Current Biology, Volume *26*

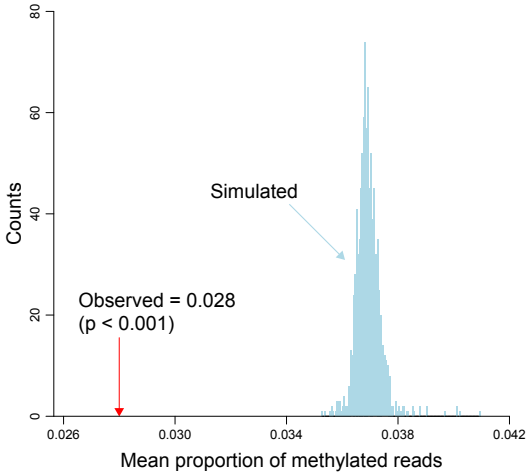
Supplemental Information

Robust DNA Methylation in the Clonal Raider Ant Brain

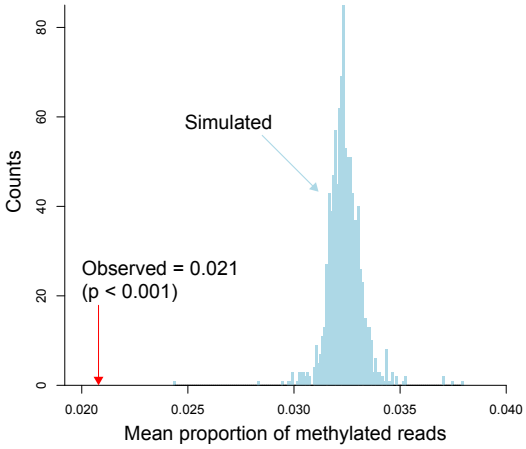
Romain Libbrecht, Peter Robert Oxley, Laurent Keller, and Daniel Jan Christoph Kronauer

Figure S1 – Related to Figure 2

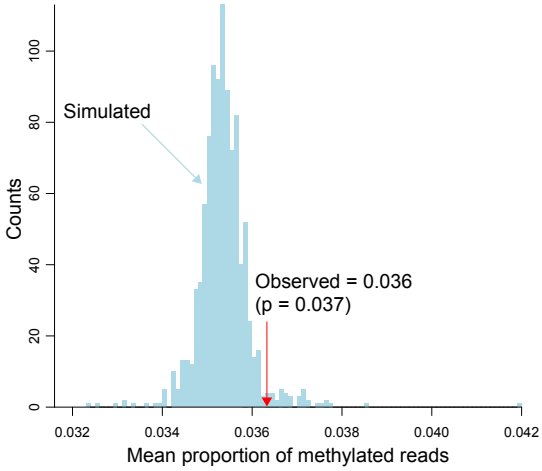
A/ Skipped exons



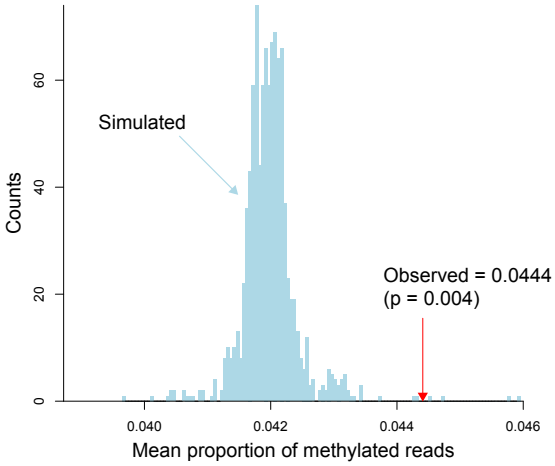
B/ Mutually exclusive exons



C/ Retained introns



D/ Alternative 3' splice sites



E/ Alternative 5' splice sites

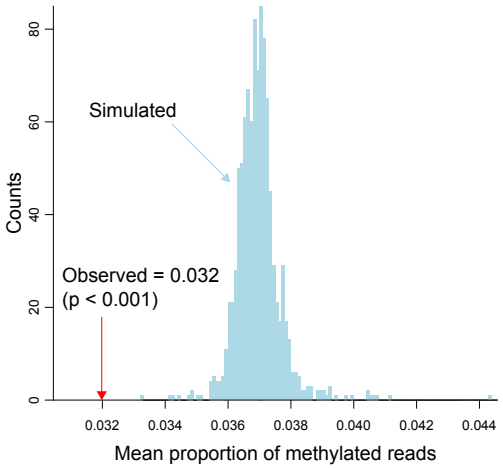


Figure S1 – Alternative splicing is associated with altered levels of DNA methylation.

For each of five types of alternative splicing (skipped exons, mutually exclusive exons, retained introns, alternative 3' splice sites and alternative 5' splice sites), we generated a list of all alternatively spliced exons (Supplemental Experimental Procedures). For each exon, we randomly selected another exon in the genome that had the same position in a gene with similar expression (Supplemental Experimental Procedures). We repeated this process to generate 1,000 random lists of exons for each type of alternative splicing. Then we compared the mean proportion of methylated reads per exon calculated from the empirical list of exons to the random distribution generated from the random lists of exons. The analyses revealed that skipped exons (A), mutually exclusive exons (B) and alternative 5' splice sites (E) were associated with lower levels of DNA methylation than expected, while retained introns (C) and alternative 3' splice sites (D) were associated with higher levels of DNA methylation than expected.

Figure S2 – Related to Figure 1

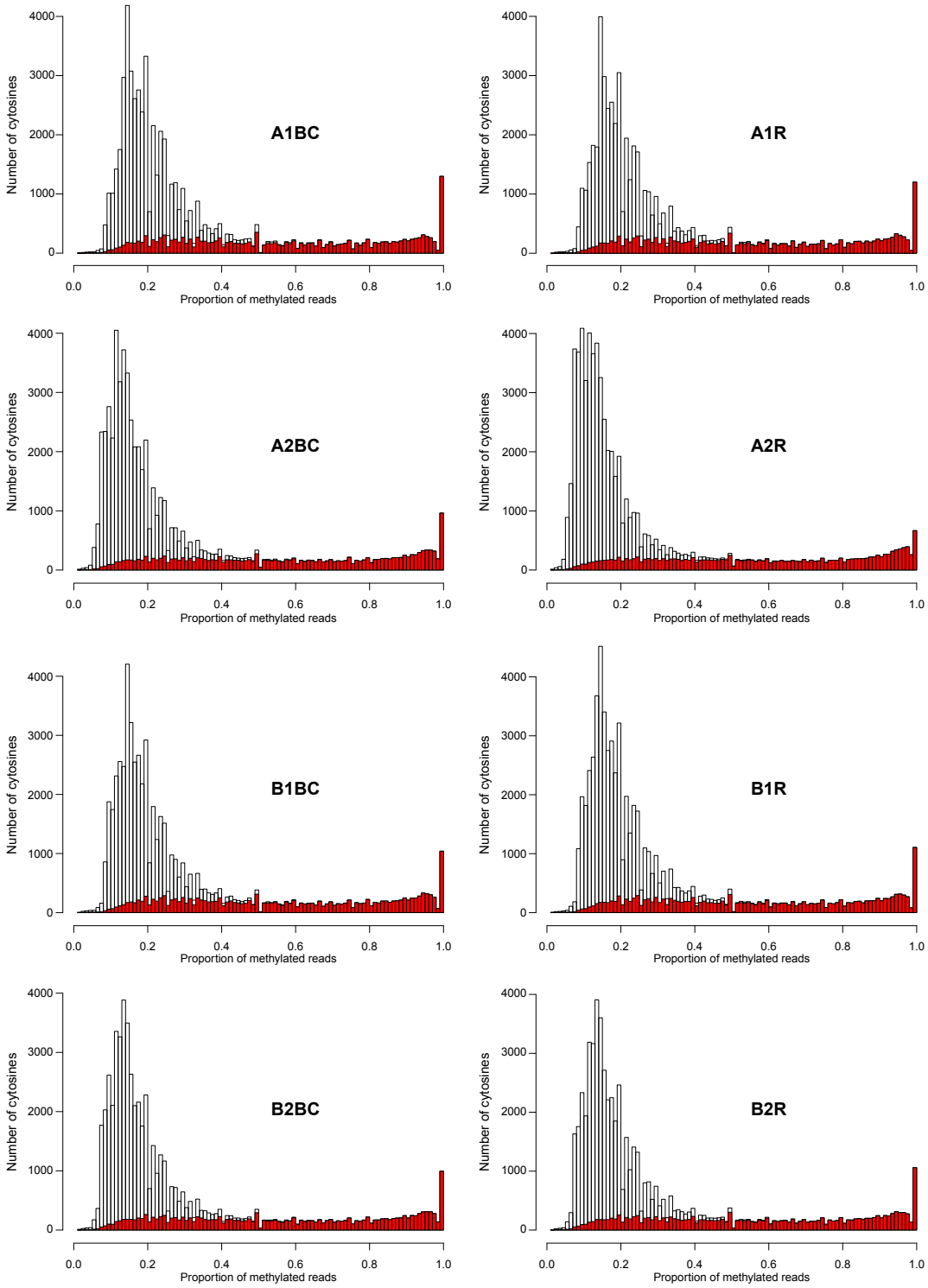


Figure S2 – Cytosines that were robustly methylated across samples also showed robust methylation within samples.

This graph shows for each sample the distribution of the proportion of methylated reads for the methylated cytosines. Cytosines that were methylated in all eight samples (in red) had a higher proportion of methylated reads compared to cytosines that were only methylated in a subset of samples (in white). Almost all cytosines with more than 60% methylated reads were methylated in all eight samples. Robust methylation across samples is thus associated with robust methylation within samples. For each graph, the first two symbols (“A1”, “A2”, “B1”, “B2”) indicate the source colony, “BC” stands for brood care phase, and “R” for reproductive phase. White and red bars are stacked.

Figure S3 – Related to Figure 3

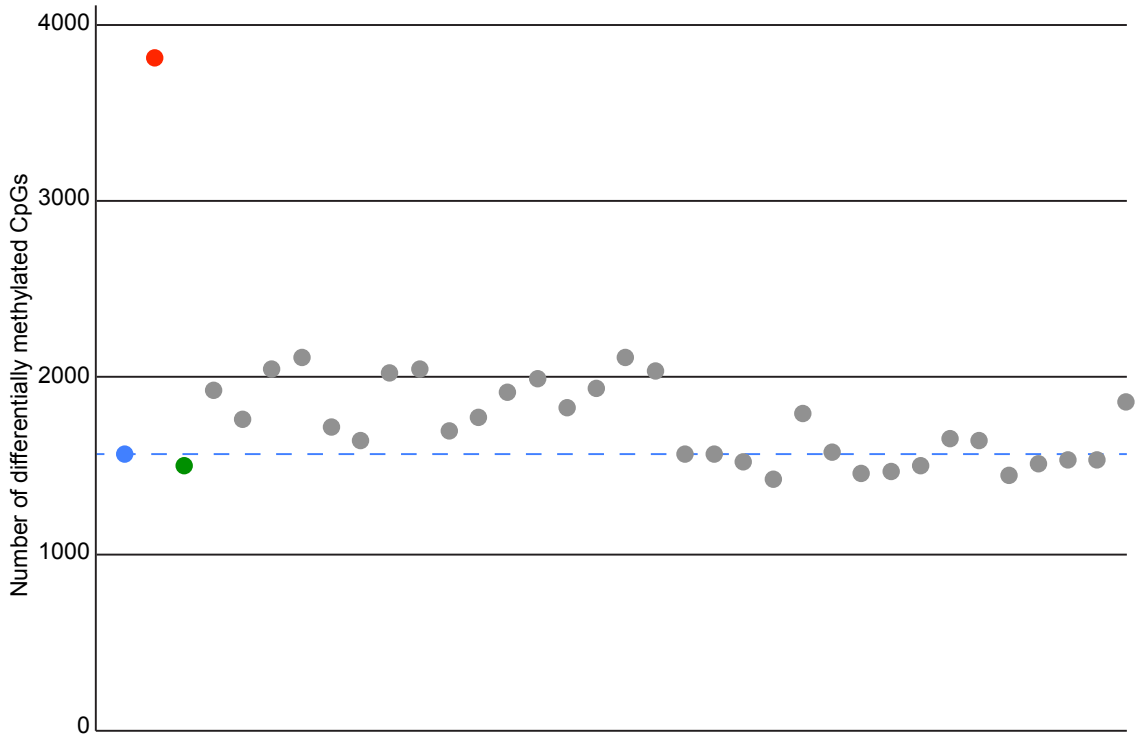


Figure S3 – The comparison between the reproductive phase and the brood care phase did not return more differentially methylated CpGs than expected by chance

This graph shows the number of differentially methylated CpGs obtained when comparing a set of four samples (e.g., the four samples collected in the reproductive phase) to another set of four samples (e.g., the four samples collected in the brood care phase). A given CpG was considered differentially methylated if it was methylated in all the samples of one set but unmethylated in all the samples of the other set. Each dot corresponds to one comparison of two sets of samples, and there are as many dots as there are possible combinations of four and four samples. The blue dot is the comparison between the reproductive phase and the brood care phase. The red dot is the comparison between the first and the second batch of sequencing. The green dot is the comparison between the clonal lineage A and the clonal lineage B. The grey dots are the thirty-two random comparisons that do not have any biological basis.

Neither the comparison between the reproductive phase and the brood care phase (blue dot) nor the comparison between clonal lineage A and clonal lineage B (green dot) had more differentially methylated CpGs than the random comparisons (grey dots), i.e. what could be expected by chance.

The finding of more differentially methylated CpGs when comparing the two batches of sequencing (red dot) compared to random comparisons (grey dots) shows that there is a batch effect in our bisulfite sequencing data, and suggests that differential methylation could in fact have been detected by this analysis if it had existed at appreciable levels.

We performed the same analysis to look for differential methylation in non-CpGs. We found 1,053 non-CpGs that were differentially methylated between the two phases, while random comparisons returned an average of 1,196 differentially methylated non-CpGs (median = 1,189; ranging from 942 to 1,482). Thus, as in CpGs, we did not detect significant differential methylation in non-CpGs between the two phases of the *C. biroi* colony cycle.

Table S1 – Related to Figure 1

	Species name	ms-AFLP methylation rate
Ants	***Cerapachys biroi***	**0.76**
	Aphaenogaster albisetosa	0.04
	Camponotus festinatus	0.07
	Messor pergandei	0.06
	Pogonomyrmex barbatus	0.38
	Pheidole obtusospinosa	0.09
Bees	*Apis mellifera*	0.05
	Melipona bicolor	0.08
	Trigona spinipes	0.01
Wasps	*Liostenogaster flavolineata*	0.03
	Metapolybia cingulata	0.06
	Polistes dominulus	0.19
	Polybia sericea	0.11
	Vespula pensylvanica	0.11
Termite	*Coptotermes lacteus*	0.07

Table S1 – Methylation sensitive AFLP (ms-AFLP) analysis is consistent with the genome of *C. biroi* being more methylated than the genomes of other previously studied social insects [S2-S4]. The methods, enzymes and primers used to perform ms-AFLP in *C. biroi* were the same as in [S3]. The ms-AFLP methylation rate is the estimated percentage of methylated CCGG sites (see [S3] for details). We performed ms-AFLP on eight DNA samples each extracted from a pool of eight heads. On average the ms-AFLP methylation rate was 0.76 ± 0.03 (mean ± sd) in *C. biroi*.

Table S2 – Related to Figure 2

		Mean proportion of methylated cytosines	SE
	Genome	**0.0206**	**0.0010**
LTR Retrotransposons	BEL ***	0.0119	0.0003
	Copia *	0.0158	0.0002
	Gypsy **	0.0141	0.0002
	Others ***	0.0130	0.0005
Non-LTR Retrotransposons	CR1 *	0.0155	0.0006
	Jockey	0.0169	0.0008
	R1 *	0.0142	0.0004
	SINE	0.0201	0.0054
	Others **	0.0147	0.0004
DNA transposons	hAT ***	0.0099	0.0003
	Helitron *	0.0154	0.0006
	Marine Tc1 *	0.0154	0.0003
	Sola	0.0185	0.0005
	Transib ***	0.0126	0.0008
	Others ***	0.0117	0.0002

Table S2 – Transposable elements were hypomethylated compared to the genome baseline.

Eighty percent (12 out of 15) of the transposable element classes included in the analysis had a significantly lower proportion of methylated cytosines compared to the whole genome (*** for $P < 0.001$, ** for $P < 0.01$, * for $P < 0.05$; Supplemental Experimental Procedures).

Table S3 – Related to Figure 1

GO term	Ontology	Description	q value
GO:0003723	MF	RNA binding	< 0.0001
GO:0005515	MF	protein binding	< 0.0001
GO:0003735	MF	structural constituent of ribosome	< 0.0001
GO:0005488	MF	binding	< 0.0001
GO:0005524	MF	ATP binding	< 0.0001
GO:0000166	MF	nucleotide binding	< 0.0001
GO:0004812	MF	aminoacyl-tRNA ligase activity	< 0.0001
GO:0004672	MF	protein kinase activity	< 0.0001
GO:0001104	MF	RNA polymerase II transcription cofactor activity	< 0.001
GO:0008026	MF	ATP-dependent helicase activity	< 0.0001
GO:0005737	CC	cytoplasm	< 0.0001
GO:0005634	CC	nucleus	< 0.0001
GO:0005622	CC	intracellular	< 0.0001
GO:0005840	CC	ribosome	< 0.0001
GO:0016592	CC	mediator complex	< 0.0001
GO:0006886	BP	intracellular protein transport	< 0.0001
GO:0006396	BP	RNA processing	< 0.0001
GO:0006397	BP	mRNA processing	< 0.0001
GO:0008033	BP	tRNA processing	< 0.0001
GO:0006281	BP	DNA repair	< 0.001
GO:0016192	BP	vesicle-mediated transport	< 0.0001
GO:0006511	BP	ubiquitin-dependent protein catabolic process	< 0.0001
GO:0006412	BP	translation	< 0.0001
GO:0006418	BP	tRNA aminoacylation for protein translation	< 0.0001
GO:0006468	BP	protein phosphorylation	< 0.0001
GO:0006457	BP	protein folding	< 0.0001

Table S3 – List of Gene Ontology (GO) terms significantly enriched in genes with robust methylation (n = 6929) compared to genes without robust methylation (n = 3502).

Twenty-six GO terms were significantly enriched in genes with robust methylation. The three ontology categories are molecular function (MF), cellular component (CC) and biological processes (BP). The q values were obtained by correcting the p values for multiple testing [S1]. To determine whether such GO term enrichment could be expected by chance, we randomly generated 10,000 lists of 6929 genes. No GO term was significantly enriched in any of those random lists (all q values > 0.05).

Supplemental Experimental Procedures

DNA extraction

DNA was extracted from pools of 20 brains using the standard protocol of the QIAamp® DNA Micro Kit (Qiagen) with a final elution in 40 μl of buffer AE.

RNA extraction

RNA was extracted using a modified Trizol/phenol chloroform protocol. RNA was extracted using Trizol (Invitrogen) followed by RNeasy (Qiagen) purification with DNAse I (Qiagen) on-column digestion.

Library preparation and sequencing

Library preparation, whole-genome bisulfite sequencing and RNA sequencing were performed at the Epigenomics Core at Weill Cornell Medical College as follows:

Whole-Genome Bisulfite Sequencing (WGBS)

Briefly, 100 ng of DNA were bisulfite converted using the EZ DNA Methylation-Gold Kit (cat # D5005, Zymo Research Corporation, 17062 Murphy Ave. Irvine, CA 92614). The single stranded DNA obtained was processed for library construction using the EpiGenome Methyl-Seq kit EGMK81324 as per the manufacturer's protocols (Illumina Madison, 5602 Research Park Blvd., Suite 200 Madison, WI 53719). 5' tagged random hexamers were annealed to single-stranded DNA and subsequently 3' tagged with a terminal-tagging oligo. The di-tagged DNA was enriched using 10 cycles of PCR, with PCR primers compatible with Illumina sequencing. Each library was made with a unique index sequence and each batch of four libraries was pooled together. The pools were clustered at 7 pM on a paired-end read flow cell and sequenced for 100 cycles on an Illumina HiSeq 2500.

RNA-Seq

RNA-Seq library preparations were done using established Illumina methods for mRNA-Seq (Part #RS-122-2001). Briefly, poly A+ RNA was purified from 200 ng of total RNA with oligo-dT beads. Purified mRNA was fragmented with divalent cations at elevated temperature to ~200bp. First strand cDNA synthesis was performed with random hexamer priming and reverse transcriptase. Second strand cDNA synthesis was performed using RNAseH and DNA PolI. Following dscDNA synthesis, the double stranded products were end repaired, followed by addition of a single 'A' base and ligation to the Illumina TruSeq adaptors. The resulting product was amplified with 15 cycles of PCR. Each library was made with a unique index sequence and each batch of four libraries was pooled together. The pools were clustered at 6.5pM on a paired-end read flow cell and sequenced for 100 cycles on an Illumina HiSeq 2500.

Data processing

Primary processing of sequencing images was done using Illumina's Real Time Analysis software (RTA). CASAVA 1.8.2 software was then used to demultiplex samples and generate raw reads and corresponding quality scores. The WGBS raw data was quality filtered, adapter trimmed, aligned to the *Cerapachys biroi* genome (Official Gene Set version 2.0.1), and methylation calls were generated using the in-house bisulfite sequencing analysis pipeline in the Epigenomics Core at Weill Cornell Medical College [S5]. RNA-Seq reads passing Illumina's purity filter were adapter trimmed and aligned to the *Cerapachys biroi* genome using STAR aligner [S6]. Aligned read counts for each gene were calculated using HTSeq, and a variance-stabilized transformation applied using DESeq2.

Alternative splicing

The *Cerapachys biroi* genome was annotated using NCBI's eukaryotic genome annotation pipeline, identifying 5,112 genes with more than one isoform. These isoforms were searched for skipped exons, mutually exclusive exons, retained introns, and alternative 5' and 3' splice sites. Alternative splicing events were classified using gff_make_annotation.py from the rnaseqlib package (http://yarden.github.com/rnaseqlib), using the "commonshortest" flanking rule. For each alternatively spliced exon, we identified a list of equivalent exons in the *Cerapachys biroi* genome. Using the variance-stabilized transformed expression level from the brain RNA-Seq data, we identified genes expressed between 0.8 and 1.2 times the expression level of the alternatively spliced genes. We then removed those

genes that had alternative splicing in the same manner as the target exon. In the remaining genes, we selected the exon in the same position in the gene (first exon, second exon, etc.) as the target exon. For each of the five types of alternative splicing events, we generated 1,000 lists of exons drawn randomly from the list of equivalent exons, such that each list contained a single non-alternatively spliced equivalent exon for each of the alternatively spliced exons. For each type of alternative splicing, we calculated the mean proportion of methylated reads for the empirical list of exons and for the random lists of exons, and then compared the observed value to the expected distribution.

Transposable elements

Transposable elements in the *C. biroi* genome assembly were identified using RepeatModeler (http://www.repeatmasker.org) and the RepBase [S7] database of repeat elements. All elements that had more than 400 fragments identified in the genome were grouped independently, while all remaining elements were categorized as "others" in the LTR retrotransposon, Non-LTR retrotransposon or DNA transposon classes. SINEs were also grouped independently to enable comparisons with the literature. The RepeatModeler GFF output was used to define the positions of all transposable elements for methylation analysis.

For each transposable element class in each sample, we calculated the proportion of methylated cytosines with a minimum coverage of 10x in the focal sample (Table S2). We built a linear model to compare the mean proportion of methylated cytosines in the transposable element classes to the genome, using the eight samples as replicates (Table S2). Additionally, we used a Wilcoxon rank sum test to compare the mean proportion of methylated cytosines across all classes of transposable elements to the genome (Main text).

Testing the statistical method used in previous studies to detect differentially methylated genomic regions without biological replicates

Most studies that have reported caste-specific differential methylation in social insects used the same statistical method to detect such differences [S8-S11]. We applied this method to our data. To make our analysis comparable to these previous studies, instead of performing one analysis with four replicates, we performed four analyses with one replicate each. Each analysis compared the reproductive phase and the brood care phase for one source colony. For each exon with more than three CpGs and less than 100 CpGs, we built a generalized linear model (binomial family) that explained the proportion of methylated reads by the caste and the position. We used all the CpGs (minimum coverage = 3x) in the focal exon as replicates in the model, as did previous studies that used this method. We analyzed each exon separately and then corrected the p values for multiple testing [S1].

As presented in the main text, the analyses detected several hundred differentially methylated exons ($p < 0.05$) for the four source colonies, but none of those exons were differentially methylated in all four comparisons, revealing that the lists of exons reported by the analyses were actually random or colony-specific lists of exons. The main problem of this statistical method is that it does not involve biological replicates, but uses the different CpGs in a given genomic region as replicates in the model. This makes this statistical method prone to false positives arising from individual variation or experimental noise.

Experimental procedures for the study of DNA methylation in the social insect literature

In the main text we focus on previous studies that performed whole-genome bisulfite sequencing (WGBS) to compare DNA methylation between the queen and worker castes in social insect species. WGBS is the only method providing the genome-wide single nucleotide resolution of DNA methylation comparable to the data we collected in *C. biroi*. In this part of the supplement, we provide a wider review of the experimental procedures used to study DNA methylation in social insects.

Using the whole body to perform whole-genome bisulfite sequencing is problematic

As detailed in the main text and above, most social insect studies that have compared DNA methylation between castes using WGBS used the same statistical method, which does not require biological replicates but is prone to producing false positives stemming from sample-specific DNA methylation [S8-S11]. Additionally, two of those studies used the whole body to extract DNA [S9, S10], which is problematic when comparing social insect castes that differ in morphology and allometry, such as queens and workers [S8]. If tissues that show between-caste differences in their relative proportion to the whole body (e.g., the abdomen in queens and workers) have specific patterns of DNA methylation, this would result in apparent

differences in DNA methylation between castes. In fact, comparing honeybee queen and worker brains [S8, S12] may also be problematic, because a queen brain is structurally different from a worker brain, thus not directly comparable. In our study, the use of morphologically and genetically identical individuals in the reproductive phase and in the brood care phase allowed us to circumvent this problem.

Whole-genome bisulfite sequencing to compare nurses and foragers

To our knowledge there is only one WGBS study of DNA methylation that uses a replicated design to test whether methylation differs between honeybee queens and workers, and between nurses and foragers [S12]. No significant differences in DNA methylation were detected between queen and worker brains, in contrast with a previous study that did report differential methylation [S8]. Those two studies differed in the age and maturity of the individuals used for WGBS, which might help account for some of the discrepancies. Arguably more importantly, however, the two studies also differed with respect to the use of biological replication. Interestingly, Herb et al (2012) reported an association between differences in DNA methylation and behavioral differences within the worker caste, as differentially methylated regions were detected when comparing honeybee nurses and foragers [S12]. Removing nurses prompted some of the foragers to revert to nursing, and differentially methylated regions were also detected when comparing such reverted nurses to the workers that remained foragers. Because there was a significant overlap between the two comparisons, Herb et al (2012) concluded that their data provided evidence for dynamic DNA methylation that can switch back and forth depending on behavior. However, such dynamic patterns of DNA methylation are not required to explain their data, as stable individual differences in DNA methylation may have affected the response threshold to forage, and thus the likelihood of switching back from foraging to nursing tasks when all the nurses were removed. In addition, the workers in each replicate were produced by at least five different queens, resulting in individual genetic differences that could have independently affected DNA methylation [S13, S14] and the transitions between nursing and foraging [S15, S16].

Candidate gene approach

A recent study using a candidate gene approach revealed that larval DNA methylation in the gene *Egfr* is associated with worker size variation in the ant *Camponotus floridanus* [S17]. Another study reported that downregulating *Dnmt3* (coding for the enzyme responsible for *de novo* DNA methylation) in honeybee larvae triggered the development of queens rather than workers [S18]. This finding still awaits confirmation, as the decrease of DNA methylation after *Dnmt3* knockdown was not investigated genome-wide but in only ten CpGs spanning three exons of a single gene [S18].

Methylation sensitive AFLP

Methylation sensitive AFLP provides an estimation of the proportion of methylated sites at the level of the genome. It has been used to assess how common DNA methylation is in a variety of social insect species [S2], but only two studies used it to compare DNA methylation between castes: one in ants [S4] and one in termites [S3]. In *Pogonomyrmex* ants, the analysis revealed a higher proportion of methylated loci in virgin queens than in workers [S4]. In *Coptotermes* termites, the proportion of methylated loci did not differ significantly between sexes or between castes [S3].

Supplemental References

S1. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 289-300.

S2. Kronforst, M., Gilley, D., Strassmann, J., and Queller, D. (2008). DNA methylation is widespread across social Hymenoptera. Curr. Biol. *18*, 287-288.

S3. Lo, N., Li, B., and Ujvari, B. (2012). DNA methylation in the termite *Coptotermes lacteus*. Insectes soc. *59*, 257-261.

S4. Smith, C.R., Mutti, N.S., Jasper, W.C., Naidu, A., Smith, C.D., and Gadau, J. (2012). Patterns of DNA methylation in development, division of labor and hybridization in an ant with genetic caste determination. PLoS ONE *7*, e42433.

S5. Garrett-Bakelman, F.E., Sheridan, C.K., Kacmarczyk, T.J., Ishii, J., Betel, D., Alonso, A., Mason, C.E., Figueroa, M.E., and Melnick, A.M. (2015). Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. J. Vis. Exp. *96*, 52246.

S6. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinfo. *29*, 15-21.

S7. Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. *9*, 418-420.

S8. Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol. *8*, e1000506.

S9. Glastad, K.M., Hunt, B.G., Soojin, V.Y., and Goodisman, M.A. (2014). Epigenetic inheritance and genome regulation: is DNA methylation linked to ploidy in haplodiploid insects? Proc. R. Soc. B *281*, 20140411.

S10. Bonasio, R., Li, Q., Lian, J., Mutti, N.S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., Ding, Y., et al. (2012). Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Curr. Biol. *22*, 1755-1764.

S11. Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S.E., Robinson, G.E., and Maleszka, R. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc. Natl. Acad. Sci. USA *109*, 4968-4973.

S12. Herb, B.R., Wolschin, F., Hansen, K.D., Aryee, M.J., Langmead, B., Irizarry, R., Amdam, G.V., and Feinberg, A.P. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. Nat. Neurosci. *15*, 1371-1373.

S13. Gutierrez-Arcelus, M., Lappalainen, T., Montgomery, S.B., Buil, A., Ongen, H., Yurovsky, A., Bryois, J., Giger, T., Romano, L., Planchon, A., et al. (2013). Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife *2*, e00523.

S14. Wedd, L., Kucharski, R., and Maleszka, R. (2015). Differentially methylated obligatory epialleles modulate context-dependent LAM gene expression in the honey bee *Apis mellifera*. Epigenetics. *27*, online first.

S15. Calderone, N.W., and Page, R.E. (1988). Genotypic variability in age polyethism and task specialization in the honey bee, *Apis mellifera* (Hymenoptera: Apidae). Behav. Ecol. Sociobiol. *22*, 17-25.

S16. Siegel, A.J., Fondrk, M.K., Amdam, G.V., and Page Jr, R.E. (2013). In-hive patterns of temporal polyethism in strains of honey bees (*Apis mellifera*) with distinct genetic backgrounds. Behav. Ecol. Sociobiol. *67*, 1623-1632.

S17. Alvarado, S., Rajakumar, R., Abouheif, E., and Szyf, M. (2015). Epigenetic variation in the *Egfr* gene generates quantitative variation in a complex trait in ants. Nat. Commun. *6*, 6513.

S18. Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. Science *319*, 1827.