TITLE

Diversity of cytosine methylation across the fungal tree of life

AUTHORS

Adam J. Bewick[1], Brigitte T. Hofmeister[2], Rob Powers[3], Stephen J. Mondo[4], Igor V. Grigoriev[4,5], Timothy Y. James [3], Jason E. Stajich[6], Robert J. Schmitz[1]

AFFILIATIONS

[1]Department of Genetics, University of Georgia, Athens, Georgia, USA. [2]Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA. [3]Department of Ecology and Evolutionary Biology, University of Michigan, Michigan, USA. [4]US Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [5]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, California, USA. [6]Department of Microbiology and Plant Pathology, University of California, Riverside, Riverside, California, USA.

CORRESPONDING AUTHORS

Robert J. Schmitz schmitz@uga.edu and Adam J. Bewick bewickaj@uga.edu

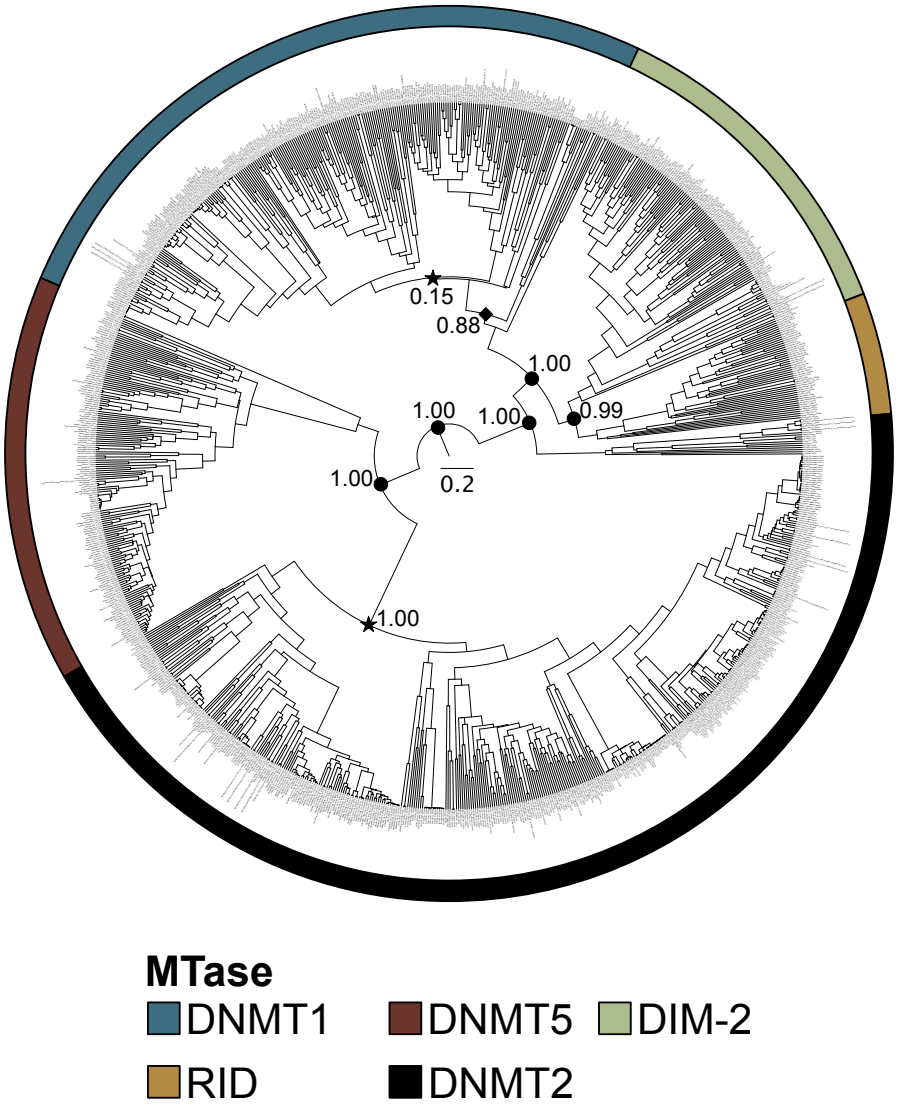
SUPPLEMENTARY INFORTMATION

Supplementary Figure 1

5mC DNA and tRNA MTase phylogeny of fungi.

Evolution of DNA and tRNA methyltransferases across fungi. Values at selected nodes indicate posterior probability. Nodes with a star specify duplications, and the single node with a diamond specifies the clade containing '*DnmtX*' (29). Branch lengths are in units of amino acid substitutions per amino acid site. Phyla abbreviations: Ascomycota (Asco-), Basidiomycota (Basidio-), Blastocladiomycota (Blastocladio-), Chytridiomycota (Chytridio-), Mucoromycota (Mucoro-), and Zoopagomycota (Zoopago-).

Supplementary Fig. 1



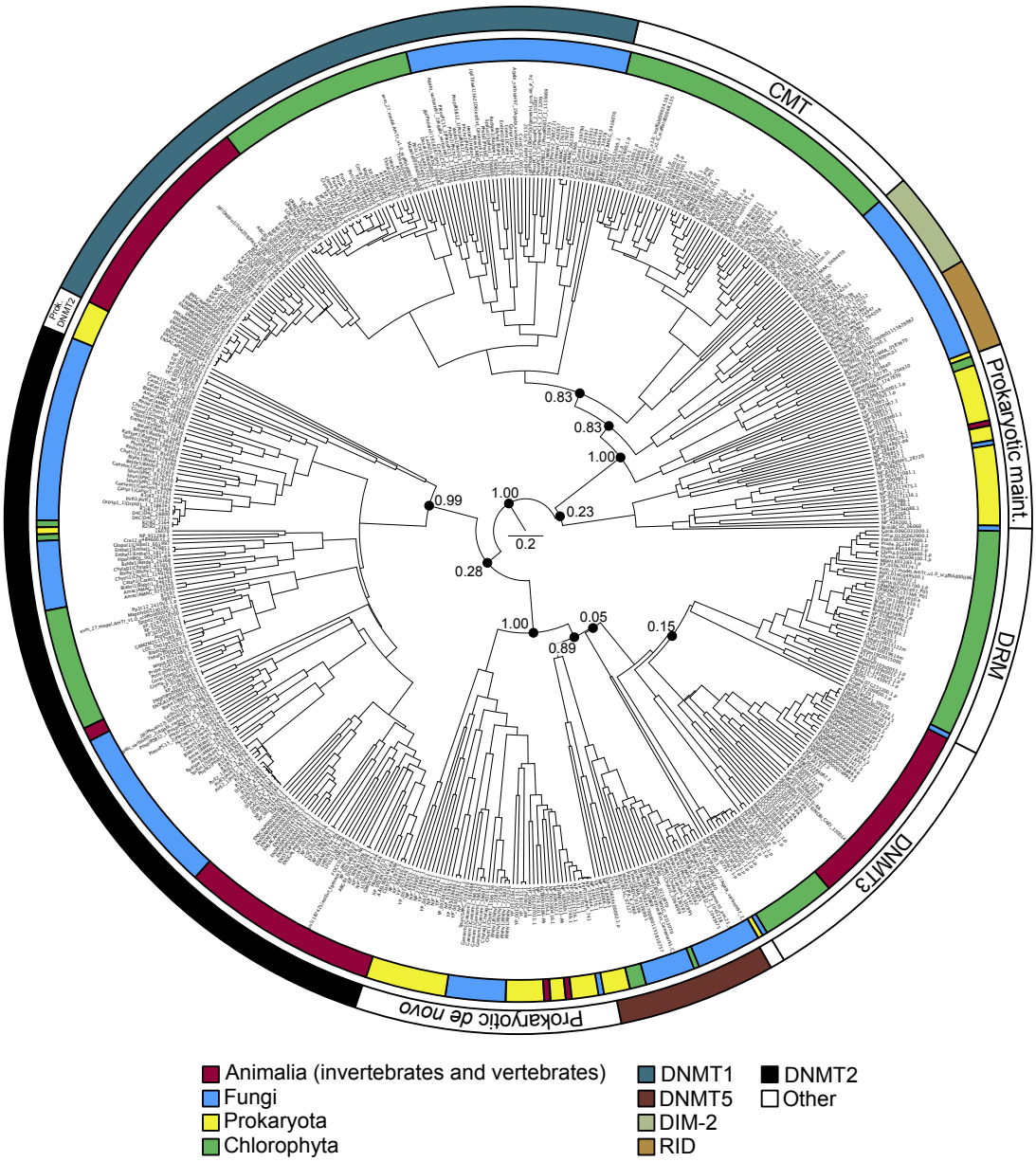
63
64
65
66
67

Supplementary Figure 2

5mC DNA and tRNA MTase phylogeny of animalia, chlorophyta, fungi, and prokaryota.

Evolution of DNA and tRNA methyltransferases across the tree of life. Values at selected nodes indicate posterior probability. Branch lengths are in units of amino acid substitutions per amino acid site. Phyla abbreviations: Ascomycota (Asco-), Basidiomycota (Basidio-), Blastocladiomycota (Blastocladio-), Chytridiomycota (Chytridio-), Mucoromycota (Mucoro-), and Zoopagomycota (Zoopago-).

Supplementary Fig. 2



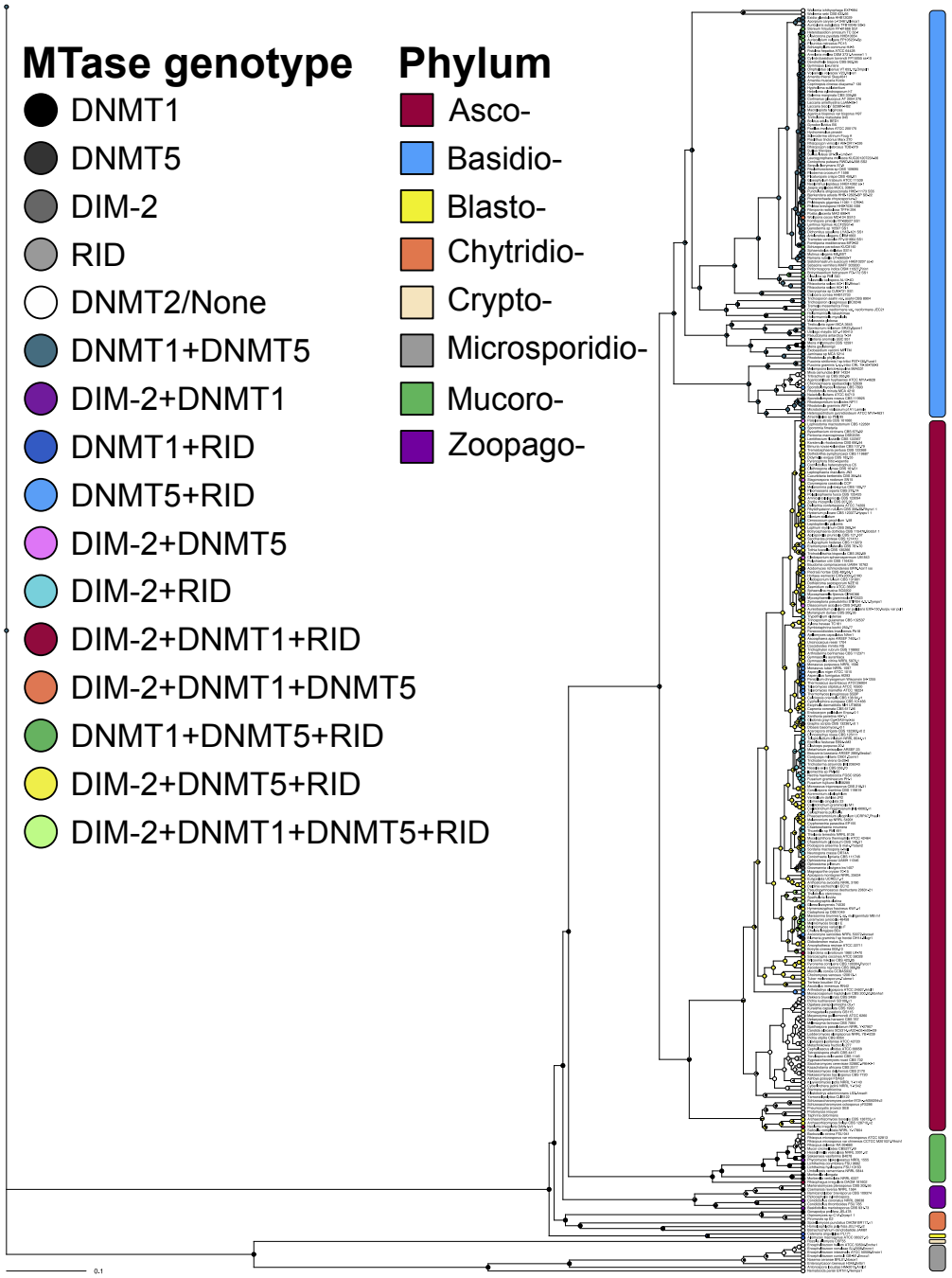
99
100
101
102
103

Supplementary Figure 3

Stochastic character mapping of sixteen observed DNA MTase genotypes.

The colored bars indicate phylum of encompassing fungal species. Phyla abbreviations: Ascomycota (Asco-), Basidiomycota (Basidio-), Blastocladiomycota (Blastocladio-), Chytridiomycota (Chytridio-), Mucoromycota (Mucoro-), and Zoopagomycota (Zoopago-). The colored circle adjacent to each species corresponds to the observed genotype. Pie charts at nodes indicate empirical Bayesian posterior probabilities of the genotypes in a common ancestor. Branch lengths are in units of amino acid substitutions per amino acid site.

Supplementary Fig. 3

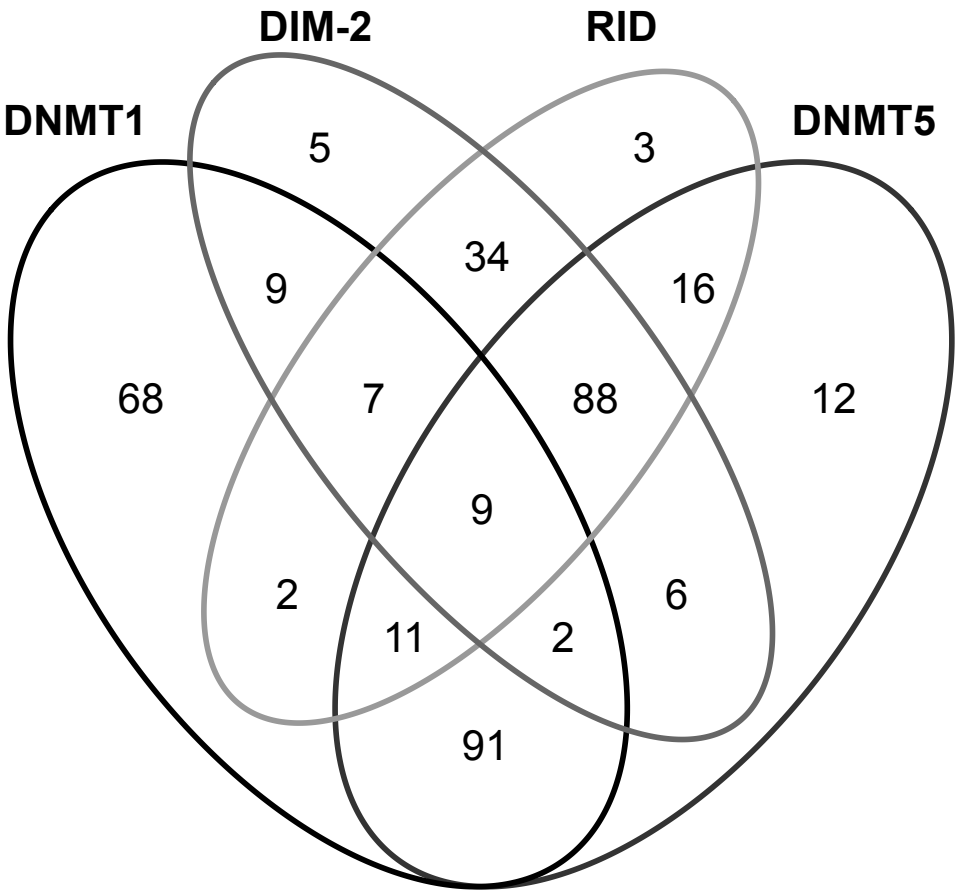


Supplementary Figure 4

Individual and combinations of DNA MTases.

The values within the areas of the Venn diagram indicate the number of fungal species/strains with the corresponding DNA MTase genotype.

Supplementary Fig. 4



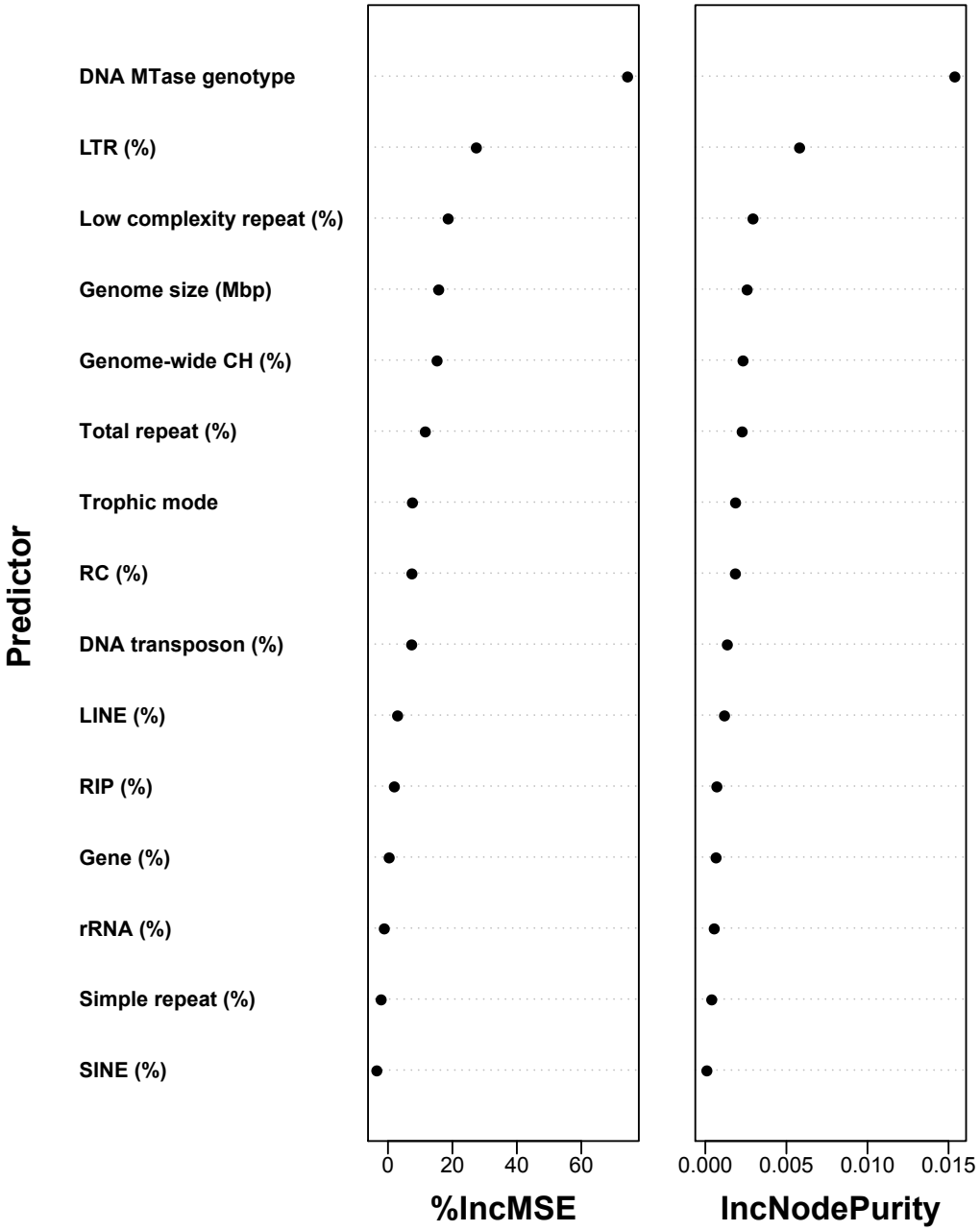
171
172
173
174
175

Supplementary Figure 5

Predictors of methylation.

Rank-order from best to worse predictors of genome-wide CG methylation.

Supplementary Fig. 5



207
208
209
210
211

212 **Supplementary Figure 6**

213

214 **Explanations for discrepancies between DNA MTase genotype and**
215 **methylation levels.**

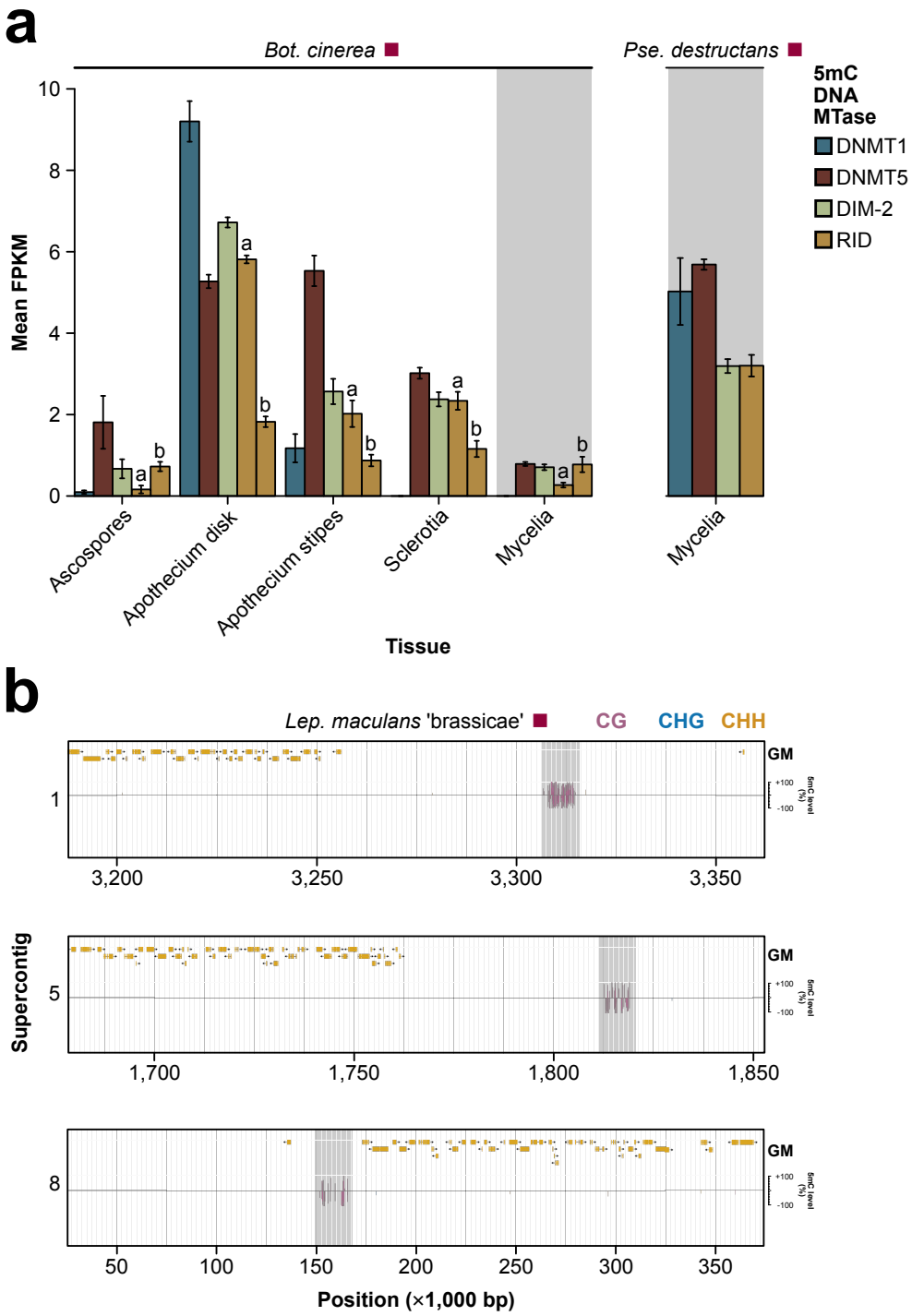
216

217 (**a**) Tissue-specific expression of DNA MTases in *Bot. cinerea* and *Pse.*
218 *destructans*. Grey-highlighted boxes specify tissues used for methylation
219 analyses. *a* and *b* above bars in *Bot. cinerea* specify RID in-paralogs. Error bars
220 correspond to the standard error of the mean (SEM). (**b**) Methylation in *Lep.*
221 *maculans* 'brassicae' is localized to short, gene-poor regions at the beginnings
222 and ends of supercontigs.

223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242

243

Supplementary Fig. 6



244

245

246

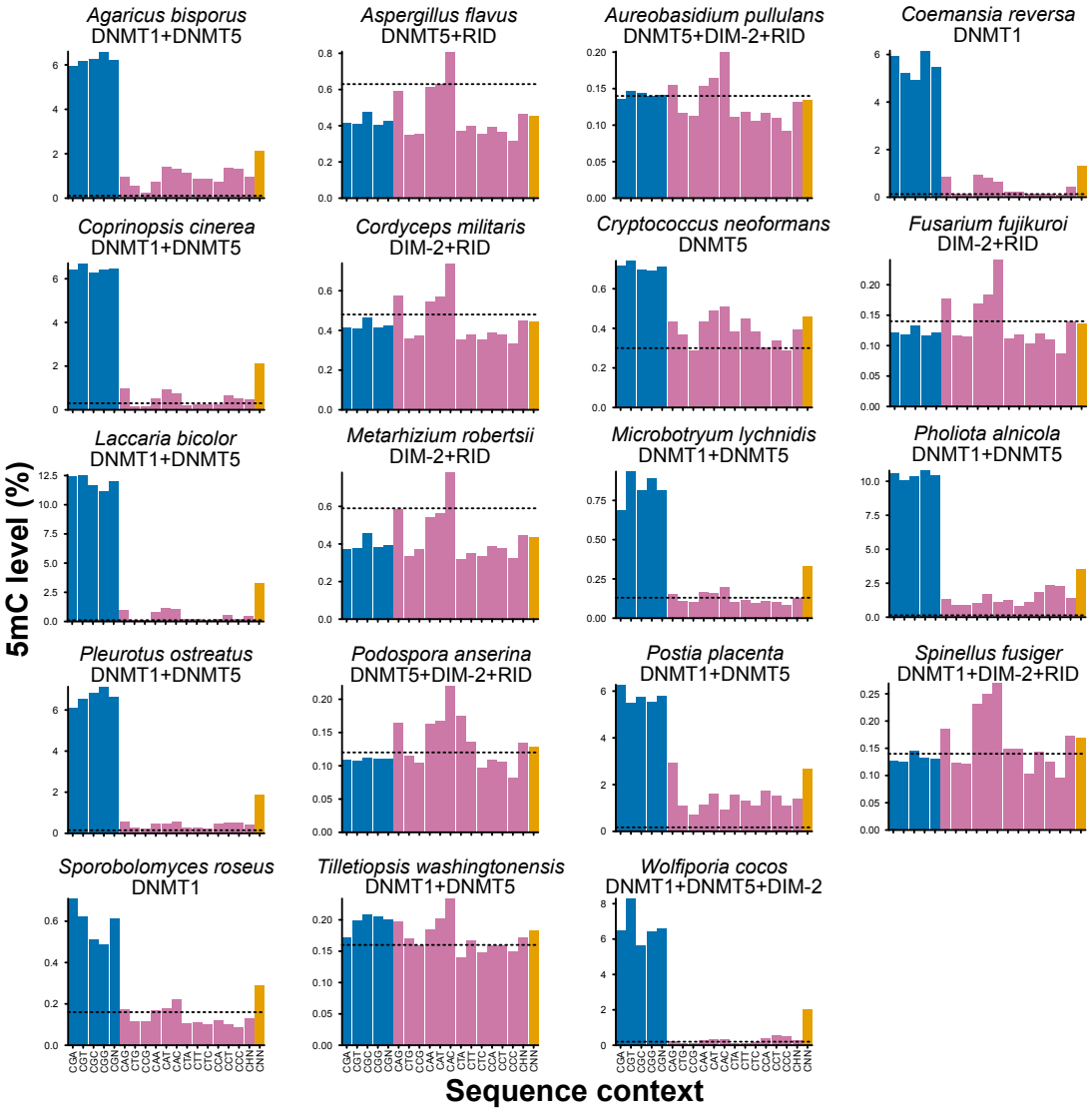
247

Supplementary Figure 7

Context-specific methylation.

Weighted methylation levels at all sequence contexts found across the genome for fungal species not included in Fig. 2.

Supplementary Fig. 7



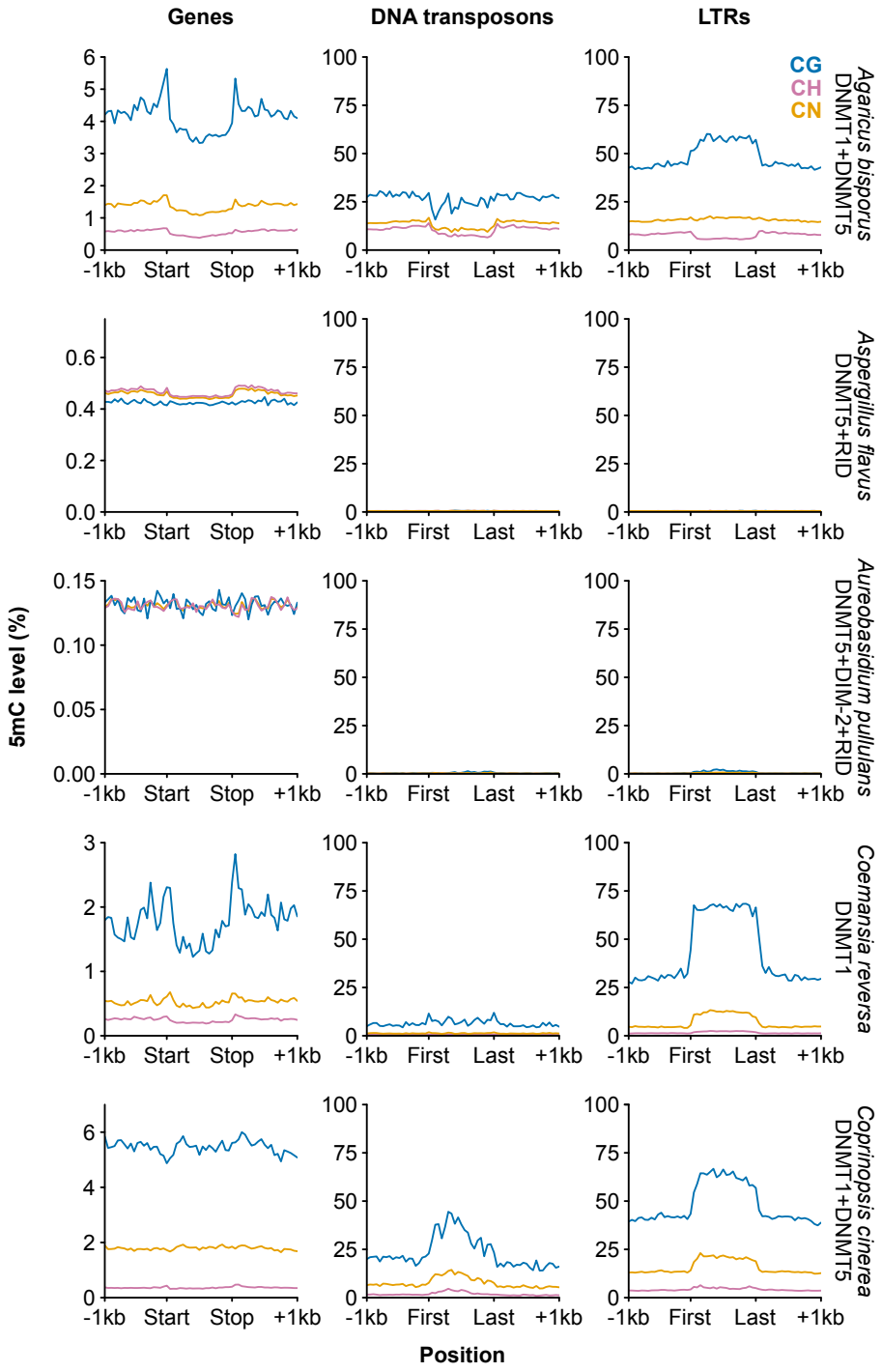
Supplementary Figure 8

Genic and repeat and transposon methylation profiles.

Weighted methylation at CG, CH, and CN sites upstream, within, and downstream of all genes for fungal species not included in Fig. 4.

315

Supplementary Fig. 8



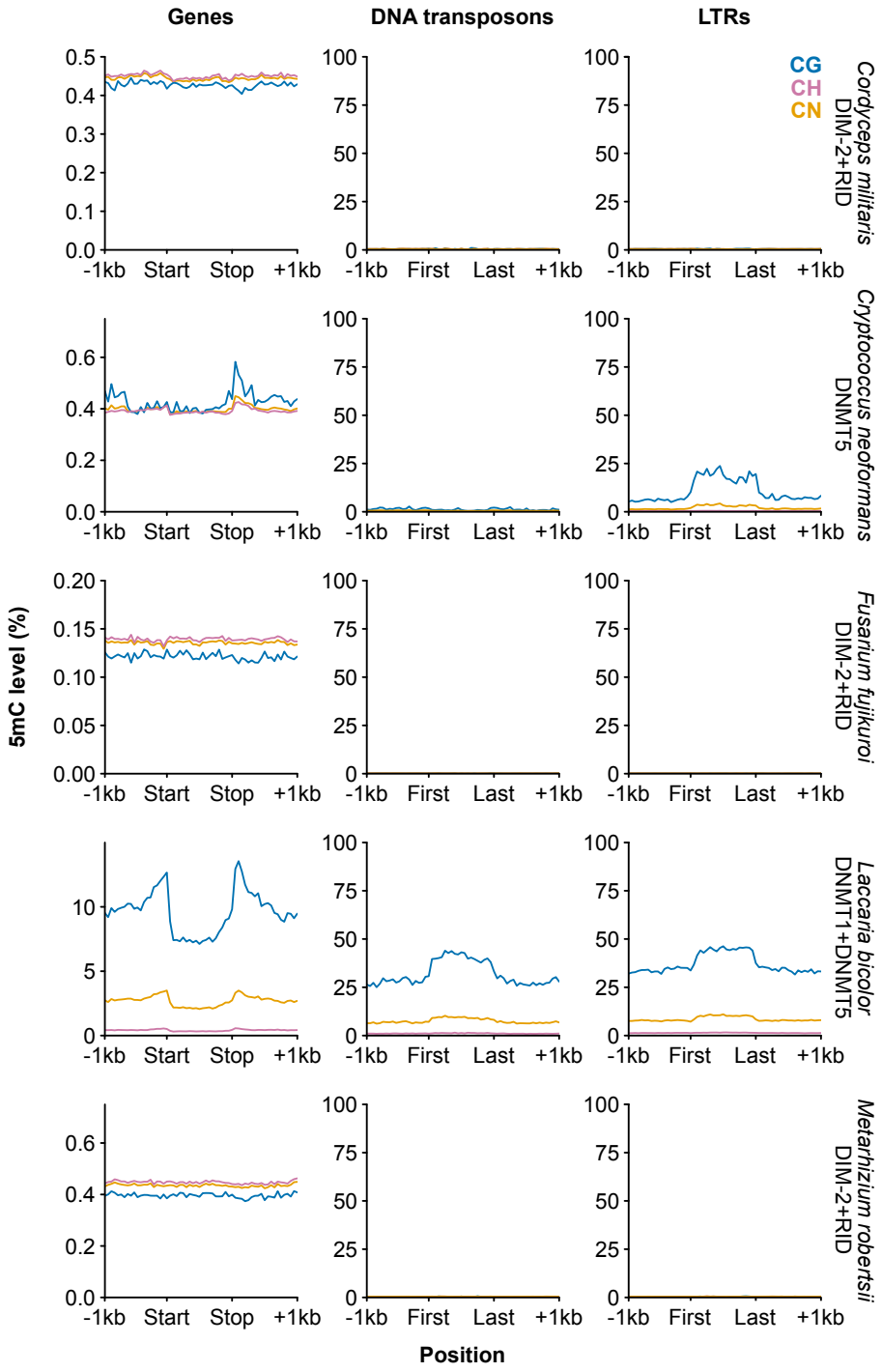
316

317

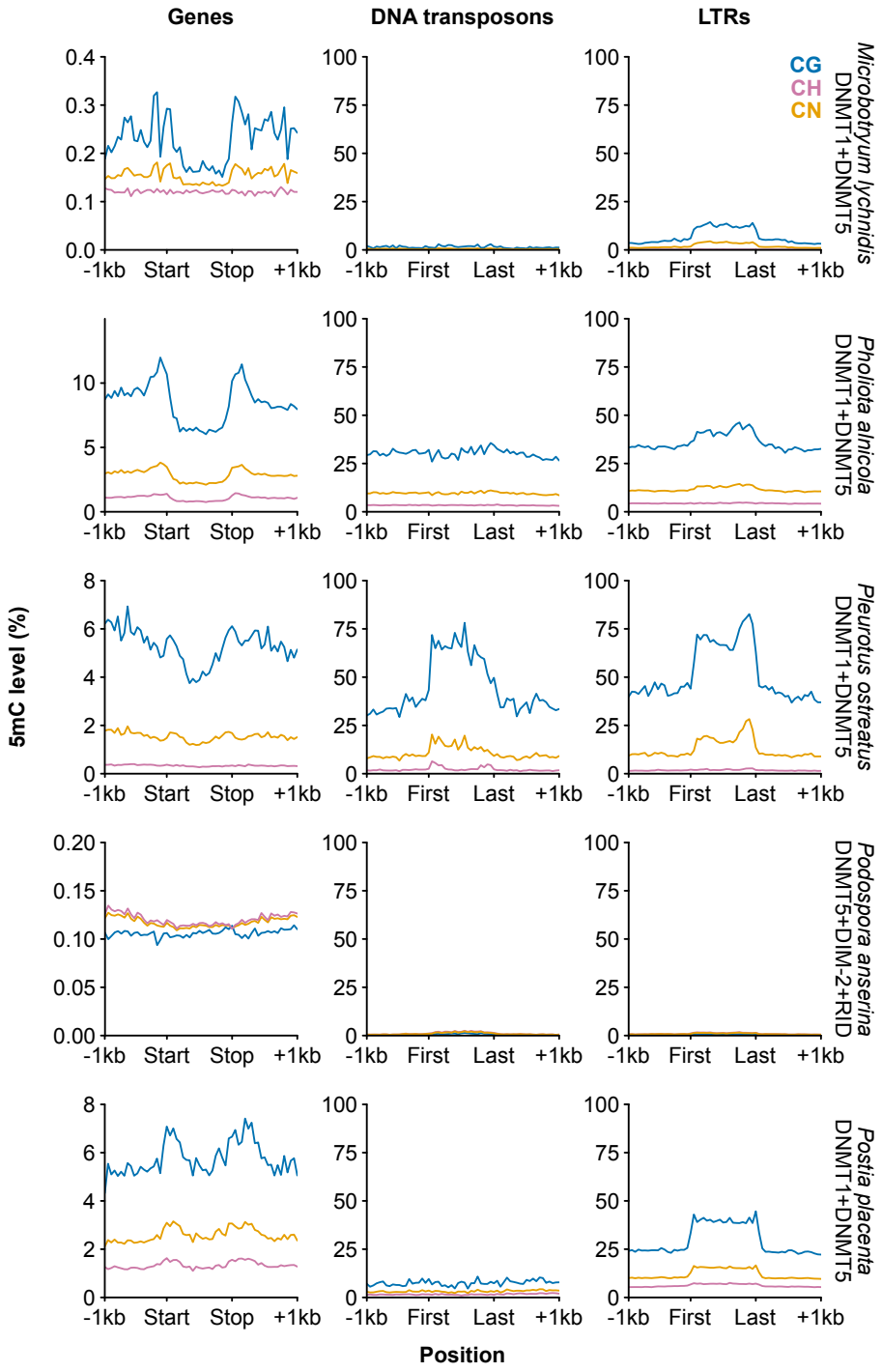
318

319

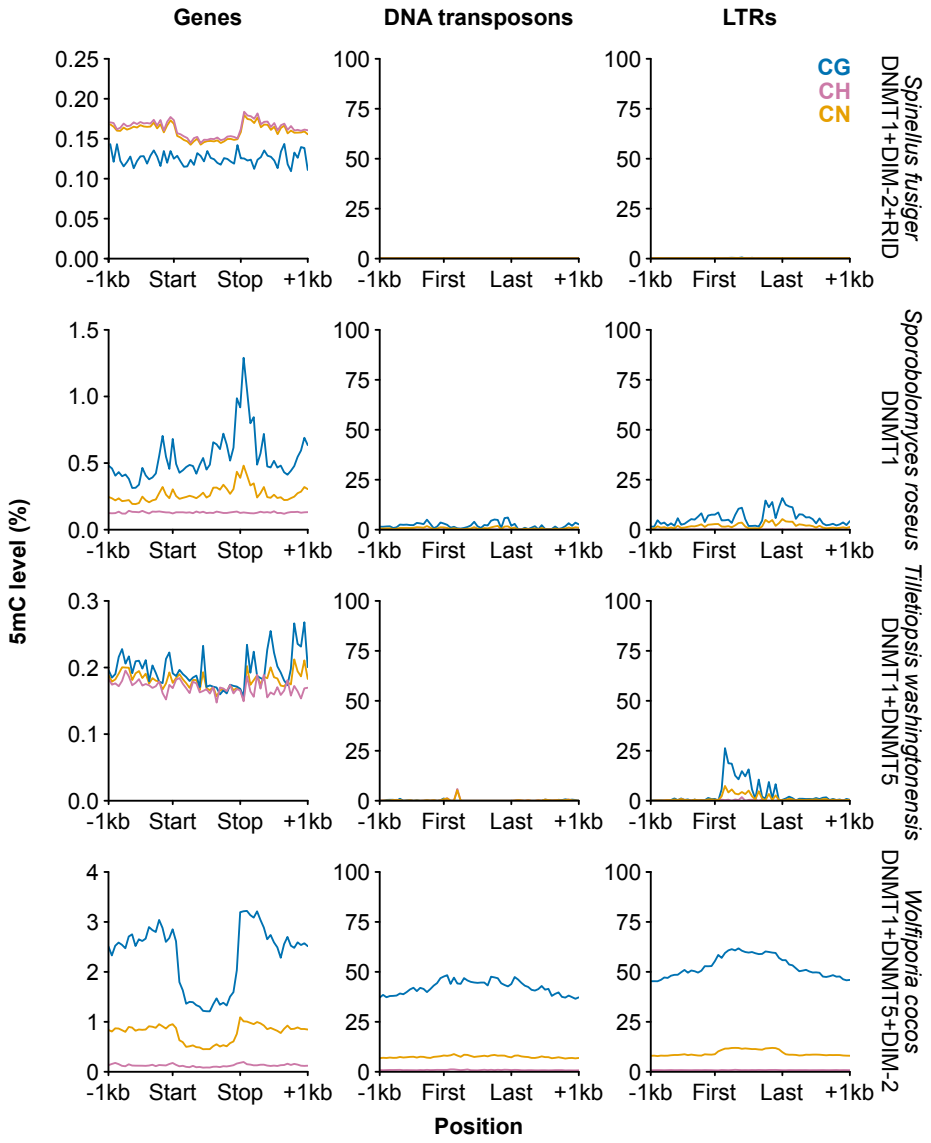
Supplementary Fig. 8



Supplementary Fig. 8



Supplementary Fig. 8



330

331

332

333

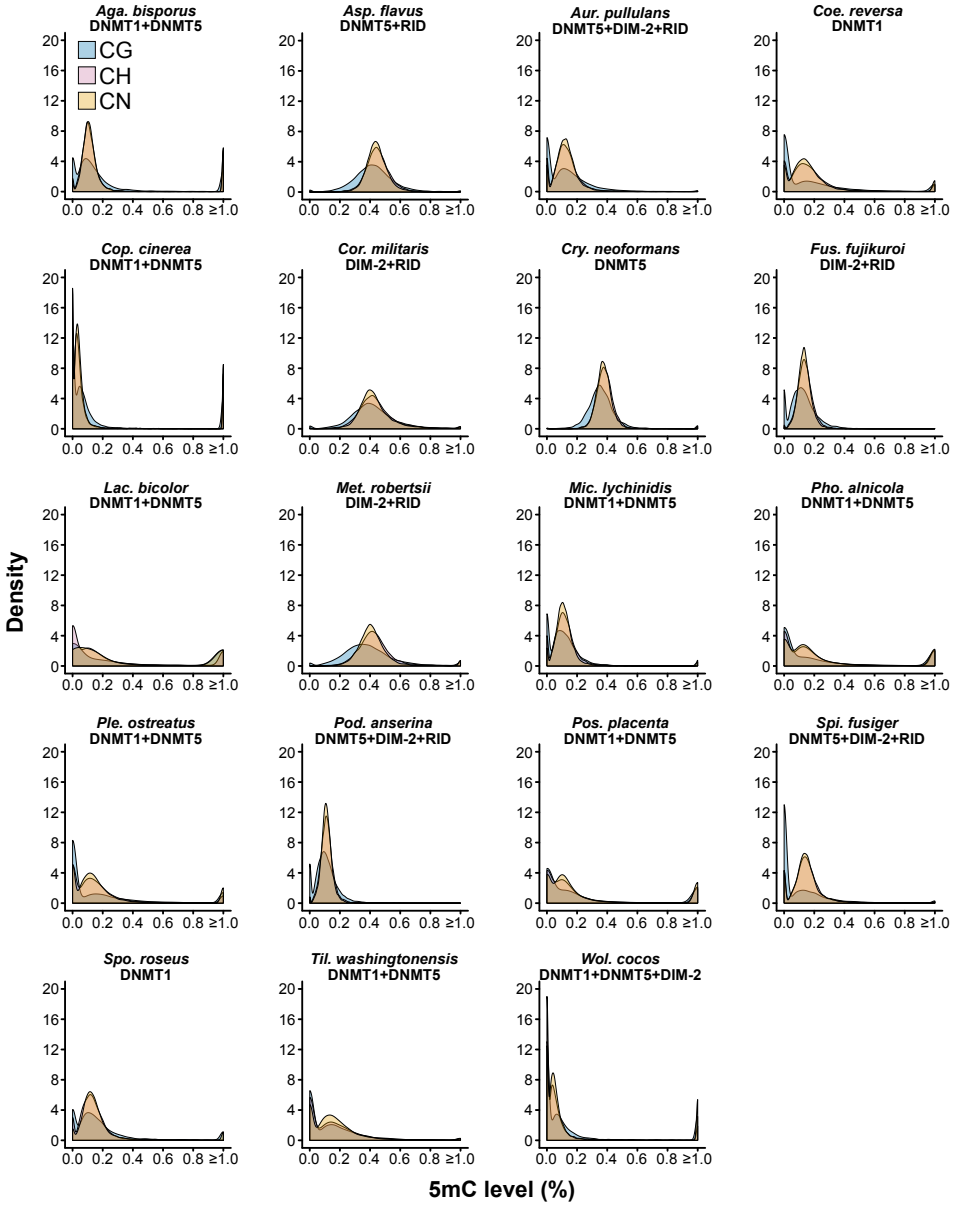
334

Supplementary Figure 9

Genic 5mC density.

Distribution of 5mC levels for genes for species not used in Fig. 3b.

Supplementary Fig. 9



366

367

368

369

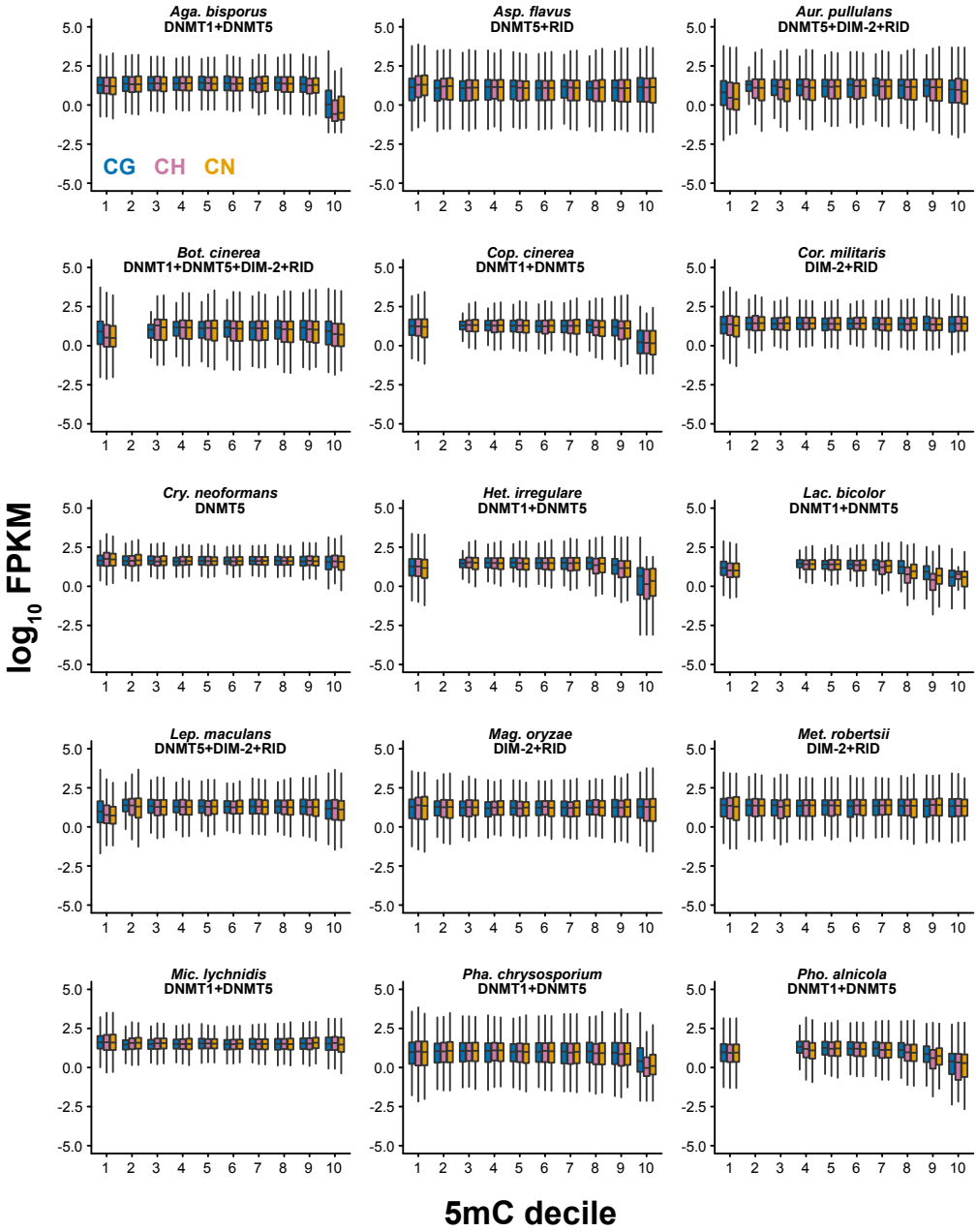
370

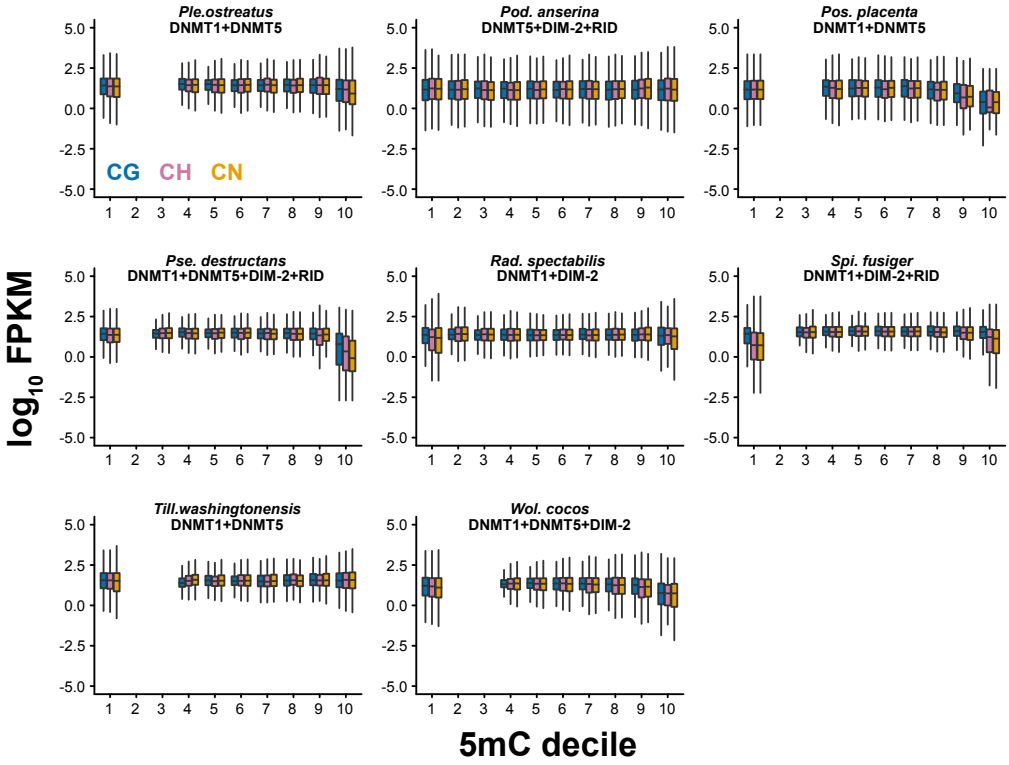
Supplementary Figure 10

Relationship between gene expression (Fragments Per Kilobase of transcript per Million mapped reads [FPKM]) and genic 5mC levels.

Relationship between gene expression measured as FPKM and 5mC levels. Empty deciles correspond to missing genic data for those 5mC levels. Boxplot elements: center line, median; upper and lower "hinges", first and third quartiles (the 25th and 75th percentiles), respectively; whiskers, 1.5× interquartile range.

Supplementary Fig. 10





406
407
408
409
410

411 **Supplementary Figure 11**

412

413 **DNA transposon and Long Terminal Repeat (LTR) 5mC density.**

414

415 Distribution of 5mC levels for DNA transposons and LTRs for species not used in
416 Fig. 4b and c.

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

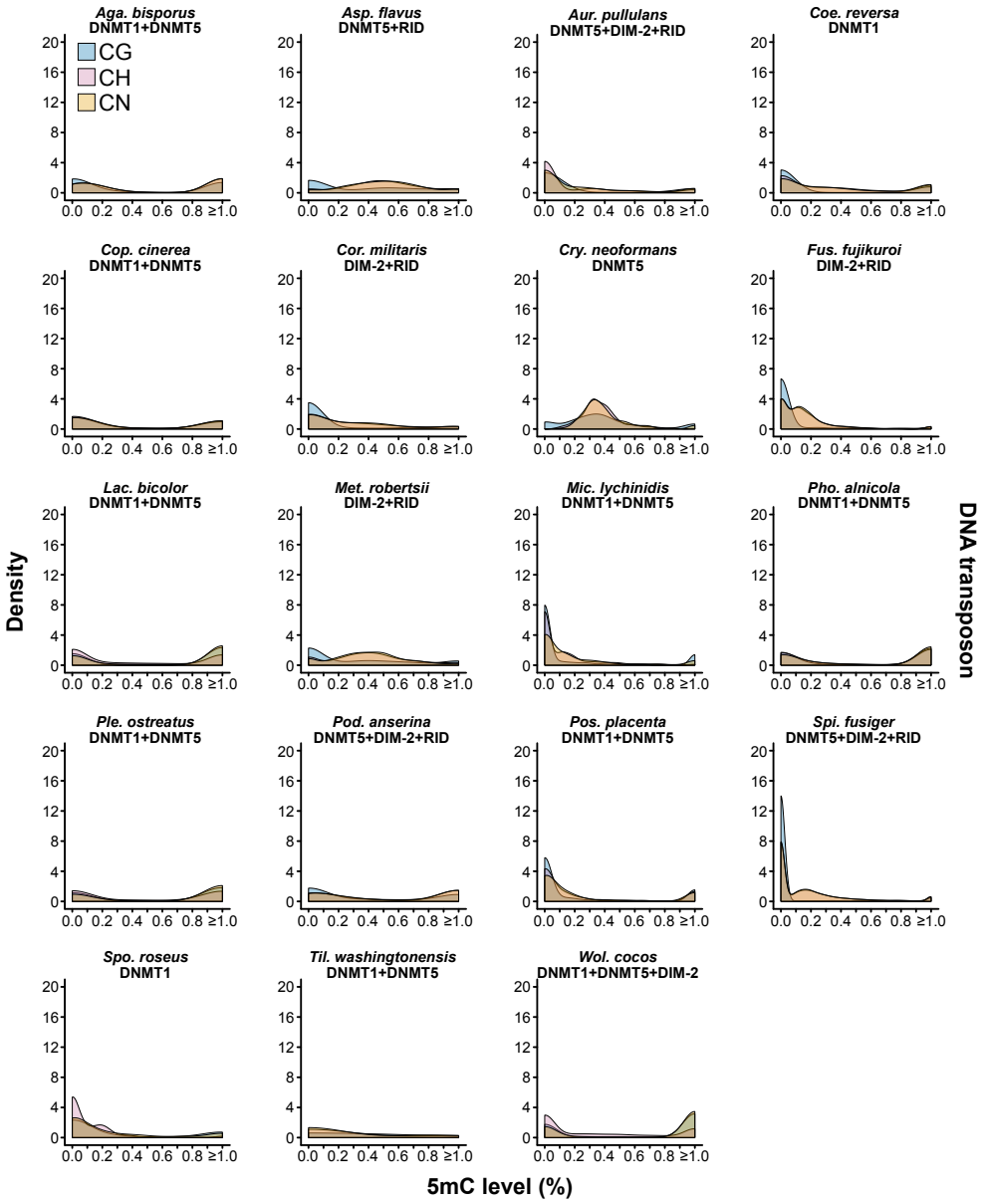
438

439

440

441

Supplementary Fig. 11

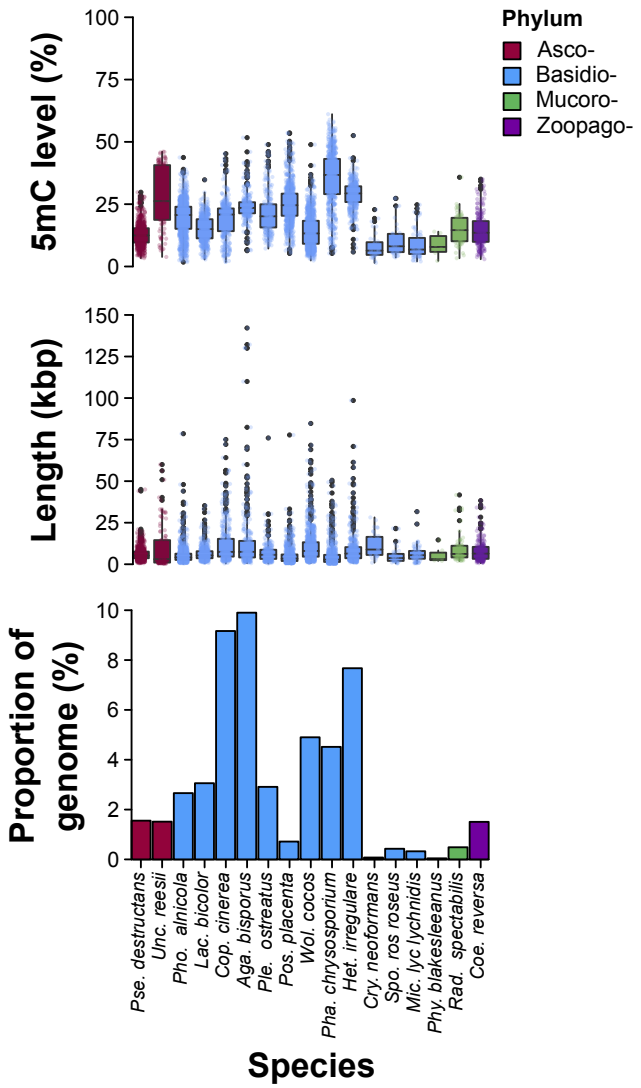


Supplementary Figure 12

Methylated Cytosine Clusters.

Methylation level across all contexts and length distributions of Methylated Cytosine Clusters (MCCs), and proportion of genome occupied by methylated clusters for 17 fungal species. Boxplot elements: center line, median; upper and lower "hinges", first and third quartiles (the 25th and 75th percentiles), respectively; whiskers, 1.5× interquartile range; large points, outliers.

Supplementary Fig. 12



478

479

480

481

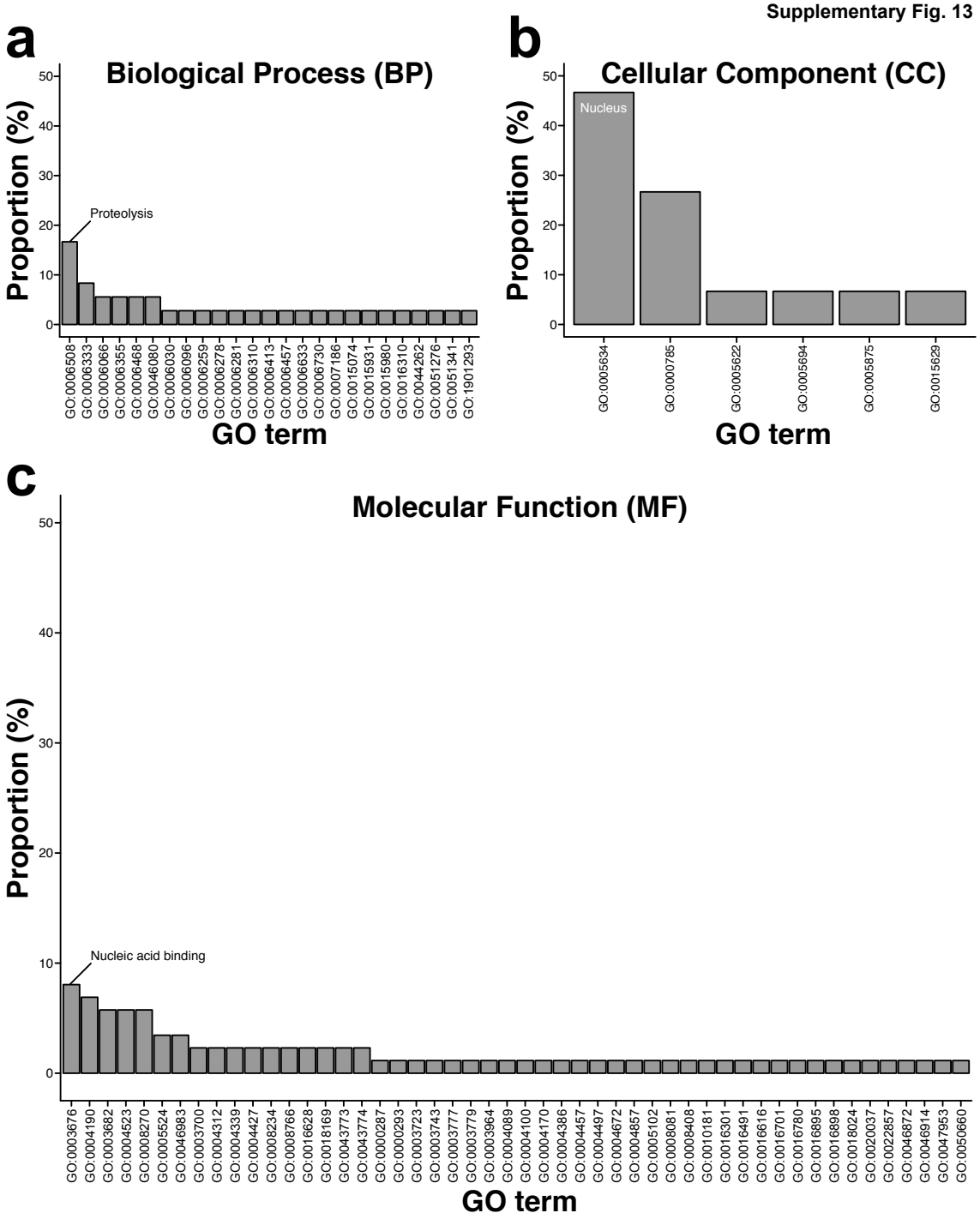
482

Supplementary Figure 13

Gene Ontology (GO) term enrichment of genes located in MCCs.

Proportion of statistically significant (*P* value <0.05) GO terms for genes located in MCCs.

Supplementary Fig. 13

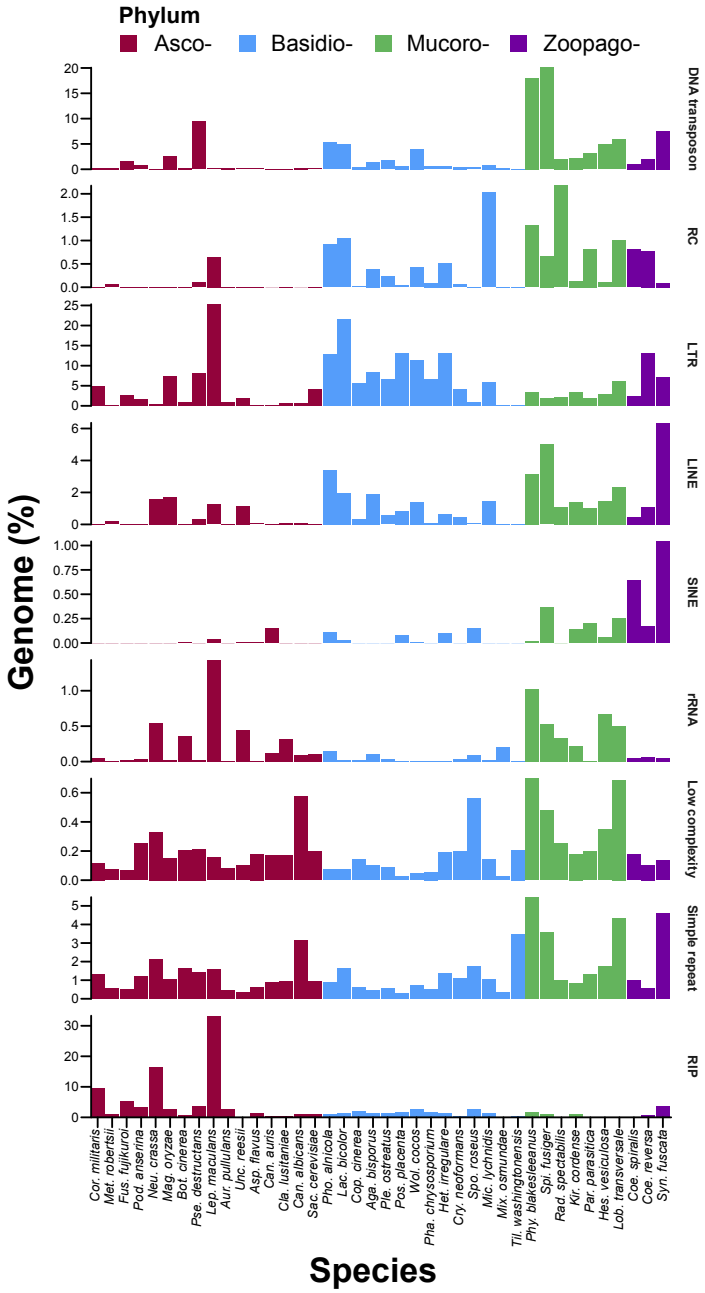


Supplementary Figure 14

Repeat and transposon content.

The genomic proportion of repeats and transposons for 40 fungal species. Repeats were detected and annotated using REPET v2.5. RIP loci were detected and annotated using the RIP product index, the RIP substrate index, and composite RIP index. See **ONLINE METHODS** for further details on repeat and transposon detection and annotation.

Supplementary Fig. 14



550

551

552

553

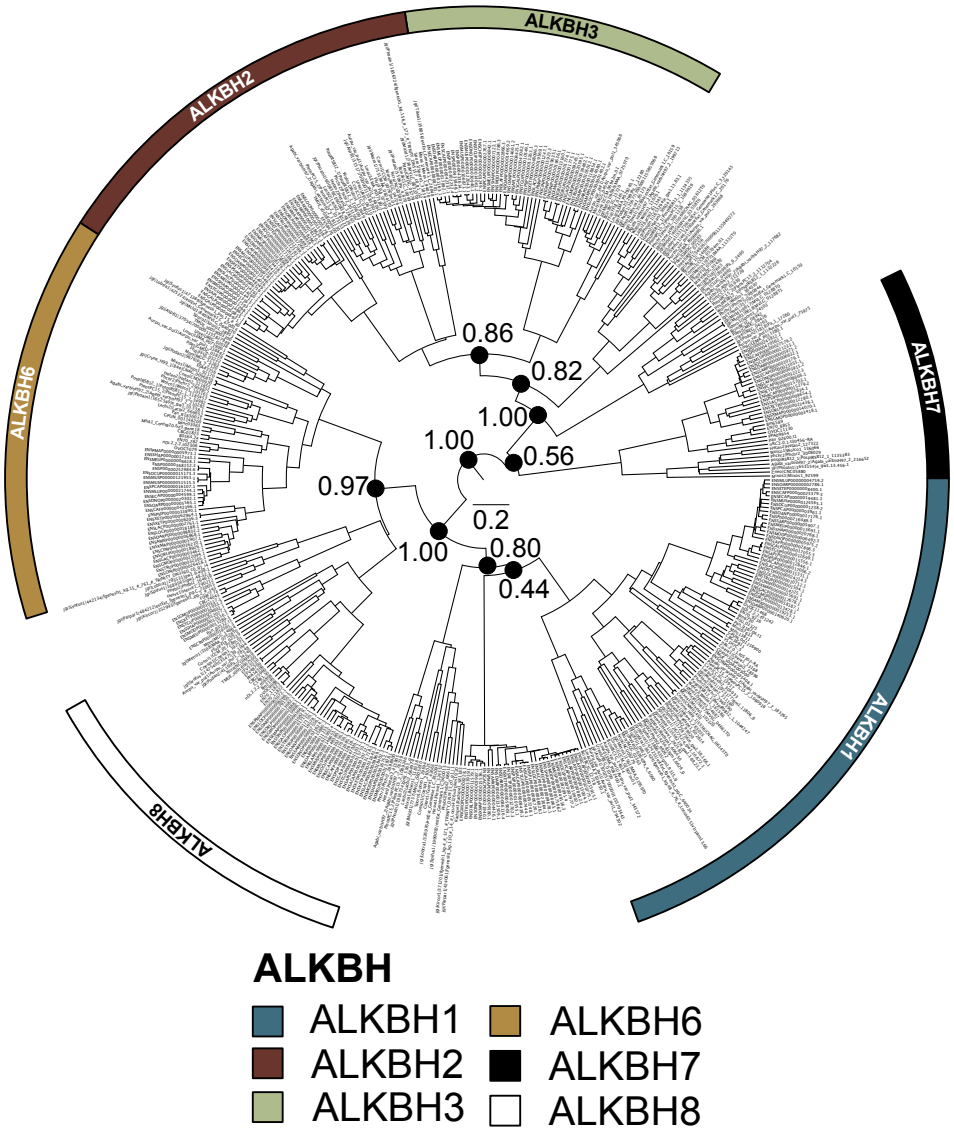
554

Supplementary Figure 15

ALKBH phylogeny.

Phylogenetic relationships and ALKBHs of Chordata, fungi, and Nematoda. Values at selected nodes indicate posterior probability. Branch lengths are in units of amino acid substitutions per amino acid site. Unlabeled clades represent taxa-specific clades of ALKBHs.

Supplementary Fig. 15



586

587

588

589

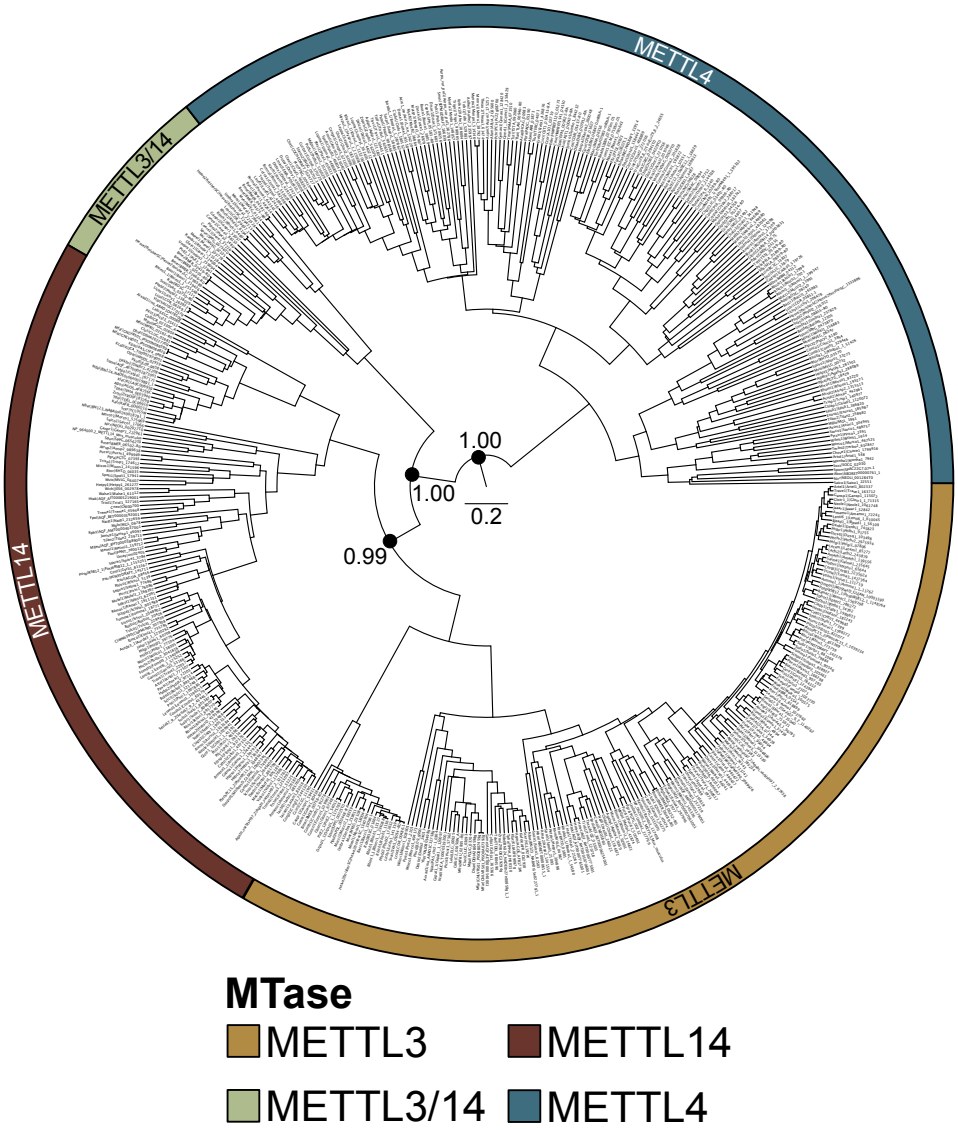
590

Supplementary Figure 16

METTL phylogeny.

Phylogenetic relationships and METLLs of fungi. Values at selected nodes indicate posterior probability. Branch lengths are in units of amino acid substitutions per amino acid site.

Supplementary Fig. 16



622
623
624
625
626

627 **Supplementary Table 1 | 5mC DNA MTase and the tRNA methyltransferase**
628 **DNMT2 annotations for 528 fungal species investigated in this study.**
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657

Supplementary Table 2 | WGBS and mapping statistics for 40 fungal species investigated in this study.

Supplementary Table 3 | Number of species per phylum for each observed 5mC MTase genotype.

Supplementary Table 4 | 5mC DNA MTase and the tRNA methyltransferase DNMT2 annotations for a subset of animalia, chlorophyta, fungi, and prokaryota. Protein models correspond to those used in Supplementary Fig. 2.

749 **Supplementary Table 5 | Number of CG-, CH, and CN-enriched genes**
750 **across fungal species investigated.**
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778

Supplementary Table 6 | ALKBH annotations for Chordata, fungi, and Nematoda investigated in this study. Protein models correspond to those used in Supplementary Fig. 15.

Supplementary Table 7 | Results from Pagel's test for correlated evolution.

Supplementary Table 8 | Results from Phylogenetic Generalized Least Squares (PGLS).

870 **Supplementary Table 9 | Annotated proteins from fungal species**
871 **containing the N-6 DNA Methylase domain (PF02384) as identified by**
872 **Interproscan v5.23-62.0 (45).**
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900

Supplementary Table 10 | METTL annotations for fungal species investigated in this study. Protein models correspond to those used in Supplementary Fig. 16.

930 **Supplementary Table 11 | Annotated proteins for fungal species containing**
931 **the domain the methyltransferase small domain (i.e., N6AMT1 proteins) as**
932 **identified by Interproscan v5.23-62.0 (45).**