The heterotic hybrid offspring of *Arabidopsis* accessions C24 and Landsberg erecta have altered methylomes. Changes occur most frequently at loci where parental methylation levels are different. There are context-specific biases in the nonadditive methylation patterns with *mCG* generally increased and *mCHH* decreased relative to the parents. These changes are the result of two main mechanisms, Trans Chromosomal Methylation and Trans Chromosomal deMethylation, where the methylation level of one parental allele alters to resemble that of the other parent. Regions of altered methylation are enriched around genic regions and are often correlated with changes in siRNA levels. We identified examples of genes with altered expression likely to be due to methylation changes and suggest that in crosses between the C24 and Ler accessions, epigenetic controls can be important in the generation of altered transcriptional levels that may contribute to the increased biomass of the hybrids.

In the formation of a hybrid, the genome and epigenome of each of the parents are brought together within the one nucleus. The interactions of these two sets of genetic instructions result in the unique characteristics of the hybrid, including superior performance. Both the level and pattern of expression of many genes are altered in hybrids (1–3). Altered methylation levels have mostly been explained by the interaction between the alleles of a gene delivered by the parents involving a range of interactions such as dominance, overdominance, and epistatic interactions between loci (1). Despite these genetic analyses, there is a lack of understanding of the molecular mechanisms underpinning heterosis. It has been suggested that the magnitude of hybrid vigor is positively correlated with the genetic distance or amount of sequence variation between the parental genomes (4, 5). However, crosses between genetically similar parents such as *Arabidopsis* accessions or subspecies of rice can produce hybrids displaying significant heterosis, apparently breaking down the relationship between genetic distance and extent of hybrid vigor (6, 7). It has been reported that the epigenome evolves at a significant rate than the genetic sequence (8–10). Both parental genomes have similar frequencies of cytosines (Cs) in the genome (Dataset S1, Table S1). Both parental genomes have similar frequencies of highly methylated C residues (*mCs*) in each of the three *mC* contexts—~26% *mCG*, ~6% *mCHG*, and ~2% *mCHH* (Dataset S1, Table S1)—that are similar to levels found in Columbia (21). These frequencies increased to ~30% *mCG*, ~15% *mCHG*, and ~8% *mCHH* (Dataset S1, Table S2) when accounting for all *mCs* with a level of methylation above the false-positive methylation rate, which was determined by correcting against the unmethylated chloroplast genome sequences (*SI Materials and Methods*). Both hybrids showed slight increase in the frequency of *mC* (above parental levels) to ~57% *mCG*, ~19% *mCHG*, and ~10% *mCHH* (Dataset S1, Tables S1 and S2), consistent with the hybrids inheriting *mCs* common to both parents and sites unique to one of the parents.

In both parents and hybrids, the majority of *mCG* sites are highly methylated (>80%), whereas *mCHG* and *mCHH* sites display a broader distribution of methylation (Fig. S1A). Chromosomal distributions of *mC* mirrored the patterns of 24-nt siRNA density with the highest levels associated with the pericentromeric regions of the chromosomes (Fig. L4 and Fig. S2). Ingeneric regions, the frequency of *mCG* sites is high in the gene body and lower in the flanking regions, whereas *mCHG* and *mCHH* sites are highest in flanking regions and lower throughout the

In this study, we showed that the methylation patterns of the parental lines are largely inherited faithfully in the hybrids (additive methylation), with exceptions applying mostly to those regions of the genome where the parents have unequal levels of DNA methylation (nonadditive methylation). The differences in the nonadditive methylation frequencies apply to all three cytosine sequence contexts (*mCG*, *mCHG*, and *mCHH*) with *mCG* predominantly showing an increase and *mCHH* a general decrease. In some cases, we were able to track the parental chromosomes by virtue of SNPs in the DNA sequence. We identified processes causing the nonadditive methylation levels in the hybrids that we termed Trans Chromosomal Methylation (TCM) and Trans Chromosomal deMethylation (TCdM). These processes are frequently associated with the action of 24-nt siRNAs resulting in changes of methylation according to the contributions of siRNAs from both parental sources. We describe several examples showing the effects the altered methylome of the hybrids can have on patterns of gene expression.

**Results**

We compared the methylomes of C24, Ler, and their reciprocal F1 hybrids by generating >90 million reads on bisulphite-treated DNA from each genome. The data provided at least two-read coverage of >68% of cytosines (Cs) in the genome (Dataset S1, Table S1). Both parental genotypes have similar frequencies of highly methylated C residues (*mCs*) in each of the three *mC* contexts—~30% *mCG*, ~8% *mCHG*, and ~10% *mCHH* (Dataset S1, Tables S1 and S2), consistent with the hybrids inheriting *mCs* common to both parents and sites unique to one of the parents.

In both parents and hybrids, the majority of *mCG* sites are highly methylated (>80%), whereas *mCHG* and *mCHH* sites display a broader distribution of methylation (Fig. S1A). Chromosomal distributions of *mC* mirrored the patterns of 24-nt siRNA density with the highest levels associated with the pericentromeric regions of the chromosomes (Fig. L4 and Fig. S2). In generic regions, the frequency of *mCG* sites is high in the gene body and lower in the flanking regions, whereas *mCHG* and *mCHH* sites are highest in flanking regions and lower throughout the
gene body (Fig. S1B). The bulk of the chromosomal distribution and levels of methylation are similar in parents and hybrids, paralleling the data for siRNA in this hybrid system (18). Localized differences were observed where siRNA expression in the hybrid deviated from the expected midparent value (MPV; average of the two parents), most frequently at loci with large differences in siRNA levels between parents. To examine localized changes in mCs, we used a cluster approach identifying regions containing several mCs within localized segments (SI Materials and Methods). Two 1°C cluster datasets were produced for each of the parents and F1 hybrids; the first used a more stringent cluster criterion (SI Materials and Methods), defining a cluster to have a minimum of six mCs with the distance between two adjacent mCs no greater than 20 nt. The second dataset involved more of the 1°C-seq data by defining clusters containing at least three mCs with no greater than 20 nt between neighboring mCs. Both approaches yielded similar results (see Dataset S2 for a comparison) and, as such, the 3°C cluster dataset was used for further analyses so as to use the greatest 1°C-seq coverage. Within this dataset, the parental and hybrid samples contained an average of ∼152,000 clusters (Dataset S1, Table S3) located in transposons (TEs), gene bodies (transcribed regions of genes), flanking regions of gene bodies (±1 kb), or in intergenic regions (Fig. S3A and SI Materials and Methods). TEs have the most extensive stretches of methylation (Fig. S3B), whereas the flanking regions show elevated 1°C densities for both intergenic and TE clusters (Fig. S3B). Intragenic regions generally contain comparatively short and sparsely methylated clusters (Fig. S3B).

C24 and Ler Accessions Have Localized Differences in Cytosine Methylation. To compare parental methylation we identified ∼114,000 1°C clusters with coverage in both parents (Dataset S1, Table S4). Each cluster was allocated a methylation score (0 = Low to ≥5 = High) dependent on its average methylation for all mCs and for each of the methylation contexts (Dataset S1, Table S5; see SI Materials and Methods). Clusters were considered differentially methylated if their scores differed by two or more units (≥2).

Between the two parents, ≥23% of mC clusters were differentially methylated based on total 1°C levels (Dataset S1, Table S4). When individual contexts were analyzed independently, 1°C CHH was most frequently differentially methylated (35% of clusters) with 1°C CG and 1°C CHG showing 21% and 18%, respectively (Dataset S1, Table S4). The preponderance of CG methylation is sufficient in many instances to mask the changes in 1°C CHH and 1°C CHG in clusters. Clusters differentially methylated between parents were enriched in gene bodies and their flanking regions and were underrepresented in TEs (Fig. 1B and Fig. S3C). These findings are consistent with an earlier tiling array analysis, which showed that methylation differed between the Col and Ler accessions in gene bodies but not in TEs (13).

F1 Hybrids Show Locus Specific Nonadditive Methylation. We identified 76,496 C24 1°C clusters that had coverage in Ler and at least one hybrid and 68,893 Ler 1°C clusters that had coverage in C24 and at least one hybrid (Dataset S1, Table S6; see SI Materials and Methods). Hybrid clusters were assigned a methylation score for both the expected MPV and observed 1°C level with nonadditive methylation identified if scores between the MPV and observed value differed by ≥2 (Dataset S1, Table S5 and SI Materials and Methods). The majority of clusters identified as either additive or nonadditive showed a similar state in both reciprocal hybrids (Dataset S1, Table S7). Seven percent of the clusters showed nonadditive methylation in total 1°C levels, and 20% showed nonadditive methylation in at least one 1°C context (Dataset S1, Table S6). 1°C CHH levels were nonadditive in ≥22% of clusters. 1°C CHG in ≈8%, and 1°C CG in ≈5% (Dataset S1, Table S6). Nonadditive 1°C clusters were enriched in gene bodies and flanking regions and underrepresented in TEs (Fig. 1C and Fig. S3C). Differential methylation between parents was also enriched in gene bodies and flanking regions (Fig. 1B), suggesting a correlation between differences in parental methylation and nonadditive methylation in hybrids. To determine whether this correlation was the case, hybrid clusters were divided into 10 groups based on differences in methylation between parents (Fig. 2 and Dataset S2). The frequency of nonadditive methylation in the hybrids increased when parental methylation levels were different. Clusters showing nonadditive 1°C CHG had approximately equal frequencies above and below the MPV, whereas nonadditive 1°C CG clusters were predominantly above MPV and nonadditive 1°C CHH clusters predominantly below MPV.

Alterations to the Hybrid Methyleome Occur Through Trans Chromosomal Interactions. Nonadditive methylation in the hybrids indicates alterations to the 1°C levels of one or both parental epialleles. Nonadditive methylation in the hybrid may be achieved in a number of ways, for instance, a nonadditive increase in methylation can arise either through an increase in methylation of the low parent, the high parent, or both parental epialleles. To determine what patterns of change in allelic methylation most frequently lead to nonadditive methylation in the hybrid, we tracked 5,479 1°C clusters where the methylation pattern for each parental allele could be followed by using SNPs within the sequenced reads of the C24 × Ler hybrid (Dataset S1, Table S8;
see SI Materials and Methods). Consistent with the general cluster analysis, the majority of SNP clusters showing nonadditive methylation in the hybrid occurred most frequently when parents had different levels of methylation (Fig. 3). Additive methylation in the hybrids was almost exclusively due to parental epialleles retaining their methylation pattern in the hybrid (Fig. 3 and Fig. S4). Of the possible changes to allelic methylation that would lead to nonadditive methylation, two predominant patterns emerged (Fig. 3 and Fig. S4). The first involved the high parent mC epiallele retaining its methylation pattern, whereas the low parent epiallele increased in methylation, leading to an overall gain in methylation levels. We called this Trans Chromosomal Methylation (TCM; Fig. 3 and Fig. S4). The second involves a decrease in methylation of the high parent epiallele, causing a reduction in overall methylation, which we called Trans Chromosomal deMethylation (TCdM; Fig. 3 and Fig. S4). These two trans chromosomal events accounted for 681 of the 789 (86%) nonadditive mC allelic inheritance events and represent the dominant processes leading to nonadditive methylation in the hybrids (Fig. 3).

TCM events often lead to increases in methylation levels above MPV, whereas TCdM can also be observed without SNPs in loci where the high parent methylation level drops significantly in the hybrids (Fig. S4B). Often a corresponding TCM can be scored in the allele from the low methylation parent. Among ≈5,500 mC clusters where one parent was methylated and the other unmethylated, ≈2,700 gained the observed mC levels through TCM or TCdM (Fig. S4B). All clusters showing nonadditive methylation are included within this subset of clusters. Extrapolation from our dataset provides an estimate of ≈8,500 TCM/TCdM events genome wide (Dataset S1, Table S9; see SI Materials and Methods).

Bisulphite PCR followed by Sanger sequencing showed examples of TCM and TCdM over segments longer than could be tracked through the shorter deep sequencing reads around SNPs (Fig. 4 and Fig. S5). Locus A was methylated in all C residues in C24 and had only low CG methylation in Ler (Fig. 4 and Fig. S5A). In both reciprocal hybrids, the C24 allele retained the level and distribution of methylation of the C24 parent. The Ler allele showed a dramatic change in its methylation state, the mC pattern resembling the C24 epiallele with an increase to existing mCs and methylation of previously unmethylated Cs. In the low methylation parent. Among ≈5,500 mC clusters where one parent is highly methylated, the other parent has no methylation, and the hybrid a higher than MPV mC level (Fig. S4B). Conversely, TCdM can be tracked through the shorter deep sequencing reads around SNPs (Fig. 4 and Fig. S5). Locus A was methylated in all C residues in C24 and had only low CG methylation in Ler (Fig. 4 and Fig. S5A). In both reciprocal hybrids, the C24 allele retained the level and distribution of methylation of the C24 parent. The Ler allele showed a dramatic change in its methylation state, the mC pattern resembling the C24 epiallele with an increase to existing mCs and methylation of previously unmethylated Cs.
reduction of methylation in the hybrids is associated with a loss of siRNA in the corresponding region (Fig. 4 and Fig. S5). At Locus F, the adjacent TCM and TCdM events are associated with localized increases and decreases in hybrid siRNA levels, respectively (Fig. 4 and Fig. S5A). In only one TCM region (Fig. 4, Locus F) a SNP in a siRNA sequence was present, revealing that the previously low siRNA producing Ler parent allele had increased siRNA expression in the hybrids (Fig. S5C).

siRNAs Direct Nonadditive Methylation in the Hybrids. siRNAs have a role in establishing de novo methylation in its three contexts and maintaining mCHH methylation (22). We have reported that loci with reduced levels of siRNAs in the hybrids show reduced levels of DNA methylation (TCdM; ref. 18).

We frequently noted that the distribution of methylation often extends beyond the distribution of siRNAs in our datasets. For example, Fig. 4 Locus A has an extensive region of TCM spreading into AT3G43340, which appears not to be associated directly with siRNAs but is within 250 nt of the siRNA region (Fig. 4, Locus A). When we examined all siRNA clusters in the genome normalized to 500 nt (Fig. S6A and SI Materials and Methods), we found that DNA methylation in all contexts can occur for a distance of 350–400 nt on either side of the siRNA region. In our dataset, DNA methylation in a siRNA region or within ±400 nt was considered siRNA-associated methylation.

We used the ≈5,800 mC clusters where one parent is methylated and the other unmethylated to examine the association of nonadditive methylation and siRNAs. Very few siRNA independent mCHG and mCHH clusters were found, consistent with the known close association between these methylation contexts and siRNAs (23). In agreement with previous results (13, 24, 25), only a small proportion of gene body CG methylation was associated with siRNA, but the proportion was greater in clusters showing a nonadditive increase in methylation (P < 2.6×10^{-15}; Dataset S1, Table S10). Nonintragenic clusters showing an increase in mCG were also more frequently associated with siRNA (P < 1.65×10^{-08}; Dataset S1, Table S10). As a population, the siRNA-associated mC clusters deviate above the expected MPV, whereas siRNA-independent clusters coincide with the MPV (Fig. S6B), suggesting that increases in methylation outside of genes is a result of siRNAs.

Alterations in DNA Methylation Correlate with Changes in Gene Expression. TCM and TCdM events have the potential to affect gene expression, especially if they are located close to or within a gene. Among the ≈4,900 mC clusters showing nonadditive methylation in our dataset, 398 were associated with genes and had substantial stretches of mCs (SI Materials and Methods). Fifty-four of these genes showed a ≥1.2-fold difference in expression from MPV, and the same correlative pattern between methylation and gene expression held true in parent–parent and hybrid–MPV comparisons (Dataset S2). Of these genes, 22 had methylation changes in the gene body and 32 in the flanking region. The majority (38 of 54) showed an inverse correlation between methylation levels and gene expression, whereas the remaining 16 genes showed a positive correlation. Limitations imposed by our dataset and filtering methods are likely to mean these genes are only a proportion of those in the hybrids, owing their differential expression to novel methylation patterns. However, this gene list
serves to highlight that nonadditive methylation in the hybrids apparently leads to alterations in gene expression.

Nine examples are shown in Fig. 5 and Fig. S6C, which were matched against Columbia transcriptome and methylome data from wild-type (WT) and mutants with altered DNA methylation patterns (24, 26, 27). At2g32160, At3g43340, and At5g47590 show TCM extending over the promoter region and into the gene body and having an extensive siRNA distribution correlated with a decrease in gene expression (Fig. 5). At5g38720, At4g19690, and At4g22310 show TCM in either the downstream region or in the last exon of the gene correlated with a decrease in expression (Fig. S6C). Consistent with a role for DNA methylation in controlling gene expression in At2g32160, the Columbia WT epiallele resembles the unmethylated Ler epiallele. In the rdd mutant, which lacks demethylases, this allele becomes methylated in a pattern similar to C24 and the hybrids, and gene activity is suppressed (Dataset S1, Table S11). At3g43340 is transcriptionally active in a met1 Col mutant, and shows a loss of methylation in the upstream region and first exon (Dataset S1, Table S10). In the hybrids, methylation is gained on the Ler allele in this region of At3g43340 and is consistent with a decreased transcript level of the Ler allele and of the overall expression level (Figs. 4 and 5). At5g38720 shows a similar up-regulation in Col met1 and ddc mutants with the affected methylated region corresponding to the TCM region in the hybrid and which presumably causes the decrease in gene expression (Fig. S6C and Dataset S1, Table S11).

At4g15920, At4g09490, and At3g26612 all showed increased expression levels in the hybrid correlating with localized TCdM events and are associated with reductions in levels and distribution of siRNAs (Fig. 5). A loss of methylation in the region upstream of At4g09490 in the Col met1 mutant coincides with increased expression of this gene (Dataset S1, Table S11), paralleling the TCdM event identified in the hybrid over this same region (Fig. 5).

Discussion

In our analysis of the epigenome in the progeny of crosses between the Arabidopsis accessions C24 and Lerr, we explored the relationship between two epigenetic systems, DNA methylation and the production of siRNAs. We examined the mC distribution in the genetically similar C24 and Ler accessions and in their hybrid progeny and asked whether alterations in the methylation influenced the hybrid transcriptome, which ultimately must be involved in the characteristic increase in biomass of the F1 hybrid. The parents showed significant differences between their methylomes with one-quarter of mC regions differentially methylated. These regions were enriched for genes and their flanking regions and underrepresented in transposable elements consistent with comparisons between other Arabidopsis accessions (13, 14). The hybrid had significant changes in methylation levels at localized regions of the genome, frequently corresponding to those chromosomal segments where methylation levels were markedly different in the C24 and Ler parents. Seven percent to 20% of mC clusters had nonadditive changes either above or below the expected MPV in either Total mC or at least one mC context. We identified two major processes associated with nonadditive methylation, TCM, which led to methylation levels greater than the MPV, and TCdM, which produced methylation levels lower than the MPV. Both of these processes are involved in a change in methylation of one parental allele to resemble that of the other parental allele with the de novo methylation or demethylation events superimposed on the efficiently and accurately maintained background inherited from the parental genotypes. These processes affected all mC contexts, with TCM-generated increases in methylation more frequently associated with 5mCG and TCdM-generated decreases in methylation most obvious in 5mCHH.

Transallelic interactions have also been reported in para-mutation (28, 29), certain cases of self-incompatibility (30), si-RNA-directed flowering Wageningen A (FWA; refs. 31 and 32), P4H locus transmethylation (33), and remethylation and demethylation of loci in epiRILs (34–36). In most cases, siRNAs appear to play a central role. Many alterations to the hybrid methylome through TCM and TCdM are associated with siRNAs. The previously reported reduction in 24-nt siRNA levels in the C24/Ler hybrid system correlates with the decrease in 5mCHH levels (18), which depends on siRNAs for continued maintenance. The reduction of siRNA levels in the hybrids has a lesser effect on 5mCG and 5mCHG, which have siRNA-independent maintenance of methylation via METHYLTRANSFERASE 1 (MET1; ref. 37) and CHROMOMETHYLASE 3 (CMT3; ref. 38). MET1 and CMT3 activity may also explain the more frequent nonadditive increases in CG and CHG methylation compared with CHH, because even low levels of de novo RNA-directed DNA methylation in the hybrid would be efficiently maintained by these methyltransferases.

Many siRNA-associated loci show additive inheritance of allelic methylation, suggesting that other factors may determine the ability of a locus to undergo TCM or TCdM; the chromatin state of a region may impose the level of siRNAs needed to establish or maintain DNA methylation. TCM events occur when the threshold is exceeded and TCdM events occur below threshold levels, possibly resulting from a dilution of siRNAs over the two alleles in the hybrid (39). We also found a large subset of 5mC clusters exhibiting TCM/TCdM in the absence of
siRNAs, these being primarily associated with gene body methylation. Mechanisms independent of siRNAs also may function to alter the hybrid methylome. In large part, the chromosomal segments with altered methylation levels and patterns in the hybrid were concentrated in genes and in their flanking regions. We were able to identify a number of genes where altered levels and distribution of siRNAs and resulting DNA methylation corresponded with altered gene transcription. Most of these examples fit the general rule that methylation is suppressive of transcription. Although we described only a small number of examples, these paralleled results in mutants affecting methylation levels. It is not clear from our limited data what proportion of the altered hybrid transcriptome is directly caused by changes to siRNA and methylation levels. Limitations of our study are that we sampled only one developmental time point and that the data were from a mix of tissues.

Other reports indicate that alterations to epigenetic systems occur in hybrids (7, 17, 18). If these systems contribute to heterosis, then a loss of maximum vigor in generations beyond the F1 would be due not only to the segregation of alleles and epialleles but also to possible alterations to the epialleles in the F2 and beyond. For example, a proportion of loci we examined had low levels of TCM, which may be either lost or increased in subsequent generations reminiscent of transgenerational changes in epiRLs (17, 34–36).

Our analysis indicates that changes in gene expression in a hybrid are influenced not only by differences in DNA sequence between parents, but also by variation in gene-associated siRNA and DNA methylation marks. It is likely that TCM and TCDM occur in all hybrid systems. These epigenetically controlled changes may be among the initial events in the cascades of altered gene expression that contribute to hybrid heterosis. It will be necessary in different hybrid combinations to analyze specific tissues at a number of developmental stages to gain a better understanding of how the epigenome contributes to hybrid vigor.

Materials and Methods

Experimental designs and plant material used are described in SI Materials and Methods. MethylC-seq libraries were prepared from immature floral buds with sequencing performed by Geneworks on the Genome Analyzer Ix (Illumina). miRNA-seq was carried out by the Biomolecular Resource Facility at The Australian National University on the Genome Analyzer Ix (Illumina). Processing of siRNA, methylation, transcription, and methylated regions are described in SI Materials and Methods. Raw and processed mapped MethylC-seq sequences are deposited in GEO with accession number GSE35542.

ACKNOWLEDGMENTS. We thank Alhua Wang and Limin Wu for technical assistance in the laboratory. This work was supported by National Collaborative Research Infrastructure Strategy and Science and Industry Endowment Fund funded by the Australian Government and by the Agricultural Genomics Centre, New South Wales Government BioFirst Initiative.

**Supporting Information**

Greaves et al. 10.1073/pnas.1201043109

**SI Materials and Methods**

**Plant Material.** Hybrids between *Arabidopsis thaliana* C24 and Landsberg erecta (Ler) were generated by hand pollination. Seeds were sterilized, stratified for 2 d at 4 °C, and sown onto MSN medium, pH 7 (MS medium supplemented with 0.6% wt/vol Noble agar). At 18 d, plants were transferred to soil (Debco Seed Raising Mix supplemented with 1 g/L Osmocote Extract Mini patterned release fertilizer pellets) and grown in a 16-h day cycle at 22 °C/18 °C (day/night). Light intensity was maintained between 120-190 μmol·m⁻²·s⁻¹ by using fluorescent tubes and incandescent bulbs. Immature floral buds were collected from several plants per sample, frozen in liquid N₂, and stored at −80 °C. For seedling analysis at 15 days after sowing whole seedlings were frozen in liquid N₂ and stored at −80 °C.

**Bisulfite PCR.** Five hundred nanograms to 2 μg of genomic DNA from immature floral buds was isolated using the Plant DNeasy Minikit (Qiagen). Isolated DNA was bisulfite converted by using MethylEasy Xceed kit (Human Genetic Signatures). Bisulfite PCR was carried out (1) by using unbiased bisulfite primers, allowing the amplification of both methylated and unmethylated fragments (Dataset S1, Table S12). Annealing temperatures for primers were set at either 50 or 55 °C. At least eight clones were sequenced for each sample, and the level of methylation was analyzed and visualized by using Microsoft Excel 2007. For allelic discrimination of hybrid methylation, at least seven clones of each allele were analyzed.

**MethylC-seq.** MethylC-seq was carried out as described in Lister et al. (2). Immature floral bud genomic DNA was extracted by using the plant DNeasy minikit (Qiagen) and stored at −20 °C. DNA was sonicated to a size of 200–500 bp. DNA was end-repaired and ligated with Illumina methylated adapters. DNA was run on a gel and size selected by gel purification (≈300 bp in size) before undergoing bisulfite conversion by using the MethylEasy Xceed Kit (Human Genetic Signatures). Libraries were amplified, and single 75-bp reads were obtained on a Genome Analyzer II by Geneworks. The bisulfite conversion efficiency was checked by analyzing the conversion of chloroplast DNA using standard Bisulfite PCR and Sanger sequencing (3).

**mRNA-seq.** Total RNA was extracted from 15-d-old seedlings for C24, Ler, C24 × Ler, and Ler × C24 (24 plants per sample) by using the RNeasy minikit (Qiagen). mRNA-seq libraries for Illumina sequencing were then prepared by the Biomolecular Resource Facility, The Australian National University. Approximately 25 million single 75-bp reads were obtained for each sample and mapped to the *Arabidopsis* genome (TAIR 9) by using TopHAT. Expression levels of genes were validated by using qRT-PCR, except for At3g34330 and At5g45790. At3g43340 is near identical to At2g39140 not allowing a reliable comparison of the gene expression level via qRT-PCR. At5g45790 had low expression levels and could not be reliably analyzed. qRT-PCR results were not in full agreement with mRNA-seq data for At4g22310 and At5g38720.

**qRT-PCR.** Total RNA from 15-d-old seedlings was extracted by using RNeasy Plant Mini Kit (Qiagen) and treated with DNase I amplification grade (Invitrogen). Two micrograms of total RNA was used to perform cDNA synthesis (20 μL) by using oligo dT primers and SuperScript III RNaše Reverse transcriptase (RT; Invitrogen). Each qRT-PCR was performed by using SYBR green and 0.25 μL of the RT product by using platinum Taq (Invitrogen). All gene expression levels were compared with the reference gene At4g26410 (Dataset S1, Table S12). qRT-PCR conditions were as follows: step 1: 94 °C for 10 min; step 2: 95 °C for 20 s, 60 °C for 20 s, and 72 °C for 20 s (45 cycles). The Corbett RG-6000 machine was used to run all qRT-PCR reactions.

**Mapping and 13C Cluster Formation.** Seventy-five nucleotide MethylC-seq reads were trimmed and mapped to the TAIR9 *A. thaliana* Columbia reference genome by using BSMAP (4). Mapping was accomplished by using a reiterative single nucleotide trimming of the read until either a successful hit was found or 30 nt had been removed, resulting in >90% of the reads mapped with >60% mapped uniquely. All reads with identical genomic location and sequences were removed before analysis of the mapped data. An error rate was determined by the number of uniquely mapped methylated Cs to the number of uniquely mapped Cs on the Chloroplast genome. A bisulfite conversion rate of 98% for all samples was determined. True methylated Cs were determined by a P value under the binomial distribution $P = \text{binomial} (n, m, \text{error rate})$, where $m = \text{number of methylated reads}$, $n = \text{number of unmethylated reads}$; $x = m + n$, which is the sequencing depth on a single C, “error rate” was the error rate for nonconversion of the samples. True 13C cytosines were determined as any C with a P value under the binomial distribution of 0.0025 (p cutoff). This P value depended on the coverage of reads over that given C. 13C clusters were formed by using MergeBed from the BEDTools package version 2.25 (5). Clusters were defined by using multiple combinations of the minimal number of 13C required (ranging from 3 to 6) and the minimal distance between each 13C (ranging from 15 to 25 nt). We settled on examining two representative cluster sets: (i) clusters formed when at least three 13C were within 20 nt of the adjacent 13C; and (ii) clusters formed when at least six 13C were within 20 nt of the adjacent 13C (Dataset S2). The first approach, yielded ≈150,000 13C clusters for each sample. Cluster methylation levels were calculated as the average methylation level of the 13C within the cluster, with unique values attained for total (All 13C), 13CG, 13CHG, and 13CHH contexts. Depending on the level of coverage, or the presence of 13C, cluster lengths and individual 13C cluster methylation levels were different for each sample. As a result, a two-way analysis was undertaken first, comparing regions that were methylated in C24 (referred as C24-identified), and second, comparing regions that were methylated in Ler (referred as Ler-identified). This approach resulted in two datasets for the parental comparison and four datasets (two for each hybrid) for the hybrid analysis that were analyzed independently. Numbers through the paper are represented as the average from the different analyses with the number of clusters in each class found in Dataset S1, Tables S3, S4, and S6. All analyses gave consistent results and patterns. For comparative analysis between parents or between parents and hybrids, clusters with 75% coverage of the same C were used. In such cases, where clusters had between 75 and 99% coverage of the same C, the Cs not covered in all samples were trimmed from the analysis and not used to establish a cluster methylation level. Any cluster with <75% of the same C covered between samples was not used in any comparative analysis.

Clusters were allocated a methylation index value ranging from 0 (low/unmethylated) to ≥5 (highly methylated) based on the average methylation of the 13C's contained within (Dataset S1, Table S12).
SNP Methylation Analysis. SNPs within the MethylC-seq reads were identified by using the genomic sequences for the C24 and Ler accession obtained from the 1,001 genomes website (www.1001genomes.org). This analysis resulted in 220,778 SNP containing reads through which the mC cluster was identified by using a script invoking various commands from the BEDTools package version 1.2 (5). Assigning Clusters to a genomic feature was prioritized by TE > Gene body > ≥ 1 kb from gene > Intergenic. If a mC cluster overlapped two genomic features, such as gene body and upstream, the mC cluster was assigned based on the feature that the mC cluster overlap was the largest. For profiling purposes, clusters that had different levels of methylation between parents, or hybrids and MPV, yet still had overall high levels of methylation (mC index value ≥ 4) were not considered differentially methylated because these regions would still have levels of methylation likely to result in a repressive state. To determine whether a statistical enrichment for differentially methylated clusters in particular features occurred the results were subjected to a Fisher exact test (P < 0.05) by using the R software package.

Identification of Candidate Genes That May Be Regulated Through DNA Methylation. Flanking regions. We identified 1,155 clusters that were nonadditively methylated and in the ±1-kb flanking regions of genes. These genes were then filtered for the number of mC (≥ 5), differential expression between parents (≥ 1.1-fold difference), and hybrid gene expression different from MPV (≥ 1.2-fold difference). This analysis resulted in 75 candidate genes that could be potentially altered in hybrids because of nonadditive methylation. This list was then manually filtered by examining the relationship of methylation to gene expression. Only those genes where the relationship between methylation and gene expression was consistent between both parent–parent and hybrid–MPV were analyzed resulting in 32 candidate genes with altered expression levels in the hybrid that may be due to nonadditive changes in methylation.

Gene body. To study the effect of methylation changes in the gene body, we selected all genes that contained a mC cluster (C24 × Ler = 13,125; Ler × C24 = 13,542). All mC clusters within the same gene were merged and an overall methylation level produced for the gene as a whole. These original clusters were then filtered for the number of mC [10–20 depending on the context present, as each context was weighted differentially because of presumed differences in suppressive activity in a gene (i.e., CHG is strongly suppressive, whereas CG is less so)], hybrid gene expression different from MPV (≥ 1.2 fold difference), and hybrid differential methylation, resulting in 323 candidate genes. These genes were then manually filtered by studying total methylation levels, numbers of C’s, and the relationship between methylation and gene expression in the parents and between the parents and hy-
brid. This filtering left 23 candidate genes with differential expression in the hybrid that may be due to the associated changes in gene body methylation.

Software. R-script, BEDTools, and AWK programming language were used for all bio-informatics analysis done in this paper. Final graphs were generated in Microsoft Excel 2007.


Fig. S1. Genomic methylation. (A) Histogram of methylation frequencies on the y axis for different levels (%) of CG (blue), CHG (green), and CHH (red) methylation (x axis). (B) Methylation levels across genic regions (Gene body and ±1 kb) for CG (blue), CHG (green), and CHH (red).
Fig. S2. Distribution of 23/24nt siRNA clusters, \(^{30}C\) read coverage, and \(^{30}C\) frequencies across all five *Arabidopsis* chromosomes of both accessions and the reciprocal hybrids. Gene, TE, and repeat densities are represented as a heat map underneath each chromosome.
Fig. S3. (A) Genomic distribution of methylated clusters (for details of the method of assigning a $\text{mC}$ cluster to genomic features, see SI Materials and Methods). (B) Length and density (methylated C's / total C's in cluster) of methylated clusters in different genomic features. (C) Regions of differential methylation between the parental ecotypes C24 and Ler. (D) Genomic distribution of nonadditive methylation in hybrids. Two graphs each are shown for C24-identified and Ler-identified clusters. Arrows indicate statistical enrichment or underrepresentation of differentially methylated $\text{mC}$ clusters in different genomic features (Fisher exact test; $P < 0.05$). (E) Genomic distribution of methylated clusters analyzed in the hybrids.
Fig. S4.  (A) Average methylation levels of clusters shown in the permutation tables in Fig. 3. Red graphs denote allelic combinations expected to result in increased methylation, blue graphs denote allelic combinations expected to result in decreased levels of methylation, and gray graphs denote allelic combinations expected to result in additive levels of methylation. Error bars = SEM.  (B) Average methylation levels of hybrid clusters when one parent is highly methylated, the other parent has no methylation, and the changes in hybrid methylation levels can only be attained by TCM or TCdM.
Fig. S5. Examples of altered allelic methylation in the hybrids through TCM or TCdM. (A) Allelic methylation levels for Ler x C24 hybrid (see Fig. 4 for parents and C24 x Ler). (B) Example of a low TCM event (Locus C) and a decrease in $^{m}$CG at several sites in the C24 x Ler hybrid (Locus E). $^{m}$CG (blue), $^{m}$CHG (red), and $^{m}$CHH (green) methylated residues are plotted along the x axis (each tick-mark equalling 100 nt), with the y axis representing the methylation level. The distribution and abundance of siRNA reads are plotted in red above each graph and are duplicated between the two alleles of the hybrids. Regions undergoing TCM and TCdM events are underlined in red and blue, respectively. (C) The number of siRNA reads over a SNP present at Locus F (arrow).
Nonadditive methylation is associated with siRNAs. (A) The frequency of given cytosine being methylated within a siRNA clusters and ±1 kb flanking regions. Red, the cluster region; Green, adjacent flanking 200 nt; Blue, remaining 800 nt. Average methylation levels are indicated in the top graphs. n: Clusters = 18,000; Flanking regions = 9,000 (refer to SI Materials and Methods). (B) Average mCG levels of siRNA-associated and siRNA-independent hybrid mC clusters plotted against the expected MPV showing an above expected trend in mCG in siRNA-associated clusters. Dot colors denote categorization of mC status in the hybrid: red, nonadditive increase; blue, nonadditive decrease; gray, additive methylation. (C) Examples of genes correlating with altered methylation levels in the hybrids. mC clusters were located in or within ±1 kb of the described genes. Gray histograms show average methylation level of the mC cluster (highlighted in yellow) and levels of gene expression obtained from a mRNA-seq data. The next set of graphs depict the distribution and abundance of siRNAs (red) and mC (blue) in both parents and each reciprocal hybrid. Arrows at the bottom of each example show gene orientation or location of the neighboring gene. Numbers at each extreme of the regions show genomic position. IC, insufficient coverage.

Fig. S6.
Fig. S7.  (A) Allelic inheritance of clusters containing SNPs that were reported as nonadditively methylated in the whole dataset. mC clusters with a nonadditive increase in methylation are enriched for allelic contributions that would result in an increase in overall hybrid methylation levels. mC clusters with a nonadditive decrease in methylation are enriched for allelic contributions that would result in an overall decrease in hybrid methylation.  (B) Average allelic methylation levels of clusters reported in A. Red denotes allelic combinations that would result in an increase in methylation, blue allelic combinations that would result in a decrease in methylation, and black allelic combinations that would result in additive methylation levels in the hybrid.

Dataset S1, Table S1.  Overview of MethylC-seq data

Dataset S1 (XLSX)

Dataset S1, Table S2.  Contextual methylation level differences

Dataset S1 (XLSX)

All C's were divided into unmethylated [not significantly different from false methylation rate (FMR; as determined by the chloroplast conversion rate)], semimethylated (statistically different from FMR and <80%) and highly methylated (>80%). CG methylation is either unmethylated or highly methylated. CHG methylation is evenly distributed between semimethylated and highly methylated. Most CHH sites are considered semimethylated.
Dataset S1, Table S3. Shows the number of clusters formed for each sample

A cluster was formed when at least three \(^{m}C\) were within 20 nt of each other. Cluster characteristics are similar between all four samples.

Dataset S1, Table S4. Differential methylation between C24 and Ler accessions

Dataset S1, Table S5. Methylation index was used to categorize cluster methylation levels into different categories

The methylation index values differed for each context as CG, CHG, and CHH methylation have different frequencies of methylation levels (Fig. S1).

Dataset S1, Table S6. The number of nonadditively methylated clusters in Arabidopsis hybrids

Only clusters with 75% or greater of the same \(^{m}C\)'s covered in parents and at least one hybrid were analyzed. Nonadditive methylated clusters were determined if the methylation index score differed by \(\geq 2\) from the MPV. *Nonadditive \(^{m}C\) clusters in the "Any context" category refers to clusters having at least 1 context with altered methylation.

Dataset S1, Table S7. Both hybrids show similar trends in either additive or nonadditive methylation at the same locus

Dataset S1, Table S8. Raw data on \(^{m}C\)'s trackable via SNPs contained in the same reads, number that matched our clusters and the number of \(^{m}C\)s that past the filter and were used for analysis (Fig. S8)

Dataset S1, Table S9. Number of possible TCM and TCdM events in an intraspecific hybrid cross between C24 and Ler accessions

Dataset S1, Table S10. Numbers of \(^{m}C\) clusters associated with siRNAs and independent of siRNAs together with whether they were nonadditively or additively methylated

Averages are given for all clusters and then clusters divided into intragenic and nonintrageneric.

Dataset S1, Table S11. Pattern of candidate gene's methylation level and siRNA level in Col

Data based on ref. 2. Table shows siRNA level in Col, the level of methylation, and which accession (C24 or Ler) it was similar to. If Columbia was unmethylated, then there is no way it can be altered in mutants of genes in the methylation pathways (N/A). *At2g32160 was down-regulated in a rdd mutant.
Dataset S1, Table 12. Primer sequences

Dataset S1 (XLSX)

Dataset S1, Table 13. Filtering table for SNP analysis

Dataset S1 (XLSX)

This table shows the original number of C's present in a cluster and the number required to be covered in the SNP analysis for us to study a particular cluster. This filtering system enabled a good correlation between original cluster methylation levels and SNP cluster methylation levels (Fig. S9) and resulted in 5,479 clusters that were analyzed during the SNP analysis.

Dataset S2. The MethylC-seq analysis was done by using two different clustering approaches

Dataset S2 (XLSX)

The first produced clusters with a minimum of three mC (each had to be within 20 bp of the next mC), while the second approach produced clusters with at least six mC (each had to be within 20 bp of the next mC). (A–H) Hybrid clusters were divided into 10 groups based on the similarity of methylation levels between parents (H, highly methylated; SH, semi-high; SL, semi-low; L/U, low or unmethylated—refer to Dataset S1, Table S5 for respective index scores). (A–D) The three mC-derived clustering method. (E–H) The six mC-derived clustering approach. Both clustering approaches show the frequency of nonadditive methylation correlates with increased differences in parental methylation levels. (I and J) Profiling of the three mC clustering approach and the six mC clustering approach. Both approaches gave consistent results concerning genomic locations of additive and nonadditive mC clusters. (K) A list of 54 candidate genes whose altered hybrid expression levels may be a result of altered DNA methylation.