

Supplemental Online Material

Tables S1 – S5

Supplemental Methods

Supplemental Results

Supplemental References

Table S1. List of genes which were differentially expressed between reproductive and nonreproductive workers, together with functional annotation information. The different columns mention the NCBI gene names and access codes, microarray probe IDs, *Drosophila* orthologs, KEGG orthologs, information on the presence of conserved PFAM protein domains, average loess and quantile-normalised log₂ expression ratios in reproductive vs. nonreproductive workers (M), average fold change (FC), relative order of FC, average intensity (A), raw and Benjamini-Hochberg *FDR*-adjusted *p* values, log₂ expression ratios measured from the 8 arrays of both colonies (colour scale as in Figures 2 and 3), chromosomal positions in the *Amel 4.0* honeybee genome assembly (as given in the NCBI and Beebase databases), whether or not the differential genes lied within QTL regions known to be linked to worker reproductive potential and foraging predisposition (O1-O4: QTLs linked to the worker anarchy phenotype (Oxley *et al.* 2008); L1-L2: QTLs linked to variation in worker ovariole numbers (Linksvayer *et al.* 2009); P1-P4: QTLs linked to pollen foraging and sucrose responsiveness (Hunt *et al.* 2007)) and whether or not they were up or downregulated in the previous, related microarray studies of Grozinger *et al.* 2007 (on differences in brain gene expression in laying vs. non-laying bees), Thompson *et al.* 2006, 2008 (on differences in gene expression in the brains and abdomens of anarchistic vs. wild-type bees) and Kocher *et al.* 2008 (on differences in gene expression in honeybee queens with or without active ovaries) (NA=not included in microarray or spot not significantly above background). In addition, we provide biological process, molecular function and cellular component gene ontology information, and the source from where the GO information was collected from (apis: honeybee-specific GO information from Uniprot and DAVID; droso: GO information of *Drosophila melanogaster* orthologs included in Flybase and Uniprot; blast2go: GO information extracted using a Blast2Go search; pfam: GO information inferred from the presence of conserved PFAM protein domains) and lists

of interesting candidate genes (genes annotated with at least one oogenesis-related GO term, genes coding for neuropeptide precursors, neuropeptide and hormone receptors, odorant and gustatory receptors and odorant binding proteins, G-protein coupled receptors (GPCRs) and GPCR-signalling pathway related genes, genes for which the *Drosophila* ortholog was known to be 20-hydroxyecdysone (20E)-regulated (Beckstead *et al.* 2005), and genes which were differentially expressed in queen mandibular pheromone (QMP) or brood pheromone (BP) treated bees, (Grozinger *et al.* 2003; Alaux *et al.* 2009a)). Finally, we mention which genes are displayed in Figures 2 and 3 and provide additional information and notes about their likely function.

Table S2. Association tests between our differentially expressed gene sets and various other candidate gene sets, calculated using exact hypergeometric one-way Fisher exact tests or one-way Spearman rank correlation tests. The tests shown report the significance of the overlap between our set of differentially expressed genes and those identified in the microarray studies of Grozinger *et al.* 2007 and Thompson *et al.* 2006, 2008, on differences in gene expression in reproductive vs. nonreproductive and anarchistic vs. wild-type bees, that of Kocher *et al.* 2008 on differences in gene expression in honeybee queens with or without active ovaries, and those of Grozinger *et al.* 2003 and Alaux *et al.* 2009a, on gene expression changes caused by exposure to QMP and brood pheromone. In addition, we report the association between our sets of differentially expressed genes and whether genes were contained within various QTL loci linked to worker reproductive potential (OvA1-OvA4 linked to the anarchistic worker phenotype (Oxley *et al.* 2008) and L1-L2 linked to variation in worker ovariole numbers, Linksvayer *et al.* 2009) and foraging predisposition (Pln1-Pln4 linked to pollen foraging and sucrose responsiveness (Hunt *et al.* 2007)), genes for which the *Drosophila* orthologue is known to be 20-hydroxyecdysone (20E)-regulated (Beckstead *et al.* 2005) and neuropeptides that are known to modulate foraging behaviour (Alaux *et al.* 2009b; Brockmann *et al.* 2009; Grozinger *et al.* 2007; Hummon *et al.* 2006; Toth *et al.* 2010; Whitfield *et al.* 2003). The QTL confidence intervals were defined as the region contained within the 1.5 LOD (logarithm of odds) value of the LOD-score peak; chromosomal positions are with respect to the Solignac 3.0 map. Significant associations ($p < 0.05$) are highlighted in bold red; nearly significant associations ($0.05 < p < 0.1$) are shaded in pink.

Table S3. Biological processes gene ontology (GO) terms overrepresented among the sets of genes that were upregulated in reproductive and nonreproductive workers, calculated using *TopGO*. The different columns mention the GO biological process (BP) identifiers, GO names, the number of differentially expressed genes ("Significant") and total number of annotated genes ("Total Annotated") mapping to a particular GO term, fold enrichment, *p*-values calculated using *TopGO* method "elim" (Alexa *et al.* 2006), and the differentially expressed genes (Beebase identifiers) and gene names which mapped to each GO BP term (or one of its subnodes). As the gene universe, we used all the genes that were represented on the array and which were significantly expressed above background in at least one of the channels of one of the 16 hybridisations and which were annotated with functional GO information ($n=8,888$). Significantly enriched terms ($p < 0.05$) are highlighted in red; nearly significantly enriched terms are shown in grey ($0.05 < p < 0.1$).

Table S4

Results of technical (same samples of the microarray analysis) and biological (different samples) validation experiment by means of quantitative Real time Polymerase Chain Reaction (qRT-PCR), using protocols described in Supplemental Methods. The target gene identifiers are displayed together with the results from the microarray, including M (\log_2 expression ratios in reproductive vs. nonreproductive workers), FC (average fold change) and Benjamini-Hochberg *FDR*-adjusted *p* values; all shaded in green. The results of the technical and biological validation experiments are, respectively, shaded in pink and purple, displaying the \log_2 normalised signal intensity from both reproductive and nonreproductive workers and the M-value, FC and *p*-value. In addition, the Pearson correlation *R* in the Log ratios obtained using the microarray platform and qRT PCR analysis are displayed. The differential expression of the four genes in the microarray analysis has been confirmed on identical samples and on biological independent samples of bees which were only 13 days old, captured in a different season and a different year.

Table S5

Primers for the validation experiments using qRT-PCR, specific for the candidate reference genes and the target genes. Target genes have found to be significantly differentially expressed in the microarray analysis, whereas the candidate reference

genes were not (in order to normalise target gene signals). The GB number and full name are displayed, as well as the 5'-3' sequence of both the forward and reverse primer and corresponding amplicon length. The three best reference genes, selected by means of GeNorm, which were used for normalisation are displayed in bold red.

Supplemental Methods

MicroArray Design

For our microarray analyses, a new honeybee microarray was designed for the Agilent platform by the MicroArray Facility (MAF, Flemisch Institute for Biotechnology, Belgium). This array was based on the latest Prerelease 2 version of the honeybee Official Gene Set (OGSPrls2), available on Beebase (genomes.arc.georgetown.edu/downloadFASTA.html). Long oligos for the array were designed using OligoArray 2.1 (berry.engin.umich.edu/oligoarray2_1, Rouillard *et al.* 2003). To identify unique 60-mer oligonucleotide probes and to ensure that the probes had optimal melting temperatures and did not contain repetitive sequences or other anomalies, the probes were designed in an iterative process. This resulted in a total of 15,187 unique probes, representing all but 177 of the 11,062 coding sequences represented in the OGSPrls2. For 4,308 of the genes, two unique probes could be designed.

Microarray validation

To validate our microarray, additional samples were collected from two colonies, each headed by a naturally inseminated, multiple-mated queen. Following standard procedures, broodframes were kept in an incubator at 34°C and inspected daily. Bees were paint marked, according to their age, within 24 h after emergence and introduced into their colony of origin. Queen and open brood from both colonies were removed at the same time. Workers from two colonies (13 to 14 days old) were collected 13 and 16 days after dequeening and immediately snap frozen in liquid nitrogen. Dissection was carried out on thawed abdomens (head and thorax remained frozen) and ovary activation was scored on the same scale as described in the main text. Thawing was reduced to the minimum (max. 1 min.). RNA-extraction was performed on single whole bodies as described in the main text. Concentration and sample quality was determined using Nanodrop ND-1000 (Nanodrop Technologies). cDNA was produced *in duplo*, each starting with 4 µg RNA, using the SuperScript III Reverse Transcriptase

(Invitrogen) in equal volumes (12 μ l) according to manufacturers' guidelines. Both batches were subsequently pooled. Primers (Tab. S5) were designed using Primer Express 2.0 (Applied Biosystems) and Vector NTI and were validated by standard curve and dissociation protocols. Quantitative Real Time Polymerase Chain Reaction (qRT-PCR) was performed on 28 individuals (14 reproductive and 14 nonreproductive) using StepOne plus Cycloer (Applied Biosystems) for 4 target genes: GB16065 (mblk-1-like) and GB14418 (serine/threonine protein kinase 6), which were upregulated in reproductive workers; and GB11256 (similar to CG17292) and GB11943 (cytochrome P450 305D1), which were upregulated in nonreproductive workers (Table S4). In addition 10 candidate reference genes (Tab. S5) were tested and GeNorm (Van Hiel *et al.* 2009; Vandesompele *et al.* 2002) was used to determine the most stable reference genes and calculate normalized target gene expression levels for every bee, using the comparative Ct method and the geometric mean expression level of the 3 best reference genes (GB10903 (ribosomal protein L32), GB16844 (elongation factor 1-alpha) and GB12747 (eukaryotic translation initiation factor 3 subunit C)). Normalized expression levels were subsequently Log₂ transformed and differences in expression were tested for using a General Linear Model in Statistica 8.0 (Statsoft), coding ovary activation as a fixed factor and colony as a random factor. In addition, we also validated the same samples as were used in the microarray analysis using qRT-PCR, following identical protocols.

Functional Gene Ontology Annotation

In order to be able to carry out a functional analysis of the set of differentially expressed genes we annotated as many genes as possible of the official gene set prerelease 2 with gene ontology (GO) information collected from various databases, including GO information for honeybee genes included in Uniprot and DAVID (289 genes), GO information of *Drosophila melanogaster* orthologs included in Flybase and Uniprot (6,693 genes) (identified as reciprocal best blastx hits, using default parameters, but implementing soft filtering and Smith-Waterman alignment (Moreno-Hagelsieb and Latimer 2008)), GO information extracted using a Blast2Go search with default settings (Gotz *et al.* 2008; 8,579 genes) and GO terms derived from the identification of conserved PFAM domains (identified using a search at pfam.sanger.ac.uk/search) followed by the use of a PFAM2GO conversion list (www.geneontology.org/external2go/pfam2go) (4,937 genes). Overall, this strategy

resulted in the functional GO annotation of 9,388 out of 11,062 (85%) genes included in the honeybee OGSPrls2. For any one gene, GO terms from the different databases were subsequently merged, and any duplicate terms were removed. As further aids in the functional analyses, we also identified KEGG orthologs (Kyoto Encyclopedia of Genes and Genomes) using the KEGG automatic annotation server (www.genome.jp/tools/kaas), and occasionally made use of known protein-protein interactions of *Drosophila* orthologs, as catalogued in the *Drosophila* interactions database (www.droidb.org).

Supplemental Results

Microarray validation

All 4 target genes (GB11256, GB16065, GB14418 and 11943) which were differentially expressed in the microarray analysis were also significantly differentially expressed when analysed using qRT PCR, and this was true both when the same 32 samples were used as in the microarray analysis, or when an independent set of 28 samples of younger bees (13 days old) were used from two additional colonies headed by a naturally inseminated, multiple-mated queen (Table S4). The Pearson correlation R in the Log ratios obtained using the microarray platform and qRT PCR analysis were 0.99 ($p=0.01$) and 0.98 ($p=0.02$) when the same samples and an independent set of samples were analysed, respectively. These correlations are similar to and even exceed those obtained in other larger-scale Agilent microarray validation studies, carried out using Taqman or SYBR Green qRT-PCR analysis (Pearson $R = 0.90-0.94$, Arikawa *et al.* 2008).

Supplemental references

- Alaux C, Le Conte Y, Adams HA *et al.* (2009a) Regulation of brain gene expression in honey bees by brood pheromone. *Genes, Brain and Behavior*, **8**, 309-319.
- Alaux C, Sinha S, Hasadsri L *et al.* (2009b) Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proceedings of the National Academy of Sciences*, **106**, 15400-15405.
- Alexa A, Rahnenfuhrer J and Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, **22**, 1600-1607.
- Arikawa E, Sun Y, Wang J *et al.* (2008) Cross-platform comparison of SYBR(R) Green real-time PCR with TaqMan PCR, microarrays and other gene expression measurement technologies evaluated in the MicroArray Quality Control (MAQC) study. *BMC Genomics*, **9**, 328
- Beckstead RB, Lam G and Thummel CS (2005) The genomic response to 20-hydroxyecdysone at the onset of *Drosophila* metamorphosis. *Genome Biology*, **6**, R99
- Brockmann A, Annangudi SP, Richmond TA *et al.* (2009) Quantitative peptidomics reveal brain peptide signatures of behavior. *Proceedings of the National Academy of Sciences*, **106**, 2383-2388.
- Gotz S, Garcia-Gomez JM, Terol J *et al.* (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*, **36**, 3420-3435.
- Grozinger CM, Fan Y, Hoover SE *et al.* (2007) Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Molecular Ecology*, **16**, 4837-4848.
- Grozinger CM, Sharabash NM, Whitfield CW *et al.* (2003) Pheromone-mediated gene expression in the honey bee brain. *Proceedings of the National Academy of Sciences*, **100 Suppl 2**, 14519-14525.
- Hummon AB, Richmond TA, Verleyen P *et al.* (2006) From the genome to the proteome: uncovering peptides in the *Apis* brain. *Science*, **314**, 647-649.
- Hunt GJ, Amdam GV, Schlipalius D *et al.* (2007) Behavioral genomics of honeybee foraging and nest defense. *Naturwissenschaften*, **94**, 247-267.
- Kocher SD, Richard FJ, Tarpy DR *et al.* (2008) Genomic analysis of post-mating changes in the honey bee queen (*Apis mellifera*). *BMC Genomics*, **9**, 232
- Linksvayer TA, Rueppell O, Siegel A *et al.* (2009) The genetic basis of transgressive ovary size in honeybee workers. *Genetics*, **183**, 693-13SI.
- Moreno-Hagelsieb G and Latimer K (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, **24**, 319-324.
- Oxley PR, Thompson GJ and Oldroyd BP (2008) Four quantitative trait loci that influence worker sterility in the honeybee (*Apis mellifera*). *Genetics*, **179**, 1337-1343.
- Rouillard JM, Zuker M and Gulari E (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Research*, **31**, 3057-3062.
- Thompson GJ, Kucharski R, Maleszka R *et al.* (2008) Genome-wide analysis of genes related to ovary activation in worker honey bees. *Insect Molecular Biology*, **17**, 657-665.

- Thompson GJ, Kucharski R, Maleszka R *et al.* (2006) Towards a molecular definition of worker sterility: differential gene expression and reproductive plasticity in honey bees. *Insect Molecular Biology*, **15**, 637-644.
- Toth AL, Varala K, Henshaw MT *et al.* (2010) Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proceedings of the Royal Society B-Biological Sciences*, **277**, 2139-2148.
- Van Hiel M, Van Wielendaele P, Temmerman L *et al.* (2009) Identification and validation of housekeeping genes in brains of the desert locust *Schistocerca gregaria* under different developmental conditions. *BMC Molecular Biology*, **10**, 56
- Vandesompele J, De Preter K, Pattyn F *et al.* (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, RESEARCH0034
- Whitfield CW, Cziko AM and Robinson GE (2003) Gene expression profiles in the brain predict behavior in individual honey bees. *Science*, **302**, 296-299.