Supplementary information

Methods

Mutant screen, mapping and complementation testing:

The originally described *clark kent* mutants could not be utilized for a mutant screen because they spontaneously revert to wild type at a rate of approximately 3% [S. E. Jacobsen, E. M. Meyerowitz, *Science* **277**, 1100-3 (1997)]. However, we found that introducing additional transgenic copies of the *SUP* locus could eliminate this reversion. To create an appropriate starting strain for a mutant screen we backcrossed the *clark kent-3* allele to the the wild type Ler three times and then transformed this line with a 6.7 kb fragment of *SUP* genomic DNA [S. E. Jacobsen, E. M. Meyerowitz, *Science* **277**, 1100-3 (1997)], using *Agrobacterium*-mediated transformation. We selected a line that was homozygous for a single transgene locus. After two generations of self-pollination, this line exhibited a strong and stable *clark kent* mutant phenotype (no spontaneous revertants seen in 2200 plants). This stabilized *clark kent* line (referred to as line *clk-st*) was then mutagenized for 16 hours with 0.3% ethylmethanesulphonate and 621 individual M2 families were screened for plants with a wild type *SUP* floral phenotype. Each mutant was backcrossed twice to line *clk-st* to eliminate extraneous mutations and to confirm recessive inheritance. Complementation tests were performed by crossing plants heterozygous for the *cmt3-7* mutation with plants heterozygous for the other mutations. Lack of complementation was evident by a segregation of a 3:1 ratio of *clark kent*:wt plants in the F1 progeny.

The *cmt3-6* mutation, originally named *s604*, was mapped by crossing a plant heterozygous for *s604* in the Ler ecotype to a plant heterozygous for the *sup-2* mutation in the Columbia ecotype. Several F1 plants with a *sup* phenotype were selected, and these were screened for F2 families that segregated a 13:3 ratio of *sup*:wild type plants. F2 plants with a wild type *SUP* phenotype were used for mapping using a series of PCR based markers covering the five *Arabidopsis* chromosomes. These markers were derived from the collection of DNA polymorphisms provided by Cereon Genomics. The *s604* mutation mapped to the bottom of chromosome I, between markers CER464694 and CER448434. The *CHROMOMETHYLASE3* (*CMT3*) gene lies in this genomic interval. Using a marker that is

contained within the *CMT3* gene, no recombinants were detected, showing that *s604* and *CMT3* are closely linked. We thank Cereon Genomics for information on molecular markers.

The *cmt3-2* complementation test was performed by crossing a plant heterozygous for *cmt3-7* in the *clk-st* background with a plant homozygous for *cmt3-2*. The F1 plants were PCR genotyped to determine which ones carried the *cmt3-7* mutation. All F1 plants that carried the *cmt3-7* mutation had lost CpXpG methylation at the *SUP* locus as determined by bisulfite sequencing. However, plants that did not carry the *cmt3-7* mutation retained methylation at *SUP*. As a positive control we crossed a *cmt3-7* heterozygote to wild type and found that *SUP* methylation could be detected in the F1 plants. As a negative control we bisulphite-sequenced wild type and found that *SUP* methylation was not detected.

PCR based markers.

The molecular markers used in this study were CAPs markers [A. Konieczny, F. M. Ausubel, *Plant J* **4**, 403-10 (1993)] composed of the following combinations of oligonucleotide primers and restriction enzymes: *cmt3-4* (5'-TTAATCTTAATATAAATGACTATGAACGGGTT -3' and 5'-AAAGAGAATTAAAACTCACTTGCATGATTCC -3' with *Mnl*I), *cmt3-5* (5'-GACGAAGTAAGTTCCTTGAATCC -3' and 5'- ACAGTTTCCCATTAGCCCC -3' with *Taq*I), *cmt3-6* (5'-CAAGTGAGGAATGGAATGATGGCGGCTGGGGCTTAT -3' and 5'- TTTTGAAATGGCGTTGTGGG -3' with *Alu*I), and *cmt3-7* (5'- TTGACTACCCCGGGAATGAACCCATTTGT -3' and 5'- GATCTGCAACAAATCTCAGC -3' with *Cac*8I).

Bisulphite sequencing.

Bisulfite sequencing was performed as previously described [S. E. Jacobsen, H. Sakai, E. J. Finnegan, X. Cao, E. M. Meyerowitz, *Curr Biol* **10**, 179-186 (2000)], except that PCR products were cloned using the TOPO TA cloning kit (Invitrogen), and individual clones were sequenced. Primer sequences and cycling conditions for each experiment are available upon request. The *met1* mutant used in these experiments was previously described as line number 170 [S. E. Jacobsen, H. Sakai, E. J. Finnegan, X. Cao, E. M. Meyerowitz, *Curr Biol* **10**, 179-186 (2000)]. This mutant was induced in the Columbia ecotype and then crossed into the Landsberg *erecta* ecotype five times. The line was then allowed to self-pollinate for three generations, at which time it developed a strong hypermethylated *superman* (*clark kent*) phenotype.

Since the *SUP* transgene present in *clk-st* is identical to the endogenous gene, we could not distinguish them, and the data are a therefore a mixture of both. The complete region analyzed within the top strand of the *SUP* gene in Web Figure 1 corresponds to positions 992 to 2019 in GenBank accession AB025608. The region labeled SUPERMAN 5' region in Fig 2A and Web Table 1 corresponds to 992 to 1353, and the region labeled SUPERMAN 3' region corresponds to 1276 to 2019. The region analyzed in the *FWA* gene was the top strand from positions 1250 to 1553 in GenBank accession AF178688.

The Athila retrotransposon LTR sequence was analyzed because it showed a high frequency of CpXpG sites. Using MacVector 7.0 (Oxford Molecular Group) the Athila LTR sequence analyzed by bisulfite sequencing was found to contain 2.6 times more Cp(A/T)pG sites than would be expected based on the base composition of the sequence. Athila is a middle repetitive sequence. To avoid complications of analyzing several Athila LTR sequences simultaneously, we took advantage of a particular juxtaposition of an LTR sequence next to a 180 bp centromeric repeat sequence in clone F26M13 on chromosome V (GenBank accession AB046429). For the PCR amplification step of the bisulfite sequencing protocol, one primer was placed in the LTR sequence and the other in the centromeric

regions, region A which is the bottom strand corresponding to positions 10062 to 10336, and region B, which is the top strand corresponding to the positions 10304 to 10475 in AB046429. The region of the 180 bp repeat sequence analyzed corresponded to positions 10508 to 10652 in AB046429. To isolate a single 180 bp repeat sequence, we used the same strategy of anchoring one set of primers to the adjacent Athila LTR sequence.

Southern blots

The Athila LTR probe was derived by PCR from genomic DNA corresponding to positions 9747 to 10350 in GenBank accession AB046429. The 180 base pair centromeric repeat probe was kindly provided by Eric Richards, and is described in [A. Vongs, T. Kakutani, R. A. Martienssen, E. J. Richards, *Science* **260**, 1926-8 (1993)].

Northern blots

The Athila probe was generated by amplifying genomic DNA corresponding to positions 67757 to 70045 in BAC clone F7N22 (GenBank accession AF058825). The Athila sequence on this BAC is similar to the one analyzed by bisulfite sequencing (80-90% nucleic acid sequence identity in the most homologous regions). The Ta3 probe was generated by amplifying genomic DNA corresponding to positions 83448 to 88130 in BAC clone F28L22 (GenBank accession AC007505). Blots contained equivalent amounts of RNA as judged by ethidium bromide staining of the rRNA bands.

Table 1. Number of cytosines methylated in different sequence contexts within cloned PCR products of bisulfite treated DNA.

	CpXpG	CpG	Asym*
SUPERMAN 5' region 362 nt (15 clones/genotype).			
Total number of sites	135	15	1080
Number methylated			
Line *clk-st*	78 (58%)	0 (0%)	167 (15%)
cmt3-7	0 (0%)	0 (0%)	11 (1.0%)
met1	77 (57%)	0 (0%)	127 (12%)
SUPERMAN 3' region 744 nt (15 clones/genotype).			
Total number of sites	315	135	2115
Number methylated			
Line clk-st	157 (50%)	22 (16%)	338 (16%)
cmt3-7	0 (0%)	12 (8.9%)	167 (7.9%)
met1	141 (45%)	5 (3.7%)	324 (15%)
Athila LTR region A 275 nt (10 clones/genotype).			
Total number of sites	140	80	220
Number methylated			
Line *clk-st*	113 (81%)	80 (100%)	12 (5.5%)
cmt3-7	0 (0%)	62 (78%)	6 (2.7%)
met1	63 (45%)	11 (14%)	14 (6.4%)
Athila LTR region B 172 nt (27 clones/genotype).			
Total number of sites	135	54	621
Number methylated			
Line *clk-st*	67 (50%)	43 (80%)	20 (3.2%)
cmt3-7	2 (1.5%)	46 (85%)	23 (3.7%)
met1	47 (35%)	9 (17%)	23 (3.7%)
180 bp repeat 145 nt (12 clones/genotype).			
Total number of sites	24	24	120
Number methylated			
Line *clk-st*	9 (38%)	17 (71%)	6 (5.0%)
cmt3-7	0 (0%)	20 (83%)	11 (9.2%)
met1	3 (13%)	6 (25%)	21 (18%)

*Asym (Asymmetric) is defined by cytosines within the context

Web Figure 1: Detailed methylation pattern at the *SUP* locus for *clk-st* (red), *cmt3-7* (blue), and *met1* (green). "M's" denote methylated cytosines from 15 individual cloned PCR products of bisulfite treated genomic DNA. The 27 CpXpG sites are shown in red and highlighted by red asterisks. The 9 CpG sites are shown in blue and highlighted by blue asterisks. The 1028 nucleotide sequence shown was assembled from 30 clones arising from two separate PCR products, described in Fig. 2 and Web Table 1, which overlap by 78 base pairs. The two sequences were joined together at nucleotide 284.