

Figure S11: Methylation level and length in base pairs for **(A)** mCG regions, **(B)** mCHG regions, **(C)** and mCHH regions.

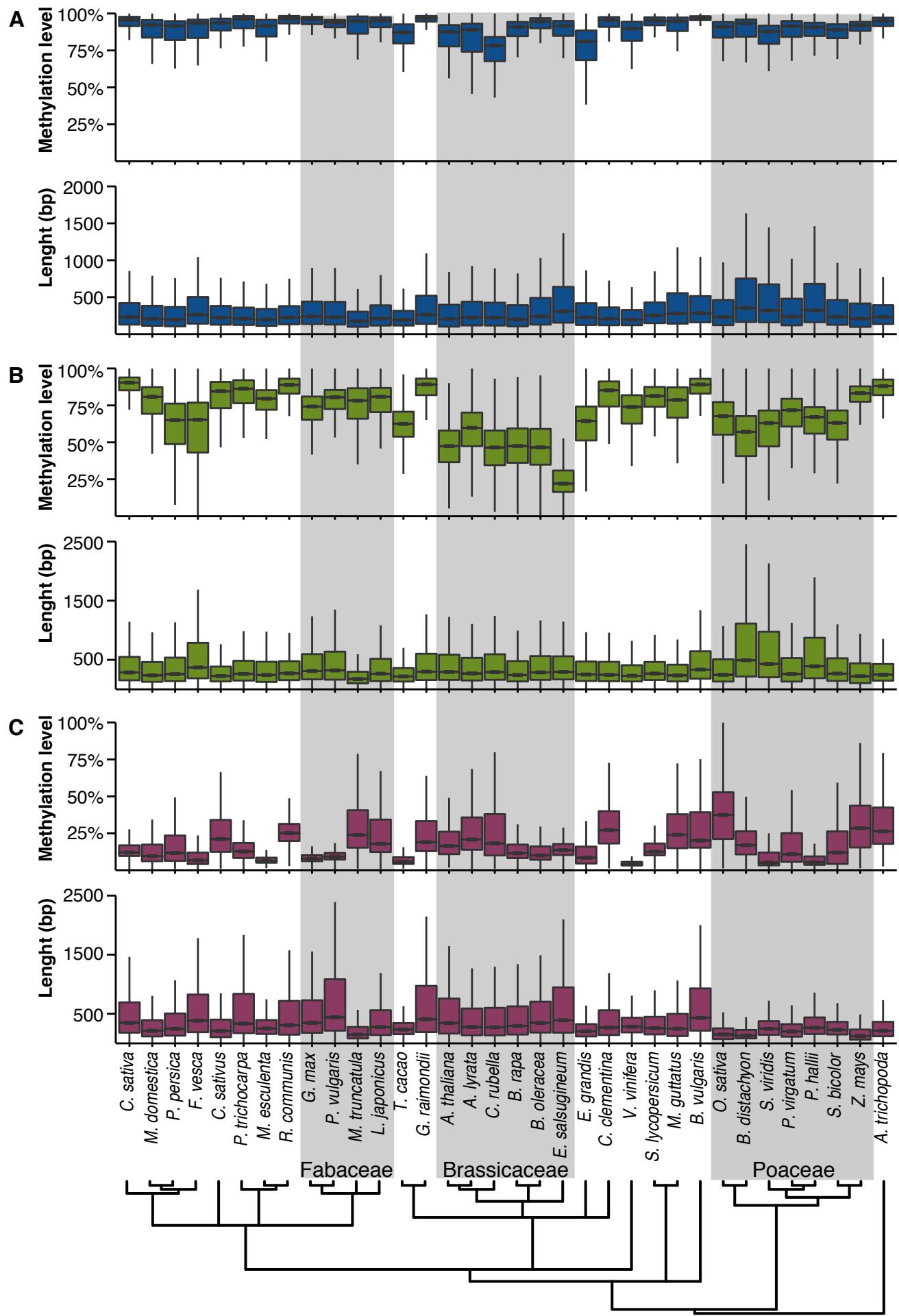
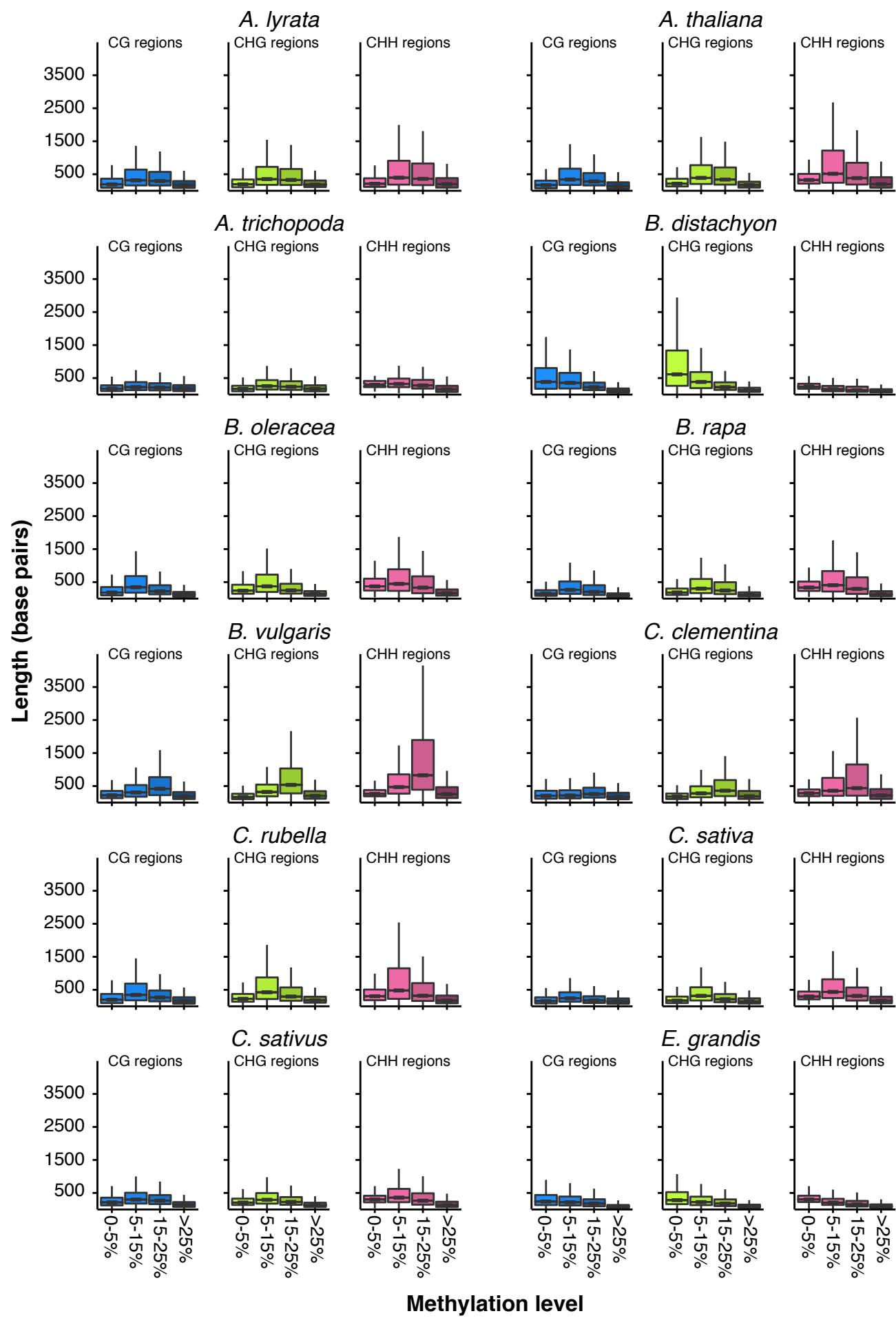
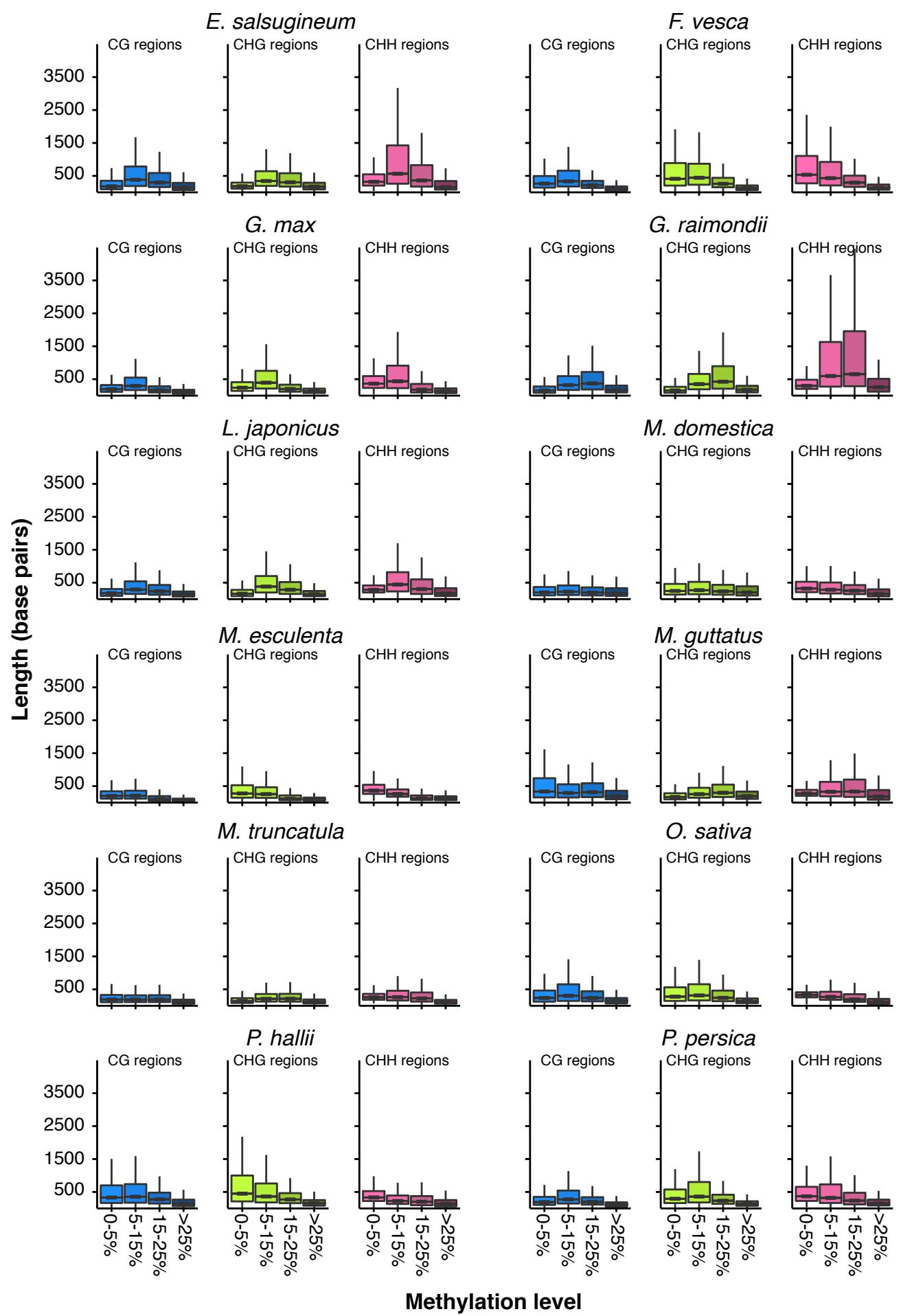


Figure S12: Size in base pairs of mCG, mCHG, and mCHH regions for each species based on the methylation levels of those regions: < 5%, 5-15%, 15-25%, > 25%.





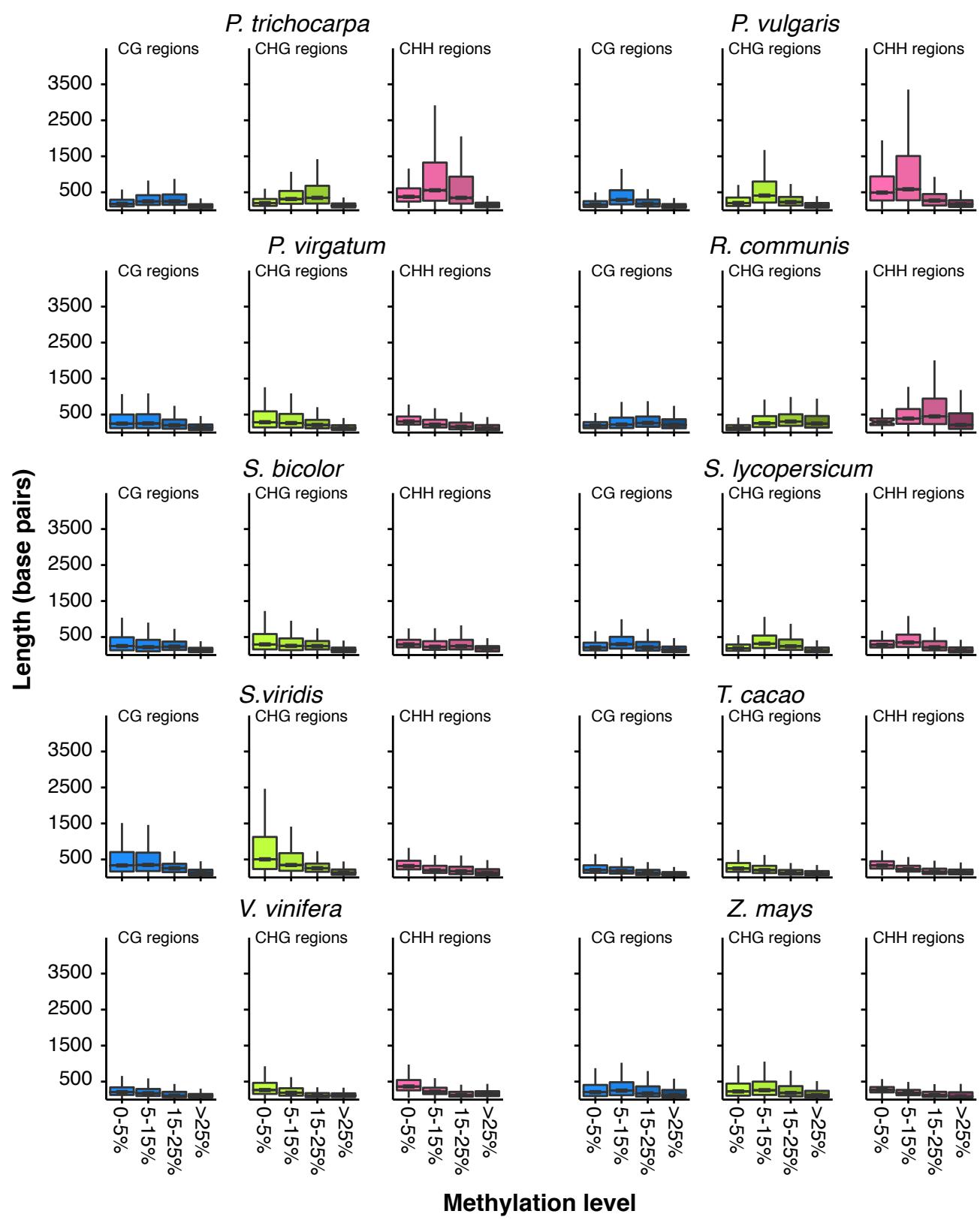
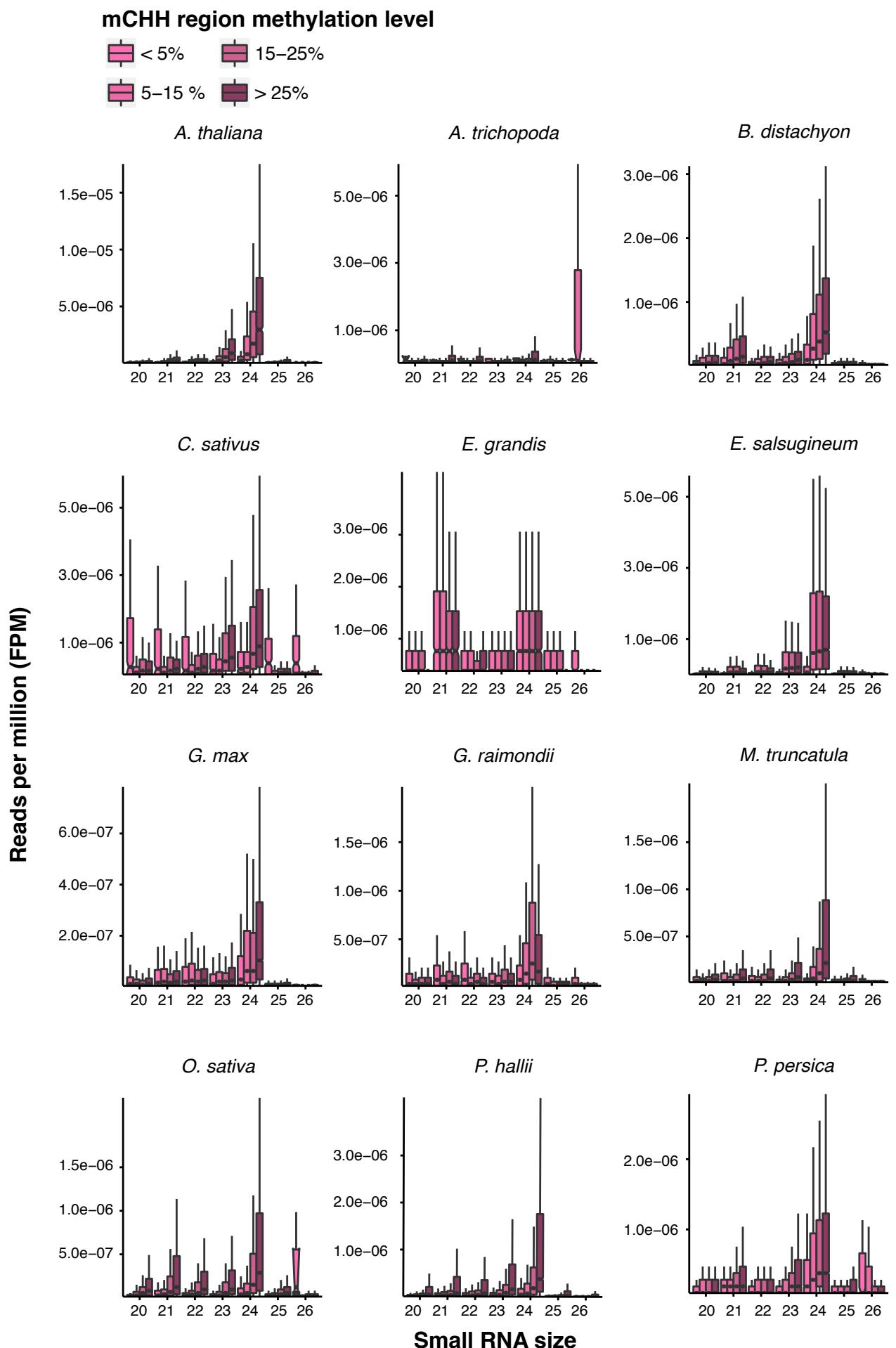


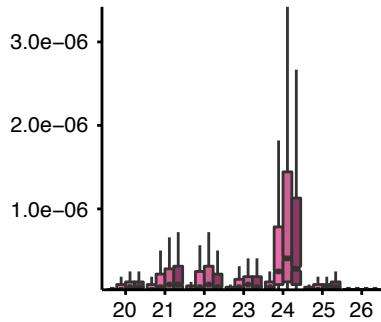
Figure S13: sRNA levels for each sRNA size class in mCHH regions for each species with sRNA data. mCHH regions are broken into four groups based on methylation level of that region: < 5% mCHH, 5-15% mCHH, 15-25% mCHH, > 25% mCHH.



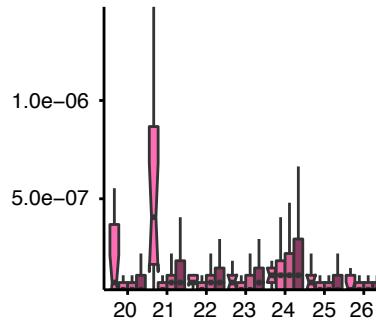
mCHH region methylation level



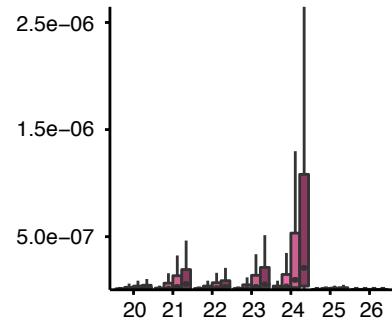
P. trichocarpa



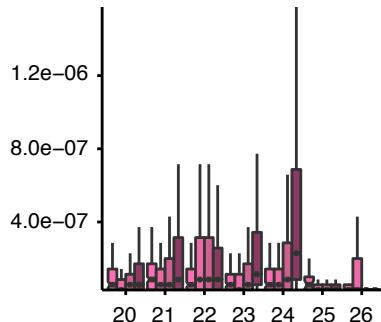
R. communis



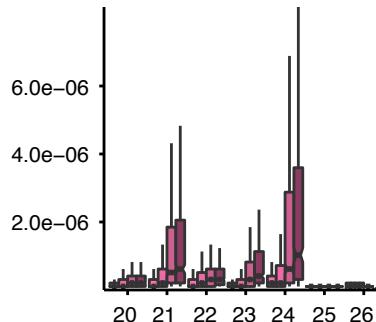
S. bicolor



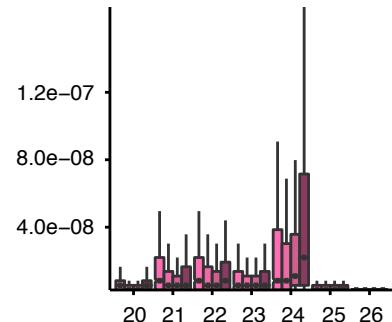
S. lycopersicum



V. vinifera

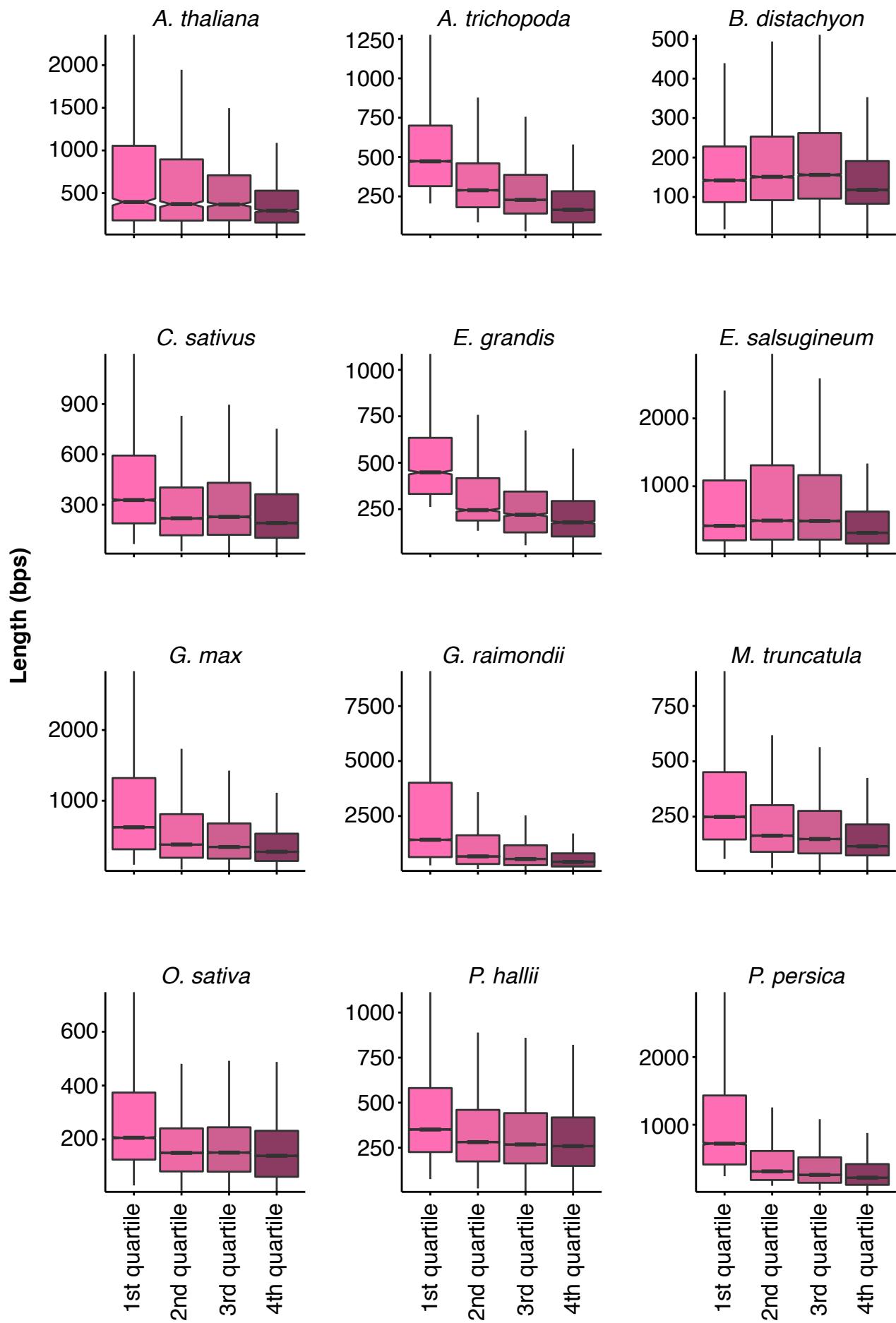


Z. mays



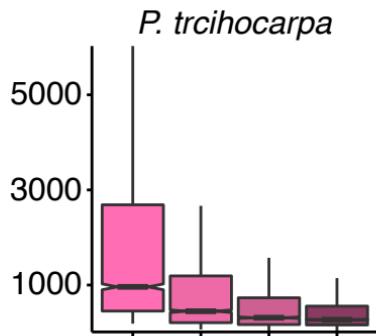
Small RNA size

Figure S14: mCHH regions were assigned to quartiles based on 24nt siRNA levels and the distribution of their lengths plotted. 1st quartile = lowest 24nt siRNA level, 4th quartile = highest 24nt siRNA levels.

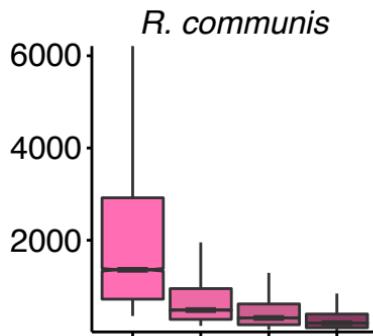


Length (bps)

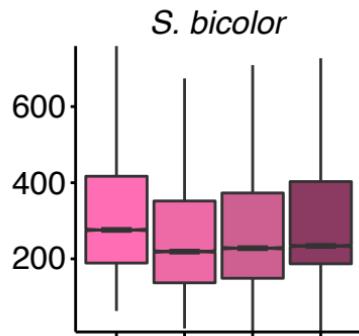
P. trichocarpa



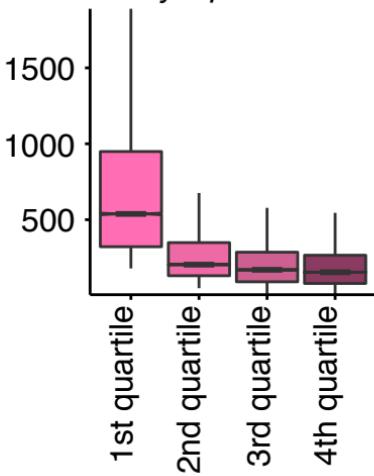
R. communis



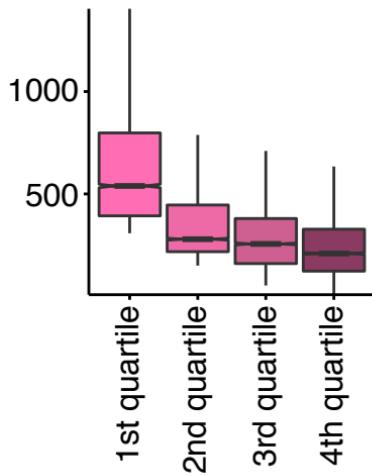
S. bicolor



S. lycopersicum



V. vinifera



Z. mays

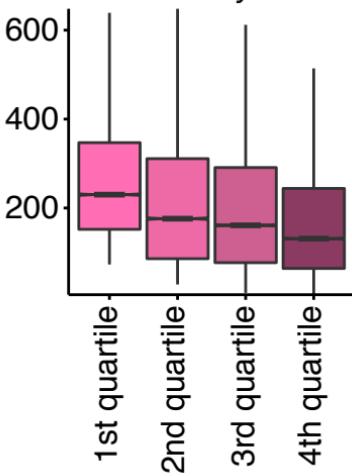


Figure S15: (A) mCG (blue), mCHG (green), mCHH (maroon) levels within CDS plotted against total number of repeats in the genome. **(B)** mCG (blue), mCHG (green), mCHH (maroon) levels within coding sequences (CDS) plotted against the percentage of genes with repeats within genes (UTR, CDS, and intron). **(C)** Percentage of genes with repeats plotted against total number of repeats in the genome.

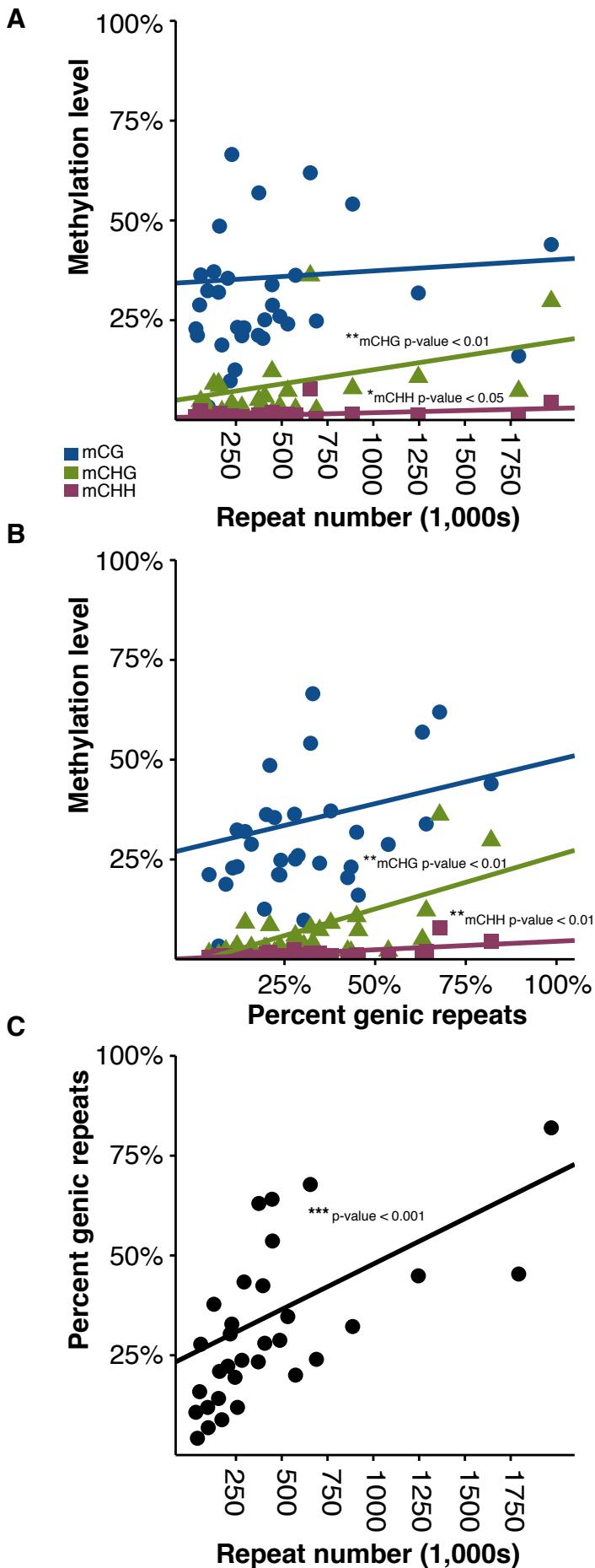
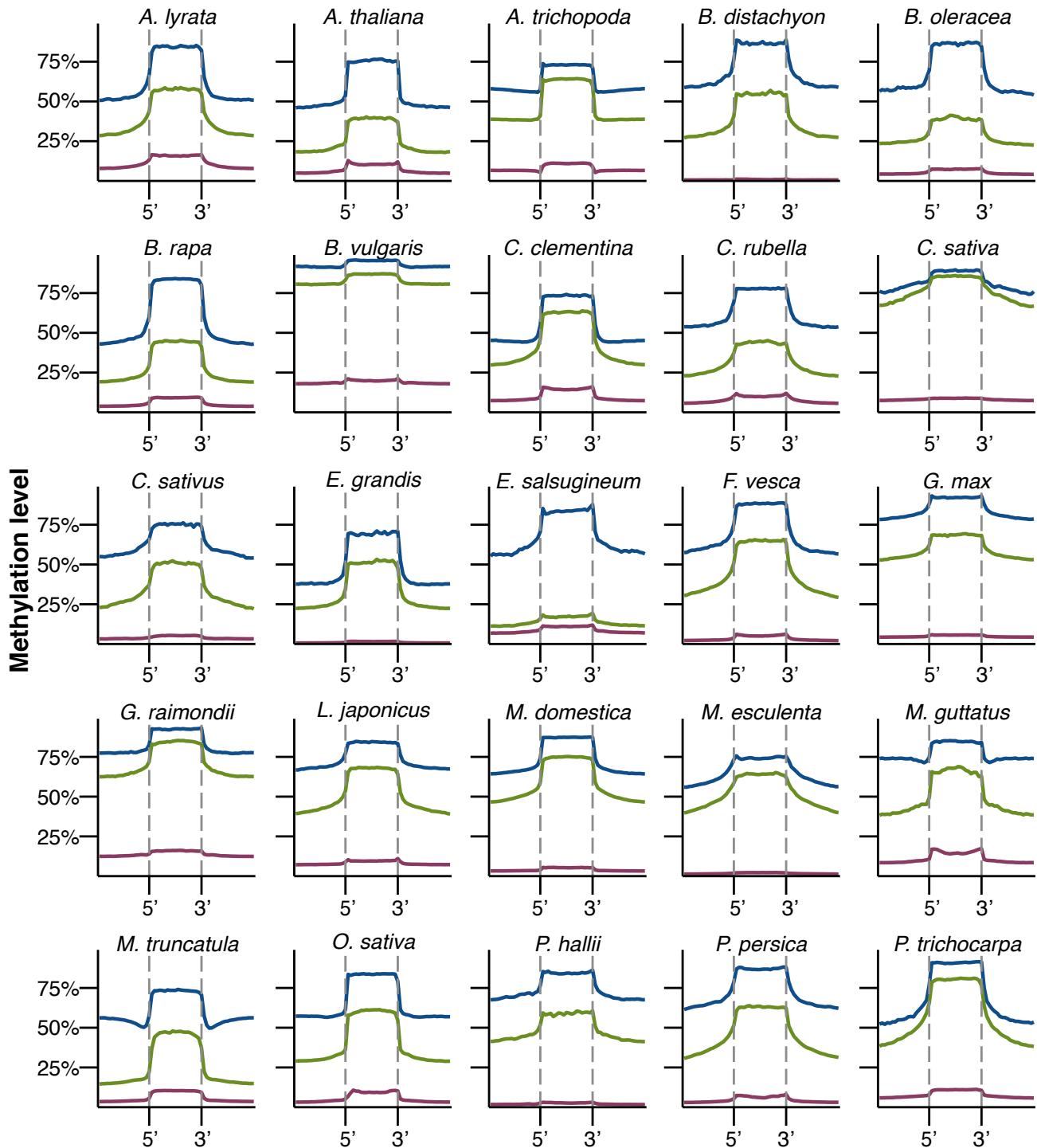


Figure S16: Methylation levels for mCG (blue), mCHG (green), and mCHH (maroon) upstream, across, and downstream of repeat sequences were plotted for all species with annotated repeats.



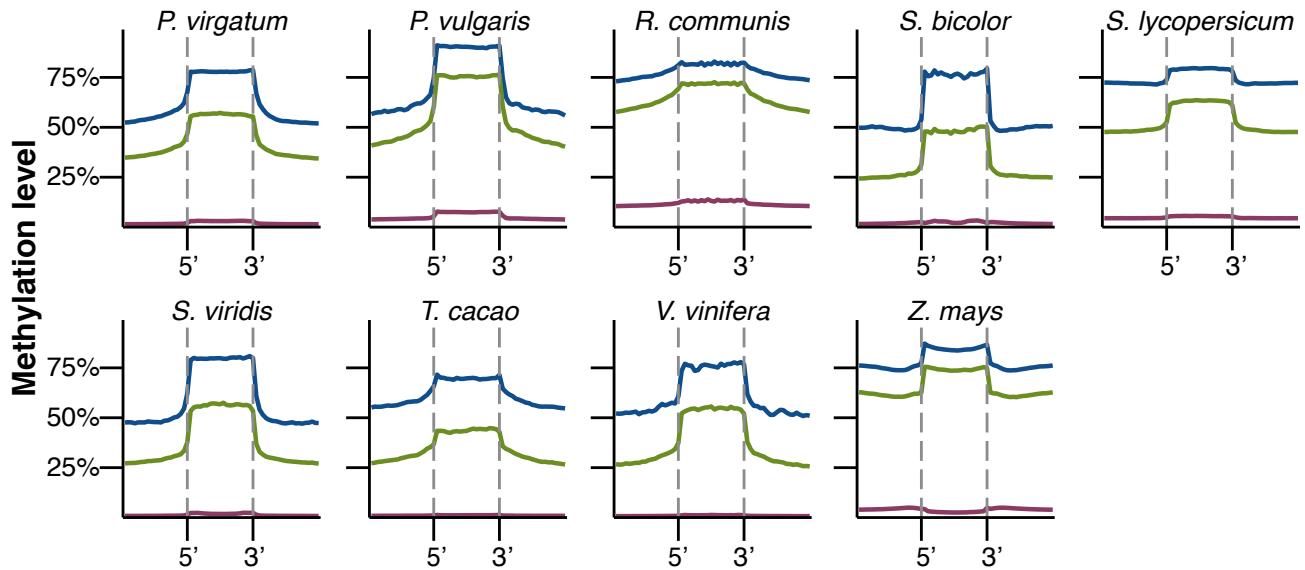


Figure S17: A binomial test was used to determine if a statistically significant number of methylated site within coding regions for each sequence context. Depending upon which contexts were methylated, genes were classified into one of four methylation classes. Browser screen shots of example genes from *F. vesca* are shown.

gbM mCG q-value < 0.05
 mCHG q-value > 0.05
 mCHH q-value > 0.05

mCHG mCG q-value = 0-1
 mCHG q-value < 0.05
 mCHH q-value > 0.05

mCHH mCG q-value = 0-1
 mCHG q-value = 0-1
 mCHH q-value < 0.05

UM mCG q-value > 0.05
 mCHG q-value > 0.05
 mCHH q-value > 0.05

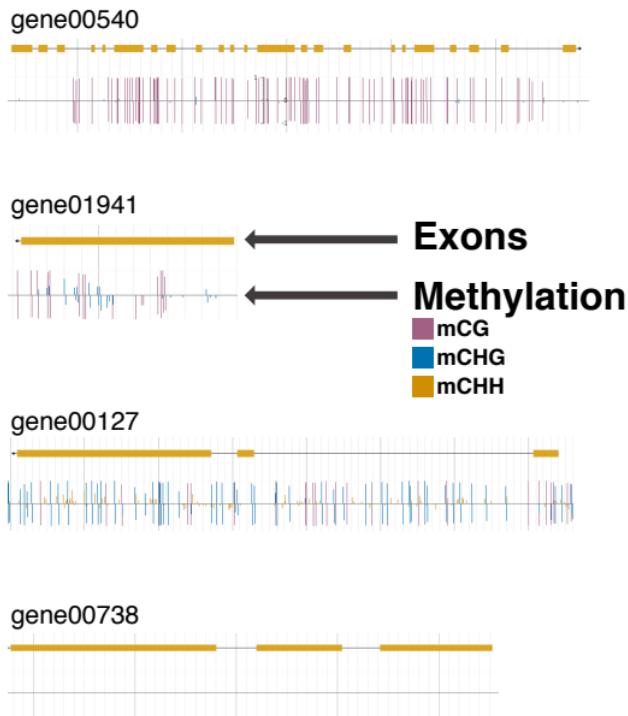
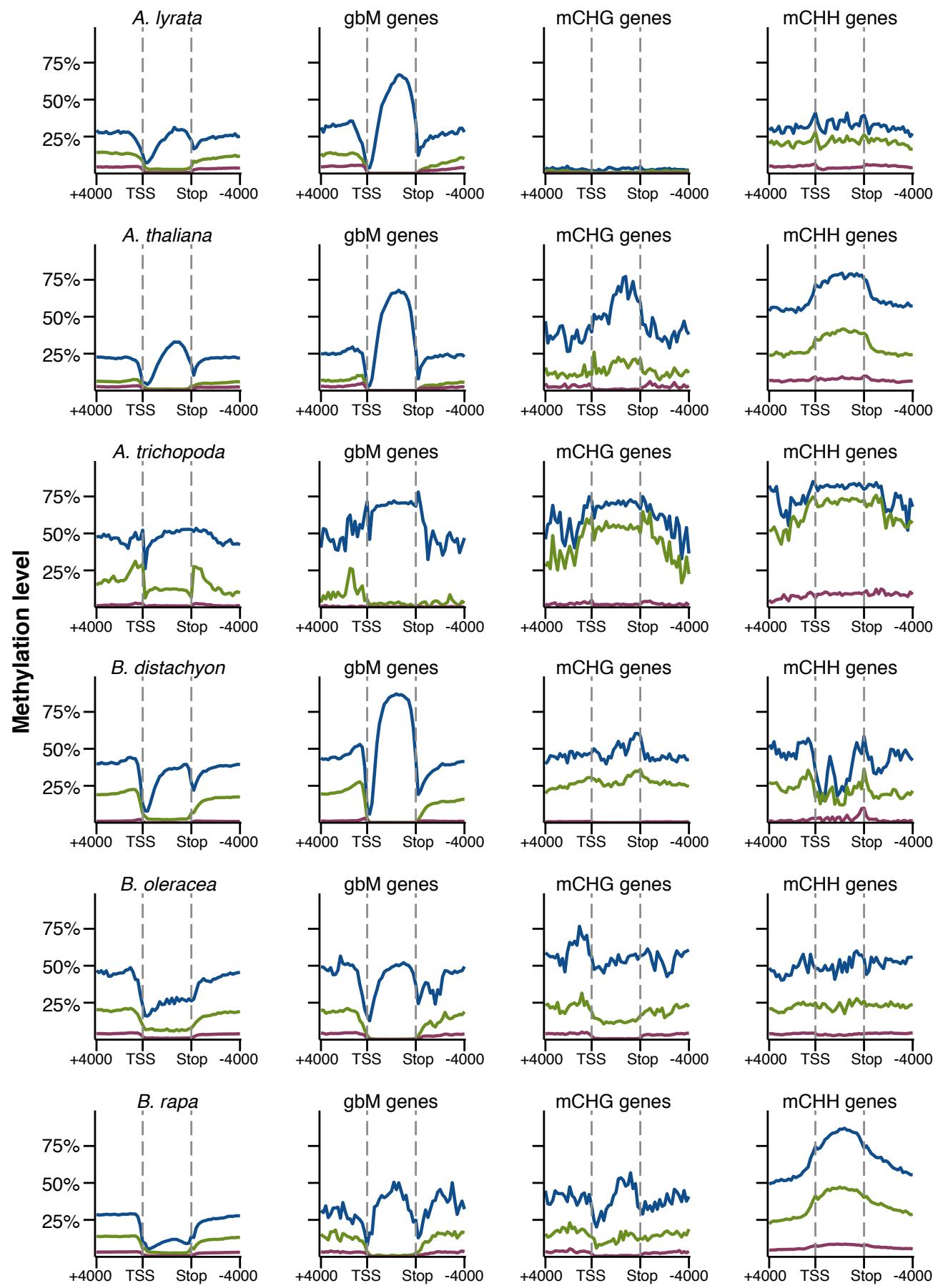
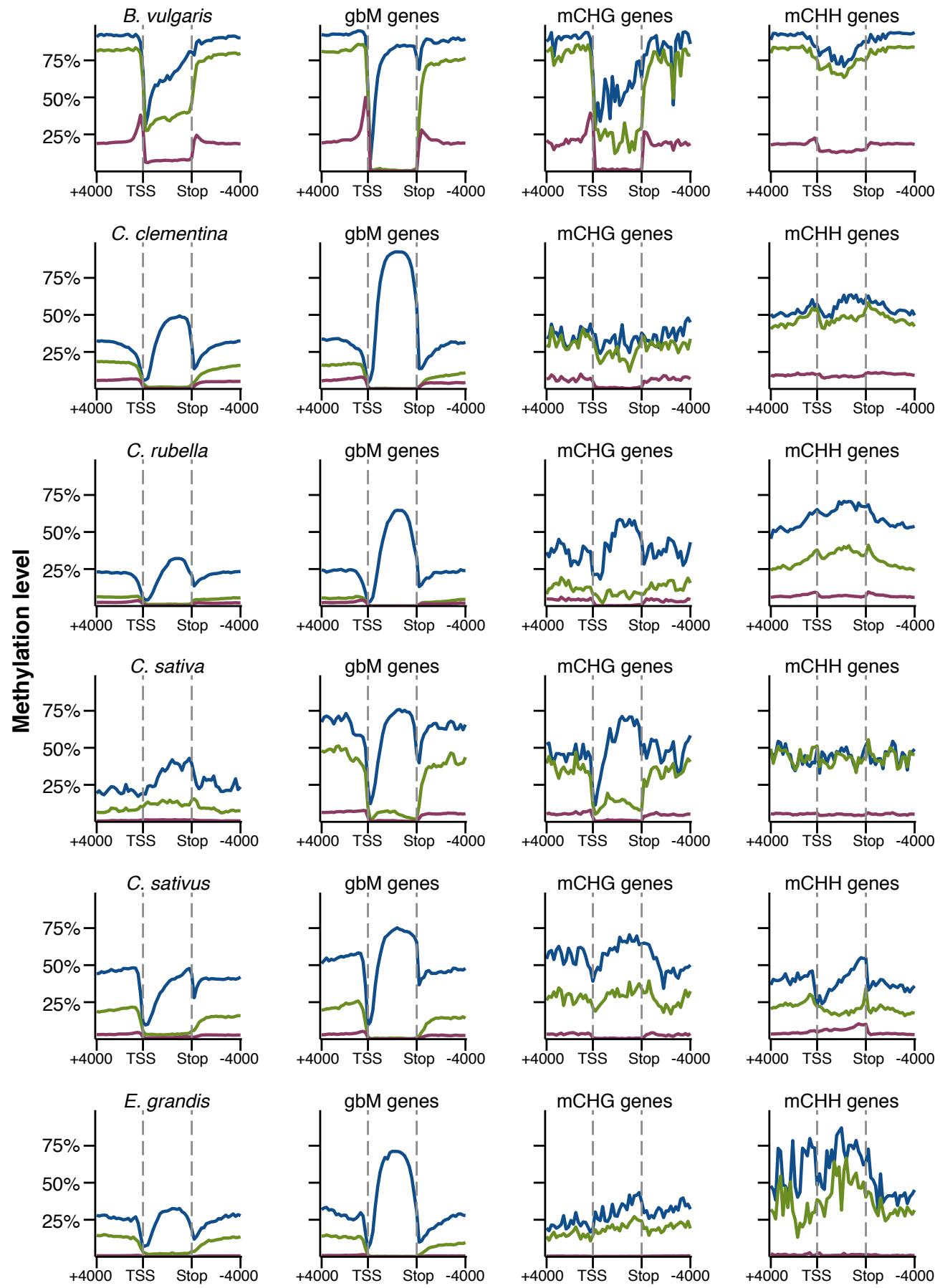
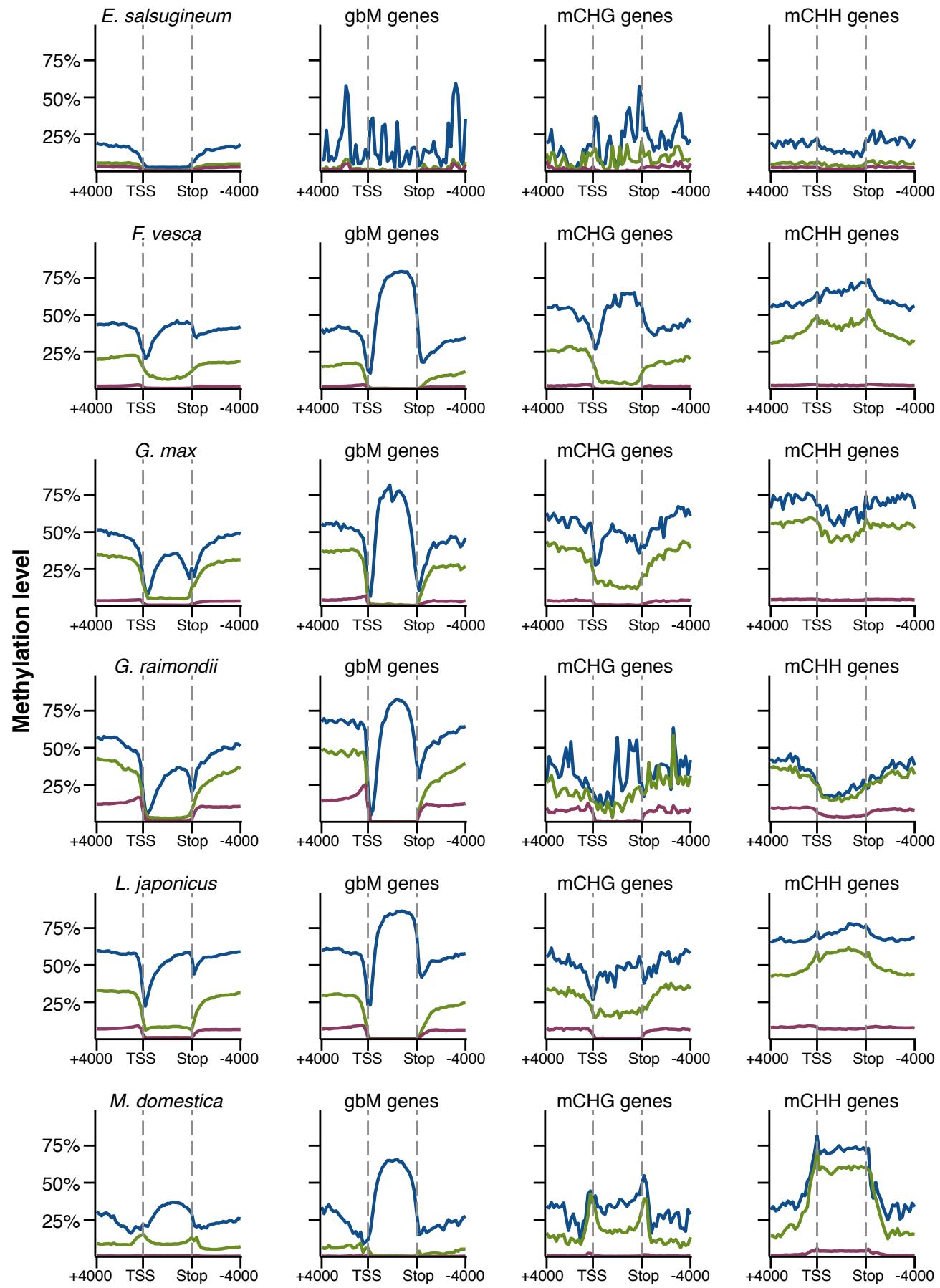
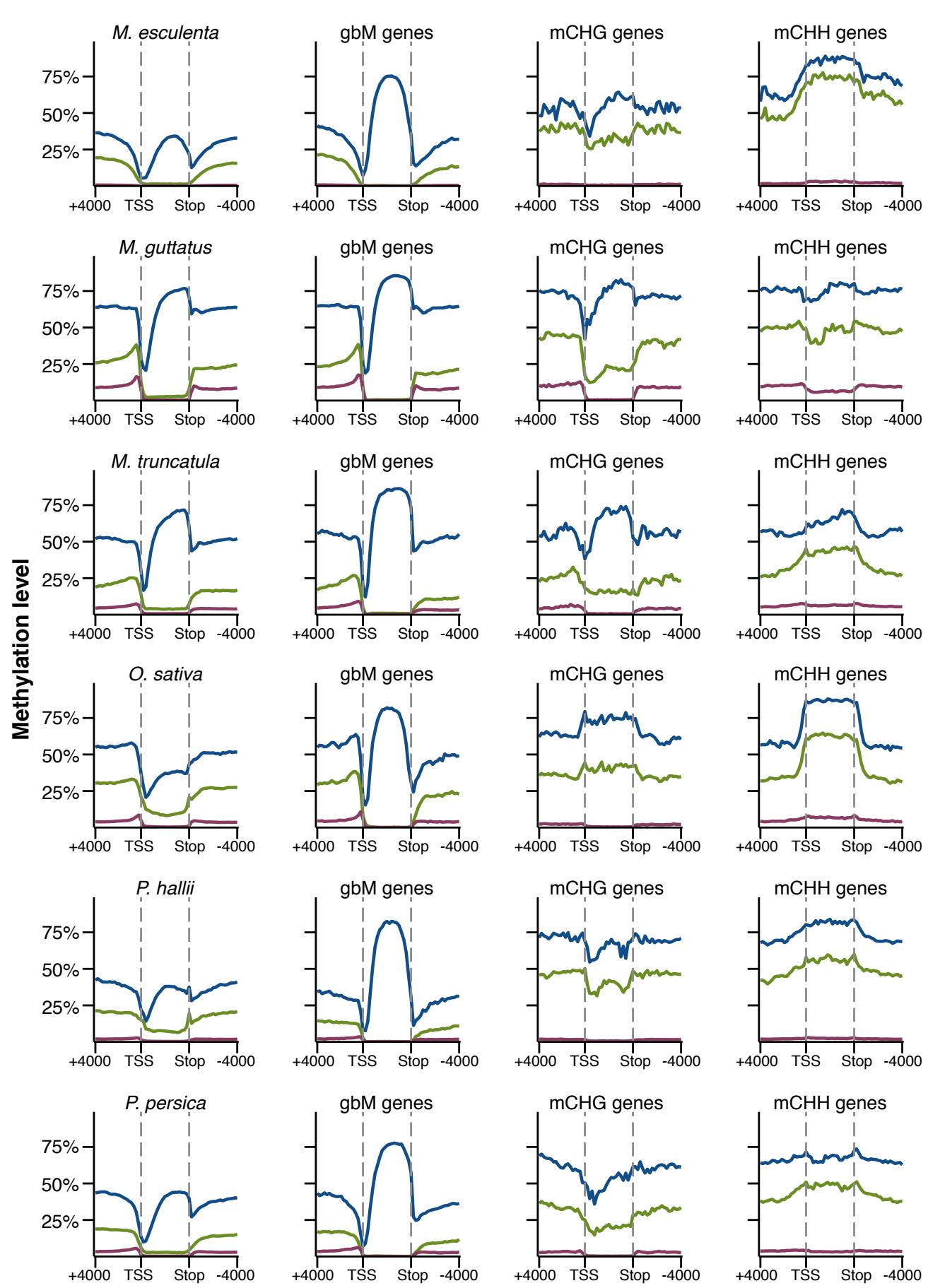


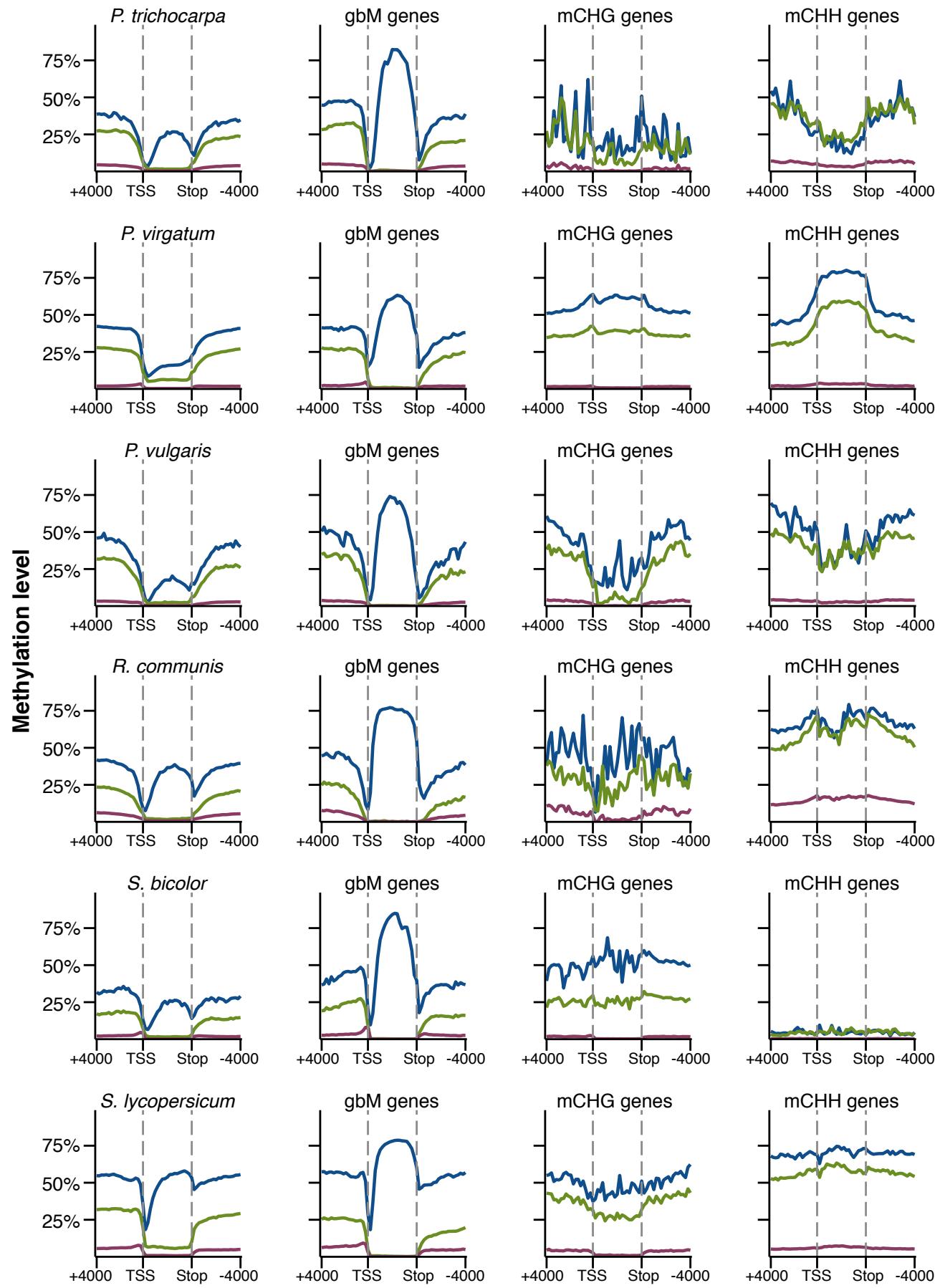
Figure S18: Methylation levels for mCG (blue), mCHG (green), and mCHH (maroon), upstream, across, and downstream for all genes, CG gbM genes, mCHG genes, and mCHH genes for all species.











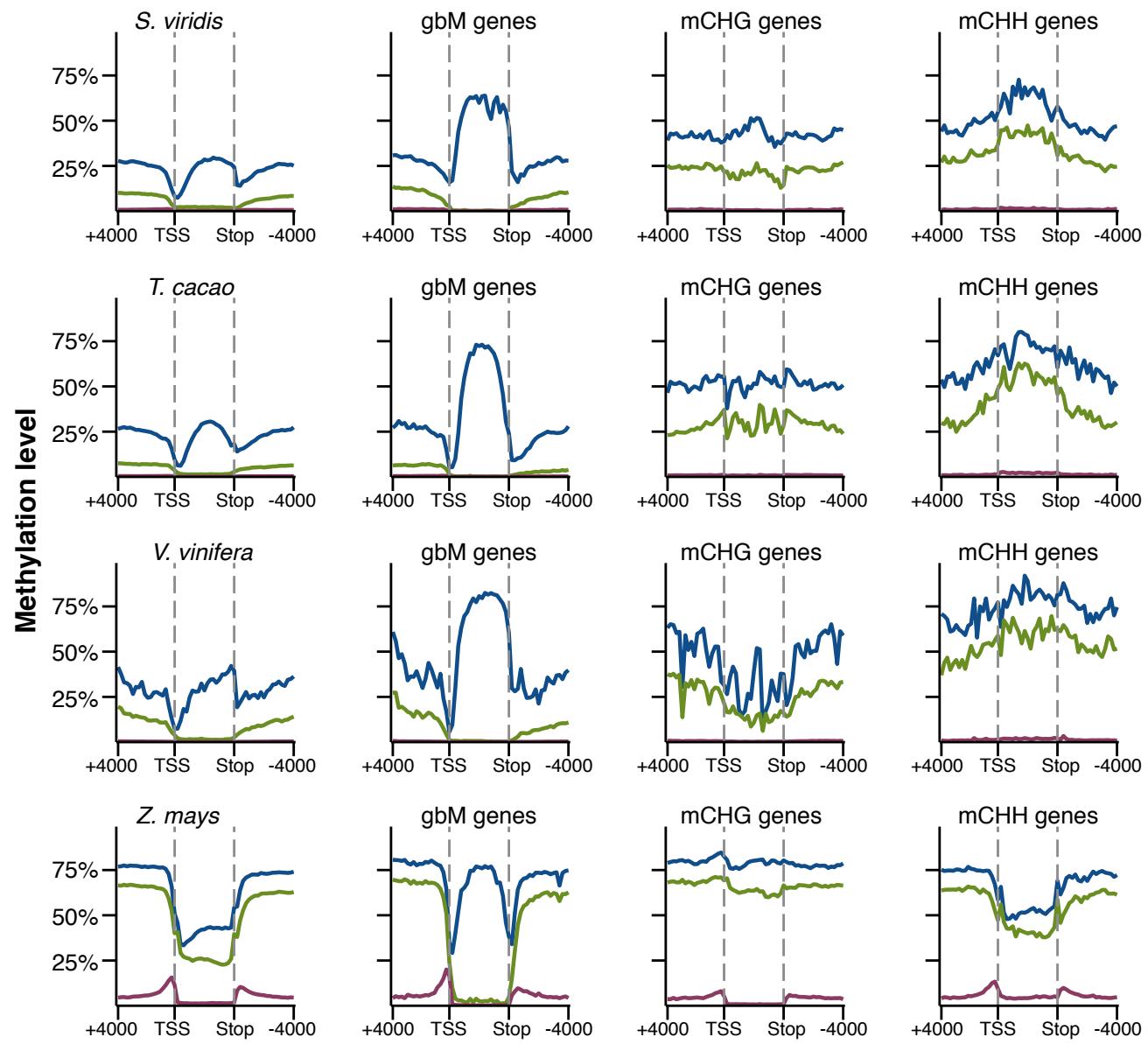


Figure S19: Gene expression levels (FPKM) for all genes, CG gbM genes, mCHG genes, mCHH genes, and mCG-TSS genes for all species with RNA-seq data.

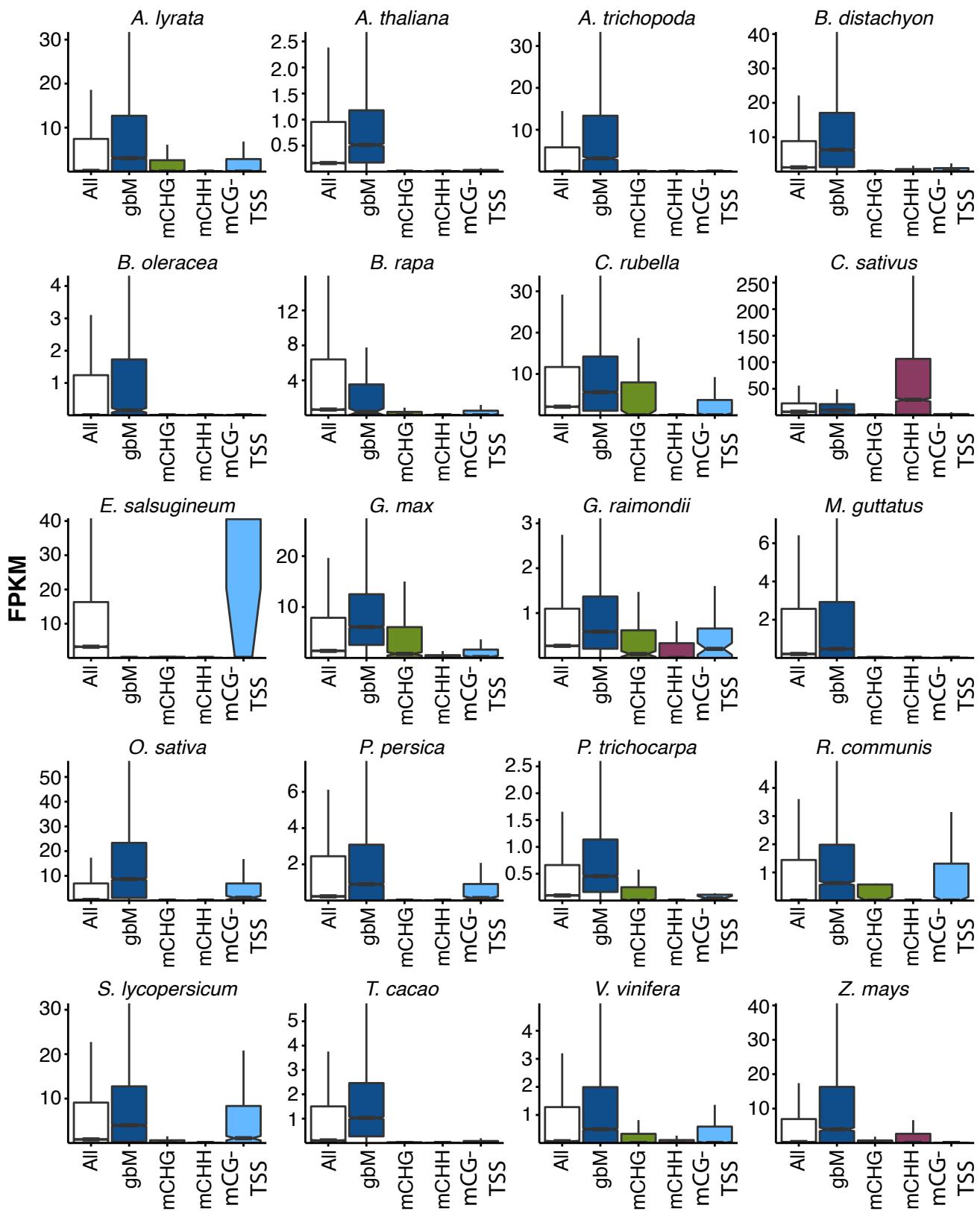


Table S5: Physical characteristics of CG gbM genes versus UM genes: gene length, exon number, and [O/E] observed number of CG dinucleotides versus expected number given GC content of the gene. The molecular evolution of CG gbM genes versus UM genes was determined using either *A. thaliana* or *O. sativa* as an outgroup. Molecular evolution characteristics showing non-synonymous mutations (dN), synonymous mutations (dS) and the ratio of dN/dS (ω). Those in red do not conform to prediction.

	Physical characteristics			Molecular evolution characteristics			
	Length	Exon #	[O/E]	Outgroup	dN	dS	ω
<i>C. sativa</i>	gbM>UM	gbM>UM	gbM~UM	<i>A. thaliana</i>	gbM~UM	gbM~UM	gbM~UM
<i>M. domestica</i>	gbM>UM	gbM>UM	gbM>UM	<i>A. thaliana</i>	gbM~UM	gbM~UM	gbM~UM
<i>P. persica</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM
<i>F. vesca</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM
<i>C. sativus</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>P. trichocarpa</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>M. esculenta</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>R. communis</i>	gbM>UM	gbM>UM	gbM>UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM
<i>G. max</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM
<i>P. vulgaris</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM
<i>M. truncatula</i>	gbM>UM	gbM>UM	gbM>UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM
<i>L. japonicus</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>T. cacao</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>G. raimondii</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>A. thaliana</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	-----
<i>A. lyrata</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>C. rubella</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>B. rapa</i>	gbM>UM	gbM>UM	gbM~UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM
<i>B. oleracea</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>E. salsugineum</i>	gbM<UM	gbM<UM	gbM<UM	<i>A. thaliana</i>	-----	-----	-----
<i>E. grandis</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM
<i>C. clementina</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>V. vinifera</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>S. lycopersicum</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM
<i>M. guttatus</i>	gbM>UM	gbM>UM	gbM>UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM>UM
<i>B. vulgaris</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	gbM<UM	gbM<UM	gbM~UM
<i>O. sativa</i>	gbM>UM	gbM>UM	gbM<UM	<i>A. thaliana</i>	-----	-----	-----
<i>B. distachyon</i>	gbM>UM	gbM>UM	gbM<UM	<i>O. sativa</i>	gbM>UM	gbM<UM	gbM>UM
<i>S. viridis</i>	gbM>UM	gbM>UM	gbM<UM	<i>O. sativa</i>	gbM~UM	gbM<UM	gbM>UM
<i>P. virgatum</i>	gbM>UM	gbM>UM	gbM<UM	<i>O. sativa</i>	gbM>UM	gbM<UM	gbM>UM
<i>P. hallii</i>	gbM>UM	gbM>UM	gbM<UM	<i>O. sativa</i>	gbM>UM	gbM<UM	gbM>UM
<i>S. bicolor</i>	gbM>UM	gbM>UM	gbM<UM	<i>O. sativa</i>	gbM>UM	gbM<UM	gbM>UM
<i>Z. mays</i>	gbM>UM	gbM>UM	gbM<UM	<i>O. sativa</i>	gbM>UM	gbM<UM	gbM~UM
<i>A. trichopoda</i>	gbM>UM	gbM>UM	gbM>UM	<i>A. thaliana</i>	gbM~UM	gbM<UM	gbM~UM

Figure S20: Gene expression levels (FPKM) for all genes, CG gbM genes, mCHG genes, mCHH, and TSS-mCG genes for second replicate of *C. sativus*.

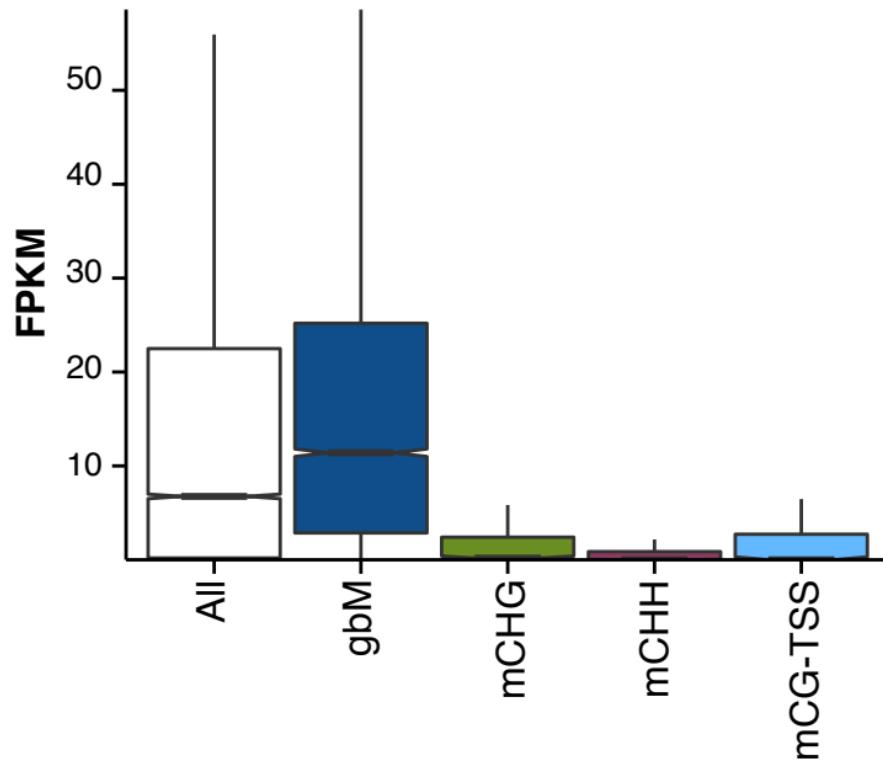


Figure S21: Methylation states (CG gbM = blue, mCHG = green, mCHH = maroon, UM = white) of genes (horizontal axis) orthologous to mCHH genes in **(A)** *L. japonicus*, **(B)** *A. lyrata*, and **(C)** *Z. mays*.

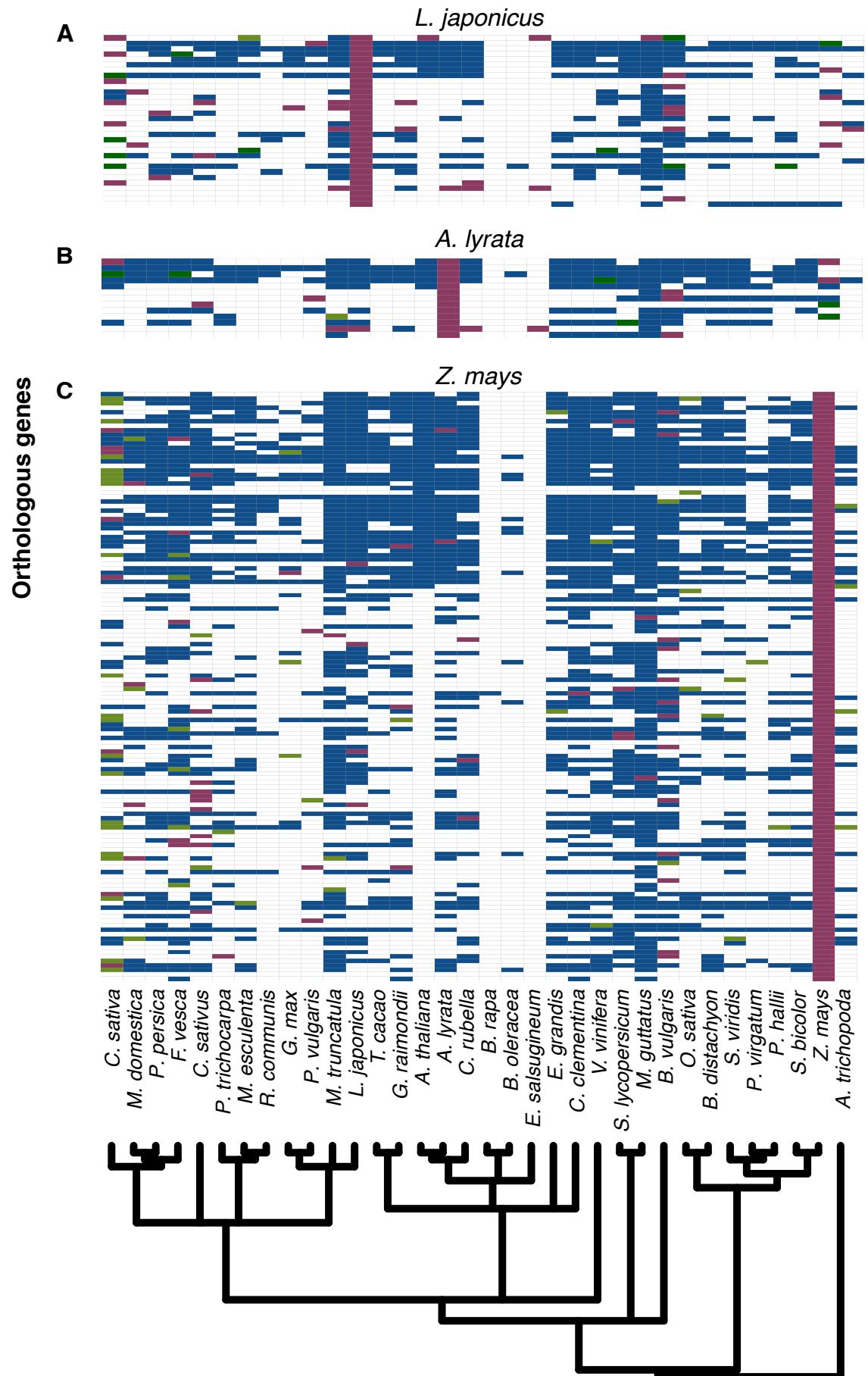


Figure S22: GO term enrichment was performed for each gene class for each species. The most commonly enriched GO terms (horizontal axis) for the Biological Process (first panel), Molecular Function (second panel) and Cellular Component (third panel) are shown for **(A)** CG gbM genes, **(B)** mCHG genes, and **(C)** mCHH genes.

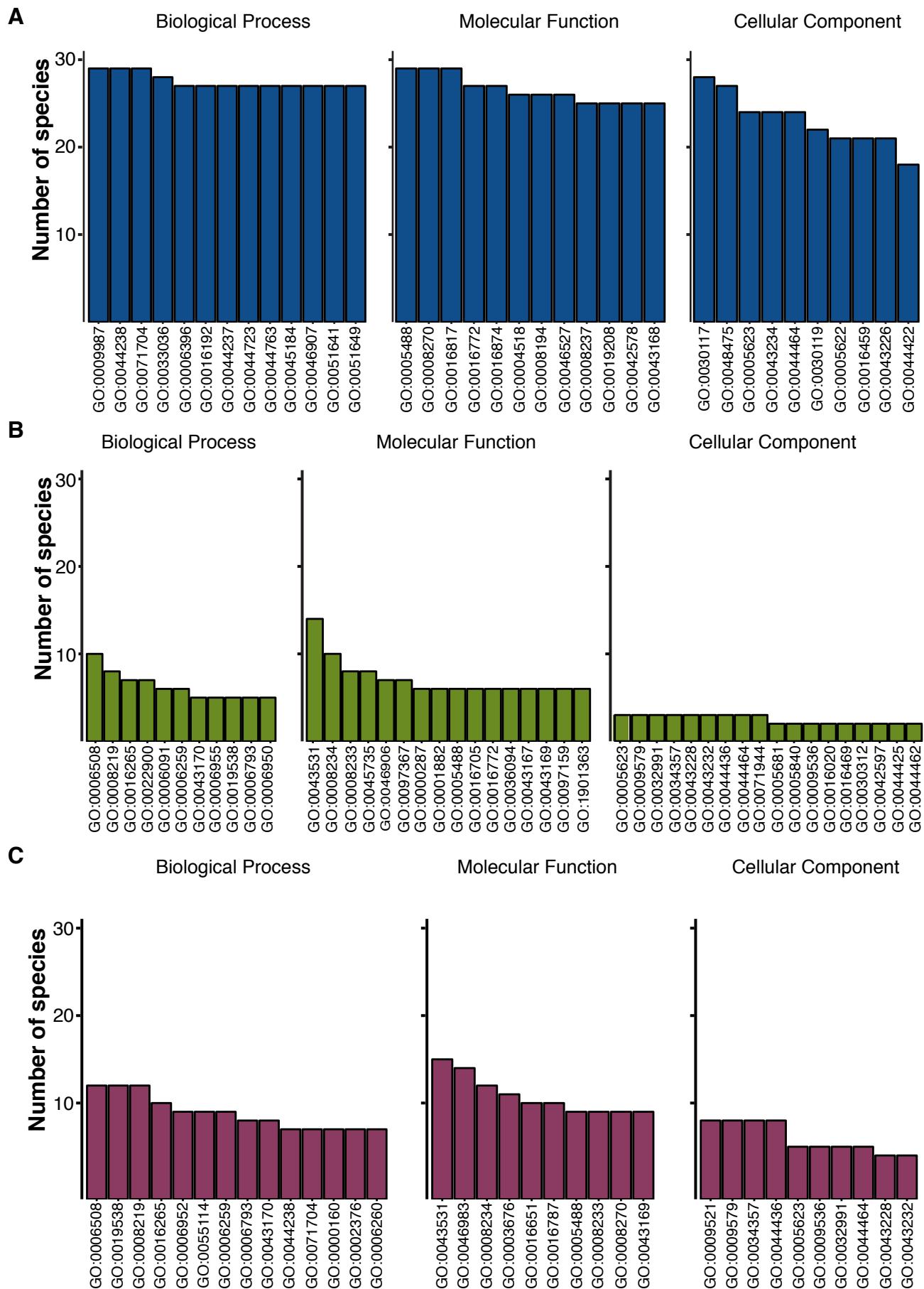
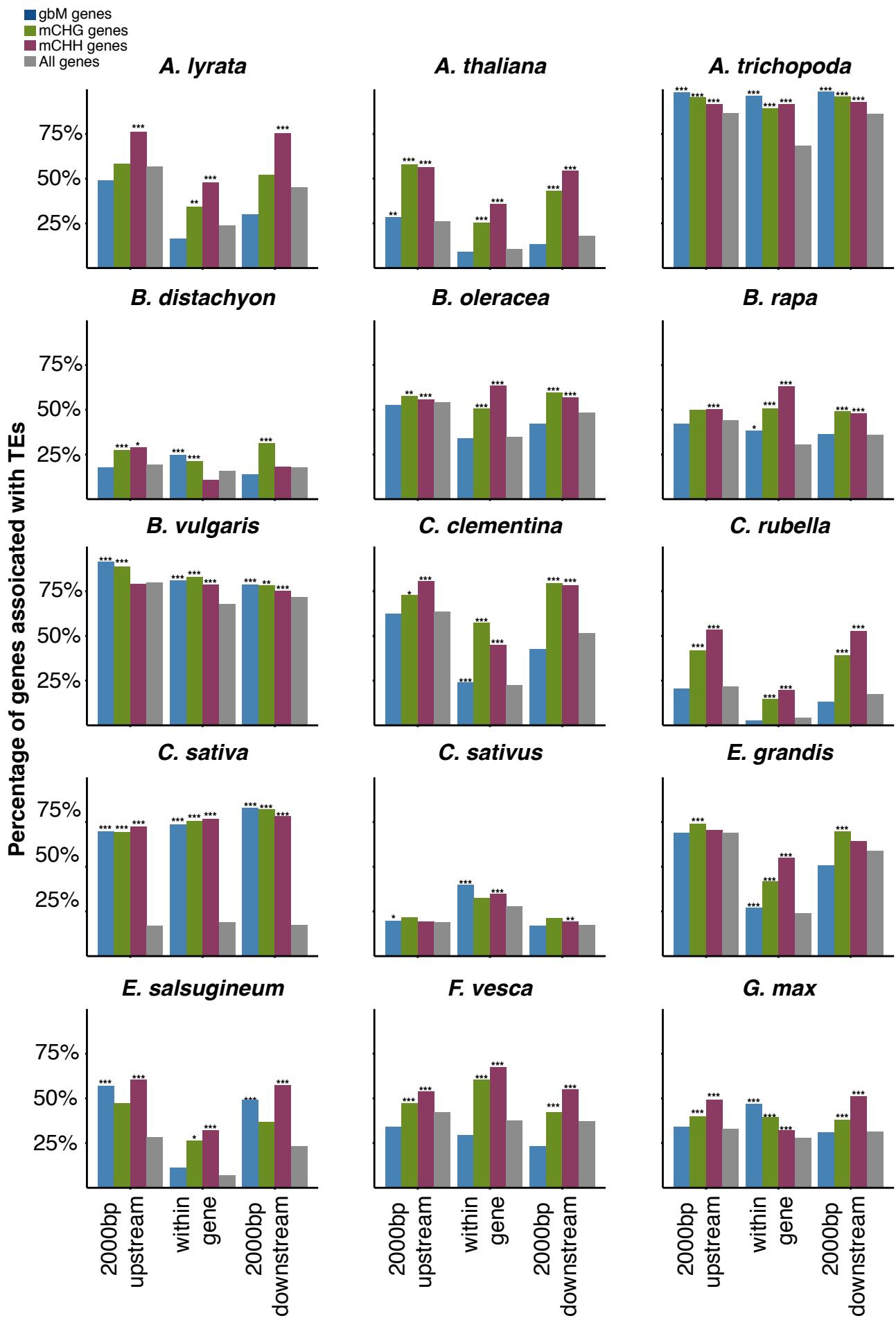
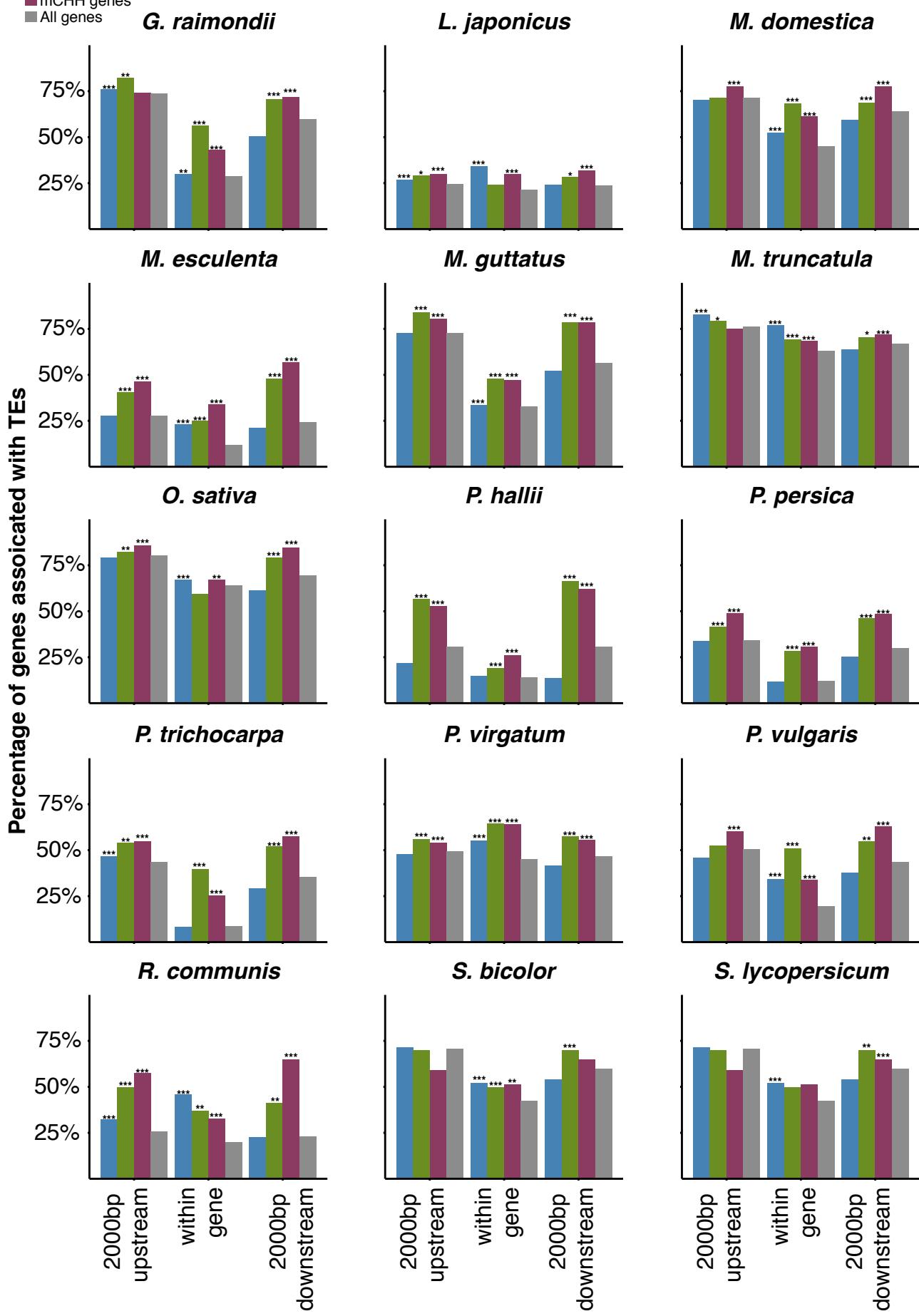


Figure S23: Association of transposable elements (TEs) 2000 bps upstream, within gene bodies, and 2000 bps downstream with gbM genes (blue), mCHG genes (green), mCHH genes (maroon), and all genes (grey). Enrichment of each TEs for each set of genes is indicated by asterisk (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001).



gbM genes
mCHG genes
mCHH genes
All genes



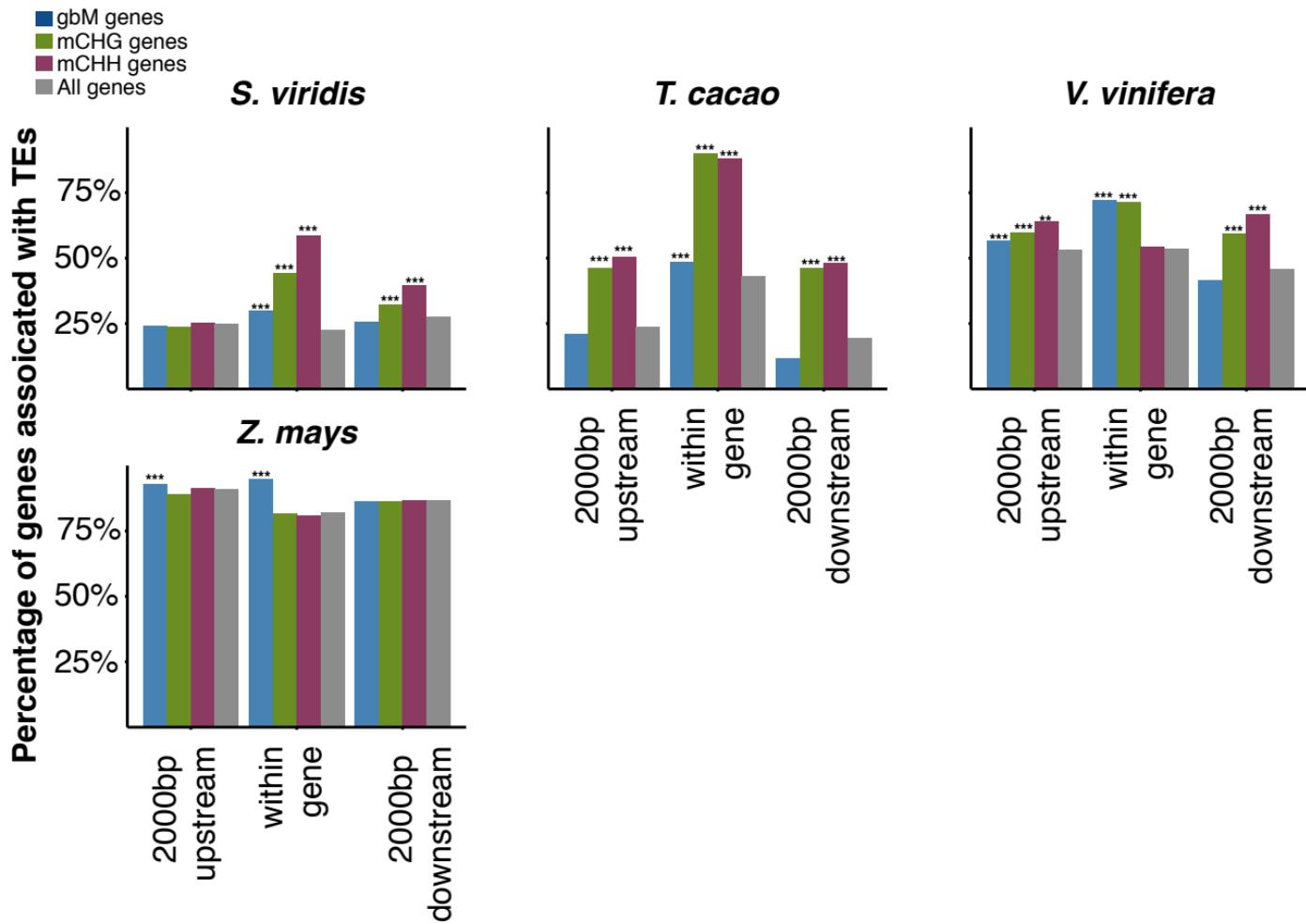
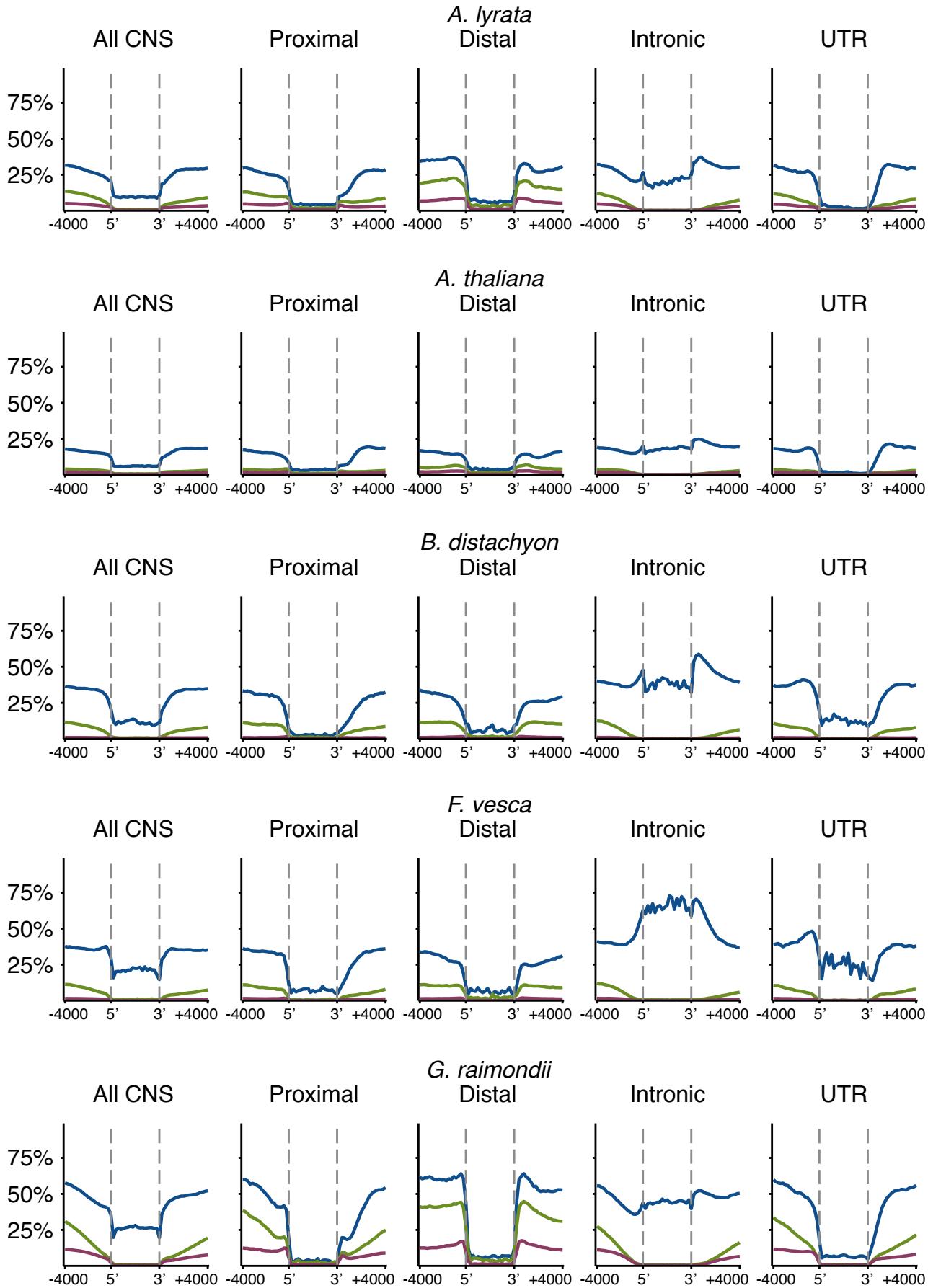


Figure S24: Methylation levels for mCG (blue), mCHG (green), and mCHH (maroon), upstream, across, and downstream for all conserved non-coding sequences (CNS), CNS proximally located to genes ($1 < 1\text{kb}$), CNS distally located to genes ($> 1\text{kb}$), and CNS located within utr and intronic sequences.



M. esculenta

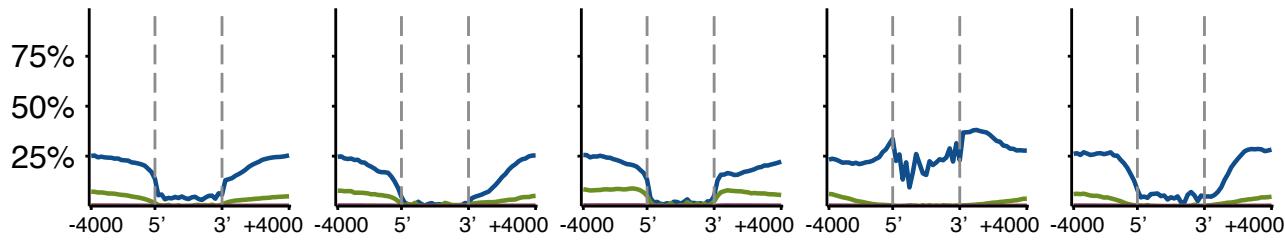
All CNS

Proximal

Distal

Intronic

UTR

*O. sativa*

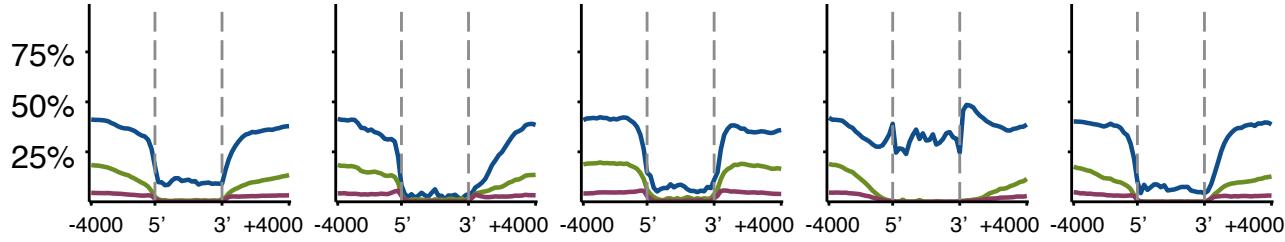
All CNS

Proximal

Distal

Intronic

UTR

*P. persica*

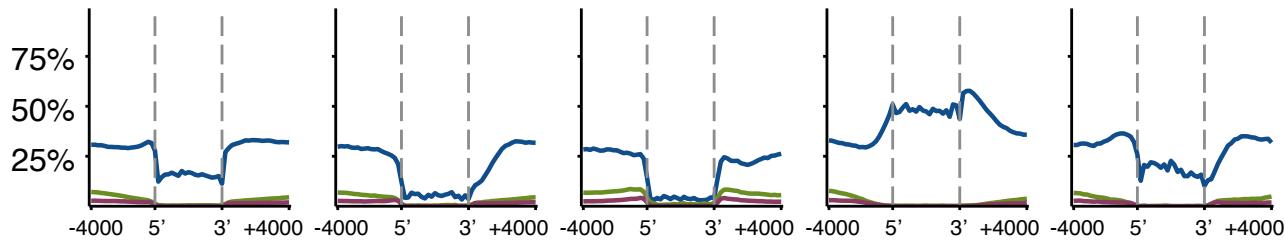
All CNS

Proximal

Distal

Intronic

UTR

*P. trichocarpa*

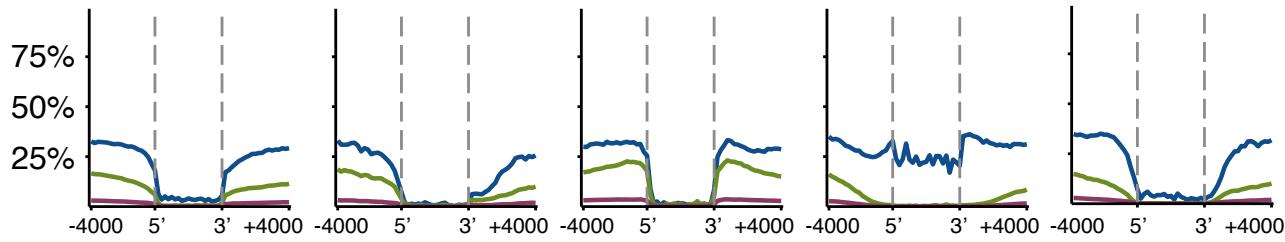
All CNS

Proximal

Distal

Intronic

UTR

*T. cacao*

All CNS

Proximal

Distal

Intronic

UTR

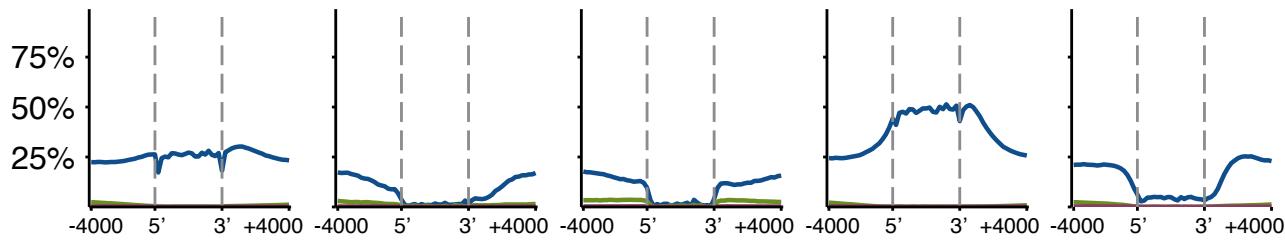


Figure S20: Association of upstream and downstream mCHH islands and gene expression. Genes are classified as NE = not expressed, 1 = first (lowest) quartile of expression, 2 = second quartile of expression, 3 = third quartile of expression, 4 = fourth (highest) quartile of expression. ** indicates a difference in proportion from the fourth quartile at $p < 0.01$.

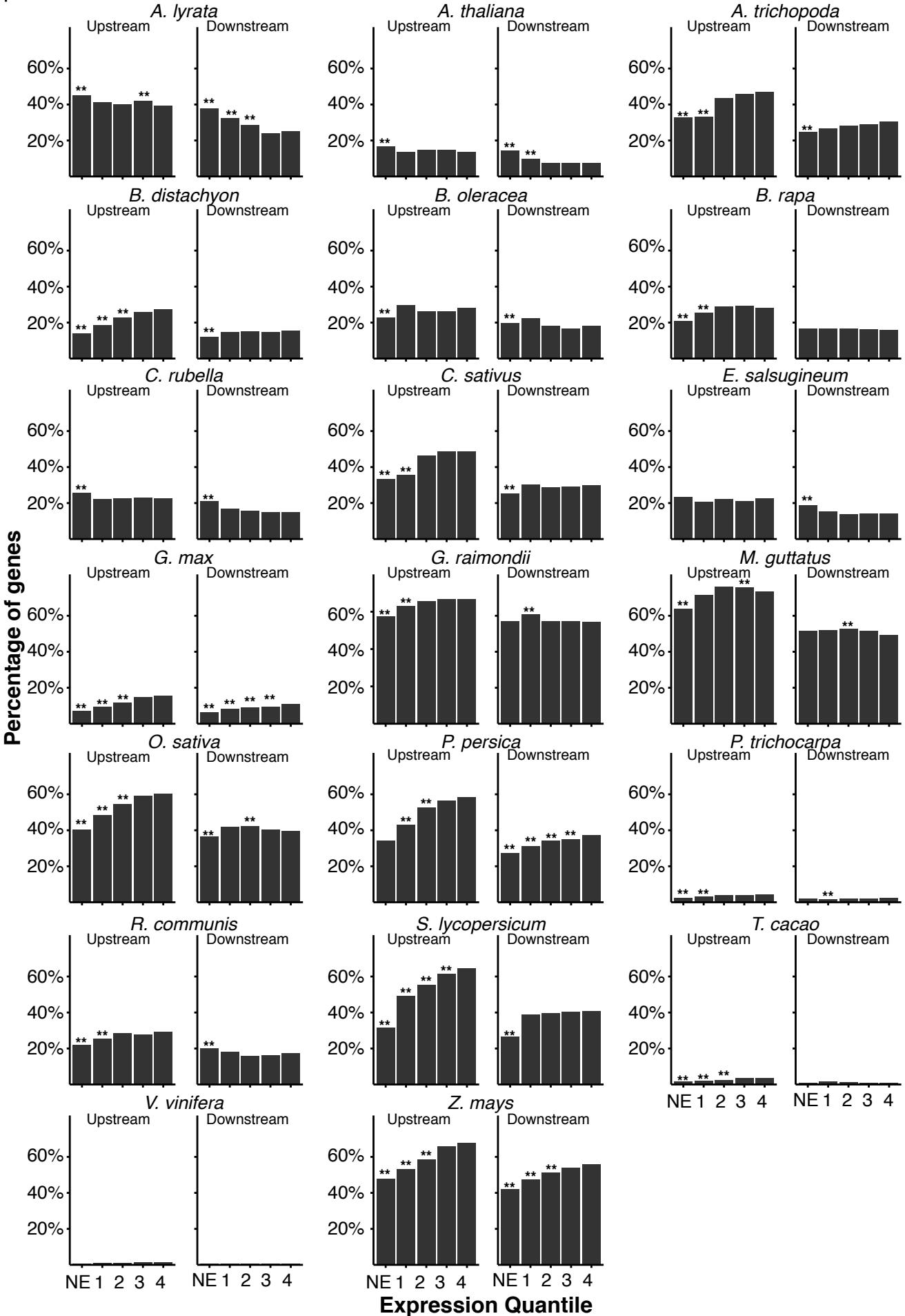
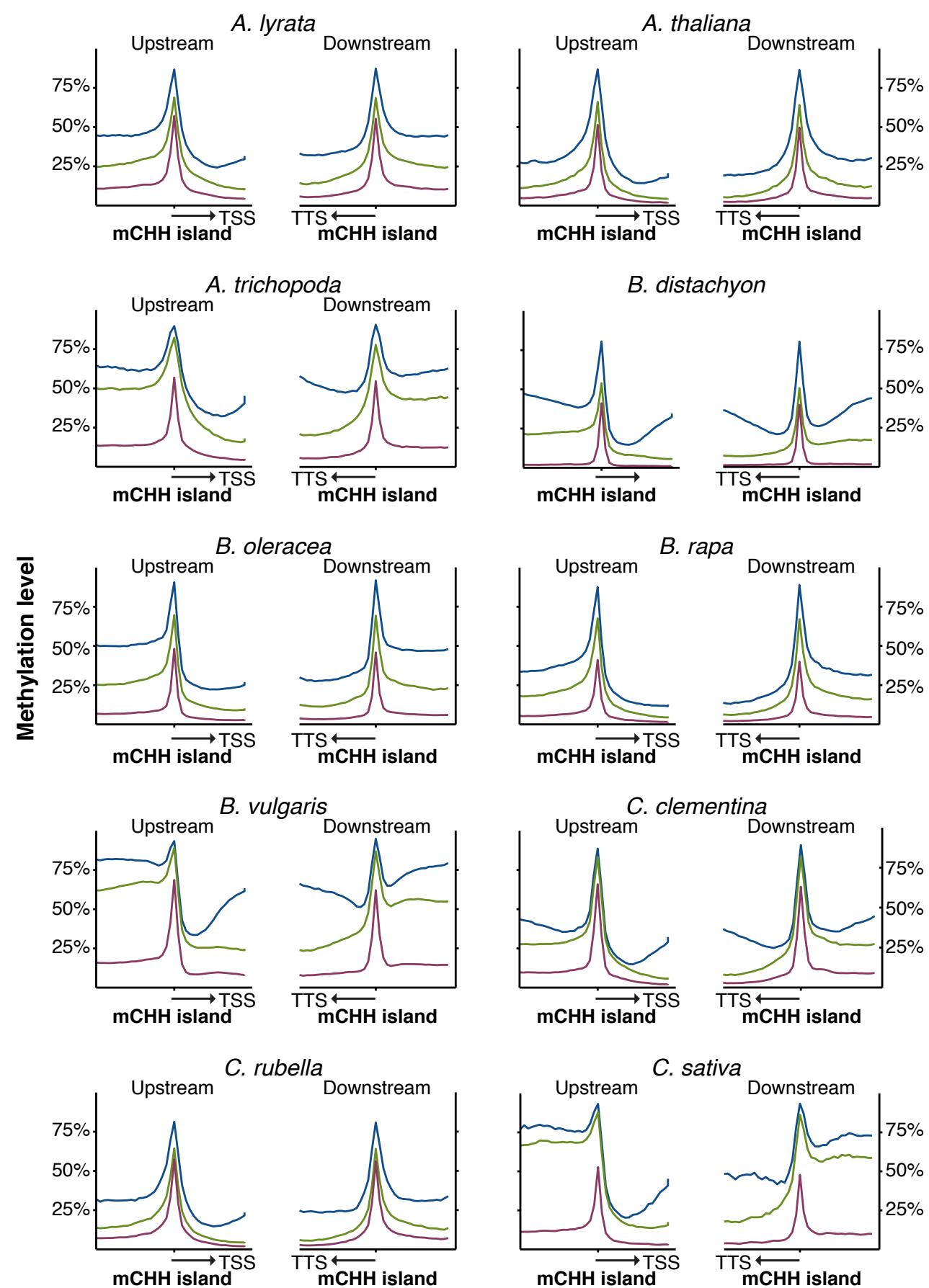
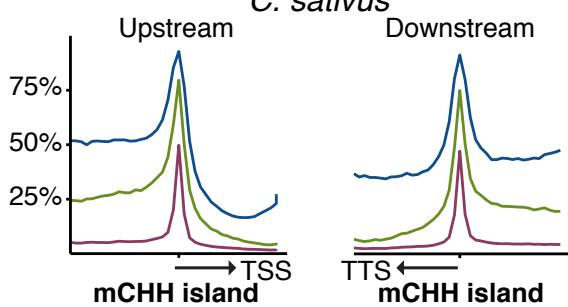


Figure S26: Methylation patterns of mCHH islands for all species. Methylation levels for mCG (blue), mCHG (green), and mCHH (maroon) were plotted for 100bp windows centered on the peak level of mCHH CHH islands found upstream of genes.

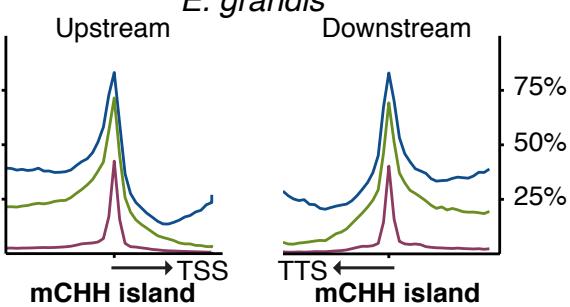


Methylation level

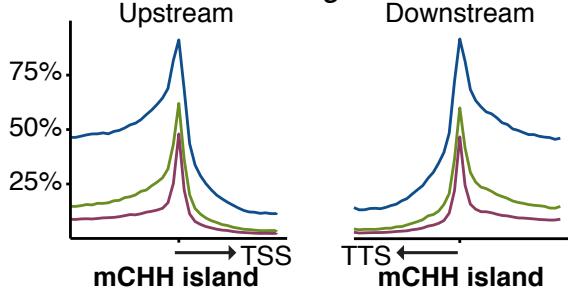
C. sativus



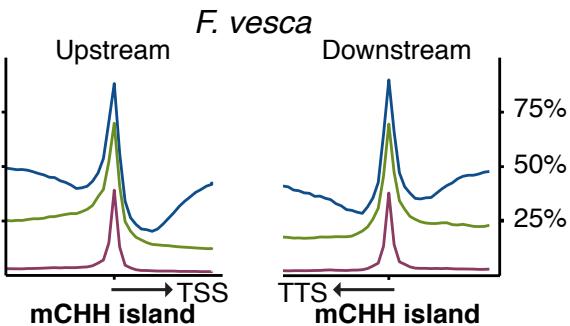
E. grandis



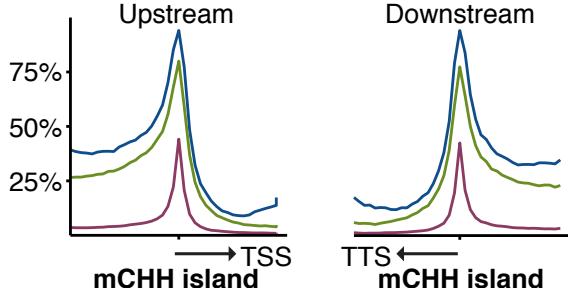
E. salsugineum



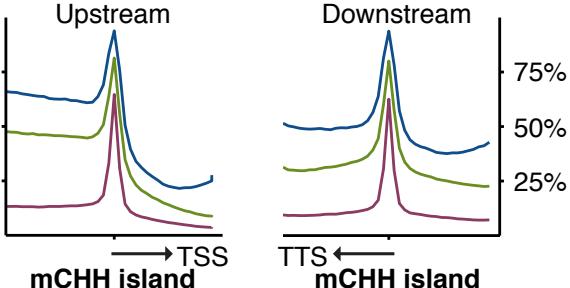
F. vesca



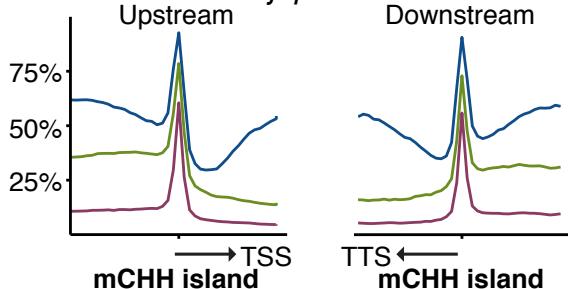
G. max



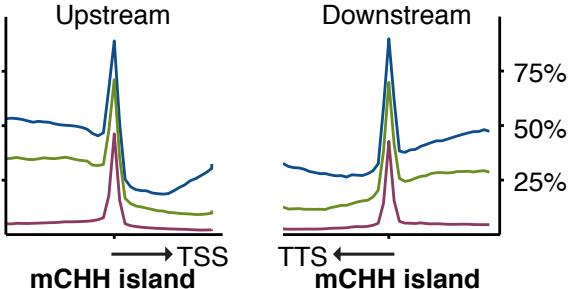
G. raimondii



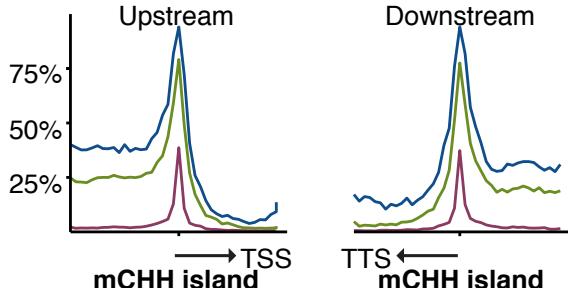
L. japonicus



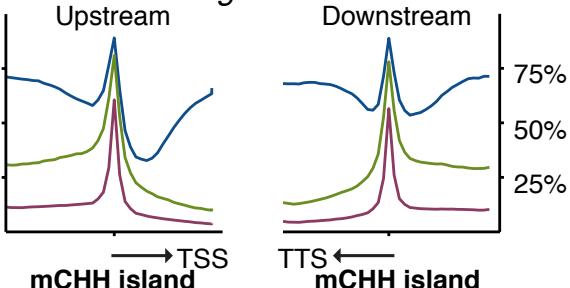
M. domestica



M. esculenta

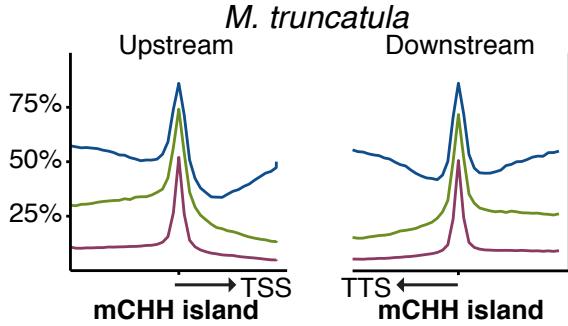


M. guttatus

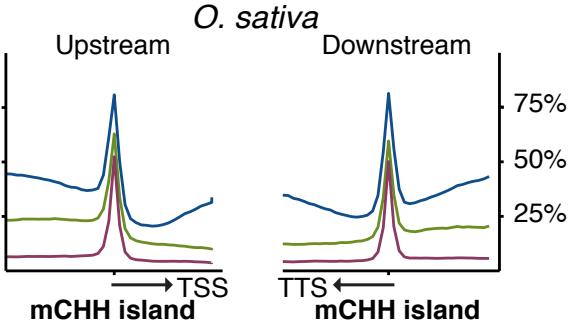


Methylation level

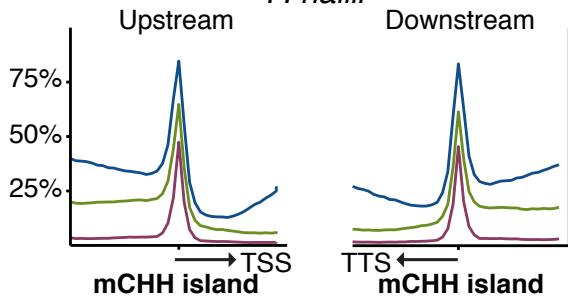
M. truncatula



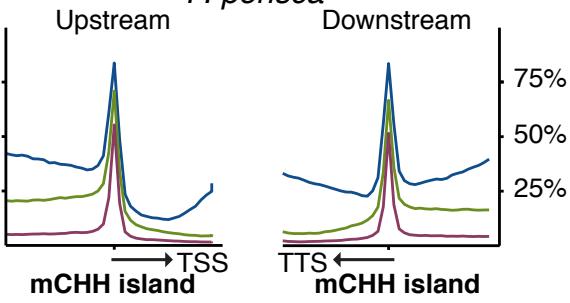
O. sativa



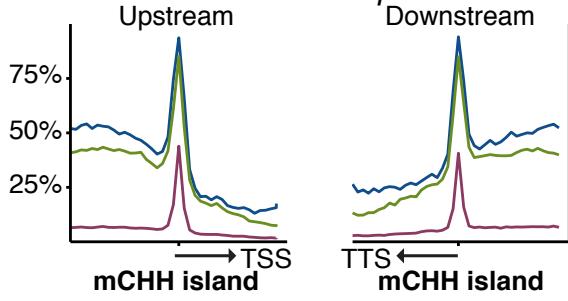
P. hallii



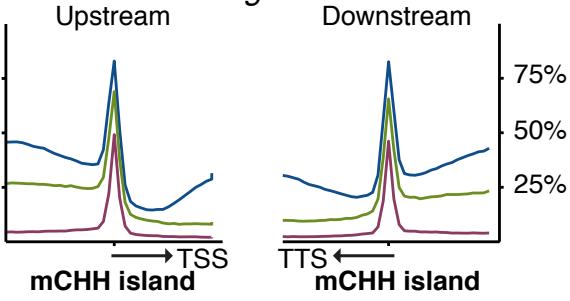
P. perisca



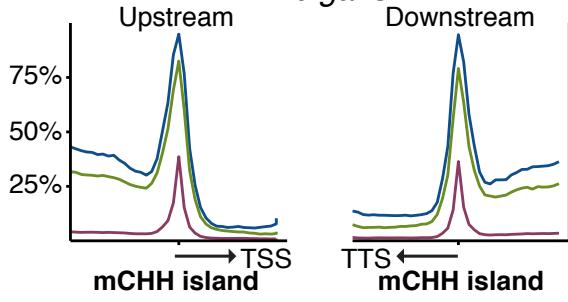
P. trichocarpa



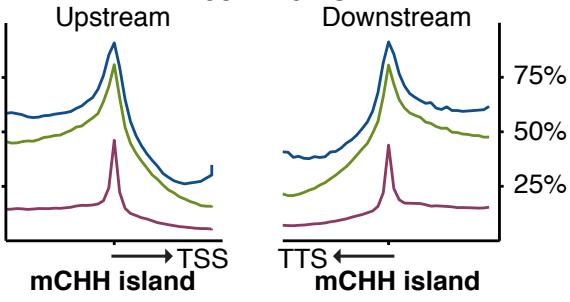
P. virgatum



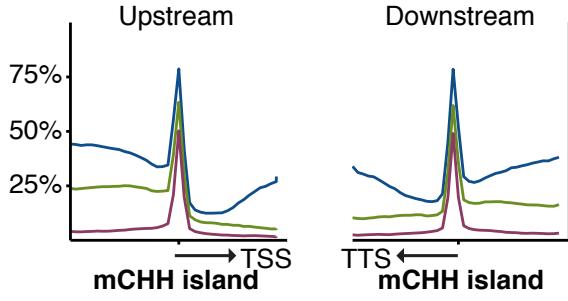
P. vulgaris



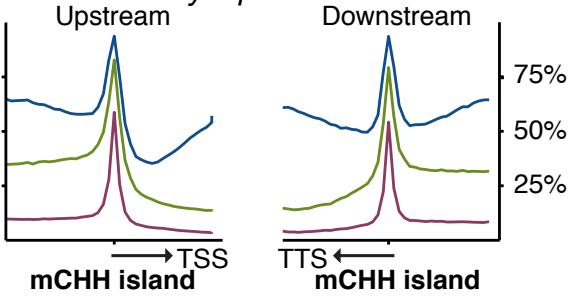
R. communis



S. bicolor



S. lycopersicum



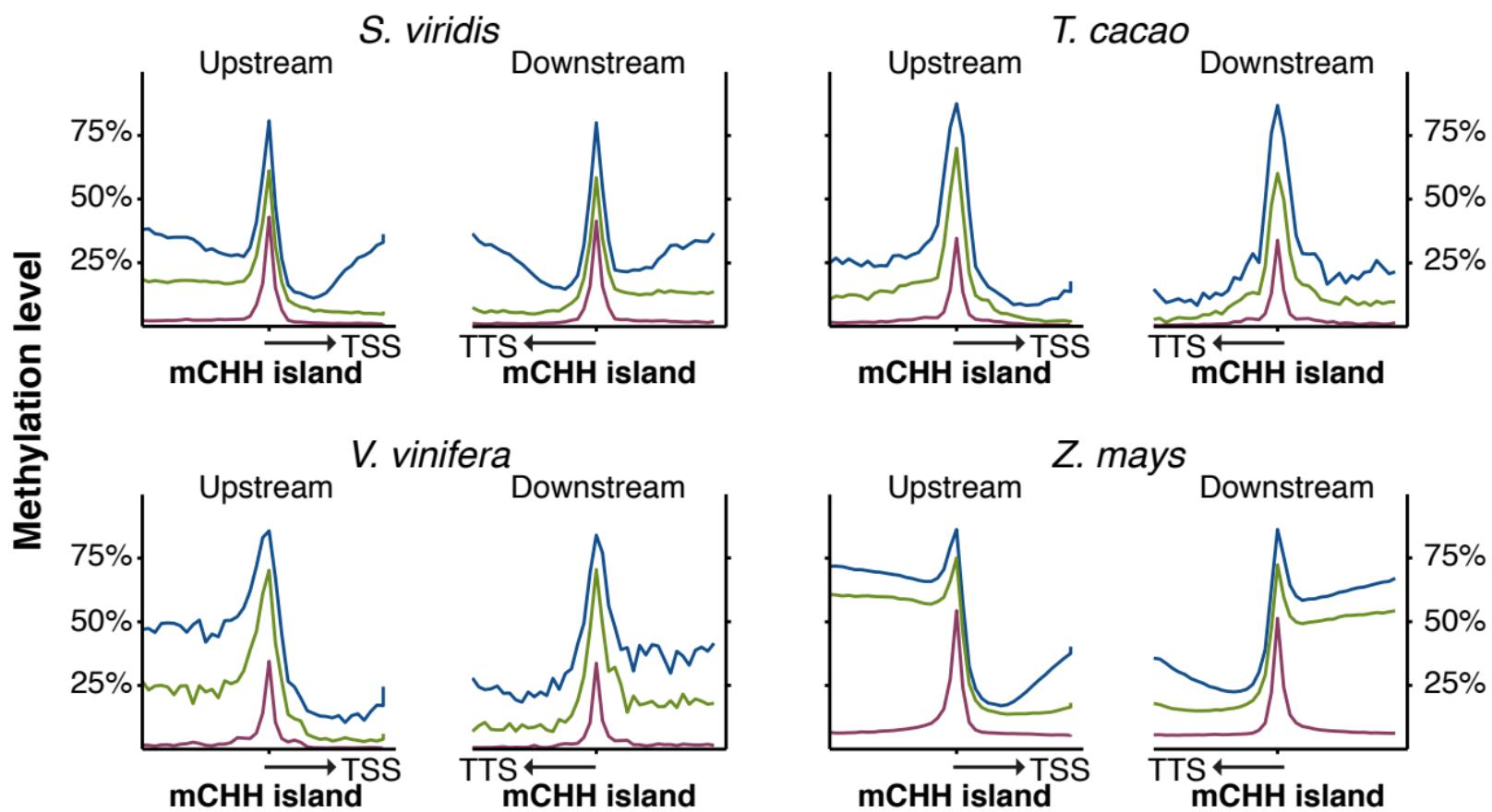


Figure S27: (A) Percentage of genes with upstream (blue) or downstream (orange) mCHH-islands plotted against total number of repeats. **(B)** Upstream and downstream mCHH islands are correlated with percentage upstream (blue) and percentage of downstream repeats (orange). **(C)** Percentage of genes with upstream (blue) or downstream (orange) repeats plotted against total number of repeats.

