

Supporting Information

Ausin et al. 10.1073/pnas.1618019113

Ptaedav.28490/1-895 1 MSPAKRTRRQTAGIETPTLENGSAVKKRKSEKESATPKENGNTVKENGNTLKENGNI VKENGTVKE 68
AT1G69770_CMT3/1-839 1 MAP-----KRRKPATKDDTTK-----SIPK 20

Ptaedav.28490/1-895 69 NGGTVKENGNTVKENGGI VKENPSSAPKAKVGAARLAGGDRP SGGPAAKTKLPGEDRLLGAPMPKAE 136
AT1G69770_CMT3/1-839 21 PKKRAPKRAKTVKEEPTVVEEKEKHA-----RFLDEPIPESE 59

Ptaedav.28490/1-895 137 AQRWP LRYE--KKKNAQNK SNGSAGDDEEQVVLNVKAHY LRAQVDG- ELYNLGDCASVKGEDGKADY 201
AT1G69770_CMT3/1-839 60 AKSTWPDYKPIEVQPPKAS SRKKT KDDEKVEIIRARCHY RRAIVDERQIYELNDAYVQS GEGKDPF 127

Ptaedav.28490/1-895 202 IGSILEFFETT DGKQYFTTQWFYRAEDTAIKTEASFHDKKRVFYSEIKDDNLL ECTISKLK-----S 263
AT1G69770_CMT3/1-839 128 ICKIIEFMFEGANGKLYFTARWYRPSDTVMKEFEILIKKR VFFSEIQDTNELGLLEKKNILMIPLN 195

Ptaedav.28490/1-895 264 ERKES SIPP--CDYYYDMGYNLAYTTFTYLPAGKSKNVAASS--DSTSTVCDSEENKADNDTW SGN 326
AT1G69770_CMT3/1-839 196 ENTKETIPATENCDFECDMNYFLPYDTFEAIQETMMAISESTISSD TDI REGAAAISEIGEC SQET 263

Ptaedav.28490/1-895 327 NCSKSELTL DLYSGCGMSTGLCFGANLSGVNLVTKWAVDLN EFACKSLKHNPETEV RNELADDFL 394
AT1G69770_CMT3/1-839 264 ECHK-KATLLDLYSGCGAMSTGLCMGAQLSGLNLVTKWAVDMNAHACKSLQHNPETNVRNMT AEDFL 330

Ptaedav.28490/1-895 395 ELLKHWKLY-----QKYCGSDGTGNTKAAETKNQKEEDDDSEI SEEF EVELIGTRYKQ 451
AT1G69770_CMT3/1-839 331 FLLKEWELCIHFS LRNSPNS EYANLHGLNNVEDNEDVSEES ENEDDGEV----FTVDKIVGIFS CV 394

Ptaedav.28490/1-895 452 ATKSDESGLQFK----GYDESEDSWEPVEGLGDCEESMKEFVMK GAKAKLLPLPGD VDVICGGPPCQG 515
AT1G69770_CMT3/1-839 395 PKKLLKRGLYLVWRWLYDDSHDTWEP IEGLSNCRGKIEEFVKLGYSGLPLPGGD VDVICGGPPCQG 462

Ptaedav.28490/1-895 516 ASGFNFRNTEAPLED SKNQIIVYMDIVDFLKP RYVLMENVVDILKFAGGV LGRYALSRLVHMSYQA 583
AT1G69770_CMT3/1-839 463 ISGHNRFRNLDP LEDQKNQLLVYMNIVEY LKPKFVLMENVVDM LKMAKGYLARFAVGRLLQMNYQV 530

Ptaedav.28490/1-895 584 KLGMMVAGCYGLPQFRMR-----KLPQYLP LTHDVVQRGGVPNEWERNMVAYDENHTVKLEKA 641
AT1G69770_CMT3/1-839 531 RNCMMAAGAYGLAQFRLRFFLWGALPSEIIPQFPLP LTHDLVHRGNIVKEFQGNIVAYDEGHTVKLADK 598

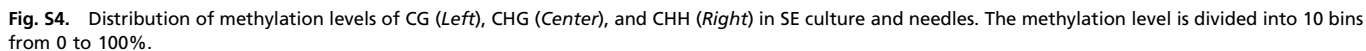
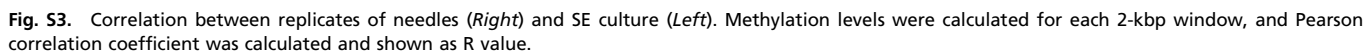
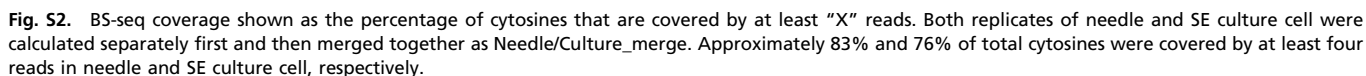
Ptaedav.28490/1-895 642 LILGDAISDLPEIANSEQRDEM QYGA PRTEFQQYIRMPKEVMNGRMLPSGSASKRASQKAILYDHRP 709
AT1G69770_CMT3/1-839 599 LLLKDVISDLPAVANSEKRDEITYDKDPTTFQKFIRLRKDE-----ASGSQSKSKSKKHVLYDHHP 660

Ptaedav.28490/1-895 710 LQLNEDDYQRVCRIPKNKGANFRDLP GVI IREDNVVELDTSMERILLPSGKPLIPDYAISFVKGRSLK 777
AT1G69770_CMT3/1-839 661 LNLNINDYERVCQVPKRKGANFRDFPGVIVGPGNVVKLEEGKERVKLESGKTLVPDYALTYD GKSCK 728

Ptaedav.28490/1-895 778 PFGRLWDETVPTVVTRAEPHNQAVLHP EQDRVLSIRENARLQGFDPYKYLHGTVKERYIQVGNVAV 845
AT1G69770_CMT3/1-839 729 PFGRLWDEIVPTVVTRAEPHNQVI IHP EQNRVLSIRENARLQGFDPDYKYLFGPPKQYIQVGNVAV 796

Ptaedav.28490/1-895 846 PVARALGFALGMAIQKLCT-DEPVVKLP EKFP LCFDNQQNEDGAMDVGEQT 895
AT1G69770_CMT3/1-839 797 PVAKALGYALGTAFAQGLAVGKDP LLLTLP EGF AFMKPTLPSELA----- 839

Fig. S1. Sequence alignment of *Arabidopsis* CMT3 and a putative CMT3 ortholog of *P. taedav*. Both protein sequences were extracted by reciprocal best hit of BLASTP, and then were aligned by Muscle.



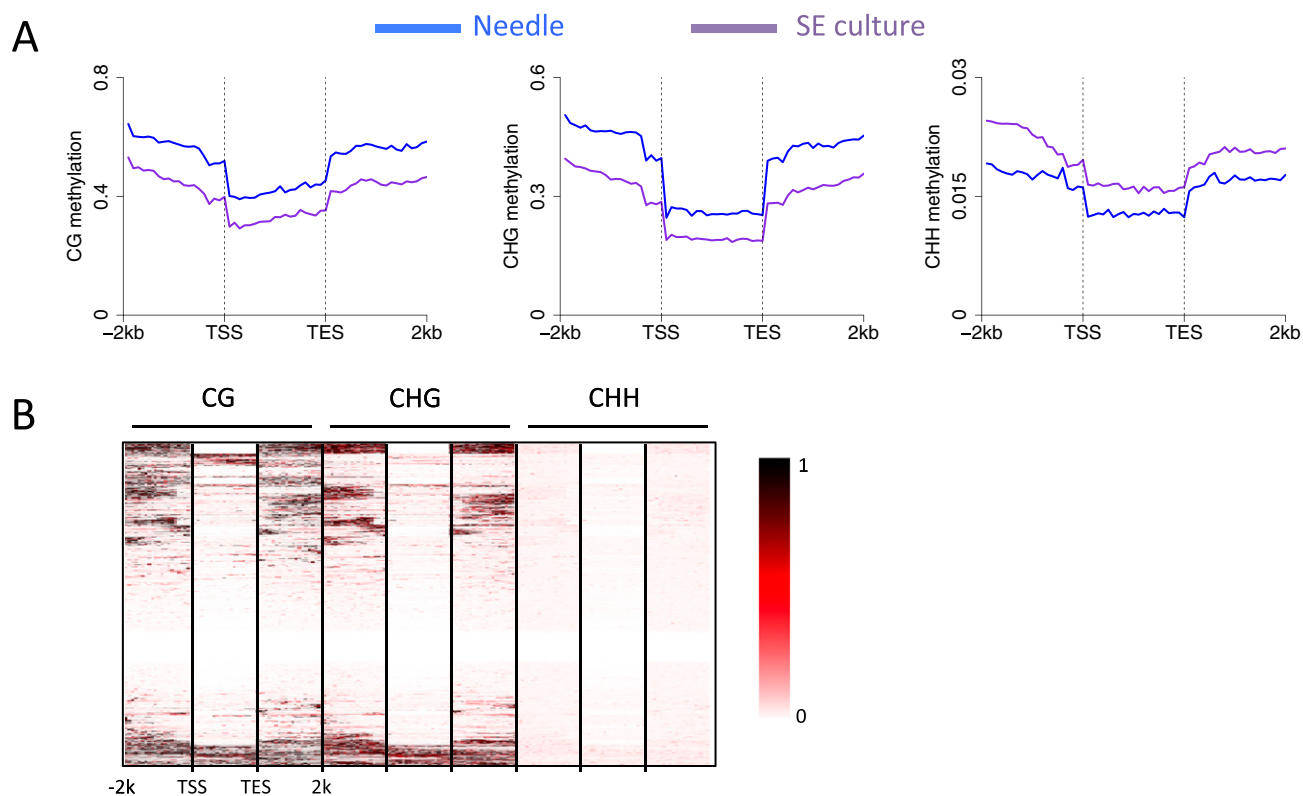


Fig. S5. Methylation patterns of genes excluding intronic TE insertions. (A) Metaplot of methylation in genes excluding intronic TE insertions. (B) Heatmap of methylation in genes excluding intronic TE insertions.

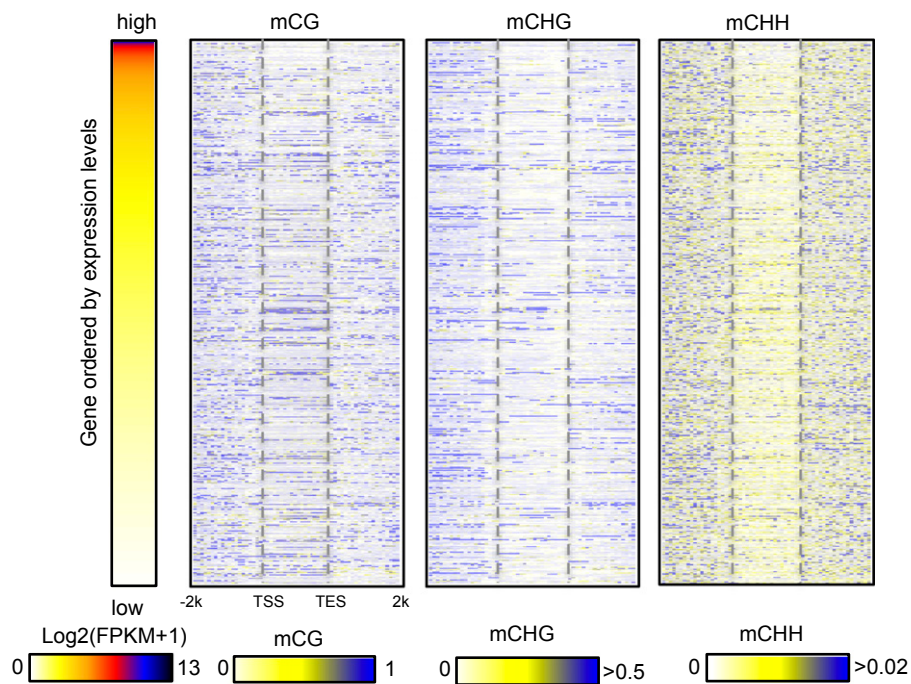


Fig. S6. Heatmap of DNA methylation in genes ordered by expression. Gene abundance was estimated by FPKM. In the case of zero value, we used $\log_2(\text{FPKM}+1)$ to order gene expression. DNA methylation in each sequence context is shown correspondingly.

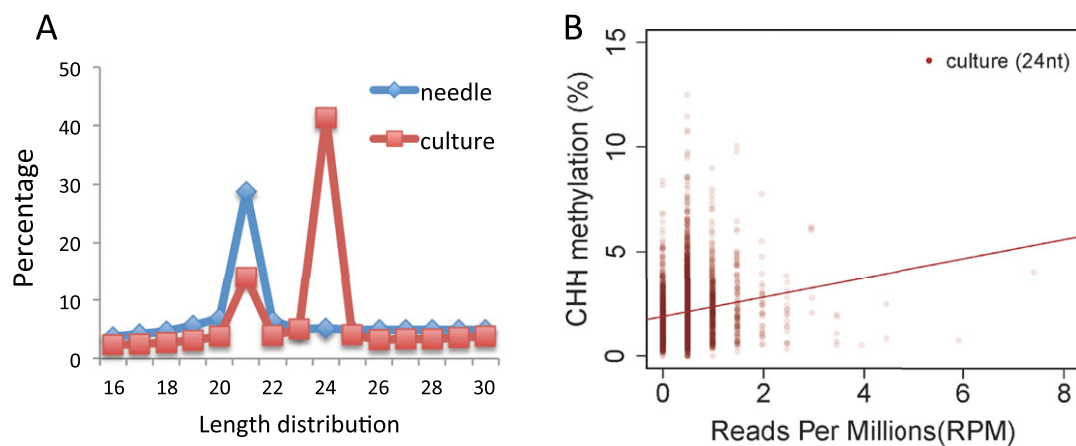


Fig. S8. Correlation between 24-nt siRNA abundance and CHH methylation levels. (A) Comparison of the length distributions of siRNA between needle and SE culture. (B) The correlation between CHH methylation and 24-nt siRNA abundance of supercontig1 in SE culture. Each bin represents a 500-bp window. Red line represents a fitted line from a linear model (lm function in R software) for SE culture.

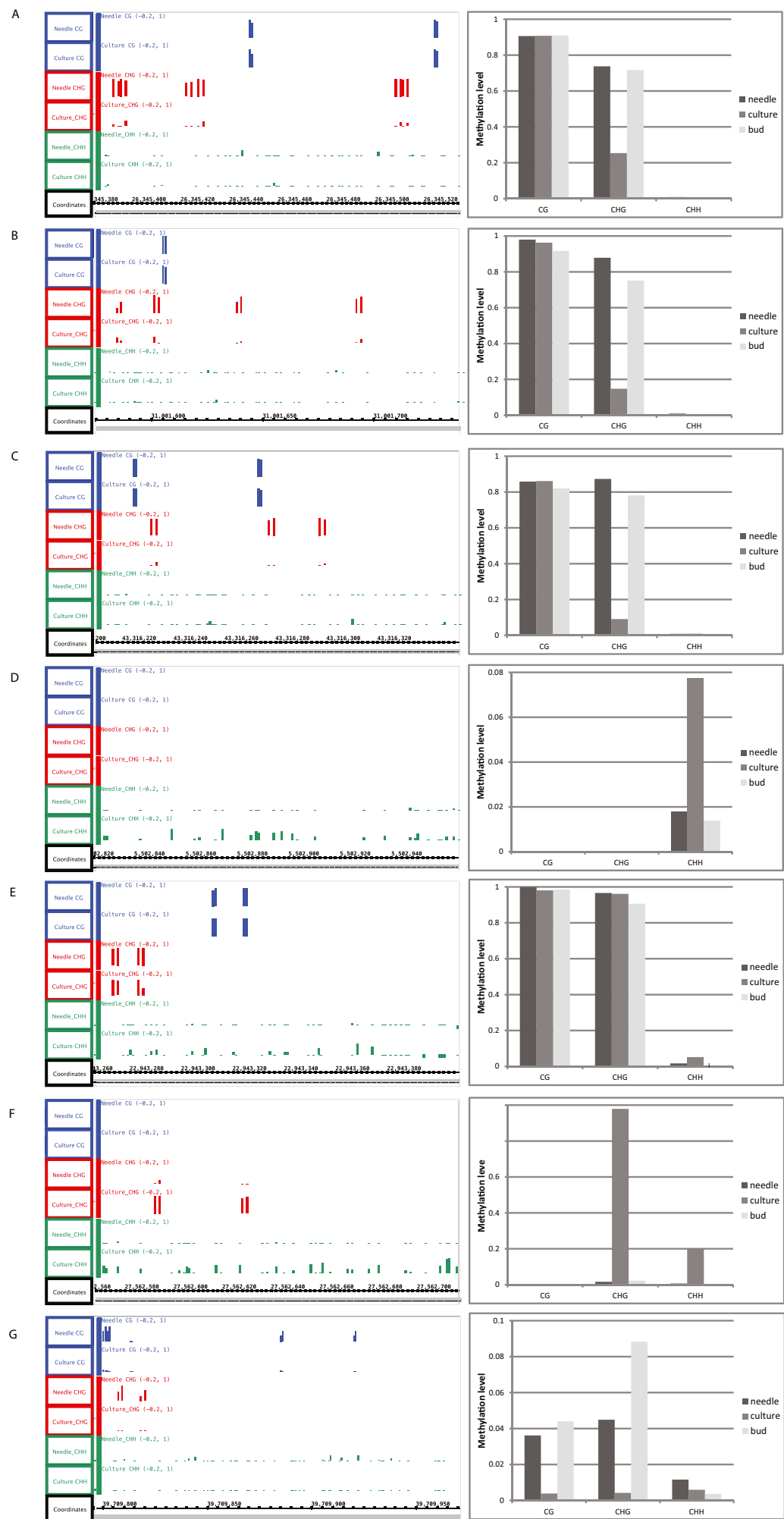


Fig. S9. Detection of methylation differences between needles and SE culture by BS-seq and validation of BS-seq data by traditional bisulfite sequencing. BS-seq data [screenshot from Integrated Genome Browser (IGB) browser] from selected regions in supercontig 1 (see *Experimental Procedures* for supercontigs) are shown on the left and traditional bisulfite data (see Table S2 for PCR primers used) are shown on the right. In BS-seq data, CG, CHG, and CHH are shown in blue, red, and green, respectively. Regions in A–C lose CHG methylation in SE culture; regions in D and E gain CHH methylation in SE culture; region in F gains CHG and CHH methylation in SE culture; and region in G loses all three types of methylation in SE culture. Traditional bisulfite sequencing data obtained from buds is also included for comparison purpose.

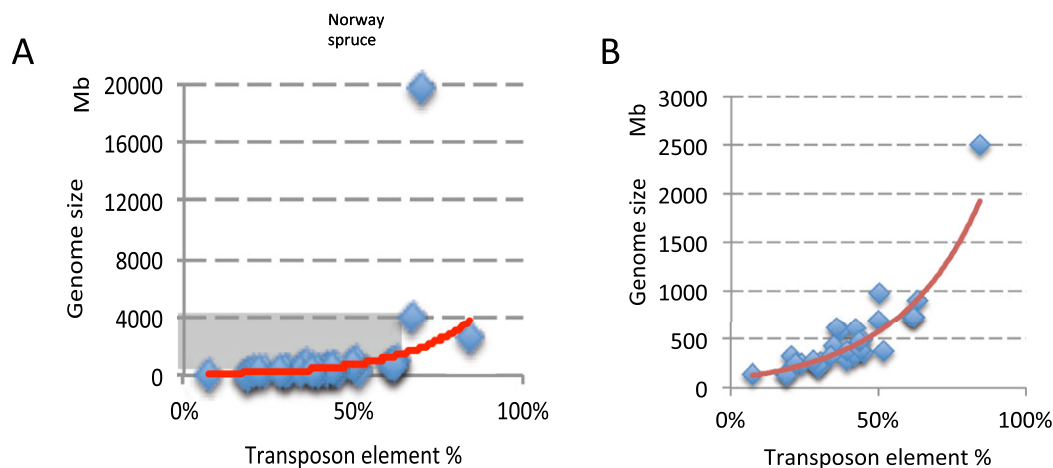


Fig. S10. Correlation between transposon element content and genome size. (A) Genome size is correlated with TE content. Thirty-three angiosperm species were used to show this correlation. (B) This is a zoomed-in view of the gray area in A. Red line is exponential regression line.

Table S1. Summary of BS-seq results and estimation of methylation levels

Samples	Read no.	Uniquely mapped	Mapped ratio, %	%mCG	%mCHG	%mCHH	Conversion rate, %	Sequencing depth, Gb
Needle rep1	4,071,099,559	2,374,889,504	58.3	74.7	69.1	1.5	94.91	118
Needle rep2	4,538,423,678	2,636,656,075	58.1	74.4	68.9	1.5	95.43	131
Culture rep1	3,776,335,406	2,102,719,947	55.7	68.6	62.8	2.6	94.49	105
Culture rep2	3,801,275,690	2,123,873,990	55.9	66.3	60.9	1.9	95.00	106

Table S2. PCR primers used in traditional bisulfite sequencing

Regions	Primer sequences
Supercontig1: 26,345,393–26,345,542	5'-TTGAGTGAAAAAATTYGAATATTATAAATTGTTTGGA-3 5'-TTTTAAACTTTTAAACACRTTCCATACCCT-3'
Supercontig1: 31,001,585–31,001,748	5'-YTTTTTGAGTGTTTTGGGTAAATTTGAG-3' 5'-CTAAAATAATCATTTAAAATACTTTTCATATTCATAAATTTATTTTA-3'
Supercontig1: 43,316,225–43,316,367	5'-TATATTATAATTTTTTTGTTTTATATTATGTTTTTTATTTTGYTTGAA-3' 5'-CATCAAAACAAAAATCTCTTTRCAAAATATATAAAAAAATAC-3'
Supercontig1: 5,502,827–5,502,968	5'-ATTTGAGTGATTGTTTTTTTTTTTAYGTATAYTGA-3' 5'-CACACTARACCTATCATACCACATAATATTTTC-3'
Supercontig1: 22,943,274–22,943,415	5'-TGTGYAATAATATAAYGAAATTGTGTGYGAATA-3' 5'-TCTTATTTCTAAAAATTTAATTAATACTCTCACTCCATA-3'
Supercontig1: 27,562,572–27,562,721	5'-TTATTATTGTTAATTTTTTAATTGAYGAGATTTTAATTTTTTAATATAA-3' 5'-AAATCTCTTCAAATACAAATARARTTAAAAATTTCTTA-3'
Supercontig1: 39,709,799–39,709,973	5'-AGAGAGGATGAAGGGAATGATTGA-3' 5'-CAATAAAAAATAAAATATAAATTAATACTAAAAATCAATACCAAAAAATAAT-3'

Table S3. Summary of sRNA-seq read alignment

Samples	Total reads*	Uniquely mapped	Mapped ratio, %
Needle	2,061,322	1,494,660	72.51
SE culture	2,261,492	2,027,354	89.65

*Nonredundant reads.

Other Supporting Information Files

[Dataset S1 \(PDF\)](#)