Monoallelic methylation and allele specific expression in a social insect

Zoë N. Lonsdale[1*] (zl107@leicester.ac.uk), Kate D. Lee[2] (kate.d.lee@gmail.com), Maria Kyriakidou[1] (maria.kiriakidu@gmail.com), Harindra E. Amarasinghe[1] (heak1@yahoo.co.uk), Despina Nathanael[1] (dn62@student.le.ac.uk), Eamonn B. Mallon[1] (ebm3@le.ac.uk)

1 Department of Genetics, University of Leicester, Leicester, U.K.

2 Bioinformatics and Biostatistics Support Hub (B/BASH), University of Leicester, Leicester, U.K.

∗ **Corresponding author**

Abstract

Background: The social hymenoptera are emerging as models for epigenetics. In mammals and flowering plants' epigenetics, methylation affects allele specific expression. There is contradictory evidence for the role of methylation on allele specific expression and monoallelic methylation in social insects. The aim of this paper is to investigate allele specific expression and monoallelic methylation in the bumblebee, *Bombus terrestris*.

Results: We found nineteen genes that were both monoallelically methylated and monoallelically expressed. A number of these genes are involved in reproduction. Fourteen of these genes express the hypermethylated allele, while the other five express the hypomethylated allele.

We also searched for allele specific expression in twenty-nine published RNA-seq libraries. We found 555 loci with allele-specific expression.

Conclusions: Genomic imprinting in mammals often involves monoallelic methylation and expression. It is tempting to associate our results with genomic imprinting, especially as a number of the genes discovered are exactly the type predicted by theory to be imprinted. Caution however should be applied due to the lack of understanding of the functional role of methylation in gene expression in insects and in the as yet unquantified role of genetic cis effects in insect allele specific methylation and expression.

Keywords: methylation, allele specific expression, hymenoptera, genomic imprinting

Background

Epigenetics is the study of heritable changes in gene expression that do not involve changes to the underlying DNA sequence [1]. Social hymenoptera (ants, bees, and wasps) are important emerging models for epigenetics [2, 3, 4, 5]. This is due to theoretical predictions for a role for an epigenetic phenomenon, genomic imprinting, in their social organisation [6], the recent discovery of parent-of-origin allele specific expression in honeybees [7], and data showing a fundamental role in social insect biology for DNA methylation, an epigenetic marker [8]. Genomic imprinting is allele specific expression in diploid individuals, where expression is dependent on the sex of the parent from which an allele was inherited [9]. In mammals and flowering plants, genomic imprinting is often associated with methylation marks passed from parents to offspring [10].

However the presence of allele specific expression and methylation does not necessarily mean an epigenetic process is involved. Allele specific expression, and DNA methylation, can be due to processes other than genomic imprinting. Allele specific expression is known to be caused by a number of genetic as well as epigenetic processes [11]. The genetic process usually involves cis effects such as transcription factor binding sites, or less often, untranslated regions which alter RNA stability or microRNA binding [12]. As well as genomic imprinting, DNA methylation is also involved in cellular differentiation [13]. Allele specific methylation can also be affected by the allele's genotype as well as epigenetics [14].

There is contradictory evidence for the role of methylation on allele specific expression and monoallelic methylation in social insects. Methylation is associated with allele specific expression in a number of loci in the ants *Camponotus floridanus* and *Harpegnathos saltator* [15]. Recently, we found evidence for allele specific expression in bumblebee worker reproduction genes [16] and that methylation is important in bumblebee worker reproduction [17]. However, other work on the honeybee *Apis mellifera* found no link between potentially imprinted loci and known methylation sites in that species [18].

The aim of this paper is to investigate allele specific expression and monoallelic methylation in the bumblebee, *Bombus terrestris*. The recently sequenced genome of the bumblebee, *Bom-*

bus terrestris displays a full complement of genes involved in the methylation system [19]. An extreme form of imprinting involves monoallelic expression (one allele is completely silenced). In the canonical mammal and flowering plant systems, this is often associated with monoallelic methylation. In this paper, we examined the link between monoallelic methylation and monoallelic expression in the bumblebee, *Bombus terrestris*, by examining two whole methylome libraries and an RNA-seq library from the same bee. MeDIP-seq is an immunoprecipitation technique that creates libraries enriched for methylated cytosines [20]. Methyl-sensitive restriction enzymes can create libraries that are enriched for non-methylated cytosines (MRE-seq) [20]. Genes found in both libraries are monoallelically methylated, with the hypermethylated allele being in the MeDIP-seq data and the hypomethylated allele in the MRE-seq data [20]. Monoallelic expression was identified in these loci from the RNA-seq library. If only one allele was expressed then we knew that these loci were both monoallelically methylated and monoallelically expressed in this bee. We confirmed this monoallelic expression in one locus using qPCR.

We then more generally searched for allele specific expression by analysing twenty nine published RNA-seq libraries from worker bumblebees [21, 22]. We identified heterozygotes in the RNA-seq libraries and measured the expression of each allele. We then identified loci that showed significant expression differences between their two alleles.

Results

In total, we found nineteen genes that were both monoallelically methylated (present in both Me-DIP and MRE-seq libraries) and monoallelically expressed (only one allele present in the RNA-seq library), for an example see *Bicaudal-D* in Figure 1. Of the nineteen genes, fourteen had the hypermethylated (MeDIP) allele expressed, while five had the hypomethylated (MRE-seq) allele expressed (see supplementary table 1).

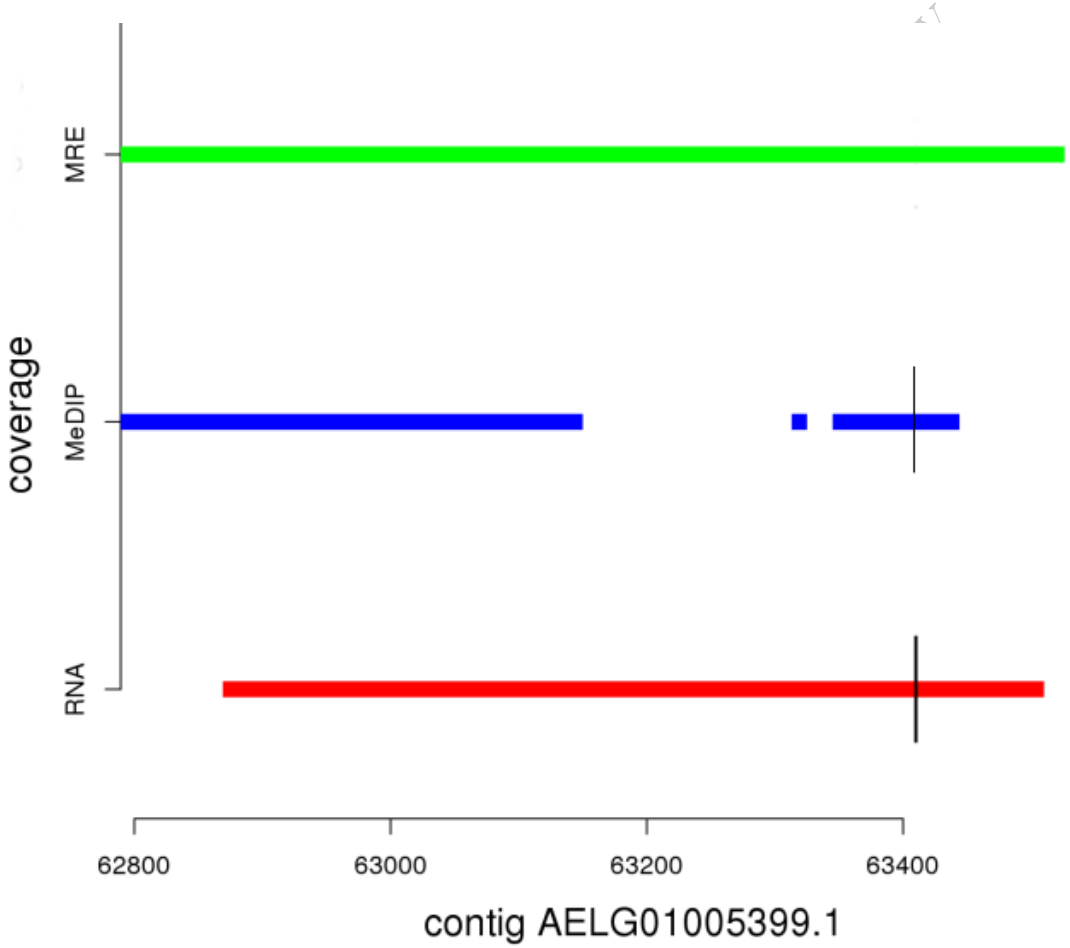

Figure 1: **Coverage of three libraries for bicaudal d.** Horizontal lines represent available reads for each library over this genomic range (x-axis). The vertical line shows the position of the snp and which genomic library shares the same allele.

Monoallelic expression was confirmed in one of these nineteen (*slit homolog 2 protein-like*

(AELG01000623.1)) by allele specific qPCR [16]. The allele with a guanine at the snp position had a mean expression of 6.04 ±8.28 (standard deviation) in four bees from three different colonies. The thymine allele was not expressed at all in these bees. This was not due to the efficiency of the primers as the DNA controls of both alleles showed similar amplification (G mean = 422.70 ±507.36, T mean = 1575.17 ±503.02). In the three other loci tested (*Ras GTPase-activating protein 1*, *Ecdysone receptor*, *methionine aminopeptidase 1-like*) we found apparent monoallelic expression, but could not dismiss primer efficiency as the cause.

The nineteen genes were blasted against the nr/nt database (blastn). Four returned no hits and a further four returned noninformative hits. A number of these genes had homologs known to be methylated in other animals (Table 1). Six of the eleven genes with informative hits have functions to do with social organisation in the social insects (Table 1).

We then looked at these nineteen genes in twenty-nine previously published RNA-seq libraries. Fifteen of these nineteen genes expressed a single allele in all twenty nine RNA-seq libraries, see supplementary table 2. The remaining four genes (AELG01000620.1, AELG01001021.1, AELG01002224.1a, AELG01002224.1b) were inconsistent; they showed expression of one allele in some *B. terrestris* workers, and expression of two alleles in other workers.

We then searched more generally for allele specific expression in the twenty-nine RNA-seq libraries. 555 loci showed allele-specific expression in ≥3 of the 29 RNA-seq libraries (supplementary table 3). Blasting (Blastn) these loci against *Bombus terrestris* returned 211 hits. To search for gene ontology terms, we blasted (blastx) against *Drosophila melanogaster*, which returned 329 hits. One hundred and fifty-one Gene Ontology(GO) terms were enriched in the 555 regions showing allele specific expression (Fishers exact test p >0.05), however none were significant at the more stringent FDR >0.05. Figure 2 shows the large number of biological functions associated with these 555 genes.

Gene	Accession	Expressed allele	Function
yippee-like 1	AELG01001021.1	MeDIP	Yippee is an intracellular protein with a zinc-finger like domain. DNA methylation of a CpG island near the *yippie-like 3* promoter in humans represents a possible epigenetic mechanism leading to decreased gene expression in tumours [23].
slit homolog 2 protein-like	AELG01000623.1	MeDIP	Slit is produced by midline glia in insects and is involved in cell projection during development [24]. All three human Slits were found to be hypermethylated in hepatocellular carcinoma cell lines [25].
methionine aminopeptidase 1-like	AELG01000544.1	MeDIP	Methionine aminopeptidases catalyse N-terminal methionine removal [26]. MAP1D in humans was found to be potentially oncogenic [26].
calmodulin-lysine N-methyltransferase-like	AELG01003672.1	MRE	Calmodulin-lysine N-methyltransferase catalyses the trimethylation of a lysine residue of calmodulin. Calmodulin is a ubiquitous, calcium-dependent, eukaryotic signalling protein with a large number of interactors. The methylation state of calmodulin causes phenotypic changes in growth and developmental processes [27].
Ecdysone receptor	AELG01000543.1	MRE	In *Drosophila melanogaster*, ecdysone receptor interacts with ecdysone to activate a series of ecdysteroid genes [28]. In honeybees, *Ecdysone receptor* is expressed in the brain mushroom bodies of both workers and queens and ovaries of queens [28].
Shaker	AELG01001021.1	MeDIP	Shaker is involved in the operation of potassium ion channel. *Shaker* expression was upregulated in sterile versus reproductive honeybee workers [29].
excitatory amino acid transporter 4-like	AELG01000969.1	MRE	Excitatory amino acid transporters are neurotransmitter transporters. *Excitatory amino acid transporter 3* expression was upregulated in sterile honeybee workers [29]. *Excitatory amino acid transporter 1* expression differences were also associated with worker - queen differentiation in the paper wasp *Polistes metricus* [30].
elongation of very long chain fatty acids protein 6-like	AELG01004467.1	MeDIP	The timing of the upregulation of fatty acid metabolism was found to be different in queen and worker honeybees [31].
ras GTPase-activating protein nGAP-like	AELG01004618.1	MeDIP	*Ras GTPase-activating protein 1* was found to be upregulated in reproductive honeybee workers [29]. It is involved in oocyte meiosis.
bicaudal D-related protein homolog	AELG01005399.1	MeDIP	Bicaudal is involved in embryonic pattern formation in *Drosophila* [32]. It is thought to be involved in the differentiation between soldiers and workers in the termite *Reticulitermes flavipes* [33]. *Bicaudal protein D* has been shown to be methylated more in eggs than sperm in honeybees [34].

Table 1: The eleven of the nineteen monoallelically methylated and expressed genes that returned informative blast hits.

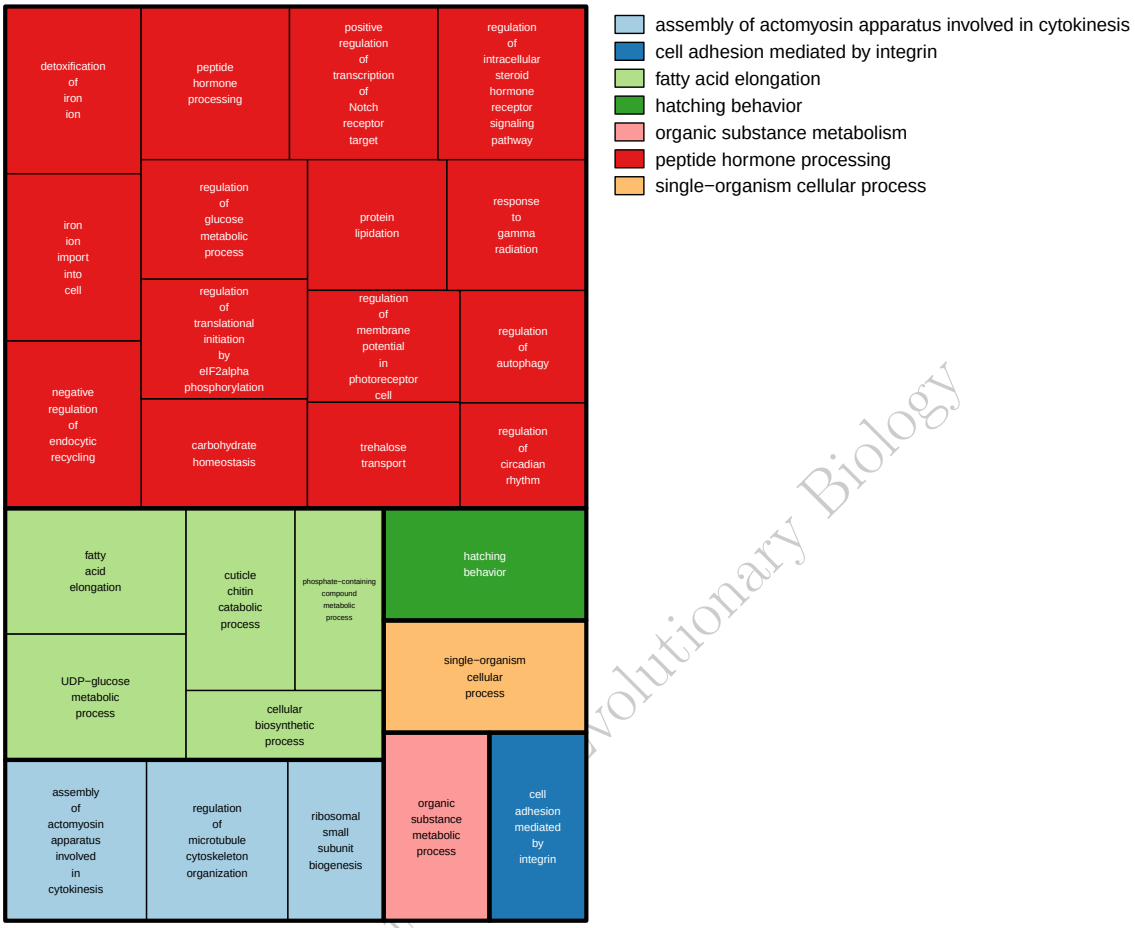

Figure 2: **GO terms associated with allele specific expression.** A summary of the enriched GO terms ($p < 0.05$, based on Blast2Go annotation) found for genes displaying allele specific expression. This figure was produced using Revigo

Discussion

Of the nineteen genes displaying monoallelic methylation and monoallelic expression, fourteen had the hypermethylated (MeDIP) allele expressed, while five had the hypomethylated (MRE-seq) allele expressed (see supplementary table 1). In ant genes with allele specific methylation, the hypermethylated allele showed more expression than the hypomethylated allele [15]. This fits with genome wide analysis that shows exonic methylation in insects associated with increased gene expression [35, 36]. Our fourteen genes with the hypermethylated allele expressed agree

with this pattern. But how to explain the five genes where the hypomethylated allele was expressed? Firstly, the role of methylation in insect gene expression is not clear cut, with the relationship between exonic methylation and expression often disappearing at the gene level [36]. Secondly, even in the canonical mammalian methylation system, the "wrong" allele has been shown to be expressed occasionally due to lineage specific effects [37, 38, 39, 40, 41].

We then looked at the expression of these nineteen genes in all twenty-nine RNA-seq libraries. If they are monoallelically expressed in these bees, we would find only one allele in a given RNA-seq library. Fifteen of these nineteen genes were confirmed to show a single allele in all twenty-nine RNA-seq libraries. We would also find only one allele if that bee was homozygous. We can not rule out that these fifteen genes just happen to be homozygous in all twenty-nine bees from five different colonies from multiple sources, although this seems unlikely. This result suggests that the finding in the monoallelic analysis can be generalized.

The remaining four genes showed inconsistent expression with one allele being expressed in some *B. terrestris* workers, and expression of two alleles in other workers. Natural intraspecific variation in imprinting has been found in other species [42]. Another explanation is that these loci are not imprinted but rather their allele specific expression is derived from genetic effects [43].

There are three main genetic, as opposed to epigenetic, affectors of allele specific expression [44]. Allele specific expression can be caused by differences in the alleles' sequence within the translated part resulting in a modified protein. A change at the alleles' cis regulatory sites, could cause differential binding of transcription factors. Transcript processing can be affected by a change in the alleles' sequence a splice site or untranslated region. This large number of possible causes of allele specific expression could explain why we see so many functions associated with the 555 genes showing allele specific expression (Figure 2).

But it is not just allele specific expression that may have genetic as well as epigenetic effects. It has been shown in humans that some allele specific methylation is determined by DNA sequence in cis and therefore shows Mendelian inheritance patterns [14]. An extreme example

of genetically controlled allele specific methylation is found in *Nasonia* wasps, where there is no evidence for methylation driven genomic imprinting, but inheritable cis-mediated allele specific methylation has been found [45]. This cis-mediated methylation has recently been suggested as being important in social insect biology [43, 46].

We have found that allele specific expression is widespread in the bumblebee. We have also found that the extreme version of allele specific expression, monoalleic expression is associated with monoallelic methylation. Genomic imprinting in mammals usually involves monoallelic methylation and expression. It is tempting to associate our results with genomic imprinting, especially as a number of the genes discovered are exactly the type predicted by theory to be imprinted [6]. Caution however should be applied due to the lack of understanding of the functional role of methylation in gene expression in insects and in the as yet unquantified role of genetic cis effects in insect allele specific methylation and expression.

Materials and Methods

Samples

Data from twenty-nine RNA-seq libraries were used for the allele specific expression analysis (six from Harrison *et al.* [21], and twenty-three from Riddell *et al.* [22]. The Riddell bees came from two colonies, one commercially reared bumblebee colony from Koppert Biological Systems U.K. and one colony from a wild caught queen from the botanic gardens, Leicester. The Harrison bees were from three commercially reared colonies obtained from Agralan Ltd. A Koppert colony worker bee was used for the MeDIP-seq / MRE-seq / RNA-seq experiment, and was from a separate Koppert colony to the bees used for the qPCR analysis. Samples are outlined in Table 2. Colonies were fed *ad libitum* with pollen (Percie du sert, France) and 50 % diluted glucose/fructose mix (Meliose Roquette, France). Before and during the experiments colonies were kept at 26°C and 60% humidity in constant red light.

Table 2: **Bees used in each experiment.** K refers to Koppert, A to Agralan and Q to the wild caught Leicester queen.

Experiment	Number	Colony
Allele specific expression RNA-seq	1	A1
	2	A2
	2	A3
	1	K1
	14	K2
	9	Q1
MeDip/MRE/RNA-seq	1	K1
qPCR	2	K3
	1	K4
	1	K5

Next generation sequencing

MeDIP-seq, MRE-seq and RNA-seq

RNA and DNA was extracted from a single five day old whole bee (Colony K2). DNA was extracted using an ethanol precipitation method. Total RNA was extracted using Tri-reagent (Sigma-Aldrich, UK).

Three libraries were prepared from this bee by Eurofins genomics. These were MeDIP-seq and MRE-seq libraries on the DNA sample and one amplified short insert cDNA library with size of 150-400 bp using RNA. Both the MeDIP-seq and MRE-seq library preparations are based on previously published protocols [20]. MeDIP-seq uses monoclonal antibodies against 5-methylcytosine to enrich for methylated DNA independent of DNA sequence. MRE-seq enriches for unmethylated cytosines by using methylation-sensitive enzymes that cut only restriction sites with unmethylated CpGs. Each library was individually indexed. Sequencing was performed on an Illumina HiSeq2000 instrument (Illumina, Inc.) by the manufacturers protocol. Multiplexed 100 base paired-read runs were carried out yielding 9390 Mbp for the MeDIP-seq library, 11597 Mbp for the MRE-seq library and 8638 Mbp for the RNA-seq library.

Previously published RNA-seq

Full details of the RNA-seq protocols used have been published previously [21, 22]. Briefly, for the Riddell bees, total RNA was extracted from twenty three individual homogenised abdomens using Tri-reagent (Sigma-Aldrich, UK). TruSeq RNA-seq libraries were made from the 23 samples at NBAF Edinburgh. Multiplexed 50 base single-read runs was performed on an Illumina HiSeq2000 instrument (Illumina, Inc.) by the manufacturers protocol. For the Harrison bees, total RNA was extracted from whole bodies using a GenElute Mammalian Total RNA Miniprep kit (Sigma-Aldrich) following the manufacturers' protocol. The six libraries were sequenced as multiplexed 50 base single-read runs on an Illumina HiSeq 2500 system in rapid mode at the Edinburgh Genomics facility of the University of Edinburgh.

Monoallelic methylation and expression - Bioinformatic analysis

We searched for genes that were monoallelically methylated (present in both methylation libraries), heterozygous and monoallelically expressed (only one allele present in the RNA-seq library).

Alignment and bam refinement

mRNA reads were aligned to the *Bombus terrestris* genome assembly (AELG00000000) using Tophat [47] and converted to bam files with Samtools [48]. Reads were labelled with the AddOrReplaceReadGroups.jar utility in Picard (`http://picard.sourceforge.net/`). The MRE-seq and MeDIP-seq reads were aligned to the genome using BWA mapper [49]. The resultant sam alignments were soft-clipped with the CleanSam.jar utility in Picard and converted to bam format with Samtools. The Picard utility AddOrReplaceReadGroups.jar was used to label the MRE and MeDIP reads which were then locally re-aligned with GATK [50, 51]. PCR duplicates for all bams (mRNA, MeDIP and MRE) were marked with the Picard utility Markduplicates.jar.

Identifying regions of interest and integrating data

Coverage of each data type was calculated using GATK DepthofCoverage [51]. Only regions with a read depth of at least six in each of the libraries (RNA-seq, MeDIP-seq and MRE-seq) was used. Heterozygotes were identified using Samtools mpileup and bcftools on each data set separately [49] and results were merged with vcf tools [52]. CpG islands were identified using CpG island searcher [53]. Regions of mRNA with overlaps of MeDIP, MRE, CpG islands and monoallelic snps were identified with custom perl scripts.

Allele specific expression - Bioinformatic analysis

We created a pipeline to search for heterozygous loci that show allele specific expression and identify the associated enriched gene ontology (GO) terms in twenty-nine previously published RNA-seq libraries [21, 22].

Each RNA library was mapped to the *Bombus terrestris* reference genome (Bter 1.0, accession AELG00000000.1) [19] using the BWA mapper [49]. The mean GC content of the 29 libraries was 42.34%, with individual libraries having a similar GC content ranging from 40-46%. GC content differed with run (Nested ANOVA: $F = 20.302$, $df = 1$, $p < 0.001$), but not by colony (Nested ANOVA: $F = 1.763$, $df = 4$, $p = 0.171$). The mean coverage of the 29 libraries was 13.29, with mean library coverage ranging from 9.84 to 17.61. Run had an effect on coverage (Nested ANOVA: $F = 7.554$, $df = 1$, $p = 0.011$), as did colony (Nested ANOVA: $F = 6.962$, $df = 4$, $p < 0.001$).

Therefore, the combat method in the R package SVA (version 3.20.0) was used to remove any batch effects and control for original differences in coverage [54, 55]. The success of this control was confirmed by the R package edgeR (version 3.14.0) [56, 57]. The SVA adjustment reduced the edgeR dispersion value from 3.9994 (BCV=2) to 0 (BCV=0.0003) (supplementary figure 1).

Bcftools (version 0.1.19-44428cd), bedtools (version 2.17.0), and samtools (version 0.1.19-44428cd) were used to prepare the RNA libraries and call the SNPs, before the SNPs were filtered based on mapping quality score [49, 58]. Only SNPs with a mapping quality score of $p < 0.05$ and a read depth of ≥ 6 were included in the analyses. The R package, QuASAR, was then used to identify genotypes (according to the Hardy-Weinberg equilibrium) and locate any allele specific expression at heterozygous sites [59]. QuASAR removes snps with extreme differential allele expression from the analyses, thus controlling for any base-calling errors. The loci (the snp position +/- 2900bp) identified as showing ASE in at least three of the thirty libraries, were blasted (Blastx) against *Drosophila melanogaster* proteins (non-redundant (nr) database) [60]. The blast results were annotated using Blast2Go [61]. Fisher's exact test was implemented to identify enriched GO terms, which were then visualised using REVIGO [62]. To identify which bumblebee genes the snps were located in, the snp position +/- 25 bp was blasted (Blastn) against the *Bombus terrestris* genome [19].

Candidate gene allele specific qPCR

DNA was extracted from four bees from three Koppert colonies using the Qiagen DNA Micro kit according to manufacturer's instructions. RNA was extracted from samples of the heads of the same worker bees with the QIAGEN RNeasy Mini Kit according to manufacturer's instructions. cDNA was synthesized from a 8µl sample of RNA using the Tetro cDNA synthesis Kit (Bioline) as per manufacturer's instructions.

We amplified numerous fragments of the 19 candidate genes. Sanger sequencing results were analyzed using the heterozygote analysis module in Geneious version 7.3.0 to identify heterozygotic nucleotide positions. It was difficult to identify snps in exonic regions of the 19 loci, which could be amplified with primers of suitable efficiency. We managed to identify a suitable region in *slit homolog 2 protein-like* (AELG01000623.1 exonic region 1838-2420).

The locus was run for 3 different reactions; T allele, G allele and reference. Reference primers were designed according to [63]. A common reverse primer (CTGGTTCCCGTCCAATCTAA) was used for all three reactions. A reference forward primer (CGTGTCCAGAATCGACAATG) was designed to the same target heterozygote sequence, upstream of the heterozygote nucleotide position. The reference primers measure the total expression of the gene, whereas the allele specific primers (T allele: CCAGAATCGACAATGACTCGT, G allele: CAGAATCGACAATGACTCGG) measure the amount of expression due to the allele. Thus the ratio between the allele specific expression and reference locus expression would be the relative expression due to the allele.

Three replicate samples were run for each reaction. All reactions were prepared by the Corbett robotics machine, in 96 well qPCR plates (Thermo Scientific, UK). The qPCR reaction mix (20µl) was composed of 1µl of diluted cDNA (50ng/µl), 1µl of forward and reverse primer (5µM/µl each), 10µl 2X SYBR Green JumpStart Taq ReadyMix (Sigma Aldrich, UK) and 7µl ddH_20. Samples were run in a PTC-200 MJ thermocycler. The qPCR profile was; 4 minutes at 95°C denaturation followed by 40 cycles of 30s at 95°C, 30s at 59°C and 30s at 72°C and a final extension of 5 minutes at 72°C.

Forward primers are different, both in their terminal base (to match the snp) and in their length. It is entirely possible that they may amplify more or less efficiently even if there was no difference in amount of template [64]. To test for this we repeated all qPCRs with genomic DNA (1µl of diluted DNA (20ng/µl) from the same bees as the template. We would expect equal amounts of each allele in the genomic DNA. We also measured efficiency of each reaction as per [65].

Median C_t was calculated for each set of three technical replicates. A measure of relative expression (ratio) was calculated for each allele in each worker bee as follows:

$$ratio_{allele} = \frac{E_{allele}^{-Ct_{allele}}}{E_{reference}^{-Ct_{reference}}} \quad (1)$$

E is the median efficiency of each primer set [64, 65]. All statistical analysis was carried out using R (3.1.0) [66].

Ethical declaration

The protocol reported here conforms to the regulatory requirements for animal experimentation in the United Kingdom.

Data accessibility

All sequence data for this study are archived at European Nucleotide Archive (ENA); Accession no. PRJEB9366 (http://www.ebi.ac.uk/ena/data/view/PRJEB9366), x, x. GO-analysis results and lists of differentially expressed transcripts are available as Supporting Information.

Competing interests

The authors declare they have no competing interests.

Author contributions

ZNL analysed the data and wrote the initial draft. KDL analysed the data and was involved in the redrafting of the manuscript. HEA carried out the experiments and was involved in the redrafting of the manuscript. DN carried out the experiments and was involved in the redrafting of the manuscript. MK analysed the data was involved in the redrafting of the manuscript. EBM designed the project, analysed the data and wrote the initial draft.

Acknowledgements

This work was financially supported by NERC grant no. NE/H010408/1 and NE/N010019/1 and NERC Biomolecular Analysis Facility research grants (NBAF 606 and 829) to EBM. Illumina library preparation, sequencing and bioinformatics were carried out by Edinburgh Genomics, The University of Edinburgh. Edinburgh Genomics is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1) and BBSRC (BB/J004243/1). ZNL would like to thank UK BBSRC for its financial support via MIBTP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Goldberg, A., Allis, C. & Bernstein, E., 2007 Epigenetics: A Landscape Takes Shape. *Cell* **128**, 635–638. ISSN 0092-8674. (doi:10.1016/j.cell.2007.02.006).

[2] Glastad, K. M., Hunt, B. G., Yi, S. V. & Goodisman, M. a. D., 2011 DNA methylation in insects: on the brink of the epigenomic era. *Insect Molecular Biology* **20**, 553–565. ISSN 1365-2583. (doi:10.1111/j.1365-2583.2011.01092.x).

[3] Weiner, S. A. & Toth, A. L., 2012 Epigenetics in social insects: a new direction for understanding the evolution of castes. *Genetics research international* **2012**, 609810. ISSN 2090-3162. (doi:10.1155/2012/609810).

[4] Welch, M. & Lister, R., 2014 Epigenomics and the control of fate, form and function in social insects. *Current Opinion in Insect Science* **1**, 31–38. ISSN 2214-5745. (doi:10.1016/j.cois.2014.04.005).

[5] Yan, H., Simola, D. F., Bonasio, R., Liebig, J., Berger, S. L. & Reinberg, D., 2014 Eusocial insects as emerging models for behavioural epigenetics. *Nature Reviews Genetics* **15**, 677–688. ISSN 1471-0056. (doi:10.1038/nrg3787).

[6] Queller, D. C., 2003 Theory of genomic imprinting conflict in social insects. *Bmc Evolutionary Biology* **3**, art. no.–15.

[7] Galbraith, D. A., Kocher, S. D., Glenn, T., Albert, I., Hunt, G. J., Strassmann, J. E., Queller, D. C. & Grozinger, C. M., 2016 Testing the kinship theory of intragenomic conflict in honey bees (Apis mellifera). *Proceedings of the National Academy of Sciences* p. 201516636. ISSN 0027-8424, 1091-6490. (doi:10.1073/pnas.1516636113).

[8] Chittka, A., Wurm, Y. & Chittka, L., 2012 Epigenetics: The making of ant castes. *Current Biology* **22**, R835–R838. ISSN 0960-9822. (doi:10.1016/j.cub.2012.07.045).

[9] Haig, D., 2000 The kinship theory of genomic imprinting. *Annual Review of Ecology and Systematics* **31**, 9–32.

[10] Reik, W. & Walter, J., 2001 Genomic imprinting: Parental influence on the genome. *Nature Reviews Genetics* **2**, 21–32. ISSN 1471-0056. 1.

[11] Palacios, R., Gazave, E., Goi, J., Piedrafita, G., Fernando, O., Navarro, A. & Villoslada, P., 2009 Allele-Specific Gene Expression Is Widespread Across the Genome and Biological Processes. *PLoS ONE* **4**, e4150. (doi:10.1371/journal.pone.0004150).

[12] Farh, K. K.-H., Grimson, A., Jan, C., Lewis, B. P., Johnston, W. K., Lim, L. P., Burge, C. B. & Bartel, D. P., 2005 The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science (New York, N.Y.)* **310**, 1817–1821. ISSN 1095-9203. (doi:10.1126/science.1121158).

[13] Bird, A., 2002 DNA methylation patterns and epigenetic memory. *Genes & Development* **16**, 6–21. ISSN 0890-9369, 1549-5477. (doi:10.1101/gad.947102).

[14] Meaburn, E. L., Schalkwyk, L. C. & Mill, J., 2010 Allele-specific methylation in the human genome: implications for genetic studies of complex disease. *Epigenetics* **5**, 578–582. ISSN 1559-2308. (doi:10.4161/epi.5.7.12960).

[15] 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. *Current Biology* **22**, 1755–1764. ISSN 0960-9822. (doi:10.1016/j.cub.2012.07.042).

[16] Amarasinghe, H., Toghill, B., Nathanael, D. & Mallon, E. B., 2015 Allele specific expression in worker reproduction genes in the bumblebee Bombus terrestris. *PeerJ* **3**, e1079. (doi: https://dx.doi.org/10.7717/peerj.1079).

[17] Amarasinghe, H. E., Clayton, C. I. & Mallon, E. B., 2014 Methylation and worker reproduction in the bumble-bee (Bombus terrestris). *Proceedings of the Royal Society B: Biological Sciences* **281**, 20132502. ISSN 0962-8452, 1471-2954. (doi:10.1098/rspb.2013.2502).

[18] Kocher, S. D., Tsuruda, J. M., Gibson, J. D., Emore, C. M., Arechavaleta-Velasco, M. E., Queller, D. C., Strassmann, J. E., Grozinger, C. M., Gribskov, M. R., San Miguel, P. *et al.*, 2015 A Search for Parent-of-Origin Effects on Honey Bee Gene Expression. *G3 (Bethesda, Md.)* ISSN 2160-1836. (doi:10.1534/g3.115.017814).

[19] Sadd, B. M., Barribeau, S. M., Bloch, G., Graaf, D. C. d., Dearden, P., Elsik, C. G., Gadau, J., Grimmelikhuijzen, C. J., Hasselmann, M., Lozier, J. D. *et al.*, 2015 The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biology* **16**, 76. ISSN 1465-6906. (doi:10.1186/s13059-015-0623-3).

[20] Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., Johnson, B. E., Fouse, S. D., Delaney, A., Zhao, Y. *et al.*, 2010 Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature biotechnology* **28**, 1097–1105. ISSN 1546-1696. (doi:10.1038/nbt.1682).

[21] Harrison, M. C., Hammond, R. L. & Mallon, E. B., 2015 Reproductive workers show queen-like gene expression in an intermediately eusocial insect, the buff-tailed bumble bee Bombus terrestris. *Molecular Ecology* **24**, 121–129. ISSN 1365-294X. (doi:10.1111/mec.13215).

[22] Riddell, C. E., Garces, J. D. L., Adams, S., Barribeau, S. M., Twell, D. & Mallon, E. B., 2014 Differential gene expression and alternative splicing in insect immune specificity. *BMC Genomics* **15**, 1031. ISSN 1471-2164. (doi:10.1186/1471-2164-15-1031).

[23] Kelley, K., Miller, K. R., Todd, A., Kelley, A., Tuttle, R. & Berberich, S. J., 2010 YPEL3, a p53-regulated gene that induces cellular senescence. *Cancer research* **70**, 3566–3575. ISSN 0008-5472. (doi:10.1158/0008-5472.CAN-09-3219).

[24] Rothberg, J. M., Jacobs, J. R., Goodman, C. S. & Artavanis-Tsakonas, S., 1990 slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. *Genes & Development* **4**, 2169–2187. ISSN 0890-9369, 1549-5477. (doi:10.1101/gad.4.12a.2169).

[25] Zheng, D., Liu, B.-B., Liu, Y.-K., Kang, X.-N., Sun, L., Guo, K., Sun, R.-X., Chen, J. & Zhao, Y., 2009 Analysis of the expression of Slit/Robo genes and the methylation status of their promoters in the hepatocellular carcinoma cell lines. *Zhonghua Gan Zang Bing Za Zhi = Zhonghua Ganzangbing Zazhi = Chinese Journal of Hepatology* **17**, 198–202. ISSN 1007-3418.

[26] Leszczyniecka, M., Bhatia, U., Cueto, M., Nirmala, N. R., Towbin, H., Vattay, A., Wang, B., Zabludoff, S. & Phillips, P. E., 2006 MAP1d, a novel methionine aminopeptidase family member is overexpressed in colon cancer. *Oncogene* **25**, 3471–3478. ISSN 0950-9232. (doi:10.1038/sj.onc.1209383).

[27] Magnani, R., Dirk, L. M. A., Trievel, R. C. & Houtz, R. L., 2010 Calmodulin methyltransferase is an evolutionarily conserved enzyme that trimethylates Lys-115 in calmodulin. *Nature Communications* **1**, 43. (doi:10.1038/ncomms1044).

[28] Takeuchi, H., Paul, R. K., Matsuzaka, E. & Kubo, T., 2007 EcR-A expression in the brain and ovary of the honeybee (Apis mellifera L.). *Zoological Science* **24**, 596–603. ISSN 0289-0003. (doi:10.2108/zsj.24.596).

[29] Cardoen, D., Wenseleers, T., Ernst, U. R., Danneels, E. L., Laget, D., DE Graaf, D. C., Schoofs, L. & Verleyen, P., 2011 Genome-wide analysis of alternative reproductive phenotypes in honeybee workers. *Molecular Ecology* **20**, 4070–4084. ISSN 1365-294X. (doi:10.1111/j.1365-294X.2011.05254.x).

[30] Toth, A. L., Tooker, J. F., Radhakrishnan, S., Minard, R., Henshaw, M. T. & Grozinger, C. M., 2014 Shared genes related to aggression, rather than chemical communication, are

associated with reproductive dominance in paper wasps (Polistes metricus). *BMC Genomics* **15**, 75. ISSN 1471-2164. (doi:10.1186/1471-2164-15-75).

[31] Li, J., Wu, J., Begna Rundassa, D., Song, F., Zheng, A. & Fang, Y., 2010 Differential Protein Expression in Honeybee (Apis mellifera L.) Larvae: Underlying Caste Differentiation. *PLoS ONE* **5**, e13455. (doi:10.1371/journal.pone.0013455).

[32] Markesich, D. C., Gajewski, K. M., Nazimiec, M. E. & Beckingham, K., 2000 bicaudal encodes the Drosophila beta NAC homolog, a component of the ribosomal translational machinery*. *Development (Cambridge, England)* **127**, 559–572. ISSN 0950-1991.

[33] Scharf, M. E., Wu-Scharf, D., Pittendrigh, B. R. & Bennett, G. W., 2003 Caste and development-associated gene expression in a lower termite. *Genome Biology* **4**, R62. ISSN 1465-6906. (doi:10.1186/gb-2003-4-10-r62).

[34] Drewell, R. A., Bush, E. C., Remnant, E. J., Wong, G. T., Beeler, S. M., Stringham, J. L., Lim, J. & Oldroyd, B. P., 2014 The dynamic DNA methylation cycle from egg to sperm in the honey bee Apis mellifera. *Development* **141**, 2702–2711. ISSN 0950-1991, 1477-9129. (doi:10.1242/dev.110163).

[35] Glastad, K. M., Hunt, B. G. & Goodisman, M. A., 2014 Evolutionary insights into DNA methylation in insects. *Current Opinion in Insect Science* **1**, 25–30. ISSN 2214-5745. (doi:10.1016/j.cois.2014.04.001).

[36] Yan, H., Bonasio, R., Simola, D. F., Liebig, J., Berger, S. L. & Reinberg, D., 2015 DNA Methylation in Social Insects: How Epigenetics Can Control Behavior and Longevity. *Annual Review of Entomology* **60**, 435–452. ISSN 0066-4170. (doi:10.1146/annurev-ento-010814-020803).

[37] Dean, W., Bowden, L., Aitchison, A., Klose, J., Moore, T., Meneses, J. J., Reik, W. & Feil, R., 1998 Altered imprinted gene methylation and expression in completely ES cell-derived

mouse fetuses: association with aberrant phenotypes. *Development (Cambridge, England)* **125**, 2273–2282. ISSN 0950-1991.

[38] Pardo-Manuel de Villena, F., de la Casa-Espern, E. & Sapienza, C., 2000 Natural selection and the function of genome imprinting: beyond the silenced minority. *Trends in genetics: TIG* **16**, 573–579. ISSN 0168-9525.

[39] Onyango, P., Jiang, S., Uejima, H., Shamblott, M. J., Gearhart, J. D., Cui, H. & Feinberg, A. P., 2002 Monoallelic expression and methylation of imprinted genes in human and mouse embryonic germ cell lineages. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10599–10604. ISSN 0027-8424. (doi:10.1073/pnas.152327599).

[40] Sapienza, C., 2002 Imprinted gene expression, transplantation medicine, and the other human embryonic stem cell. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10243–10245. ISSN 0027-8424. (doi:10.1073/pnas.172384299).

[41] Zhang, Y., Shields, T., Crenshaw, T., Hao, Y., Moulton, T. & Tycko, B., 1993 Imprinting of human H19: allele-specific CpG methylation, loss of the active allele in Wilms tumor, and potential for somatic allele switching. *American Journal of Human Genetics* **53**, 113–124. ISSN 0002-9297.

[42] Pignatta, D., Erdmann, R. M., Scheer, E., Picard, C. L., Bell, G. W. & Gehring, M., 2014 Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *eLife* **3**, e03198. ISSN 2050-084X. (doi:10.7554/eLife.03198).

[43] Remnant, E. J., Ashe, A., Young, P. E., Buchmann, G., Beekman, M., Allsopp, M. H., Suter, C. M., Drewell, R. A. & Oldroyd, B. P., 2016 Parent-of-origin effects on genome-wide DNA methylation in the Cape honey bee (Apis mellifera capensis) may be confounded by allele-specific methylation. *BMC Genomics* **17**. ISSN 1471-2164. (doi:10.1186/s12864-016-2506-8).

[44] Edsgard, D., Iglesias, M. J., Reilly, S.-J., Hamsten, A., Tornvall, P., Odeberg, J. & Emanuelsson, O., 2016 GeneiASE: Detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information. *Scientific Reports* **6**, 21134. ISSN 2045-2322. (doi:10.1038/srep21134).

[45] Wang, X., Werren, J. H. & Clark, A. G., 2016 Allele-Specific Transcriptome and Methylome Analysis Reveals Stable Inheritance and Cis -Regulation of DNA Methylation in Nasonia. *PLOS Biol* **14**, e1002500. ISSN 1545-7885. (doi:10.1371/journal.pbio.1002500).

[46] Wedd, L., Kucharski, R. & Maleszka, R., 2016 Differentially methylated obligatory epialleles modulate context-dependent LAM gene expression in the honeybee Apis mellifera. *Epigenetics* **11**, 1–10. ISSN 1559-2308. (doi:10.1080/15592294.2015.1107695).

[47] Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. & Salzberg, S. L., 2013 TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36. ISSN 1465-6906. (doi:10.1186/gb-2013-14-4-r36).

[48] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–2079. ISSN 1367-4811. (doi:10.1093/bioinformatics/btp352).

[49] Li, H. & Durbin, R., 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754–1760. ISSN 1367-4811. (doi:10.1093/bioinformatics/btp324).

[50] DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M. *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491–498. ISSN 1061-4036. (doi:10.1038/ng.806).

[51] McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.*, 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303. ISSN 1088-9051, 1549-5469. (doi:10.1101/gr.107524.110).

[52] Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T. *et al.*, 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. ISSN 1367-4803, 1460-2059. (doi:10.1093/bioinformatics/btr330).

[53] Takai, D. & Jones, P. A., 2002 Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences* **99**, 3740–3745. ISSN 0027-8424, 1091-6490. (doi:10.1073/pnas.052410099).

[54] Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D., 2012 The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883. ISSN 1367-4803, 1460-2059. (doi:10.1093/bioinformatics/bts034).

[55] Johnson, W. E., Li, C. & Rabinovic, A., 2007 Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127. ISSN 1465-4644, 1468-4357. (doi:10.1093/biostatistics/kxj037).

[56] McCarthy, D. J., Chen, Y. & Smyth, G. K., 2012 Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**, 4288–4297.

[57] Robinson, M. D., McCarthy, D. J. & Smyth, G. K., 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. ISSN 1367-4803, 1460-2059. (doi:10.1093/bioinformatics/btp616).

[58] Quinlan, A. R. & Hall, I. M., 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, 841–842. ISSN 1367-4811. (doi: 10.1093/bioinformatics/btq033).

[59] Harvey, C. T., Moyerbrailean, G. A., Davis, G. O., Wen, X., Luca, F. & Pique-Regi, R., 2014 QuASAR: Quantitative Allele Specific Analysis of Reads. *Bioinformatics* p. btu802. ISSN 1367-4803, 1460-2059. (doi:10.1093/bioinformatics/btu802).

[60] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W. & Lipman, D. J., 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402. ISSN 0305-1048. 17.

[61] Gotz, S., Garca-Gmez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Taln, M., Dopazo, J. & Conesa, A., 2008 High-throughput functional annotation and data mining with the Blast2go suite. *Nucleic Acids Research* **36**, 3420–3435. ISSN 0305-1048, 1362-4962. (doi:10.1093/nar/gkn176).

[62] Supek, F., Bonjak, M., kunca, N. & muc, T., 2011 REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800. (doi:10.1371/journal.pone. 0021800).

[63] Gineikiene, E., Stoskus, M. & Griskevicius, L., 2009 Single Nucleotide Polymorphism-Based System Improves the Applicability of Quantitative PCR for Chimerism Monitoring. *The Journal of Molecular Diagnostics : JMD* **11**, 66–74. ISSN 1525-1578. (doi:10.2353/jmoldx. 2009.080039).

[64] Pfaffl, M. W., 2001 A new mathematical model for relative quantification in real-time RTPCR. *Nucleic Acids Research* **29**, e45–e45. ISSN 0305-1048, 1362-4962. (doi:10.1093/ nar/29.9.e45).

[65] Liu, W. & Saint, D. A., 2002 A New Quantitative Method of Real Time Reverse Transcription Polymerase Chain Reaction Assay Based on Simulation of Polymerase Chain Reaction

513 Kinetics. *Analytical Biochemistry* **302**, 52–59. ISSN 0003-2697. (doi:10.1006/abio.2001.
514 5530).

515 [66] Team, R.-C., 2015. R: A language and environment for statistical computing.

Figure and table legends

Table 1. The eleven of the nineteen monoallelically methylated and expressed genes that returned informative blast hits.

Figure 1. Coverage of three libraries for bicaudal d. Horizontal lines represent available reads for each library over this genomic range (x-axis). The vertical line shows the position of the snp and which genomic library shares the same allele.

Figure 2. GO terms associated with allele specific expression. A summary of the enriched GO terms ($p < 0.05$, based on Blast2Go annotation) found for genes displaying allele specific expression. This figure was produced using Revigo

Table 2. Bees used in each experiment. K refers to Koppert, A to Agralan and Q to the wild caught Leicester queen.

Supporting information legends

Table S1. Nineteen genes showing both monoallelic methylation and monoallelic expression. Blast results and genomic coordinates of the reads from the RNA-seq, MRE-seq and MeDip-seq libraries.

Table S2. Confirmation of single allele expression of nineteen monoallelically expressed genes in twenty-nine previously published transcriptomes. For each of the 19 contigs are the previously published RNA-seq libraries with associated read counts.

Table S3. 555 genes showing allele specific expression in at least three of the 29 previously published RNA-seq libraries. This table details the blast results from both the bumblebee and drosophila genomes and the GO terms associated with the drosophila hits.

538 **Figure S1. Biological coefficient of variation (BCV) of a) raw data, and b) SVA-**
539 **adjusted data for the 29 RNA-seq *Bombus terrestris* libraries**