Supporting Information

Foret et al. 10.1073/pnas.1202392109

SI Materials and Methods

Source of DNA. Larvae (96 ± 1 h old; 14 queens and 25 workers) were collected from Canberra hives in October 2008 and froze on dry ice. The average weight of a single larva was 230.6 mg for queens and 133.5 mg for workers. Total DNAs were extracted from the larval heads (≈1.6 mm of the frontal end) containing brain, optic and retinular ganglia, neurosecretory cells, glands (corpora allata, corpora cardiaca), suboesophageal ganglion, a small number of fat bodies, the maxillae, labium, and mandibles, segmented imaginal antennae developing in hypodermal pockets, the openings of silk glands ducts at the tip of the labium-hypopharynx, trachea, and cuticle. The rest of the larval body is predominantly occupied by a large digestive system filled with processed food and bacteria (larvae do not defecate), a tracheal network, and reproductive parts that at this stage of development are already large in queens and rudimentary in workers.

DNA was purified with the MasterPure DNA purification kit (Epicentre Biotechnologies, catalog no. MCD85201). Following the manufacturer's protocol, the final prep was additionally extracted with phenol–chloroform and precipitated with 2 vol of absolute ethanol and the pellet dissolved in TE buffer (10 mM Tris_HCl, 1 mM EDTA pH 8.2). The DNAs' quality was evaluated by spectral analysis. The final yield was 2 μg of DNA per worker head and 2.3 μg of DNA per queen head.

Bisulfite Sequencing (BS-Seq) of Larval DNAs. Genomic DNAs extracted from honey bee larvae were used to generate BS-Seq libraries using a previously published protocol (1). Briefly, DNAs were sheared using a Bioruptor, modified by a pair of universal adapters, treated with sodium bisulfite, and then PCR amplified. The PCR products were then digested by DpnI to remove the universal adapters. The resulting DNA was used to generate single-end Illumina sequencing libraries following Illumina's library generation protocols. The libraries were sequenced on a high-throughput sequencing machine, Illumina Genome Analyzer IIx (GAIIx), following the manufacturer's instructions. The sequences have been submitted to the National Center for Biotechnology Information Sequence Read Archive database under accession no. SRA047112.1.

Mapping and Methylation Assessment. Sequencing data were processed first by the Illumina Data Analysis Pipeline, and then bisulfite-converted sequencing reads were aligned to the honey bee genome (official release 2.0) using BS Seeker software (2). Reads containing consecutive CHN (H = A or T or C) nucleotides are the product of incomplete bisulfite conversion (1) and were discarded. To increase the accuracy of methylation calls, only those cytosines fulfilling neighborhood quality standards were counted (bases of quality 20 or more, flanked by at least three perfectly matching bases of quality 15 or more). Only the reads mapping uniquely were used. These postmapping steps were carried out using the tools included in libngs (http://github.com/sylvainforet/libngs).

The methylation status of each CpG dinucleotide and differential methylation were assessed using our previously published methods (3). In brief, the methylation status of each cytosine base was modeled by a binomial distribution, with the number of trial equal to the number of mapping reads and the probability equal to the conversion rate. A base was called methylated if the number of reads supporting a methylated status departed from this null model significantly at the 5% level after correcting for multiple testing (4). Differentially methylated genes were identified using generalized linear models of the binomial family; the response vector *CpGmeth* (number of methylated and nonmethylated reads for each CpG in a gene) was modeled as a function of two discrete categorical variables, the caste and the CpG postion: CpGmeth = caste × CpGi. *P* values were corrected for multiple testing using the Benjamini and Hochberg method (4). These tests were carried out using the R statistical environment (http://www.r-project.org).

Other Bioinformatics Analyses. Gene ontology (GO) annotations of the honey bee predicted genes were produced using blast2go software (5). Pathway annotations were retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) website (6). The annotations of the pathways presented in this article were manually refined using homology to the fly and human genes. Enrichment in GO terms and KEGG pathways was assessed using the GOstat module of the Bioconductor platform (www.bioconductor.org).

Alternative Splicing Analysis. We used SibSim4 (http://sibsim4.sourceforge.net/) to produce spliced alignments of all of the honey bee transcriptome sequences available from GenBank, including RNA-seq data [accession nos.: SRX025526 – whole body (*i*), SRX025527 – whole body (*ii*), SRX025528 – abdomen, SRX025529 – mixed antennae, SRX025530 – embryo, SRX025531 – brain and ovaries pooled, SRX025532 – testes, SRX025533 – larvae, and SRX016658 – queen ovaries], with *Apis mellifera* scaffolds assembly v.2.0. All subsequent analyses were performed using purpose-written scripts. For manual analysis alignments were visualized in GBrowse v.1.64 run on a local Linux server after converting SibSim4 output to GBrowse track files. For alternative splicing analysis false positives caused by sequencing errors or spurious alignments were avoided by only using canonical splice sites, falling into annotated gene models, and supported by at least five reads.

Molecular Methods. All molecular biology experiments were performed as previously described (3, 7, 8). A combined TRIzol/Qiagen RNeasy protocol was used for RNA purification followed by first strand synthesis (Superscript III protocol). Starting material was 2–4 pooled larval heads, 1–5 pooled brains, and 1/4–1/2 ovary, per biological replicate, respectively. Total RNA input was 2 μg for larval heads, 0.4–1.0 μg for brains, 5 μg for ovaries.

PCR Amplification. ***Cycling conditions for StepOnePlus.*** Initial denaturation at 95 °C for 5 min, followed by 40 cycles of 95 °C for 30 s, 60 °C for 30 s, 72 °C for 45 s, and a standard melt profile.
Quantitation. Quantitation data were analyzed with StepOne v.2.1 software and Excel spreadsheets using DDCt comparative quantitation method assuming 100% amplification efficiencies and ugt or Alk 2 as reference amplicons.
RNA-seq reads mapping. RNA-seq reads [accession nos.: SRX025526 – whole body (*i*), SRX025527 – whole body (*ii*), SRX025528 – abdomen, SRX025529 – mixed antennae, SRX025530 – embryo, SRX025531 – brain and ovaries pooled, SRX025532 – testes, SRX025533 – larvae, and SRX016658 – queen ovaries] were mapped to the *Apis mellifera* scaffolds assembly v.2.0 using sim4 (9) and visualized in GBrowse v.1.64 run on a local Linux server after converting data to tracks with purpose-written scripts.
Primers.

AlkA-F1 TGTATCAATGTGCAGTCAGAAAAC
AlkB-F1 GAAACGCGAGTCGCGAAGAT
Alk-R1 TGCCATTCCTCCTGTCTACA
Alk-F2 GGCAGCATGGAAATCAAGAAGTTC
Alk-R2 AAACGCATAGCCGTAGAAACTGCA

Anti-Alk-F1 TAACCAATTCTCTCGACGATAG
Anti-AlkA-R1 AACAACTTCTGTTTATATACTGGT
Anti-AlkB-R1 AAATAAAATTTGTTTGAAACTTAC

UGT-F1 CGTTGATGCTGATCAGGTTG forward exon 19/20 junction
UGT-R1 CGTCGAAATCGCTTCAAGTC reverse exon 20.

1. Cokus SJ, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452:215–219.
2. Chen PY, Cokus SJ, Pellegrini M (2010) BS Seeker: Precise mapping for bisulfite sequencing. *BMC Bioinformatics* 11:203.
3. Lyko F, et al. (2010) The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. *PLoS Biol* 8:e1000506.
4. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300.
5. Götz S, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420–3435.
6. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38(Database issue):D355–D360.
7. Foret S, et al. (2009) Epigenetic regulation of the honey bee transcriptome: Unravelling the nature of methylated genes. *BMC Genomics* 10:472.
8. Lockett GA, Kucharski R, Maleszka R (2012) DNA methylation changes elicited by social stimuli in the brains of worker honey bees. *Genes Brain Behav* 11:235–242.
9. Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res* 8: 967–974.

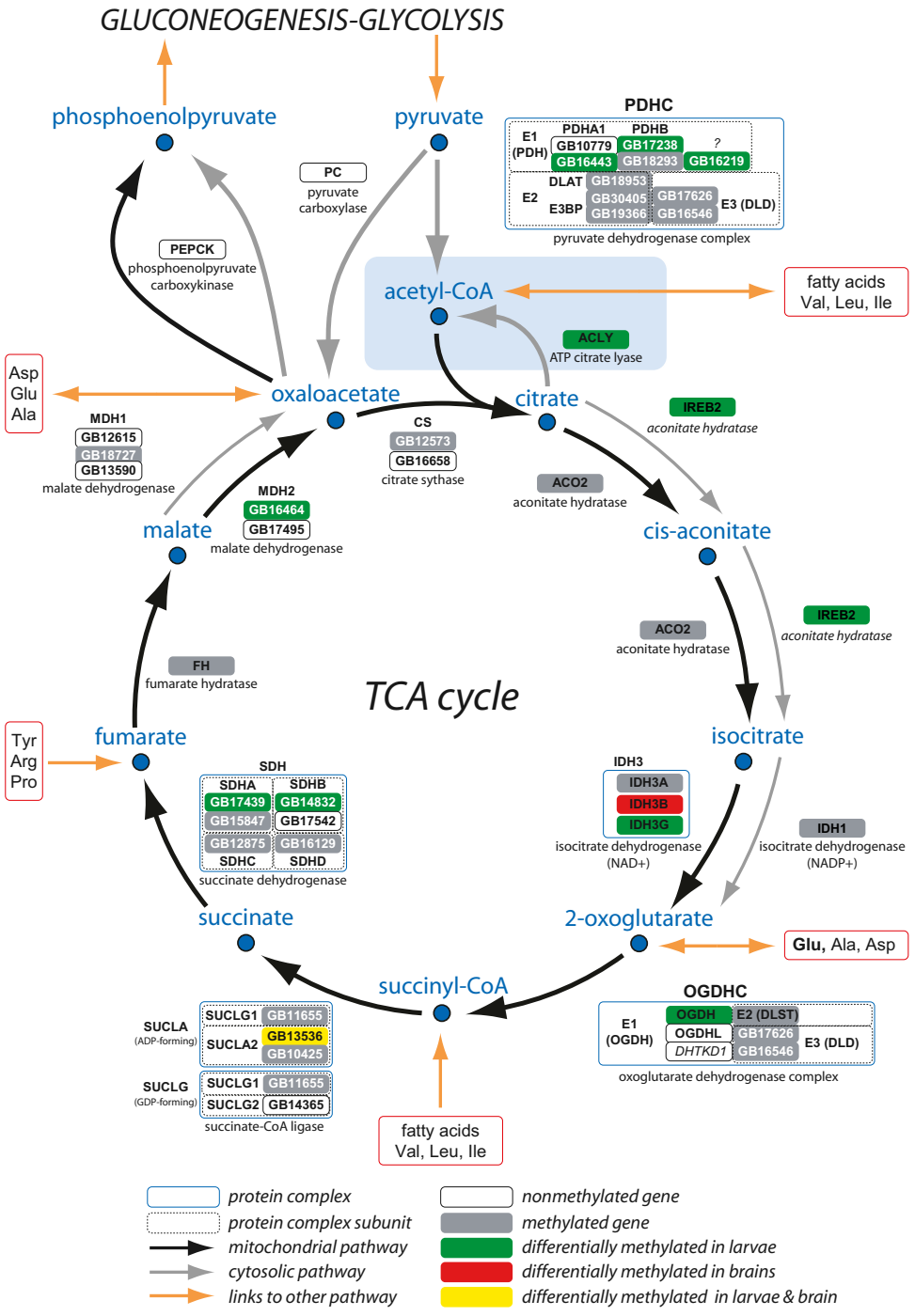

Fig. S1. Annotation of the tricarboxylic acid (TCA) cycle in *Apis* showing methylated and differentially methylated genes. In those cases in which more than one paralog has been found only one is referred to by the consensus protein name, with the others designated by the GB numbers from the official *Apis* gene list (www.beebase.org). The orange arrows indicate several exit/entry points by which the TCA cycle is connected to a larger metabolic and regulatory network at the whole-cell level. Among major metabolic pathways converging on the TCA cycle are both catabolic and anabolic reactions involved in carbohydrate and fat metabolism. More information on each gene is in Table S3.

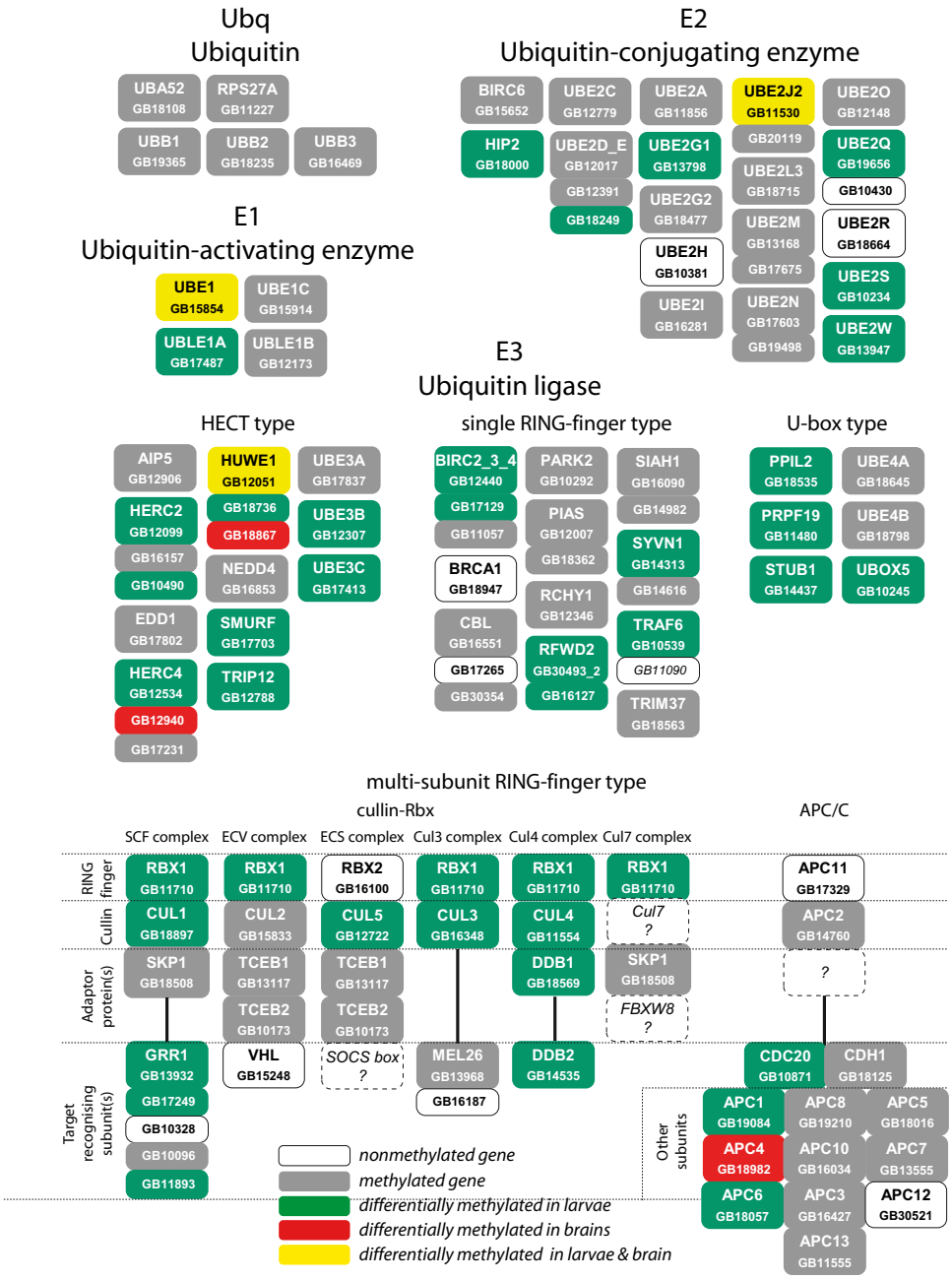

Fig. S2. Annotation of the ubiquitin/proteasome pathway in *Apis*, showing methylated and differentially methylated genes. In those cases in which more than one paralog has been found only one is referred to by the consensus protein name, with the others designated by the GB numbers from the official honey bee gene list (www.beebase.org). Detailed descriptions of each gene are in Table S5.

Table S1. List of differentially methylated genes in queen and worker larvae

Table S1 (DOC)

CpGs denotes the number of CpG dinucleotides supported by at least four reads and significantly methylated in one of the castes. A generalized linear model of the binomial family was used to identify genes that are differentially methylated between castes. The methylation level of each gene was modeled as a function of the caste and of each of its CpG dinucleotides. In the table, "Caste" indicates whether the caste is a statistically significant factor explaining differences in methylation levels, and "CpG" represents the different dinucleotides of that gene.

Table S2. List of 31 conserved domains most common in methylated genes

Table S2 (DOC)

Table S3. Detailed description of genes related to the tricarboxylic acid (TCA) cycle, the insulin/TOR network, and the ubiquitin/proteasome pathway

Table S3 (DOC)

Table S4. Examples of differentially expressed and/or methylated genes implicated in key pathways of honey bee larval development

Table S4 (DOC)

Table S5. Gene families encoding *Apis* proteins involved in defined functions and showing nonubiquitous patterns of expression

Table S5 (DOC)

This analysis was done on protein domains rather than pathways because the KEGG database is restricted to highly conserved genes (only ~3,000) and does not encompass the less conserved and typically nonmethylated domains.

Table S6. Genes differentially expressed in worker larvae after juvenile hormone treatment showing differential pattern of methylation in queens and workers

Table S6 (DOC)