%	Glycosphingolipid biosynthesis - globoseries	80	0.76%	Cytochrome P450	25	0.24%
Phenylalanine metabolism	198	1.87%	ATPases	79	0.75%	Vitamin B6 metabolism	25	0.24%
Methane metabolism	197	1.86%	Protein export	79	0.75%	Biotin metabolism	25	0.24%
Peptidases	195	1.84%	Sulfur metabolism	77	0.73%	Prion disease	23	0.22%
Glyoxylate and dicarboxylate metabolism	191	1.80%	1- and 2-Methylnaphthalene degradation	76	0.72%	Amyotrophic lateral sclerosis (ALS)	23	0.22%
Pyrimidine metabolism	188	1.78%	Photosynthesis - antenna proteins	75	0.71%	1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) degradation	23	0.22%
Reductive carboxylate cycle (CO2 fixation)	188	1.78%	MAPK signaling pathway	71	0.67%	Diterpenoid biosynthesis	22	0.21%
Naphthalene and anthracene degradation	188	1.78%	One carbon pool by folate	71	0.67%	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	22	0.21%
Valine, leucine and isoleucine biosynthesis	186	1.76%	3-Chloroacrylic acid degradation	65	0.61%	Lipopolysaccharide biosynthesis	20	0.19%
Fatty acid metabolism	185	1.75%	Polyunsaturated fatty acid biosynthesis	65	0.61%	Styrene degradation	19	0.18%
Carotenoid biosynthesis	181	1.71%	Alkaloid biosynthesis II	64	0.60%	Tetrachloroethene degradation	19	0.18%
Tryptophan metabolism	177	1.67%	Fatty acid biosynthesis	61	0.58%	Other translation proteins	18	0.17%
Glycerolipid metabolism	177	1.67%	Linoleic acid metabolism	61	0.58%	Type I diabetes mellitus	18	0.17%
Insulin signaling pathway	174	1.64%	ABC transporters	59	0.56%	Ethylbenzene degradation	17	0.16%
Glutathione metabolism	171	1.61%	Calcium signaling pathway	59	0.56%	mTOR signaling pathway	17	0.16%
Urea cycle and metabolism of amino groups	168	1.59%	Transporters	59	0.56%	Fatty acid elongation in mitochondria	16	0.15%
SNAREs	167	1.58%	Regulation of autophagy	58	0.55%	Other enzymes	15	0.14%
SNARE interactions in vesicular transport	167	1.58%	alpha-Linolenic acid metabolism	56	0.53%	Biosynthesis of ansamycins	14	0.13%
Basal transcription factors	161	1.52%	Benzoate degradation via hydroxylation	56	0.53%	C5-Branched dibasic acid metabolism	14	0.13%
Lysine degradation	160	1.51%	Melanogenesis	55	0.52%	Transcription factors	12	0.11%
Flavonoid biosynthesis	154	1.45%	Glioma	55	0.52%	Renal cell carcinoma	12	0.11%
N-Glycan biosynthesis	152	1.44%	Olfactory transduction	55	0.52%	Riboflavin metabolism	12	0.11%
Epithelial cell signaling in Helicobacter pylori infection	146	1.38%	Huntington's disease	55	0.52%	VEGF signaling pathway	11	0.10%
Propanoate metabolism	143	1.35%	Caprolactam degradation	55	0.52%	Replication complex	11	0.10%
GnRH signaling pathway	140	1.32%	Long-term potentiation	55	0.52%	Fc epsilon RI signaling pathway	11	0.10%
Sphingolipid metabolism	139	1.31%	Androgen and estrogen metabolism	54	0.51%	Long-term depression	11	0.10%
Aminosugars metabolism	137	1.29%	Aminophosphonate metabolism	54	0.51%	D-Glutamine and D-glutamate metabolism	10	0.09%
Porphyrin and chlorophyll metabolism	136	1.28%	Cellular antigens	51	0.48%	Thiamine metabolism	10	0.09%
Selenoamino acid metabolism	134	1.27%	Nicotinate and nicotinamide metabolism	47	0.44%	Lipoic acid metabolism	10	0.09%
gamma-Hexachlorocyclohexane degradation	134	1.27%	Biosynthesis of vancomycin group antibiotics	45	0.42%	Other replication, recombination and repair proteins	9	0.08%
Pentose and glucuronate interconversions	134	1.27%	Arachidonic acid metabolism	45	0.42%	Ubiquitin mediated proteolysis	8	0.08%
Arginine and proline metabolism	131	1.24%	Polyketide sugar unit biosynthesis	45	0.42%	Ubiquitin enzymes	8	0.08%
Bile acid biosynthesis	127	1.20%	Alkaloid biosynthesis I	44	0.42%	Chaperones and folding catalysts	7	0.07%
Cytoskeleton proteins	126	1.19%	Other transporters	42	0.40%	Caffeine metabolism	7	0.07%
Pantothenate and CoA biosynthesis	121	1.14%	Adipocytokine signaling pathway	42	0.40%	Protein kinases	3	0.03%
Histidine metabolism	121	1.14%	Ubiquinone biosynthesis	41	0.39%	Hematopoietic cell lineage	2	0.02%
Nucleotide sugars metabolism	119	1.12%	Brassinosteroid biosynthesis	41	0.39%	Renin - angiotensin system	2	0.02%
Streptomycin biosynthesis	116	1.10%	Peptidoglycan biosynthesis	40	0.38%	Type II secretion system	2	0.02%
Glycosyltransferases	112	1.06%	High-mannose type N-glycan biosynthesis	39	0.37%	2,4-Dichlorobenzoate degradation	1	0.01%

Supplemental Figure 5. Pathway annotation of maize genes.

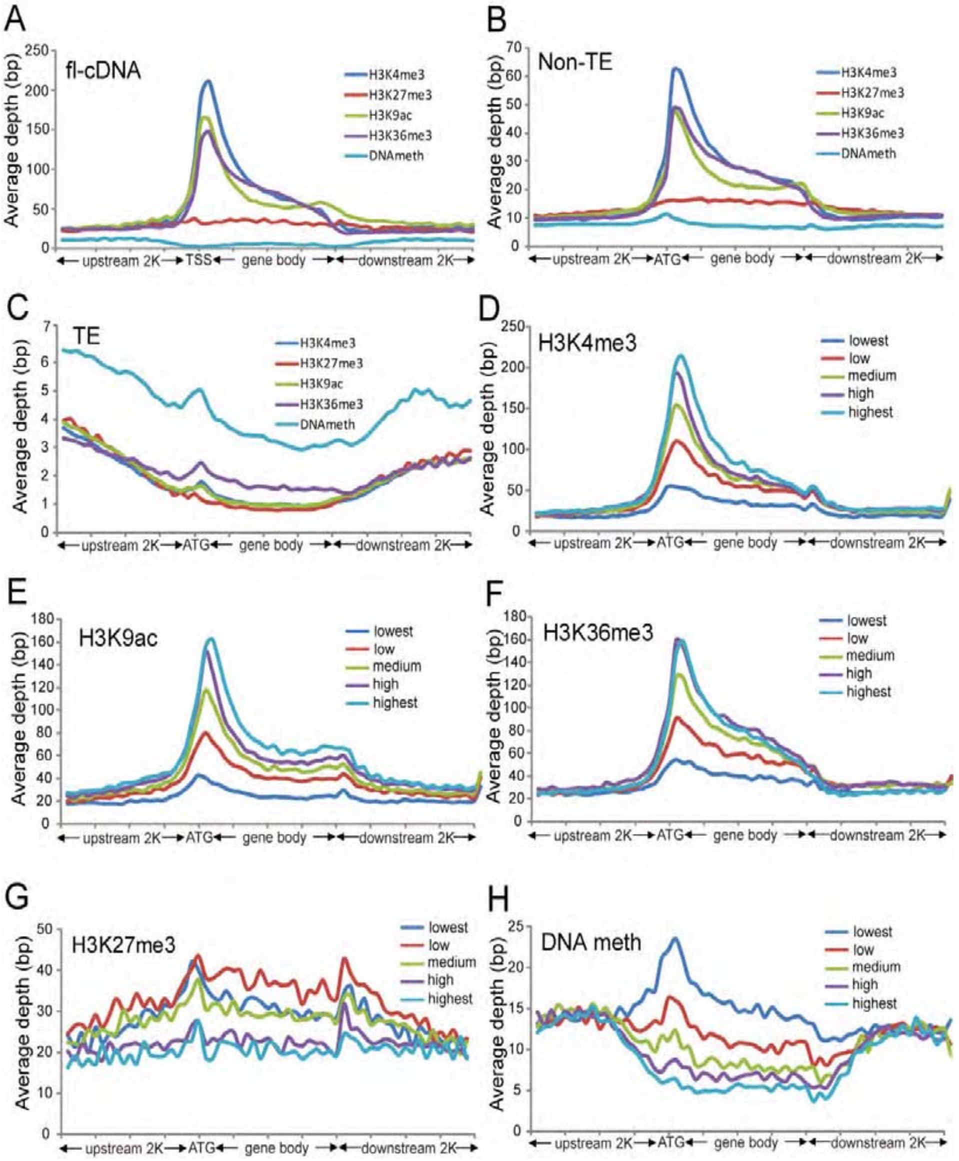
Using KOBAS, we assigned ~20,000 genes into pathways, (A) and (B) show the enrichment of genes in 175 pathways.



Supplemental Figure 6. Comparison of pathway enrichment in maize vs rice and maize vs *Arabidopsis*.

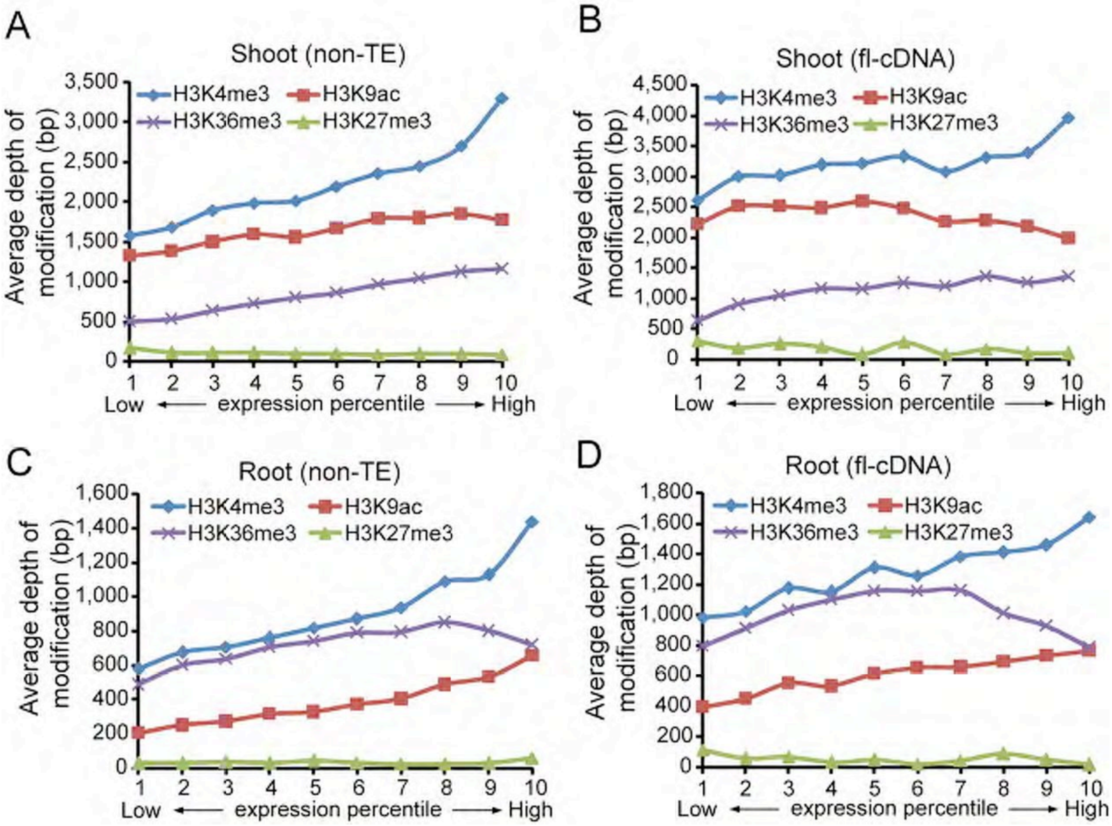


Supplemental Figure 7. Comparison of Gene Ontology enrichment between maize/rice and maize/*Arabidopsis*.



Supplemental Figure 8. Distribution of epigenetic patterns within maize genes in roots.

(A), (B) and (C) Distribution of H3K4me3, H3K27me3, H3K9ac, H3K36me3 and DNA methylation levels within flcDNAs, predicted TE-related and non-TE genes aligned from transcription start sites (TSS) and ATG, respectively. The y axis shows the average depth, which is the frequency of piled-up reads at each base divided by the size of the bin. The x axis represents the aligned genes that were binned into 40 equal portions including 2 Kb up- and down-stream regions.
(D) to (H) Distribution of H3K4me3, H3K27me3, H3K9ac, H3K36me3 and DNA methylation within five groups of genes with different expression levels summarized from validated non-TE genes.

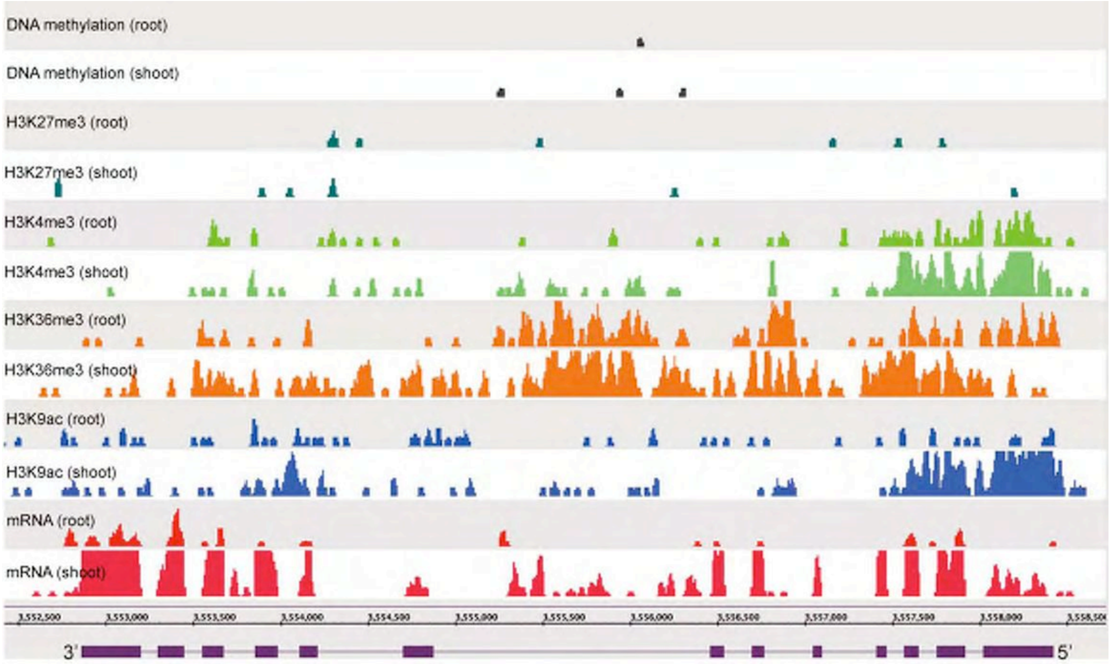


Supplemental Figure 9. Effect of modification levels on gene expression.

The *x* axis shows the expression percentiles. Genes were sorted into 10 equal percentiles based on increasing expression intensity from mRNA-seq data. The *y* axis shows the average modification level on the genes in each percentile. The modification level was represented by the sequencing depth of reads per base.

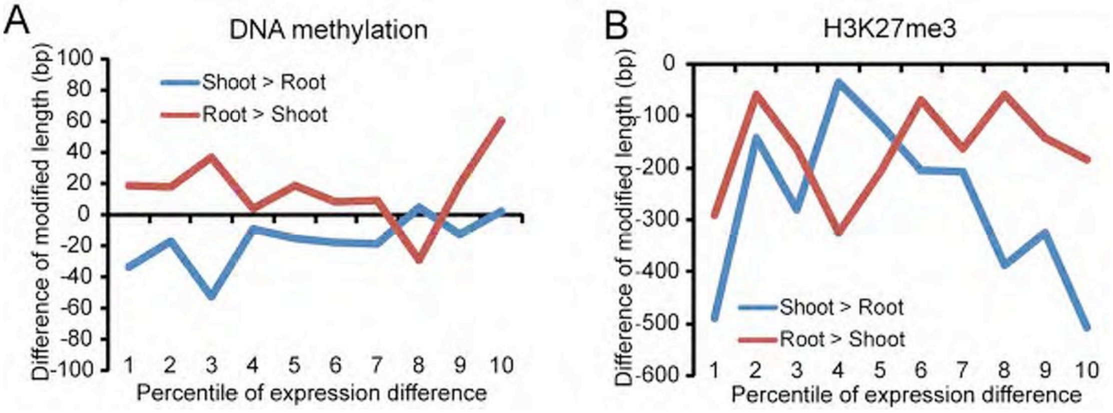
(A) and (B) Analysis using non-TE genes and flcDNA-supported genes in shoots.

(C) and (D) Analysis using non-TE genes and flcDNA-supported genes in roots.

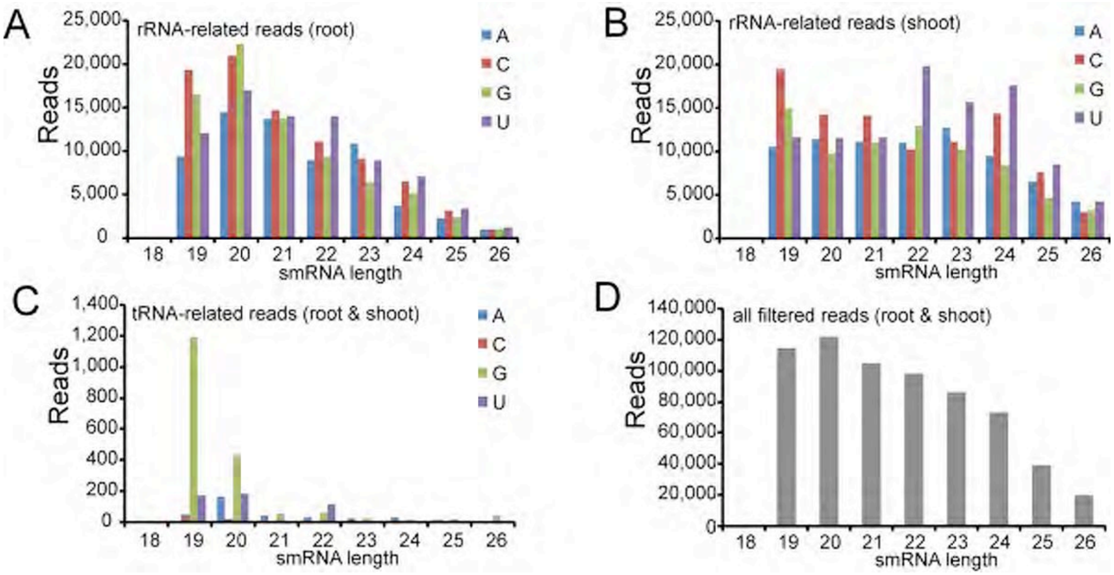


Supplemental Figure 10. A differentially expressed gene shows a different epigenetic pattern.

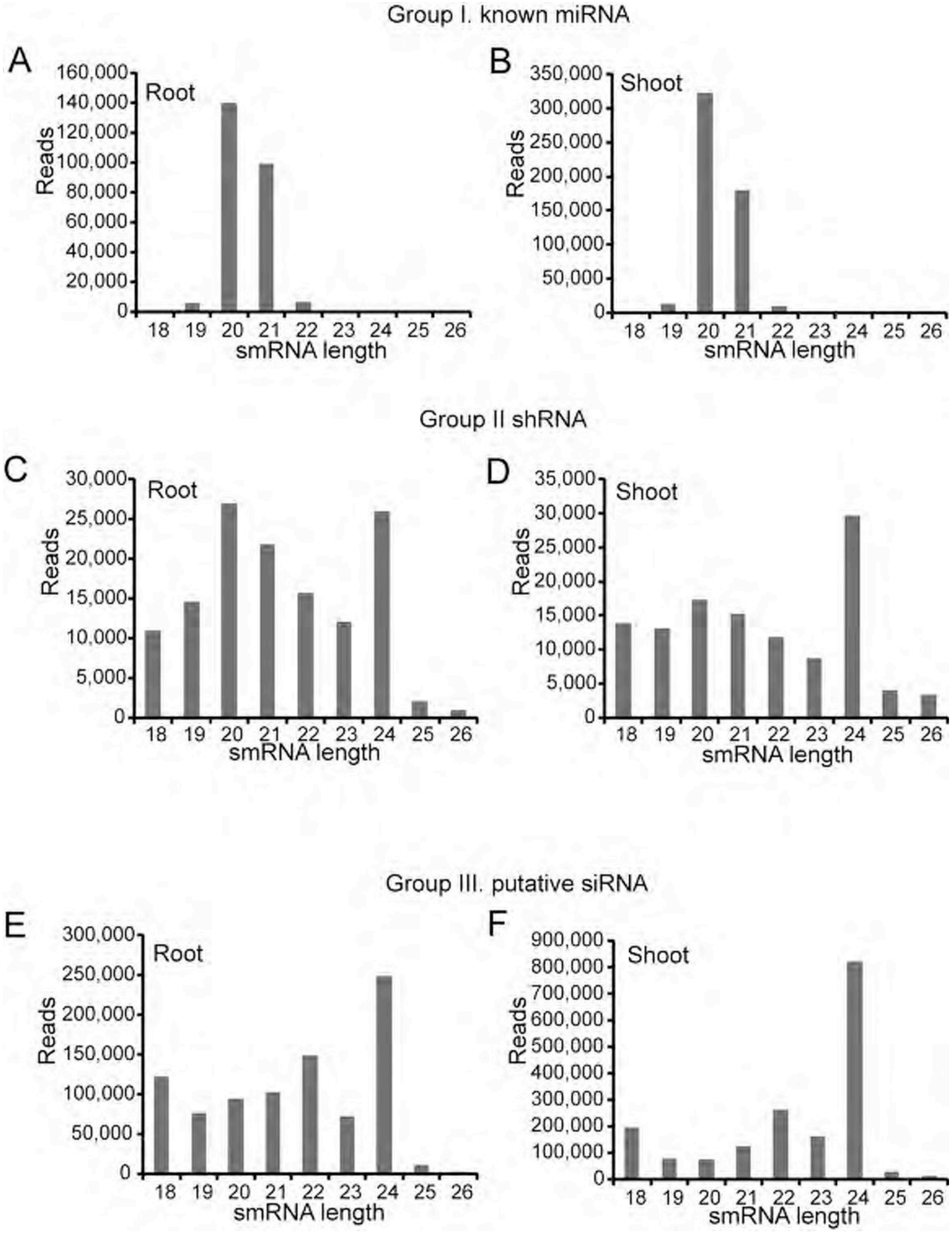
One flcDNA gene mapped within the 20Mb region displayed a higher expression level in shoots than in roots. Upon examining the three activating marks on this gene, we found that H3K4me3 and H3K36me3 were partially attenuated but that H3K9ac in roots was significantly reduced.



Supplemental Figure 11. Correlation of differential modifications of H3K27me3 and DNA methylation with differential gene expression in shoots and roots. The *y* axes show differences in the modification levels of H3K27me3 and DNA methylation in shoot and roots. The *x* axes show the differences in expression levels.



Supplemental Figure 12. Length distributions of removed smRNA reads matched with tRNA and rRNA sequences.

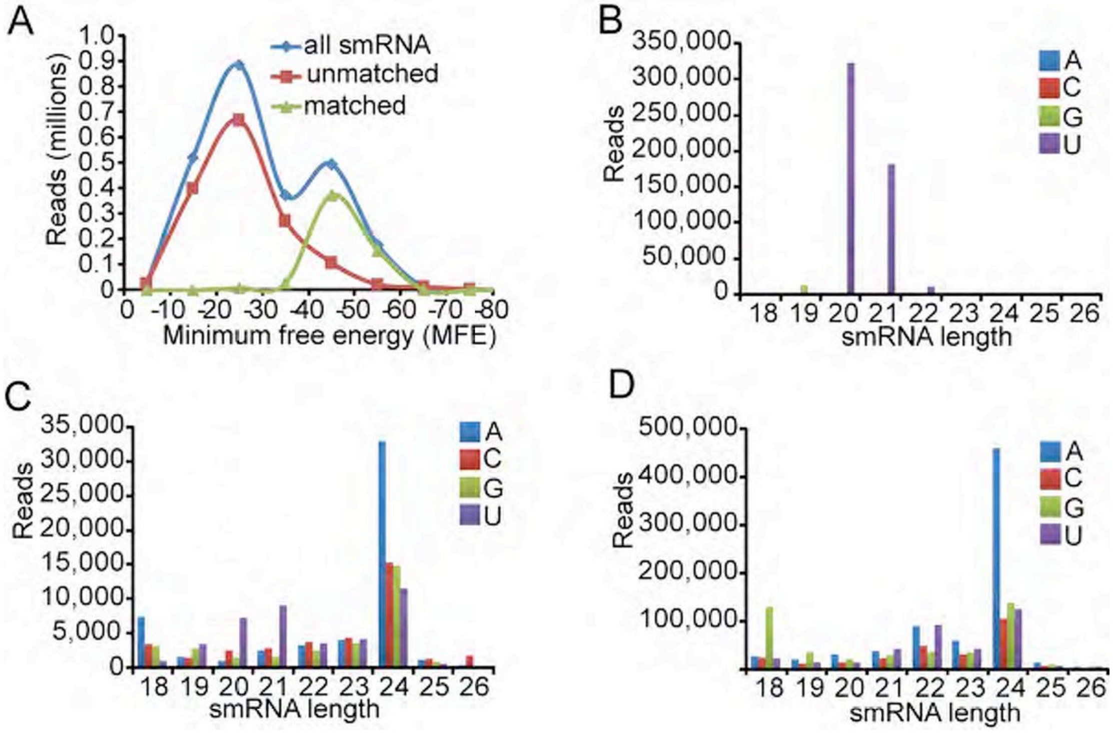


Supplemental Figure 13. Length distribution of three groups of smRNAs.

(A) and (B) Length distribution of known miRNAs (Group I) in shoots and roots.

(C) and (D) Length distribution of small hairpin RNA (Group II) in shoots and roots.

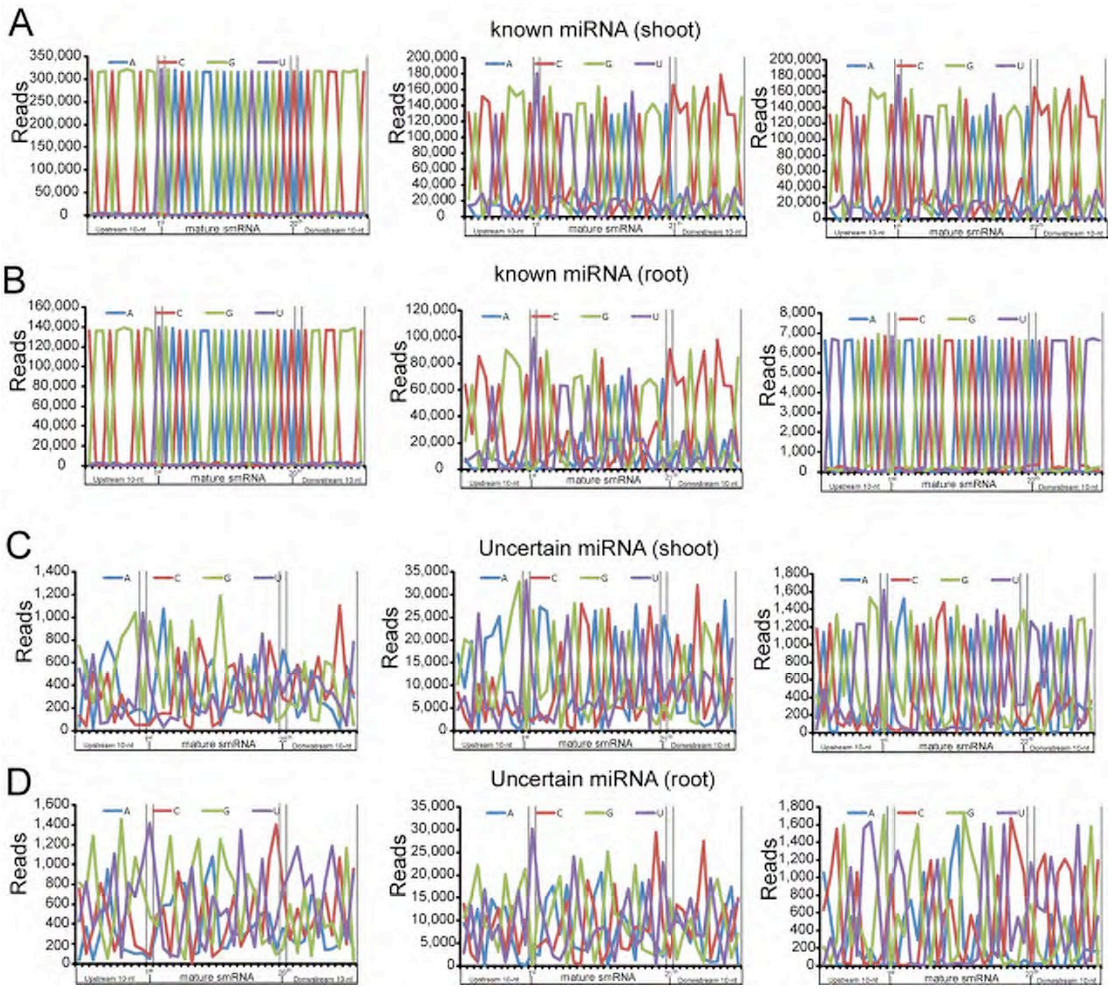
(F) and (G) Length distribution of putative siRNAs (Group III) in shoots and roots.



Supplemental Figure 14. Classification of smRNA population in shoots.

(A) Distribution of smRNAs within different MFE bins.

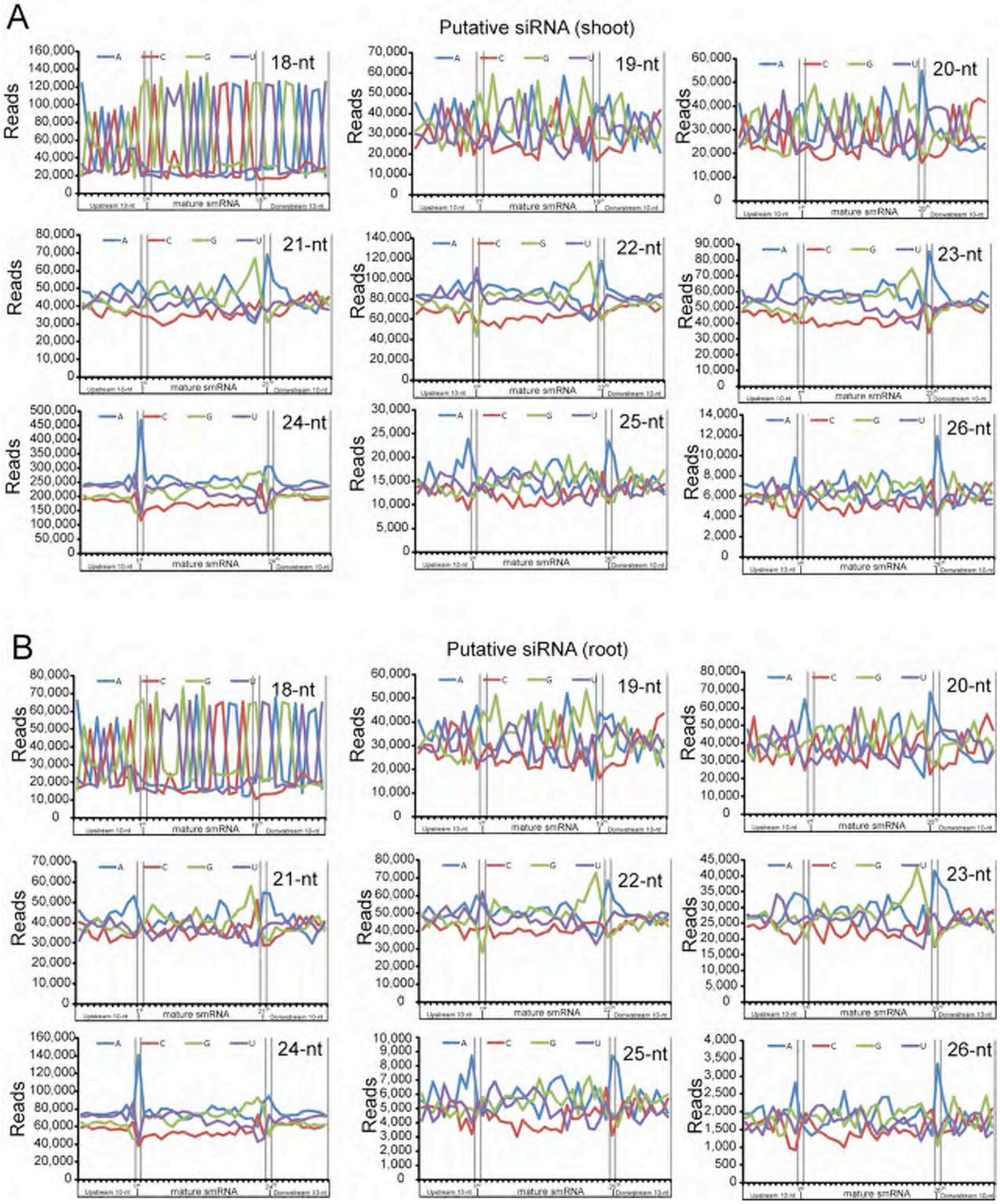
(B) , (C) and (D) Length distributions of known miRNA, shRNAs and putative siRNAs with different 5’ terminal nucleotides.



Supplemental Figure 15. Nucleotide composition of known miRNAs.

(A) and (B) Nucleotide composition of mature known miRNAs of 20 to 22 nt in shoots and roots.

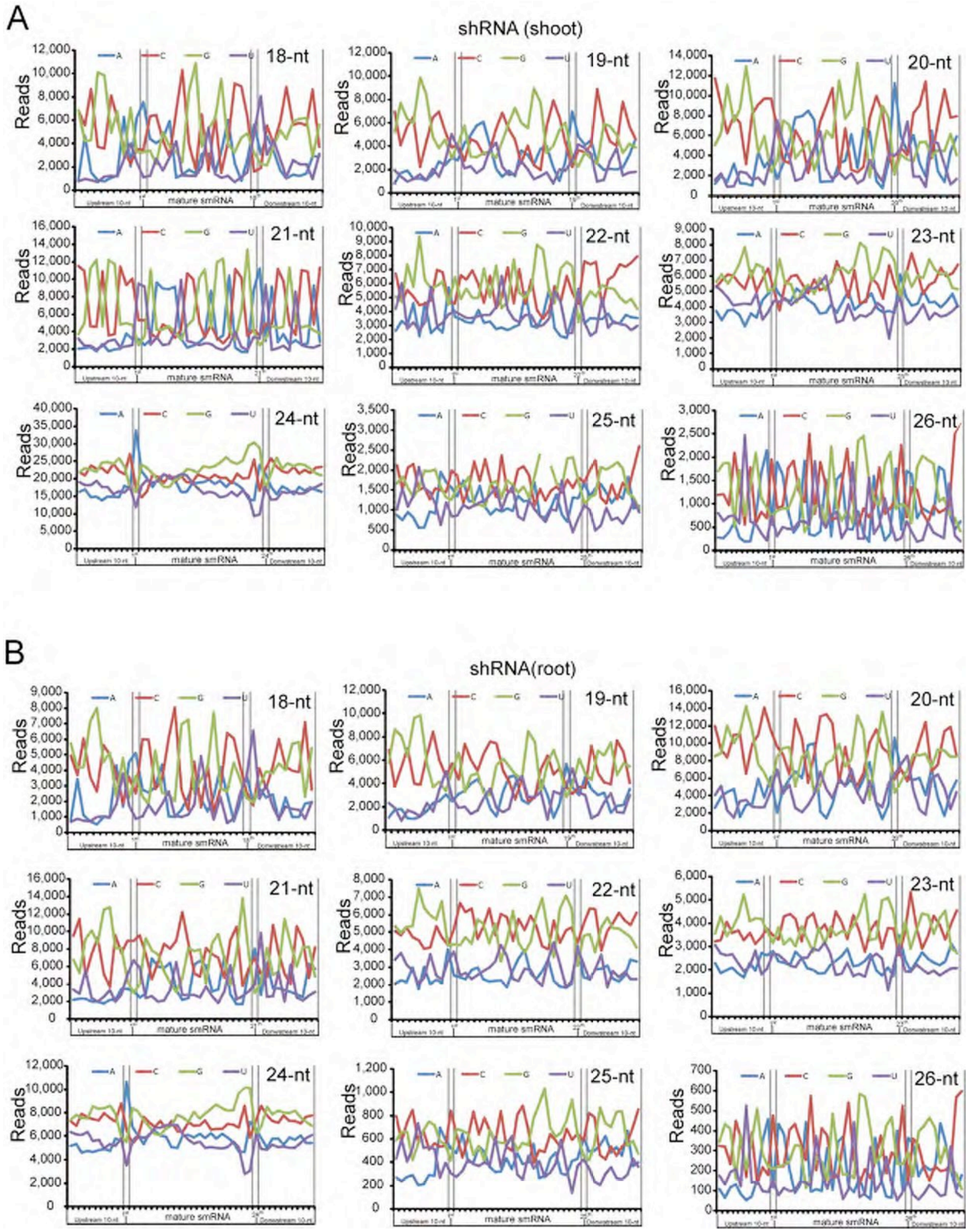
(C) and (D) Nucleotide composition of uncertain miRNAs of 20 to 22 nt in shoots and roots. Uncertain miRNAs were defined as smRNAs that did not match known miRNAs in miRBase, but whose precursor sequences failed to form a hairpin structure. Additionally, uncertain miRNAs had fewer copies than known miRNAs. Therefore, we merged uncertain miRNAs with the putative siRNA group.



Supplemental Figure 16. Nucleotide composition of putative siRNAs.

(A) Nucleotide composition of putative siRNAs from 18 nt to 26 nt in shoots.

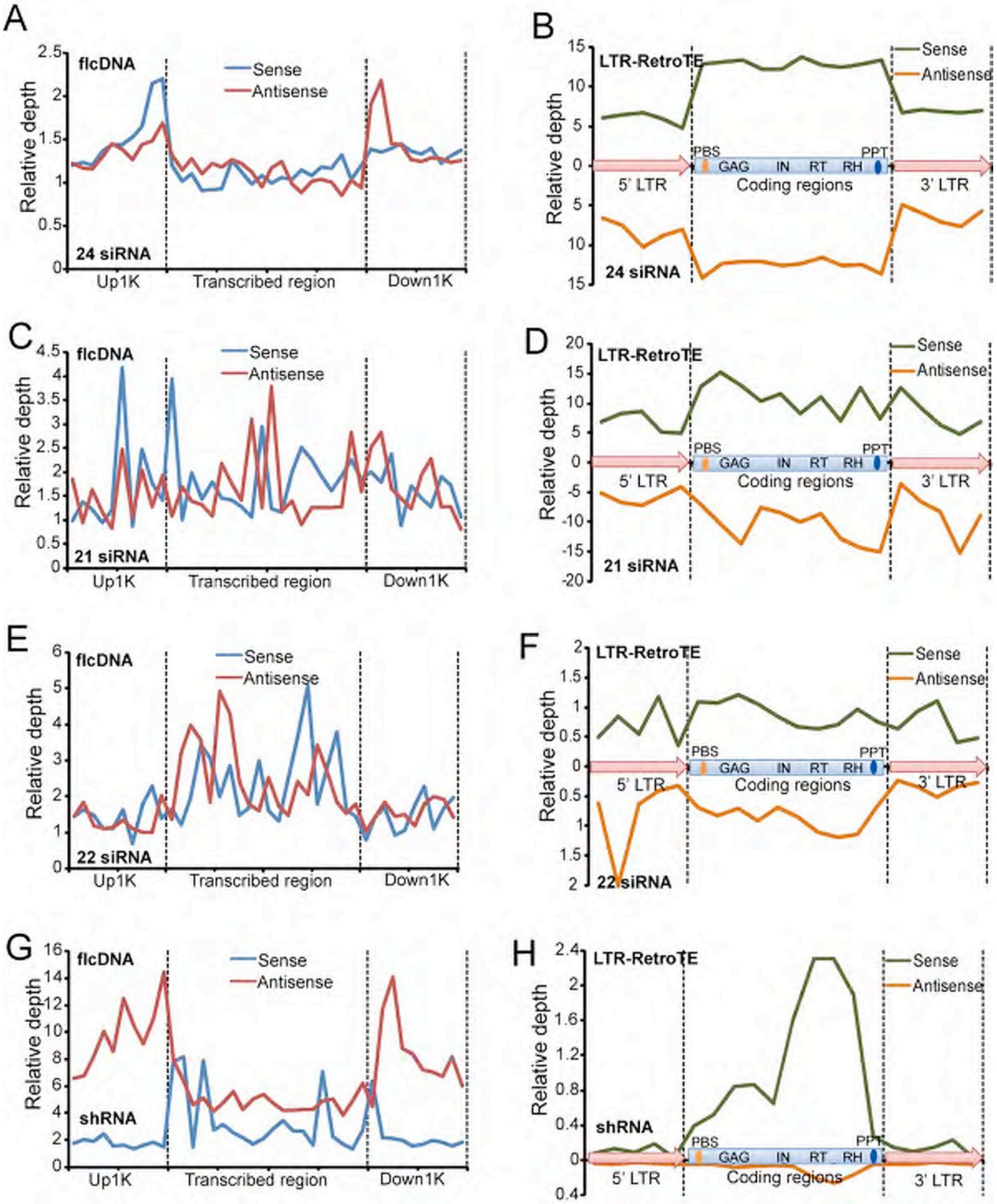
(B) Nucleotide composition of putative siRNAs from 18 nt to 26 nt in roots.



Supplemental Figure 17. Nucleotide composition of shRNAs.

(C) Nucleotide composition of shRNAs from 18 nt to 26 nt in shoots.

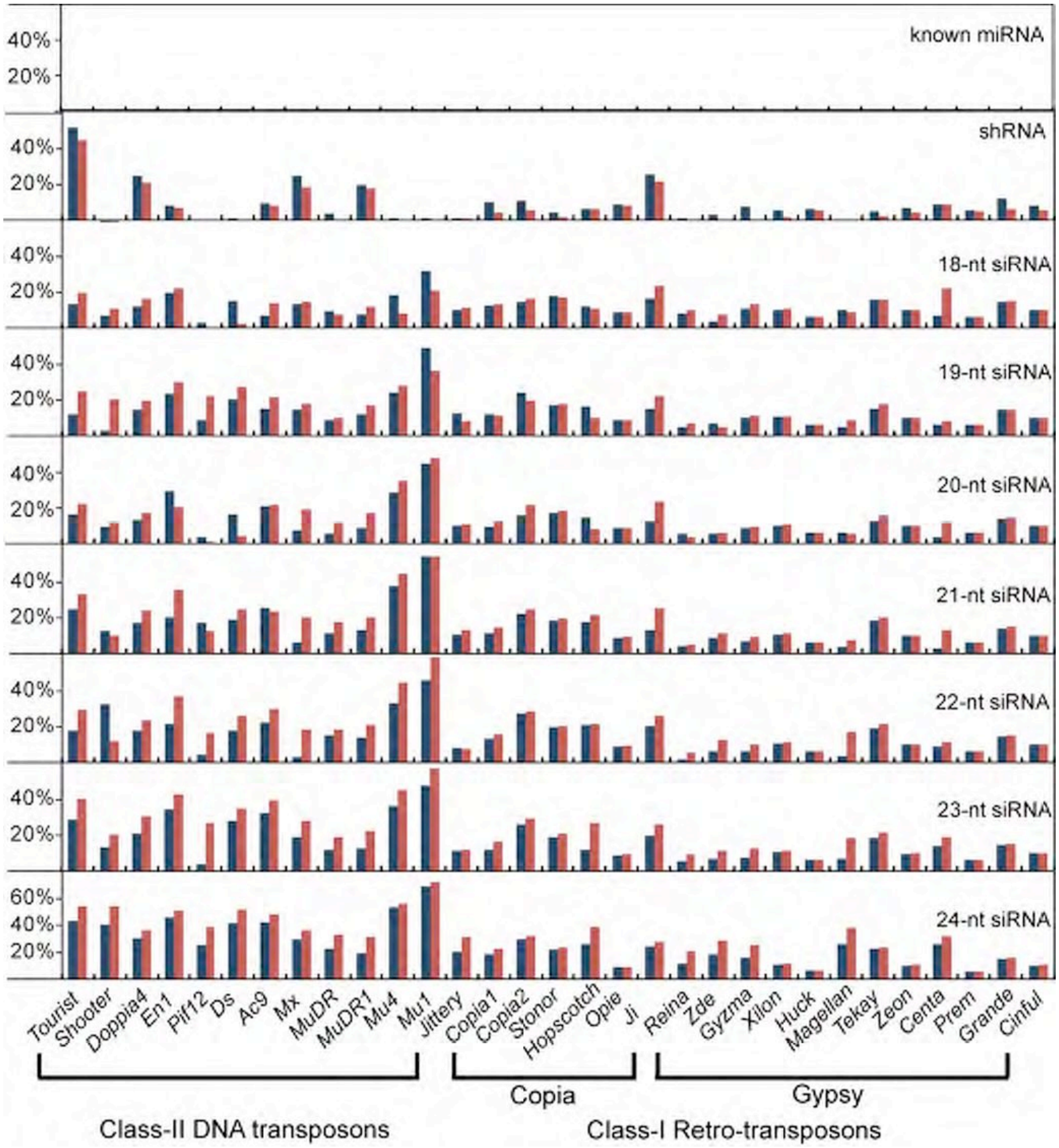
(D) Nucleotide composition of shRNAs from 18 nt to 26 nt in roots.



Supplemental Figure 18. Origin and target sites on genes and LTR-TEs for different classes of putative siRNAs in roots.

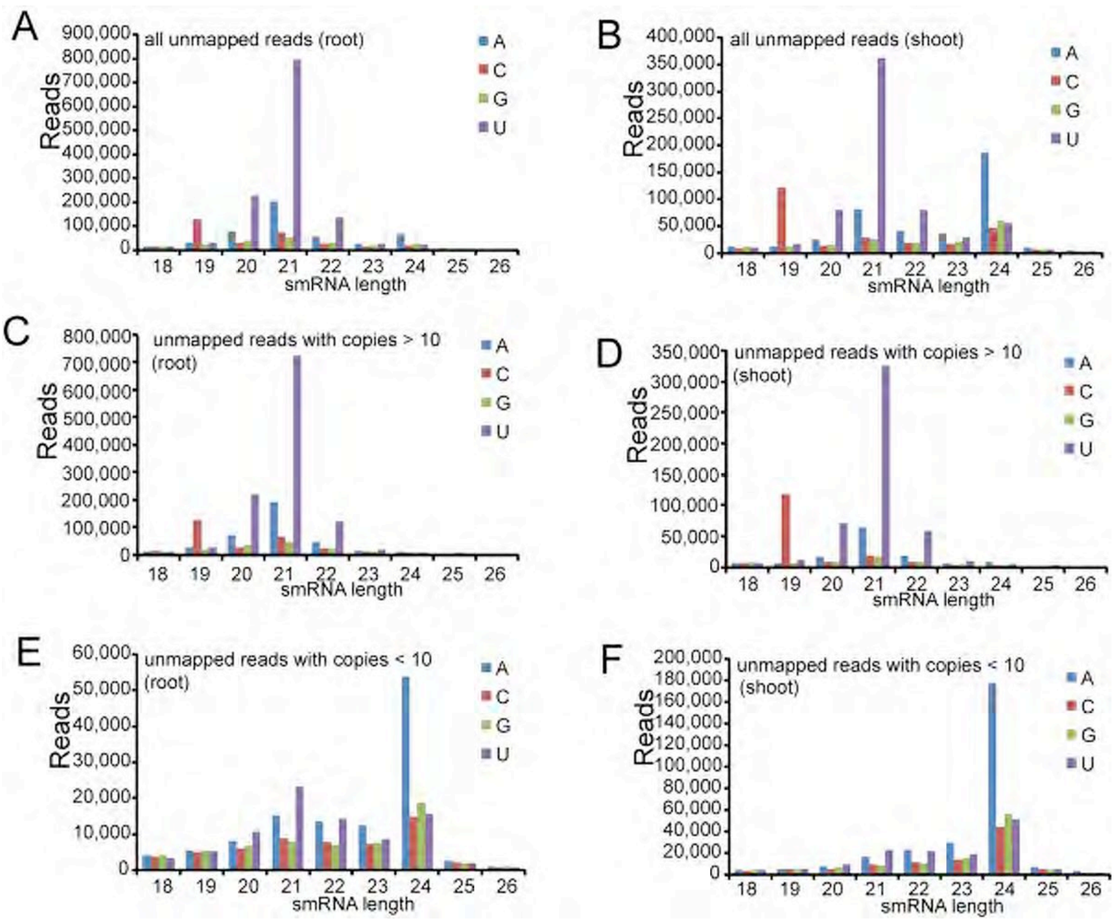
(A), (C), (E) and (G) 24, 21 , and 22 nt siRNAs and shRNAs on flcDNA genes show significant strand bias at different positions in originating and targeting strands.

(B), (D), (F) and (H) 24, 21, and 22 nt siRNAs and shRNAs on LTR-TEs. Calculation of relative depth and *de novo* identification of LTR-TEs is described in Supplemental Methods (see below).



Supplemental Figure 19. Percentages of different types of repeats generating smRNAs in different lengths.

We identified repetitive regions by RepeatMasker, and determined the percentage of smRNAs overlapping with each repetitive region.



Supplemental Figure 20. Length distribution of unmapped smRNAs classified by 5’ terminal nucleotides.

(A) and (B) Most of the 21 nt reads that could be perfectly mapped to the genome were enriched for 5’ terminal U, indicating that most of them were miRNAs. We then used the copy frequency of each unique smRNA to distinguish miRNAs from siRNAs.

(C) and (D) 21 nt smRNAs with copy frequency >=10 were enriched for 5’ terminal U indicating most of them were miRNAs.

(E) and (F) 24 nt smRNAs with copy frequency <10 were enriched for 5’ terminal A indicating most of them were siRNAs.

SUPPLEMENTAL TABLES

Supplemental Table 1. Solexa Sequencing and mapping statistics

Root	lanes	total reads	Mapped reads	Reads mapped to unique position
H3K4me3	2	8,427,836	7,311,810	2,113,484
H3k27me3	2	6,731,390	2,873,390	773,083
H3K9ac	2	11,220,867	8,509,462	2,080,436
H3K36me3	2	11,605,504	8,933,297	2,861,200
DNA methylation	5	20,188,414	19,457,976	1,685,177
Small RNA	1	3,648,212	2,728,364	813,100
Transcriptome	8	34,949,694	29,097,596	14,755,456

Shoot	lanes	total reads	Mapped reads	Reads mapped to unique position
H3K4me3	2	8,670,391	6,141,591	2,681,539
H3k27me3	2	8,662,744	3,613,140	1,208,411
H3K9ac	2	11,893,695	10,207,060	4,059,050
H3K36me3	2	10,943,618	9,650,552	3,257,569
DNA methylation	5	19,181,681	18,014,718	1,616,496
Small RNA	1	4,045,453	3,590,029	866,972
Transcriptome	8	35,919,980	30,163,104	15,946,441

Supplemental Table 2. Validation statistics of FgeneSH predicted genes

Gene Type	Validated rate		Validated genes	
	shoot	root	shoot	root
EVI	47.58%	46.17%	14,262	13,837
PRO	37.08%	37.17%	27,083	27,153
UPRO	3.77%	3.81%	3,156	3,197
TOTAL	23.82%	23.65%	44,501	44,187

Supplemental Table 3.

Number of smRNAs hitting known miRNAs in different minimum free energy (MFE) ranges.

	Shoot		Root	
	Total # of putative	# of known	Total # of putative	# of known
MFE	miRNA	miRNA	miRNA	miRNA
0~-10	0	0	0	0
-10~-20	80	24	35	17
-20~-30	8,272	156	17,815	118
-30~-40	25,118	131	14,182	101
-40~-50	372,755	94	165,786	77
-50~-60	154,206	61	86,719	49
-60~-70	1	1	0	0
-70~-80	0	0	0	0
< -80	0	0	0	0

Supplemental Table 4.

Solexa sequencing reads for miRNAs.

miRNA family	length	miRNA sequence	root	shoot	ears	*mop1-1*
miR156a/b/c/d/e/f/g/h/i	20	UGACAGAAGAGAGUGAGCAC	406,699	315,388	98	1,170
miR156j	21	UGACAGAAGAGAGAGAGCACA	51,873	0	10	148
miR156k	20	UGACAGAAGAGAGCGAGCAC	63,624	83,985	1	38
miR159a/b	21	UUUGGAUUGAAGGGAGCUCUG	56,749	13,127	3,052	9,986
miR159c/d	21	CUUGGAUUGAAGGGAGCUCCU	37,116	0	23	55
miR160a/b/c/d/e	21	UGCCUGGCUCCCUGUAUGCCA	28,268	26,398	166	147
miR160f	21	UGCCUGGCUCCCUGUAUGCCG	0	112,786	1	1
miR162	20	UCGAUAAACCUCUGCAUCCA	0	0	1	2
miR164a/b/c/d	21	UGGAGAAGCAGGGCACGUGCA	75,984	30,958	18	61
miR166a	21	UCGGACCAGGCUUCAUUCCCC	57,494	41,043	22,941	129,603
miR166/b/c/d/e/f/g/h/i	20	UCGGACCAGGCUUCAUUCCC	16,720	12,255	1,471	9,629
miR166j/k	21	UCGGACCAGGCUUCAAUCCCU	0	22,031	4,462	23,636
miR166/m	21	UCGGACCAGGCUUCAUUCCUC	22,716	13,383	9,285	60,735
miR167a/b/c/d	21	UGAAGCUGCCAGCAUGAUCUA	69,012	48,908	3,242	42,729
miR167/e/f/g/h/i	21	UGAAGCUGCCAGCAUGAUCUG	30,598	27,262	112	1,554
miR168a/b	21	UCGCUUGGUGCAGAUCGGGAC	175,193	487,810	17,608	128,980
miR169a/b	21	CAGCCAAGGAUGACUUGCCGA	38,723	6,239	10	49
miR169c	21	CAGCCAAGGAUGACUUGCCGG	166,876	15,156	1	21
miR169f/g/h	21	UAGCCAAGGAUGACUUGCCUA	24,019	24,175	3	8
miR169i/j/k	21	UAGCCAAGGAUGACUUGCCUG	2,165	23,407	5	17

Supplemental Table 4 continued.

miR171a	20	UGAUUGAGCCGCGCCAAUAU	0	0	0	1
miR171b	20	UUGAGCCGUGCCAAUAUCAC	0	3,765	0	1
miR171d/e/i/j	21	UGAUUGAGCCGUGCCAAUAUC	57,454	2,862	53	203
miR171f	21	UUGAGCCGUGCCAAUAUCACA		0	0	1
miR171h/k	21	GUGAGCCGAACCAAUAUCACU	0	0	2	3
miR172a/b/c/d	20	AGAAUCUUGAUGAUGCUGCA	11,377	18,300	14	87
miR172e	21	GGAAUCUUGAUGAUGCUGCAU	0	71,149	197	1,249
miR319a/b/c/d	20	UUGGACUGAAGGGUGCUCCC	14,420	5,868	9	32
miR393	22	UCCAAAGGGAUCGCAUUGAUCU	0	0	2	7
miR394a/b	20	UUGGCAUUCUGUCCACCUCC	81,095	2,273	33	45
miR396a/b	21	UUCCACAGCUUUCUUGAACUG	20,324	1,372	12	92
miR399a/c	21	UGCCAAAGGAGAAUUGCCCUG	0	60,025	1	3
miR399d	21	UGCCAAAGGAGAGCUGCCCUG	0	0	0	0
miR399e	21	UGCCAAAGGAGAGUUGCCCUG	28,830	88,073	2	8
miR408	21	CUGCACUGCCUCUUCCCUGGC	17,028	826	6	11
17 families, 89 microRNAs			1,554,357	1,558,824	62,841	410,312

Supplemental Methods

Generation of small RNA Solexa libraries

Total RNA was spiked with ^{32}P-labeled 19 bp and 24 bp nucleotides and was loaded on a 15% polyacrylamide/urea gel. Small RNAs (~19-24 nt) were cut out and purified. Using mutant RNA ligase, a 3' linker (AMP-5'p=5'pCTGTAGGCACCATCAATdideoxyC-3) was ligated to the small RNA fraction, and the ligated ~36-41 nt RNA product was gel purified. Using T4 RNA ligase (Ambion), a 5' adaptor (5'-GUUCAGAGUUCUACAGUCCGACGAUC-3') was ligated to these linker-containing RNAs. The resulting RNA products (~68-73 nt) were gel purified and reverse transcribed using Superscript III reverse transcriptase (Invitrogen) and a 3' RT primer (5'-ATTGATGGTGCCTACAG-3'). The obtained cDNA was amplified by PCR with the following primers: 5'-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA-3' (forward) and 5'-CAAGCAGAAGACGGCATACGATTGATGGTGCCTACAG-3' (reverse). The PCR products were purified using phenol/chloroform extraction and gel extraction.

Mapping sequencing reads to the maize reference genome by MAQ

Basic concept of MQ scores

Since Illumina's official tool ELAND can only map reads 32 nt or shorter to a reference genome, we chose another application named MAQ (Li et al., 2008a) for mapping 36 nt reads. MAQ stands for *Mapping and Assembly with Quality* and one of its features is the calculation of a "*mapping quality*" (MQ) score that measures the likelihood of a read being mapped incorrectly. MAQ integrates the uniqueness and sequencing errors of a given read. When a read can be mapped equally well to more than one position, MAQ determines the best possible location. This is especially relevant if this read includes one or two mismatched nucleotides whose positions differ between reads. A MQ score is the *phred*-scaled probability (Ewing and Green, 1998a,b) calculated as:

MQ = -10*$\log_{10}(Pr)$ where *Pr* is the probability that a read is not correctly mapped. Calculation of the *Pr* is based on a Bayesian statistical model (Li et al., 2008a). MQ=13, 20, 30 indicates probabilities of 0.05, 0.01 and 0.001 respectively, that a read is mapped to an incorrect position. Therefore, the higher the MQ score, the more stringent the mapping criteria.

To test which MQ score is a suitable cutoff for removing low quality reads, we used MQ=0, 13, 20, and 30 separately to retain reads of different quality for our pilot transcriptome analysis using a flcDNA dataset. We found that for MQ=0, more than 85% of all reads were mapped in the genome and retained for further analyses, but an increase to MQ 13 led to a loss of almost half of the reads. The MQ=30 reads mapping to the 20 Mb continuous sequence were selected to ensure that all the reads were truly derived from this region.

Dealing with MQ zero aligned reads

If a read can be mapped equally well to multiple locations without mismatch or with identical mismatches, MAQ will pick one position at random and set the MQ score as 0. Although almost half of the sequencing reads had an MQ score of zero, we found that the majority of these reads were of perfect sequencing quality but had non-unique mapping locations. After discussing this issue with Dr. Heng Li, the developer of MAQ, we decided to keep MQ=zero reads for analysis because active marks and mRNAs are associated with genic regions, which are relatively unique. Our reasoning is that by mapping ~10,000 flcDNAs to the 2.4 Gb genome, we found that ~30% of all flcDNAs had multiple best-matched locations, which indicated that even some genes have duplicated loci. Hence, the total number of reads for a duplicated gene sequence is the sum of reads resulting from all duplicated loci. For example, if we assume a gene has two duplications in the genome, and we have a total of 1,000 mRNA sequencing reads of MQ=0 associated with these two loci, then by removing all MQ=0 reads, all 1,000 reads for that gene would be lost. However, if we randomly assign all 1,000 reads to these two duplicated loci, each locus will have 500 reads, which in all likelihood better approximates the true expression of each locus. Therefore, we believe retaining all mapped reads is more suitable than removing them from further analyses.

Transforming read counts to sequencing depth

MAQ provides a function to transform read counts to sequencing depth of piled-up reads covering each base. Using depth per base as measure has several advantages: first, it facilitates defining a modification-enriched region; second, it allows a more accurate determination of exon/intron boundaries; third, it is easier and more accurate to align the sequencing depth from the genes' ATG or TSS and to split a gene into equally binned portions; fourth, sequencing depth conveniently summarizes a gene's expression level, modification level and, under some circumstances, enables normalizations.

Compilation of maize genome annotation based on 2.4 Gb of BACs

In order to compare the McrBC-seq, ChIP-seq, mRNA-seq and smRNA-seq data with different genomic elements, we conducted a basic annotation of the 2.4 Gb of BAC sequences including TE identification, flcDNA mapping, FgeneSH prediction, comparison of maize, rice and *Arabidopsis*, GO categorization, pathway annotation, *de novo* full-length LTR-retrotransposon identification, and long hairpin double-stranded RNA identification.

Mapping maize flcDNAs to the maize genome

11,742 flcDNAs were downloaded from http://www.maizecdna.org/. We first used BlastClust from NCBI to perform self-clustering of these 11,742 sequences to obtain 11,000 non-redundant sequences. We then used BLAT (Kent, 2002) to map these 11,000 nr-flcDNAs to the 2.4 Gb of BACs with identity >= 90%. We mapped 9,451 flcDNAs to the maize genome, with 7,141 flcDNAs having only one best location and the remaining 2,310 flcDNAs having multiple best locations (Supplemental Figure 2A). As examples, we show two known genes, AJ420859 (*tua5* gene for alpha tubulin) and X15642 (phosphoenolpyruvate carboxylase), with their mapped mRNA-seq reads.

Computational prediction of genes in maize genome

We prepared two sets of computationally predicted genes based on the available maize genome sequence. The first set of genes was predicted by FgeneSH and was downloaded from the official maize genome sequencing project (http://www.maizesequence.org), and the second set of genes was predicted by us using GenScan (http://genes.mit.edu/GENSCAN.html). The FgeneSH prediction set includes 508,920 gene models, composed of 322,092 TE-related genes and 186,828 non-TE protein-coding genes, whereas GenScan predicted 522,588 gene models including 329,580 TE-related and 193,008 non-TE protein-coding genes. However, after comparing the two predicted non-TE gene sets generated by these two methods, we only found ~30,000 overlapping genes, whereas the remaining 150,000 genes were unique to GenScan or FgeneSH predictions. This result suggests that both of the popular gene prediction software suites need further improvements once the maize genome sequence is completed and sufficient flcDNAs are available as a training set.

We then used BLASTP to compare the translated non-TE genes predicted by FgeneSH and GenScan with the translated rice (www.tigr.org) and *Arabidopsis* genes (www.arabidopsis.org). Using the same standards (cutoff set as 1E-5), we found that FgeneSH delivered almost twice as many homologues as GenScan (30% vs 15%). Moreover, pair-wise comparisons of the two species showed that 27- 31% of the genes shared high homology (Supplemental Figure 4). This comparative analysis led us to conclude that for maize, FgeneSH delivers relatively more reliable results.

The maize genome sequencing project divides the FgeneSH predicted gene set into four groups based on supporting evidence:

- TE-like genes (TE): genes classified as transposable elements;
- Protein-coding genes (PRO): genes having similarity to known proteins;
- Hypothetical genes (UPRO): genes having no similarity to known proteins;
- Evidence-genes: genes built from ESTs and flcDNAs.

We found that certain Evidence-genes overlapped with FgeneSH predicted genes. In these cases, we scored the FgeneSH genes as EVI since most ESTs are only partial sequences of complete coding regions.

Identification of recognizable repeats by RepeatMasker

The source code for RepeatMasker was downloaded from http://www.repeatmasker.org/ and installed on a local server. We used WU-BLAST to compare the 2.4 Gb of BAC sequences with RepBase, a database of known plant repeats (Jurka et al., 2005). The abbreviated names of different kinds of repeats (*x* axis) in Figure 1C and Supplemental Figure 19 are based on RepeatMasker classifications.

Pathway and Gene Ontology annotation based on FgeneSH predicted genes

We used the KOBAS application (Mao, 2005) to identify biochemical pathways that the products of genes predicted by FgeneSH may participate in. KOBAS assigned maize genes to pathways by comparing them to homologous genes (as determined by BLAST similarity searches with cutoff e-values <1E−5, rank <10, and sequence identity >30%) in known *Arabidopsis* pathways in the KEGG database. Using KOBAS, we were able to assign 21,553 maize genes to 175 known GO pathways (The Gene Ontology Consortium, 2000) as shown in Supplemental Figures 5A and 5B.

Moreover, we used KOBAS to rank pathways by P value, an approach designed to test whether data from a particular pathway fits the null hypothesis or the alternative hypothesis defined as

$$\mathrm{H}_0 : p_0 \leq p_1$$

$$\mathrm{H}_1 : p_0 > p_1$$

where $p_0 = m/M$, $p_1 = n/N$, m is the number of maize genes mapped to the pathway under investigation, M is the number of all maize genes with KOBAS annotation, n is the number of all genes mapped to the selected pathway, and N is the total number of genes with KOBAS annotation. The P value of a particular pathway corresponds to a test statistic following a hypergeometric distribution:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

Pathways with P values <0.05 were considered statistically significant.

We also compared the enrichment of maize genes in each pathway with rice and *Arabidopsis* by a *Chi-square* test. Using a cutoff of Q<0.001, 62 pathways including 6,786 maize genes and 1,769 rice genes were differentially enriched, while 125 pathways of 15,492 maize genes and 3,614 *Arabidopsis* genes were differentially enriched (Supplemental Figures 6A,B; Supplemental Tables 3,4).

We then downloaded rice and *Arabidopsis* genes with GO annotation from TIGRv5 (www.tigr.org) and TAIRv8 (www.arabidopsis.org), respectively. We used the same statistical model described for the pathway analysis to identify enriched GO terms. Here, *m* is the number of genes with constant expression levels that are annotated by a given GO term, *M* is the number of all genes with constant expression levels with GO annotation, *n* is the number of all genes annotated with the given GO term, and *N* is the total number of genes with GO annotation. GO terms with adjusted P values < 0.05 using Bonferroni's correction for multiple tests were considered statistically significant (Supplemental Figures 7A, B).

Pathway analysis and comparative analysis indicated that although almost 190,000 genes were predicted as non-TE protein coding genes, only ~20,000 genes could be assigned to a known GO pathway.

De novo identification of full-length LTR-retrotransposons

The traditional method for predicting transposable elements uses gene prediction software to identify ORFs in the genome, and then uses Repeat Masker to compare the predicted ORFs with TE databases and then classify each TE with a classification based on its homology with TE-related proteins. However, this method can only identify genes which are potentially related to TEs but can not determine complete LTR-retrotransposons. A full-length LTR-retrotransposon has a complicated structure: at its 5' and 3' ends there are two long terminal repeat regions, termed 5' LTR and 3' LTR, which are usually identical and oriented in the same direction. The core region of a plant retrotransposon encodes two polygenes: the *gag* gene encodes structural proteins and the *pol* gene encodes three important enzymes: IN (integrase), RT (reverse transcriptase) and RH (RNase H) essential for retrotransposons to complete their self-duplication and insertion process.

Additional signature sequences such as the TSR (target site repeat), PBS (tRNA binding site), and PPT (polypurine tract) sites are additional features that enable the prediction of functional full-length LTR-retrotransposons. We performed a *de novo* prediction of LTR-retrotransposons using LTR-finder software (Zhao and Wang, 2007) and identified 75,015 full-length LTR-retrotransposons representing 880 Mb of DNA sequence.

Visualization of epigenetic landscapes by Affymetrix' IGB

The Integrated Genome Browser (IGB) developed by Affymetrix has shown great power to visualize tiling-path based microarray data. Here, we attempted to convert our high-throughput sequencing-based data into IGB-readable formats. Since a maize pseudo-chromosome assembly is still not available, we used a high-quality continuous stretch of a 20 Mb maize sequence for an in-depth analysis. Sequencing reads from our libraries were mapped to this 20 Mb stretch without allowing any mismatch, filtered by MQ>30. mRNA-seq, ChIP-seq, McrBC-seq data were transformed into "Wiggle" format files, in which each binned 100 bp region of sequencing depth was stored; smRNA-seq data were transformed into "Bed" format and gene and TE annotations were transformed into "Psl" format. See http://genome.ucsc.edu/ for a detailed description of each file format.

Identification of long hairpin double stranded RNAs

To identify long hairpin dsRNAs, we performed a *de novo* search using *einverted*, which is a useful tool in the EMBOSS package (Rice et al., 2000) for finding inverted repeats (stem loops) in genomic DNA. We used the default parameters of *einverted* and used 80% and 90% identity of the paired stem regions >= 1Kb to identify long hairpin dsRNAs. We found 2,253 long hairpin dsRNAs with stem identity > 80% and 1,086 with identity >90%. Generally, the average length of miRNA precursors is less than 100 bases and we used a stringent criterion of at least 2Kb of dsRNAs. We therefore believe the real number of long hairpin dsRNAs with stem regions longer than 200 bp should be much higher than we estimated.

Statistical detection of epigenetically modified regions by MACS

MACS stands for “model-based analysis of ChIP-Seq data” and its function is to isolate ChIP-enriched regions from non-enriched regions based on a dynamic Poisson distribution model. Detailed algorithms and models were described by Zhang et al. (2008). We set up a *bandwidth* of 300 bp, *mfold* of 30, *p*-value of 1.00e-05 under a FDR cutoff of 1% to call peaks representing enriched epigenetic marks.

Processing of smRNA-seq data

We used a different approach to process the smRNA-seq data since smRNAs are usually enriched in 18 nt to 30 nt species and because many smRNAs are associated with repeats. We first removed the adaptor sequences from both ends of a read and then compared the trimmed smRNA reads with a plant tRNA and rRNA database from NCBI. This allowed us to remove potentially degraded rRNA and tRNA products from our dataset. Size distributions of tRNA/rRNA related smRNA reads showed a continuous decline from 19 to 26 nt (Supplemental Figure 12D). Furthermore, we did not detect any enrichment with a 5’ terminal A, which is the signature feature of siRNAs (Supplemental Figures 12A -C).

Before mapping the trimmed reads to the maize genome, we merged all reads with identical sequences. By doing so, we lowered the total number from 4.4 million to 1.6 million unique-sequence reads in shoots and from 4.0 million to 0.7 million in roots, respectively. We recorded the frequencies of each unique-sequence smRNA, which allowed us to greatly shorten the required mapping time.

Since certain kinds of siRNAs are located in heterochromatic regions enriched in repetitive sequences, we attempted to retrieve all possible positions to which a read can be mapped. We used the SOAP application (Short Oligonucleotide Analysis Package, Li et al., 2008b), to map smRNAs to the reference genome. Because we needed to retain all mapped locations, we used a stringent mapping approach without any mismatches.

Since in many cases one miRNA has several members with identical mature miRNA sequences but that come from different genomic locations with different primary and precursor miRNA sequences, keeping all mapped positions helped us to trace all the members of a given miRNA in the maize genome.

After mapping smRNA tags back to the genome, we extracted putative precursor sequences by extending 20 nt at the 5’ end and 70 nt at the 3’ end in order to predict the secondary structure using RNAfold. Using RNAfold, we calculated a minimum free energy (MFE) for each putative precursor with -40 as the cutoff to determine whether a given precursor can form a stem-loop structure (Supplemental Table 5; Figure 5C).

We then compared the smRNA sequences with miRBase, which includes all known miRNAs to date to identify known miRNAs in our data set. We used two criteria, (1) MFE of –40 and (2) whether or not a given sequence had a match in miRBase to separate miRNAs from siRNAs, and to classify all smRNA into three groups as mentioned in the main text (Supplemental Figure 10).

For Figure 6 and Supplemental Figure 17 we aligned smRNAs to the sense and antisense strands of flcDNA and LTR-TE using an averaged sequencing depth, which is the read frequency divided by the number of locations the reads can be mapped to. By doing so, we retained all the information even for repetitive regions and at the same time reduced the influence of repetitive sequences.

SUPPLEMENTAL REFERENCES

Ewing, B. and Green, P. (1998a). Base calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:** 175-185.

Ewing, B. and Green, P. (1998b). Base calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:** 186-194.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz,

J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. **110:** 462-46

Kent, W.J. (2002). BLAT-the BLAST-like alignment tool. Genome Res. **12:** 656-64.

Li, H., Ruan, J., and Durbin, R. (2008a). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18:** 1851-1858.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008b). SOAP: short oligonucleotide alignment program. Bioinformatics **24:** 713-714.

Mao, X., Cai, T., Olyarchuk, J.G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics **21:** 3787-3793.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends Genetics **16:** 276-277.

The Gene Ontology Consortium (2000). Gene ontology: tool for unification of biology. Nature Genet. **25:** 25-29.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. **9:** R137.

Zhao, X. and Wang, H. (2007). LTR-FINDER: an efficient tool for the prediction of full-length LTR retrotransposons (2007). Nucleic Acids Res. **35:** W265-W268.

Wang et al. Plant Cell (2009). Maize epigenomics. Supplemental Dataset 1. Comparisons of maize and rice pathways (chisquare test, Q<0.001)

Pathway	Rice (2514)	Maize (10589)	Rice	Maize	Pvalue	Qvalue
Chaperones and folding catalysts	93	7	3.70%	0.07%	5.98E-78	7.50593E-76
Translation factors	71	0	2.82%	0.00%	3.21E-66	2.01339E-64
General function prediction only	89	26	3.54%	0.25%	2.99E-56	1.25141E-54
Other enzymes	76	15	3.02%	0.14%	3.19E-54	1.00145E-52
Receptors and channels	50	0	1.99%	0.00%	9.23E-47	2.31924E-45
GTP-binding proteins	48	0	1.91%	0.00%	6.56E-45	1.35535E-43
Protein kinases	52	3	2.07%	0.03%	7.55E-45	1.35535E-43
Cell cycle	42	0	1.67%	0.00%	2.35E-39	3.68542E-38
Ubiquitin enzymes	51	8	2.03%	0.08%	1.53E-38	2.12912E-37
Cell cycle - yeast	38	0	1.51%	0.00%	1.18E-35	1.48428E-34
Ubiquitin mediated proteolysis	47	8	1.87%	0.08%	5.82E-35	6.65115E-34
Protein folding and associated processing	35	0	1.39%	0.00%	7.06E-33	7.39263E-32
Other translation proteins	48	18	1.91%	0.17%	9.50E-28	9.18195E-27
Other ion-coupled transporters	27	0	1.07%	0.00%	1.80E-25	1.61381E-24
Inositol phosphate metabolism	34	624	1.35%	5.89%	1.16E-20	9.72194E-20
Benzoate degradation via CoA ligation	32	600	1.27%	5.67%	3.90E-20	3.06582E-19
Wnt signaling pathway	20	0	0.80%	0.00%	4.30E-15	3.17605E-14
Starch and sucrose metabolism	69	695	2.74%	6.56%	2.91E-13	2.03269E-12
Other replication, recombination and repair proteins	22	9	0.88%	0.08%	1.23E-12	8.12461E-12
Antigen processing and presentation	16	0	0.64%	0.00%	3.24E-12	1.94067E-11
Gap junction	16	0	0.64%	0.00%	3.24E-12	1.94067E-11
Peptidases	104	195	4.14%	1.84%	7.18E-12	4.10014E-11
Function unknown	15	0	0.60%	0.00%	1.70E-11	9.28005E-11
Progesterone-mediated oocyte maturation	14	0	0.56%	0.00%	8.90E-11	4.65622E-10
Cell division	13	0	0.52%	0.00%	4.66E-10	2.24956E-09
Pores ion channels	13	0	0.52%	0.00%	4.66E-10	2.24956E-09
Phenylpropanoid biosynthesis	38	425	1.51%	4.01%	1.46E-09	6.80837E-09
Tight junction	12	0	0.48%	0.00%	2.44E-09	1.09294E-08
Colorectal cancer	10	0	0.40%	0.00%	6.66E-08	2.61551E-07
Other amino acid metabolism	10	0	0.40%	0.00%	6.66E-08	2.61551E-07
Other energy metabolism	10	0	0.40%	0.00%	6.66E-08	2.61551E-07
p53 signaling pathway	10	0	0.40%	0.00%	6.66E-08	2.61551E-07
Glycolysis / Gluconeogenesis	76	581	3.02%	5.49%	4.71E-07	1.79318E-06
Signal transduction mechanisms	8	0	0.32%	0.00%	1.82E-06	6.35028E-06
Axon guidance	8	0	0.32%	0.00%	1.82E-06	6.35028E-06
TGF-beta signaling pathway	8	0	0.32%	0.00%	1.82E-06	6.35028E-06
Cyanoamino acid metabolism	16	209	0.64%	1.97%	5.25E-06	1.78235E-05
Focal adhesion	7	0	0.28%	0.00%	9.51E-06	3.06207E-05
Prostate cancer	7	0	0.28%	0.00%	9.51E-06	3.06207E-05
Galactose metabolism	27	269	1.07%	2.54%	1.22E-05	3.84081E-05
Carotenoid biosynthesis	14	181	0.56%	1.71%	2.69E-05	8.23302E-05
Pyruvate metabolism	49	385	1.95%	3.64%	2.83E-05	8.47316E-05
Glycerophospholipid metabolism	30	275	1.19%	2.60%	3.75E-05	0.000109434
Endometrial cancer	6	0	0.24%	0.00%	4.96E-05	0.000129924
Glycan Bindng Proteins	6	0	0.24%	0.00%	4.96E-05	0.000129924
Adherens junction	6	0	0.24%	0.00%	4.96E-05	0.000129924
Other transcription related proteins	6	0	0.24%	0.00%	4.96E-05	0.000129924
Naphthalene and anthracene degradation	16	188	0.64%	1.78%	4.96E-05	0.000129924
Sphingolipid metabolism	9	139	0.36%	1.31%	7.28E-05	0.00018657
Limonene and pinene degradation	19	199	0.76%	1.88%	0.00010771	0.000270606
Biosynthesis of steroids	34	285	1.35%	2.69%	0.000120946	0.000297903
Nitrogen metabolism	29	255	1.15%	2.41%	0.000140491	0.000339388
Phosphatidylinositol signaling system	22	213	0.88%	2.01%	0.000159338	0.000377656
Glycine, serine and threonine metabolism	41	318	1.63%	3.00%	0.000198373	0.000461469
Inorganic ion transport and metabolism	5	0	0.20%	0.00%	0.000259165	0.000551793
Alzheimer's disease	5	0	0.20%	0.00%	0.000259165	0.000551793
Hedgehog signaling pathway	5	0	0.20%	0.00%	0.000259165	0.000551793
Notch signaling pathway	5	0	0.20%	0.00%	0.000259165	0.000551793
ErbB signaling pathway	5	0	0.20%	0.00%	0.000259165	0.000551793
Pentose and glucuronate interconversions	10	134	0.40%	1.27%	0.000267389	0.000559815
Fructose and mannose metabolism	34	274	1.35%	2.59%	0.000316379	0.000651524
Citrate cycle (TCA cycle)	30	248	1.19%	2.34%	0.000437965	0.00088736
Transcription factors	12	12	0.48%	0.11%	0.000653606	0.001303251
Flavonoid biosynthesis	15	154	0.60%	1.45%	0.000875104	0.001717641
Glycerolipid metabolism	19	177	0.76%	1.67%	0.000935853	0.001808618
Ribosome	223	1181	8.87%	11.15%	0.000999346	0.001902061
Bile acid biosynthesis	11	127	0.44%	1.20%	0.001133846	0.002125846
Membrane and intracellular structural molecules	4	0	0.16%	0.00%	0.001352509	0.002392953
Toll-like receptor signaling pathway	4	0	0.16%	0.00%	0.001352509	0.002392953
Chronic myeloid leukemia	4	0	0.16%	0.00%	0.001352509	0.002392953
Parkinson's disease	4	0	0.16%	0.00%	0.001352509	0.002392953
Butanoate metabolism	29	228	1.15%	2.15%	0.00152813	0.002666125
Reductive carboxylate cycle (CO2 fixation)	22	188	0.88%	1.78%	0.001671083	0.002875595
gamma-Hexachlorocyclohexane degradation	13	134	0.52%	1.27%	0.001952625	0.003314664
Methane metabolism	24	197	0.95%	1.86%	0.00204011	0.003416999

Valine, leucine and isoleucine degradation	27	212	1.07%	2.00%	0.002340917	0.003869233
Nucleotide sugars metabolism	11	119	0.44%	1.12%	0.002620608	0.004275273
Glycosphingolipid biosynthesis - globoseries	5	80	0.20%	0.76%	0.002817693	0.004537864
alpha-Linolenic acid metabolism	27	56	1.07%	0.53%	0.003103685	0.004935181
Proteasome	48	317	1.91%	2.99%	0.003699493	0.005809045
Renin - angiotensin system	5	2	0.20%	0.02%	0.003847497	0.005966859
Metabolism of xenobiotics by cytochrome P450	10	109	0.40%	1.03%	0.003926425	0.006015004
Tyrosine metabolism	29	216	1.15%	2.04%	0.004139067	0.006264362
Tryptophan metabolism	22	177	0.88%	1.67%	0.004445754	0.006648424
Lysine degradation	19	160	0.76%	1.51%	0.004553499	0.006729439
Carbon fixation	57	353	2.27%	3.33%	0.00699705	0.009330255
Basal cell carcinoma	3	0	0.12%	0.00%	0.007056098	0.009330255
Neuroactive ligand-receptor interaction	3	0	0.12%	0.00%	0.007056098	0.009330255
Non-small cell lung cancer	3	0	0.12%	0.00%	0.007056098	0.009330255
Pancreatic cancer	3	0	0.12%	0.00%	0.007056098	0.009330255
Natural killer cell mediated cytotoxicity	3	0	0.12%	0.00%	0.007056098	0.009330255
Non-enzyme	3	0	0.12%	0.00%	0.007056098	0.009330255
Acute myeloid leukemia	3	0	0.12%	0.00%	0.007056098	0.009330255
Metabolism of other cofactors and vitamins	3	0	0.12%	0.00%	0.007056098	0.009330255
B cell receptor signaling pathway	3	0	0.12%	0.00%	0.007056098	0.009330255
Long-term depression	9	11	0.36%	0.10%	0.007508065	0.00972319
Replication complex	9	11	0.36%	0.10%	0.007508065	0.00972319
Pentose phosphate pathway	30	212	1.19%	2.00%	0.008661805	0.011102859
Benzoate degradation via hydroxylation	3	56	0.12%	0.53%	0.009560775	0.012131386
GnRH signaling pathway	17	140	0.68%	1.32%	0.010057716	0.01263432
Streptomycin biosynthesis	13	116	0.52%	1.10%	0.011469621	0.014190555
Purine metabolism	42	270	1.67%	2.55%	0.011522509	0.014190555
Valine, leucine and isoleucine biosynthesis	26	186	1.03%	1.76%	0.012678422	0.015462528
Aminoacyl-tRNA biosynthesis	42	266	1.67%	2.51%	0.015098371	0.018236827
Polyketide sugar unit biosynthesis	2	45	0.08%	0.42%	0.015576159	0.018458951
Biosynthesis of vancomycin group antibiotics	2	45	0.08%	0.42%	0.015576159	0.018458951
Glyoxylate and dicarboxylate metabolism	28	191	1.11%	1.80%	0.019310089	0.022670077
Propanoate metabolism	19	143	0.76%	1.35%	0.020052855	0.023222779
Ether lipid metabolism	13	110	0.52%	1.04%	0.020150609	0.023222779
Fatty acid metabolism	27	185	1.07%	1.75%	0.020513005	0.023425513
Phenylalanine metabolism	30	198	1.19%	1.87%	0.02462118	0.027863675
Glutathione metabolism	25	171	0.99%	1.61%	0.026931296	0.030205897
Urea cycle and metabolism of amino groups	25	168	0.99%	1.59%	0.033722103	0.037280373
1- and 2-Methylnaphthalene degradation	8	76	0.32%	0.72%	0.034235585	0.037280373
mTOR signaling pathway	10	17	0.40%	0.16%	0.034568841	0.037280373
Melanoma	2	0	0.08%	0.00%	0.036800136	0.037280373
Apoptosis	2	0	0.08%	0.00%	0.036800136	0.037280373
Cell motility and secretion	2	0	0.08%	0.00%	0.036800136	0.037280373
Thyroid Cancer	2	0	0.08%	0.00%	0.036800136	0.037280373
Other carbohydrate metabolism	2	0	0.08%	0.00%	0.036800136	0.037280373
Bladder cancer	2	0	0.08%	0.00%	0.036800136	0.037280373
Electron transfer carriers	2	0	0.08%	0.00%	0.036800136	0.037280373
Small cell lung cancer	2	0	0.08%	0.00%	0.036800136	0.037280373
Other nucleotide metabolism	2	0	0.08%	0.00%	0.036800136	0.037280373
Photosynthesis - antenna proteins	8	75	0.32%	0.71%	0.03786922	0.038056501
Basal transcription factors	24	161	0.95%	1.52%	0.038682923	0.038565703
DNA polymerase	14	107	0.56%	1.01%	0.043225129	0.042420799
Lysine biosynthesis	14	107	0.56%	1.01%	0.043225129	0.042420799
Caprolactam degradation	5	55	0.20%	0.52%	0.048204218	0.046579433
Olfactory transduction	5	55	0.20%	0.52%	0.048204218	0.046579433
Histidine metabolism	17	121	0.68%	1.14%	0.051052537	0.048955168
Terpenoid biosynthesis	10	83	0.40%	0.78%	0.052292322	0.049562639
SNARE interactions in vesicular transport	26	167	1.03%	1.58%	0.052475134	0.049562639
Androgen and estrogen metabolism	5	54	0.20%	0.51%	0.053782939	0.05041877
Glycosphingolipid biosynthesis - lactoseries	1	28	0.04%	0.26%	0.055024107	0.051200211
Brassinosteroid biosynthesis	3	41	0.12%	0.39%	0.05805691	0.053264651
3-Chloroacrylic acid degradation	7	65	0.28%	0.61%	0.058090771	0.053264651
Alkaloid biosynthesis II	7	64	0.28%	0.60%	0.064276964	0.058509823
Cellular antigens	5	51	0.20%	0.48%	0.074491489	0.06732004
Pantothenate and CoA biosynthesis	18	121	0.72%	1.14%	0.076881201	0.068983407
Epithelial cell signaling in Helicobacter pylori infection	23	146	0.91%	1.38%	0.079278751	0.070630161
Alanine and aspartate metabolism	48	266	1.91%	2.51%	0.088408517	0.078209278
Glycosphingolipid biosynthesis - neo-lactoseries	2	31	0.08%	0.29%	0.089887881	0.078961905
Aminophosphonate metabolism	6	54	0.24%	0.51%	0.099567703	0.086857737
RNA polymerase	36	206	1.43%	1.95%	0.101745528	0.088145438
ABC transporters	7	59	0.28%	0.56%	0.105651861	0.090902706
Type I diabetes mellitus	8	18	0.32%	0.17%	0.108830005	0.093000188
Riboflavin metabolism	6	12	0.24%	0.11%	0.113659539	0.09647098
Bisphenol A degradation	2	29	0.08%	0.27%	0.115377505	0.097271897
Protein export	27	79	1.07%	0.75%	0.126953373	0.106317682
Selenoamino acid metabolism	22	134	0.88%	1.27%	0.128515939	0.106913503
Thiamine metabolism	5	10	0.20%	0.09%	0.143779741	0.11804804
Lipoic acid metabolism	5	10	0.20%	0.09%	0.143779741	0.11804804
High-mannose type N-glycan biosynthesis	4	39	0.16%	0.37%	0.145735727	0.118254476
Aminosugars metabolism	23	137	0.91%	1.29%	0.145913935	0.118254476
Huntington's disease	7	55	0.28%	0.52%	0.155299217	0.124700652

Photosynthesis	67	230	2.67%	2.17%	0.156046939	0.124700652
Arginine and proline metabolism	22	131	0.88%	1.24%	0.156845946	0.124700652
Type II diabetes mellitus	4	38	0.16%	0.36%	0.162529488	0.128406662
Pyrimidine metabolism	34	188	1.35%	1.78%	0.164103112	0.128839594
Ascorbate and aldarate metabolism	18	110	0.72%	1.04%	0.171729174	0.133989485
N-Glycan biosynthesis	27	152	1.07%	1.44%	0.19086652	0.145190753
Leukocyte transendothelial migration	1	0	0.04%	0.00%	0.191864459	0.145190753
Dorso-ventral axis formation	1	0	0.04%	0.00%	0.191864459	0.145190753
Circadian rhythm	1	0	0.04%	0.00%	0.191864459	0.145190753
Proteoglycans	1	0	0.04%	0.00%	0.191864459	0.145190753
Two-component system	18	108	0.72%	1.02%	0.197005545	0.148188497
Renal cell carcinoma	5	12	0.20%	0.11%	0.214576041	0.160444362
beta-Alanine metabolism	19	111	0.76%	1.05%	0.22312203	0.165847245
Phenylalanine, tyrosine and tryptophan biosynthesis	40	209	1.59%	1.97%	0.23720793	0.174804567
SNAREs	31	167	1.23%	1.58%	0.237955842	0.174804567
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	9	22	0.36%	0.21%	0.243820064	0.178071127
Glycosphingolipid biosynthesis - ganglioseries	4	34	0.16%	0.32%	0.249557166	0.180411625
Glioma	8	55	0.32%	0.52%	0.249897122	0.180411625
Glutamate metabolism	42	217	1.67%	2.05%	0.251614946	0.180613787
Oxidative phosphorylation	129	607	5.13%	5.73%	0.259094156	0.183960573
Synthesis and degradation of ketone bodies	5	39	0.20%	0.37%	0.259206272	0.183960573
Porphyrin and chlorophyll metabolism	25	136	0.99%	1.28%	0.277694035	0.195974273
Amyotrophic lateral sclerosis (ALS)	9	23	0.36%	0.22%	0.28871733	0.202615346
Methionine metabolism	20	110	0.80%	1.04%	0.320018511	0.221870855
Fc epsilon RI signaling pathway	4	11	0.16%	0.10%	0.321454219	0.221870855
VEGF signaling pathway	4	11	0.16%	0.10%	0.321454219	0.221870855
1,2-Dichloroethane degradation	3	26	0.12%	0.25%	0.329824829	0.226404352
Regulation of actin cytoskeleton	26	86	1.03%	0.81%	0.333698256	0.226695656
MAPK signaling pathway	22	71	0.88%	0.67%	0.33385848	0.226695656
Biosynthesis of siderophore group nonribosomal peptides	1	1	0.04%	0.01%	0.346928781	0.231991566
2,4-Dichlorobenzoate degradation	1	1	0.04%	0.01%	0.346928781	0.231991566
Glycosaminoglycan degradation	5	36	0.20%	0.34%	0.347198253	0.231991566
Novobiocin biosynthesis	10	28	0.40%	0.26%	0.362067201	0.240646691
Cytoskeleton proteins	36	126	1.43%	1.19%	0.37507614	0.246298903
Long-term potentiation	9	55	0.36%	0.52%	0.376453372	0.246298903
Melanogenesis	9	55	0.36%	0.52%	0.376453372	0.246298903
N-Glycan degradation	5	35	0.20%	0.33%	0.381835294	0.247995582
Peptidoglycan biosynthesis	6	40	0.24%	0.38%	0.382995049	0.247995582
Insulin signaling pathway	35	174	1.39%	1.64%	0.415342695	0.267562028
Inositol metabolism	5	34	0.20%	0.32%	0.419385833	0.2687882
Regulation of autophagy	10	58	0.40%	0.55%	0.431649627	0.275026338
Cysteine metabolism	24	82	0.95%	0.77%	0.433497865	0.275026338
D-Glutamine and D-glutamate metabolism	3	10	0.12%	0.09%	0.468126386	0.295139654
Type II secretion system	1	2	0.04%	0.02%	0.472249065	0.295139654
Hematopoietic cell lineage	1	2	0.04%	0.02%	0.472249065	0.295139654
Nicotinate and nicotinamide metabolism	8	47	0.32%	0.44%	0.481217969	0.299256078
Other transporters	13	42	0.52%	0.40%	0.503953306	0.311850737
Lipopolysaccharide biosynthesis	5	20	0.20%	0.19%	0.538814196	0.331788536
Folate biosynthesis	16	82	0.64%	0.77%	0.553214371	0.338994074
Biotin metabolism	4	25	0.16%	0.24%	0.615416245	0.375278989
Styrene degradation	4	19	0.16%	0.18%	0.669279607	0.406153089
Prion disease	4	23	0.16%	0.22%	0.739244997	0.446454878
Arachidonic acid metabolism	9	45	0.36%	0.42%	0.765657945	0.460194073
One carbon pool by folate	15	71	0.60%	0.67%	0.783434056	0.46863602
Caffeine metabolism	1	7	0.04%	0.07%	0.818175153	0.487051903
Polyunsaturated fatty acid biosynthesis	17	65	0.68%	0.61%	0.829133732	0.487051903
Photosynthesis proteins	75	305	2.98%	2.88%	0.833347205	0.487051903
Vitamin B6 metabolism	6	25	0.24%	0.24%	0.837968981	0.487051903
Cytochrome P450	6	25	0.24%	0.24%	0.837968981	0.487051903
C5-Branched dibasic acid metabolism	2	14	0.08%	0.13%	0.841361169	0.487051903
Biosynthesis of ansamycins	2	14	0.08%	0.13%	0.841361169	0.487051903
Diterpenoid biosynthesis	3	22	0.12%	0.21%	0.884219303	0.508318625
Fatty acid elongation in mitochondria	2	16	0.08%	0.15%	0.886191589	0.508318625
Ethylbenzene degradation	2	17	0.08%	0.16%	0.903909395	0.516124807
Adipocytokine signaling pathway	9	42	0.36%	0.40%	0.919091508	0.520065797
Ubiquinone biosynthesis	10	41	0.40%	0.39%	0.919091508	0.520065797
Calcium signaling pathway	13	59	0.52%	0.56%	0.924863596	0.520838649
Transporters	15	59	0.60%	0.56%	0.928749747	0.520838649
PPAR signaling pathway	18	80	0.72%	0.76%	0.937868057	0.521297712
ATPases	19	79	0.76%	0.75%	0.937868057	0.521297712
Sulfur metabolism	18	77	0.72%	0.73%	0.943112534	0.521903457
Alkaloid biosynthesis I	10	44	0.40%	0.42%	0.961519279	0.529746785
1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) degradation	2	23	0.08%	0.22%	0.966375451	0.529746785
Glycosyltransferases	27	112	1.07%	1.06%	0.970778062	0.529746785
Fatty acid biosynthesis	14	61	0.56%	0.58%	0.974154399	0.529746785
Linoleic acid metabolism	15	61	0.60%	0.58%	0.980957158	0.5311468
Tetrachloroethene degradation	1	19	0.04%	0.18%	0.985933458	0.531550092
Taurine and hypotaurine metabolism	7	27	0.28%	0.25%	0.991860877	0.532460526

Wang et al. Plant Cell (2009) Maize epigenomics. Supplemental Dataset 2. Comparisons of maize and *Arabidopsis* pathways (chisquare test, Q<0.001)

Pathway	Arabidposis (3001)	Maize (10589)	Arabidopsis	Maize	Pvalue	Qvalue
Transcription factors	216	12	7.20%	0.11%	8.42E-156	6.174E-154
Other enzymes	162	15	5.40%	0.14%	1.97E-110	7.237E-109
GTP-binding proteins	95	0	3.17%	0.00%	2.11E-74	3.8724E-73
Receptors and channels	95	0	3.17%	0.00%	2.11E-74	3.8724E-73
Translation factors	79	0	2.63%	0.00%	6.04E-62	8.863E-61
Protein kinases	68	3	2.27%	0.03%	5.53E-50	6.7551E-49
Ubiquitin enzymes	57	8	1.90%	0.08%	1.39E-36	1.4562E-35
Ubiquitin mediated proteolysis	56	8	1.87%	0.08%	7.91E-36	7.2491E-35
General function prediction only	74	26	2.47%	0.25%	1.55E-35	1.2667E-34
Cell cycle	40	0	1.33%	0.00%	1.18E-31	8.6194E-31
Protein folding and associated processing	38	0	1.27%	0.00%	4.19E-30	2.7934E-29
Cell cycle - yeast	37	0	1.23%	0.00%	2.50E-29	1.529E-28
Starch and sucrose metabolism	69	695	2.30%	6.56%	5.24E-19	2.9583E-18
Chaperones and folding catalysts	30	7	1.00%	0.07%	2.56E-17	1.3409E-16
Glycolysis / Gluconeogenesis	55	581	1.83%	5.49%	9.01E-17	4.4071E-16
Benzoate degradation via CoA ligation	66	600	2.20%	5.67%	1.18E-14	5.4194E-14
Inositol phosphate metabolism	80	624	2.67%	5.89%	2.66E-12	1.1358E-11
Other translation proteins	32	18	1.07%	0.17%	2.79E-12	1.1358E-11
Wnt signaling pathway	17	0	0.57%	0.00%	6.81E-12	2.6301E-11
Pyruvate metabolism	37	385	1.23%	3.64%	3.15E-11	1.1568E-10
Other replication, recombination and repair proteins	23	9	0.77%	0.08%	4.55E-11	1.5872E-10
Other ion-coupled transporters	15	0	0.50%	0.00%	1.41E-10	4.4917E-10
Function unknown	15	0	0.50%	0.00%	1.41E-10	4.4917E-10
Glycine, serine and threonine metabolism	28	318	0.93%	3.00%	3.19E-10	9.7433E-10
Gap junction	14	0	0.47%	0.00%	6.40E-10	1.8059E-09
Tight junction	14	0	0.47%	0.00%	6.40E-10	1.8059E-09
Butanoate metabolism	15	228	0.50%	2.15%	2.60E-09	7.0617E-09
Antigen processing and presentation	13	0	0.43%	0.00%	2.91E-09	7.1117E-09
Other amino acid metabolism	13	0	0.43%	0.00%	2.91E-09	7.1117E-09
Progesterone-mediated oocyte maturation	13	0	0.43%	0.00%	2.91E-09	7.1117E-09
Ribosome	225	1181	7.50%	11.15%	7.92E-09	1.8729E-08
Glycerophospholipid metabolism	25	275	0.83%	2.60%	9.74E-09	2.2328E-08
RNA polymerase	13	206	0.43%	1.95%	1.03E-08	2.244E-08
Galactose metabolism	24	269	0.80%	2.54%	1.04E-08	2.244E-08
Signal transduction mechanisms	11	0	0.37%	0.00%	6.00E-08	1.2226E-07
p53 signaling pathway	11	0	0.37%	0.00%	6.00E-08	1.2226E-07
Alanine and aspartate metabolism	27	266	0.90%	2.51%	1.18E-07	2.335E-07
Nitrogen metabolism	25	255	0.83%	2.41%	1.23E-07	2.3732E-07
Pores ion channels	10	0	0.33%	0.00%	2.73E-07	5.1241E-07
Prostate cancer	9	0	0.30%	0.00%	1.24E-06	2.2678E-06
Sphingolipid metabolism	8	139	0.27%	1.31%	1.66E-06	2.9775E-06
Valine, leucine and isoleucine degradation	21	212	0.70%	2.00%	1.83E-06	3.1915E-06
Cyanoamino acid metabolism	21	209	0.70%	1.97%	2.65E-06	4.5279E-06
Aminoacyl-tRNA biosynthesis	33	266	1.10%	2.51%	4.53E-06	7.5469E-06
Cell division	8	0	0.27%	0.00%	5.61E-06	8.5757E-06
Endometrial cancer	8	0	0.27%	0.00%	5.61E-06	8.5757E-06
Colorectal cancer	8	0	0.27%	0.00%	5.61E-06	8.5757E-06
Other energy metabolism	8	0	0.27%	0.00%	5.61E-06	8.5757E-06
Reductive carboxylate cycle (CO2 fixation)	19	188	0.63%	1.78%	9.61E-06	1.4386E-05
Lysine degradation	14	160	0.47%	1.51%	1.08E-05	1.5838E-05
Pentose and glucuronate interconversions	10	134	0.33%	1.27%	1.69E-05	2.4368E-05
Adherens junction	7	0	0.23%	0.00%	2.55E-05	3.3348E-05
Focal adhesion	7	0	0.23%	0.00%	2.55E-05	3.3348E-05
Other transcription related proteins	7	0	0.23%	0.00%	2.55E-05	3.3348E-05
TGF-beta signaling pathway	7	0	0.23%	0.00%	2.55E-05	3.3348E-05
Notch signaling pathway	7	0	0.23%	0.00%	2.55E-05	3.3348E-05
Carbon fixation	55	353	1.83%	3.33%	2.76E-05	3.549E-05
Citrate cycle (TCA cycle)	33	248	1.10%	2.34%	3.34E-05	4.2169E-05
Proteasome	49	317	1.63%	2.99%	6.30E-05	7.833E-05
Tyrosine metabolism	28	216	0.93%	2.04%	7.72E-05	9.4394E-05
Valine, leucine and isoleucine biosynthesis	22	186	0.73%	1.76%	7.91E-05	9.5108E-05
Bile acid biosynthesis	11	127	0.37%	1.20%	9.09E-05	0.0001075
Biosynthesis of steroids	43	285	1.43%	2.69%	9.69E-05	0.00011274
Indole and ipecac alkaloid biosynthesis	6	0	0.20%	0.00%	0.0001155	0.00012833
Glycan Bindng Proteins	6	0	0.20%	0.00%	0.0001155	0.00012833
Small cell lung cancer	6	0	0.20%	0.00%	0.0001155	0.00012833
Carotenoid biosynthesis	22	181	0.73%	1.71%	0.000141	0.00015433
Phenylalanine, tyrosine and tryptophan biosynthesis	28	209	0.93%	1.97%	0.0001662	0.00017923
Streptomycin biosynthesis	10	116	0.33%	1.10%	0.0001855	0.00019713
Glyoxylate and dicarboxylate metabolism	25	191	0.83%	1.80%	0.0002421	0.0002536
gamma-Hexachlorocyclohexane degradation	66	134	2.20%	1.27%	0.0002483	0.00025643
Fructose and mannose metabolism	43	274	1.43%	2.59%	0.0002824	0.0002876
Pyrimidine metabolism	25	188	0.83%	1.78%	0.0003364	0.00033791

Urea cycle and metabolism of amino groups	21	168	0.70%	1.59%	0.0003523	0.00034912
Aminosugars metabolism	15	137	0.50%	1.29%	0.0003817	0.00037323
Pantothenate and CoA biosynthesis	12	121	0.40%	1.14%	0.0003943	0.00038051
Ether lipid metabolism	10	110	0.33%	1.04%	0.0004053	0.00038603
Inorganic ion transport and metabolism	5	0	0.17%	0.00%	0.0005237	0.00047415
Non-small cell lung cancer	5	0	0.17%	0.00%	0.0005237	0.00047415
Chronic myeloid leukemia	5	0	0.17%	0.00%	0.0005237	0.00047415
Axon guidance	5	0	0.17%	0.00%	0.0005237	0.00047415
Tryptophan metabolism	24	177	0.80%	1.67%	0.0006575	0.00058799
Phosphatidylinositol signaling system	32	213	1.07%	2.01%	0.0007865	0.00069485
Lysine biosynthesis	11	107	0.37%	1.01%	0.0011752	0.001026
Glutamate metabolism	34	217	1.13%	2.05%	0.001308	0.0011285
Histidine metabolism	14	121	0.47%	1.14%	0.001409	0.0012015
mTOR signaling pathway	15	17	0.50%	0.16%	0.0015153	0.00127728
Nucleotide sugars metabolism	14	119	0.47%	1.12%	0.001786	0.0014883
Peptidases	83	195	2.77%	1.84%	0.0020416	0.00168218
DNA polymerase	12	107	0.40%	1.01%	0.0022258	0.00181359
Melanoma	4	0	0.13%	0.00%	0.0023742	0.00183268
Apoptosis	4	0	0.13%	0.00%	0.0023742	0.00183268
Hedgehog signaling pathway	4	0	0.13%	0.00%	0.0023742	0.00183268
Parkinson's disease	4	0	0.13%	0.00%	0.0023742	0.00183268
ErbB signaling pathway	4	0	0.13%	0.00%	0.0023742	0.00183268
GnRH signaling pathway	19	140	0.63%	1.32%	0.0026799	0.00204712
Fatty acid metabolism	29	185	0.97%	1.75%	0.0031826	0.00239259
Pentose phosphate pathway	35	212	1.17%	2.00%	0.0031974	0.00239259
Propanoate metabolism	20	143	0.67%	1.35%	0.0032456	0.00240418
Basal transcription factors	24	161	0.80%	1.52%	0.0035192	0.00258078
Epithelial cell signaling in Helicobacter pylori infection	21	146	0.70%	1.38%	0.0038952	0.00282822
Arginine and proline metabolism	18	131	0.60%	1.24%	0.0042328	0.00304321
beta-Alanine metabolism	14	111	0.47%	1.05%	0.0045333	0.00322756
Polyketide sugar unit biosynthesis	2	45	0.07%	0.42%	0.0055141	0.00385112
Biosynthesis of vancomycin group antibiotics	2	45	0.07%	0.42%	0.0055141	0.00385112
Cellular antigens	3	51	0.10%	0.48%	0.0056176	0.00388637
Purine metabolism	50	270	1.67%	2.55%	0.00596	0.00408469
Benzoate degradation via hydroxylation	4	56	0.13%	0.53%	0.0063501	0.00431178
Glycerolipid metabolism	29	177	0.97%	1.67%	0.0068027	0.00457671
Phenylpropanoid biosynthesis	155	425	5.16%	4.01%	0.0068671	0.00457804
Flavonoid biosynthesis	24	154	0.80%	1.45%	0.0070768	0.00467534
Caprolactam degradation	4	55	0.13%	0.52%	0.0073051	0.00478311
Androgen and estrogen metabolism	4	54	0.13%	0.51%	0.0084018	0.00545251
Oxidative phosphorylation	135	607	4.50%	5.73%	0.0098621	0.00634404
alpha-Linolenic acid metabolism	29	56	0.97%	0.53%	0.0107031	0.006415
Fluorene degradation	3	0	0.10%	0.00%	0.0107597	0.006415
Alzheimer's disease	3	0	0.10%	0.00%	0.0107597	0.006415
Thyroid Cancer	3	0	0.10%	0.00%	0.0107597	0.006415
Pancreatic cancer	3	0	0.10%	0.00%	0.0107597	0.006415
Toll-like receptor signaling pathway	3	0	0.10%	0.00%	0.0107597	0.006415
Natural killer cell mediated cytotoxicity	3	0	0.10%	0.00%	0.0107597	0.006415
Acute myeloid leukemia	3	0	0.10%	0.00%	0.0107597	0.006415
Metabolism of other cofactors and vitamins	3	0	0.10%	0.00%	0.0107597	0.006415
Linoleic acid metabolism	6	61	0.20%	0.58%	0.0143178	0.00846748
1- and 2-Methylnaphthalene degradation	9	76	0.30%	0.72%	0.0150318	0.00881866
Aminophosphonate metabolism	5	54	0.17%	0.51%	0.0178815	0.01036155
Two-component system	16	108	0.53%	1.02%	0.0179444	0.01036155
Methane metabolism	77	197	2.57%	1.86%	0.0186099	0.01066193
Selenoamino acid metabolism	22	134	0.73%	1.27%	0.0203607	0.01157452
Phenylalanine metabolism	76	198	2.53%	1.87%	0.0273755	0.01544255
Glycosphingolipid biosynthesis - lactoseries	1	28	0.03%	0.26%	0.0279716	0.01565835
Synthesis and degradation of ketone bodies	3	39	0.10%	0.37%	0.0314414	0.01746742
Methionine metabolism	18	110	0.60%	1.04%	0.0365501	0.02015291
Cytochrome P450	1	25	0.03%	0.24%	0.0447237	0.02447559
Nicotinate and nicotinamide metabolism	5	47	0.17%	0.44%	0.0450698	0.0244823
Membrane and intracellular structural molecules	2	0	0.07%	0.00%	0.0487506	0.02465544
Monoterpenoid biosynthesis	2	0	0.07%	0.00%	0.0487506	0.02465544
Dorso-ventral axis formation	2	0	0.07%	0.00%	0.0487506	0.02465544
B cell receptor signaling pathway	2	0	0.07%	0.00%	0.0487506	0.02465544
Other carbohydrate metabolism	2	0	0.07%	0.00%	0.0487506	0.02465544
Basal cell carcinoma	2	0	0.07%	0.00%	0.0487506	0.02465544
Neuroactive ligand-receptor interaction	2	0	0.07%	0.00%	0.0487506	0.02465544
Bladder cancer	2	0	0.07%	0.00%	0.0487506	0.02465544
Non-enzyme	2	0	0.07%	0.00%	0.0487506	0.02465544
Other nucleotide metabolism	2	0	0.07%	0.00%	0.0487506	0.02465544
Glutathione metabolism	33	171	1.10%	1.61%	0.0495311	0.02487856
3-Chloroacrylic acid degradation	9	65	0.30%	0.61%	0.0545534	0.02721478
Bisphenol A degradation	2	29	0.07%	0.27%	0.0596015	0.02953223
Metabolism of xenobiotics by cytochrome P450	19	109	0.63%	1.03%	0.060565	0.02980822

Folate biosynthesis	13	82	0.43%	0.77%	0.0634262	0.03077984
Glycosphingolipid biosynthesis - ganglioseries	3	34	0.10%	0.32%	0.0637983	0.03077984
Inositol metabolism	3	34	0.10%	0.32%	0.0637983	0.03077984
Porphyrin and chlorophyll metabolism	26	136	0.87%	1.28%	0.0772193	0.03692545
Glycosphingolipid biosynthesis - globoseries	13	80	0.43%	0.76%	0.0775436	0.03692545
Glycosyltransferases	44	112	1.47%	1.06%	0.0788837	0.03732126
Ascorbate and aldarate metabolism	20	110	0.67%	1.04%	0.0812153	0.03817805
1,2-Dichloroethane degradation	2	26	0.07%	0.25%	0.0930106	0.04344436
Brassinosteroid biosynthesis	5	41	0.17%	0.39%	0.0972228	0.0451244
Glycosaminoglycan degradation	4	36	0.13%	0.34%	0.098113	0.04525118
Riboflavin metabolism	7	12	0.23%	0.11%	0.1048928	0.04807576
ABC transporters	9	59	0.30%	0.56%	0.1059496	0.04825852
Protein export	32	79	1.07%	0.75%	0.1083404	0.0490429
Peptidoglycan biosynthesis	5	40	0.17%	0.38%	0.1102117	0.0495839
Regulation of autophagy	9	58	0.30%	0.55%	0.1179514	0.05274238
Naphthalene and anthracene degradation	67	188	2.23%	1.78%	0.1204246	0.05352195
One carbon pool by folate	12	71	0.40%	0.67%	0.1218551	0.05383146
Prion disease	2	23	0.07%	0.22%	0.14491	0.06357538
N-Glycan biosynthesis	32	152	1.07%	1.44%	0.1456457	0.06357538
Replication complex	6	11	0.20%	0.10%	0.1529673	0.06637621
Olfactory transduction	9	55	0.30%	0.52%	0.1617022	0.06975375
Limonene and pinene degradation	69	199	2.30%	1.88%	0.1657188	0.07106835
MAPK signaling pathway	28	71	0.93%	0.67%	0.1703274	0.07262008
Ubiquinone biosynthesis	6	41	0.20%	0.39%	0.1718386	0.07284087
Diterpenoid biosynthesis	11	22	0.37%	0.21%	0.1770356	0.07461254
Renal cell carcinoma	6	12	0.20%	0.11%	0.1888981	0.07915714
Terpenoid biosynthesis	16	83	0.53%	0.78%	0.1922899	0.08012062
N-Glycan degradation	5	35	0.17%	0.33%	0.2032558	0.08421126
Fatty acid biosynthesis	11	61	0.37%	0.58%	0.2101191	0.08656575
Renin - angiotensin system	2	2	0.07%	0.02%	0.2135485	0.08744527
Tetracycline biosynthesis	1	0	0.03%	0.00%	0.2208241	0.08744527
C21-Steroid hormone metabolism	1	0	0.03%	0.00%	0.2208241	0.08744527
Cell motility and secretion	1	0	0.03%	0.00%	0.2208241	0.08744527
Electron transfer carriers	1	0	0.03%	0.00%	0.2208241	0.08744527
T cell receptor signaling pathway	1	0	0.03%	0.00%	0.2208241	0.08744527
Biotin metabolism	3	25	0.10%	0.24%	0.2210805	0.08744527
Thiamine metabolism	5	10	0.17%	0.09%	0.2217935	0.08744527
Sulfur metabolism	15	77	0.50%	0.73%	0.2245431	0.08805595
SNAREs	57	167	1.90%	1.58%	0.253163	0.09875132
Long-term depression	5	11	0.17%	0.10%	0.2678708	0.10393556
Taurine and hypotaurine metabolism	4	27	0.13%	0.25%	0.3092677	0.11936623
PPAR signaling pathway	17	80	0.57%	0.76%	0.3355677	0.128839
Calcium signaling pathway	12	59	0.40%	0.56%	0.3618831	0.13821897
Lipoic acid metabolism	4	10	0.13%	0.09%	0.3748466	0.14242846
Other transporters	8	42	0.27%	0.40%	0.3854002	0.14568362
Lipopolysaccharide biosynthesis	3	20	0.10%	0.19%	0.4269696	0.16056944
Photosynthesis	73	230	2.43%	2.17%	0.4336205	0.16223862
Transporters	21	59	0.70%	0.56%	0.4435947	0.16512799
Insulin signaling pathway	56	174	1.87%	1.64%	0.450127	0.16671337
Alkaloid biosynthesis I	9	44	0.30%	0.42%	0.4646758	0.17123695
ATPases	27	79	0.90%	0.75%	0.4672243	0.17131523
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	9	22	0.30%	0.21%	0.4732631	0.1726661
Long-term potentiation	12	55	0.40%	0.52%	0.4979807	0.17901229
Huntington's disease	12	55	0.40%	0.52%	0.4979807	0.17901229
Melanogenesis	12	55	0.40%	0.52%	0.4979807	0.17901229
Type II secretion system	1	2	0.03%	0.02%	0.5269802	0.18851287
D-Glutamine and D-glutamate metabolism	3	10	0.10%	0.09%	0.572869	0.20393353
Arachidonic acid metabolism	10	45	0.33%	0.42%	0.5920048	0.20887865
Cytoskeleton proteins	40	126	1.33%	1.19%	0.592457	0.20887865
Photosynthesis - antenna proteins	18	75	0.60%	0.71%	0.609431	0.213835
Vitamin B6 metabolism	5	25	0.17%	0.24%	0.6201823	0.21657117
VEGF signaling pathway	3	11	0.10%	0.10%	0.6268751	0.21787084
Novobiocin biosynthesis	6	28	0.20%	0.26%	0.6764699	0.23399854
Photosynthesis proteins	91	305	3.03%	2.88%	0.7073265	0.24352349
Regulation of actin cytoskeleton	27	86	0.90%	0.81%	0.7246286	0.24831459
Adipocytokine signaling pathway	10	42	0.33%	0.40%	0.7419934	0.2530825
1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) degrada	5	23	0.17%	0.22%	0.7553995	0.25646226
Cysteine metabolism	21	82	0.70%	0.77%	0.7665889	0.25906177
Styrene degradation	4	19	0.13%	0.18%	0.7708387	0.259303
Fc epsilon RI signaling pathway	2	11	0.07%	0.10%	0.8173641	0.27369824
Glioma	14	55	0.47%	0.52%	0.8302339	0.27615691
Type I diabetes mellitus	5	18	0.17%	0.17%	0.8322382	0.27615691
Caffeine metabolism	1	7	0.03%	0.07%	0.8642217	0.28490903
Glycosphingolipid biosynthesis - neo-lactoseries	10	31	0.33%	0.29%	0.8663842	0.28490903
Polyunsaturated fatty acid biosynthesis	17	65	0.57%	0.61%	0.8711136	0.28518541
Type II diabetes mellitus	12	38	0.40%	0.36%	0.8754749	0.28533939

Biosynthesis of ansamycins	2	14	0.07%	0.13%	0.8979849	0.29138095
High-mannose type N-glycan biosynthesis	10	39	0.33%	0.37%	0.9119822	0.29461919
Alkaloid biosynthesis II	18	64	0.60%	0.60%	0.9165427	0.29479384
Fatty acid elongation in mitochondria	2	16	0.07%	0.15%	0.9317567	0.29837854
Ethylbenzene degradation	2	17	0.07%	0.16%	0.9443724	0.30110363
Amyotrophic lateral sclerosis (ALS)	6	23	0.20%	0.22%	0.9656479	0.30655426
C5-Branched dibasic acid metabolism	1	14	0.03%	0.13%	0.9763636	0.30862006
SNARE interactions in vesicular transport	48	167	1.60%	1.58%	0.9969836	0.31378534