Monoallelic methylation and allele specific expression in a social insect

Kate D. Lee[1†], Zoë N. Lonsdale[2†*], Maria Kyriakidou[2], Harindra E. Amarasinghe[2], Despina Nathanael[2], Eamonn B. Mallon[2]

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

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

† **Joint first authors**

∗ **E-mail: Corresponding zl107@leicester.ac.uk**

Abstract

Social insects are emerging models for epigenetics. Here we examine a link between monoallelic methylation and monoallelic expression in the bumblebee *Bombus terrestris* using whole methylome and transcriptome analysis. In mammals and flowering plants, genomic imprinting (parent of origin allele specific expression) often involves monoallelic expression and methylation. We found nineteen genes displaying monoallelic methylation and expression. They were enriched for functions to do with social organisation in the social insects. Evolutionary theory predicts that social organisation in the hymenoptera invovles genomic imprinting.

Introduction

Social hymenoptera (ants, bees and wasps) as important emerging models for epigenetics [Glastad et al., 2011, Weiner and Toth, 2012, Welch and Lister, 2014, Yan et al., 2014]. This is based on theoretical predictions for a role for genomic imprinting in their social organisation (e.g. worker reproduction) [Queller, 2003] and on data showing a fundamental role for methylation in their biology [Chittka et al., 2012]. 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 [Haig, 2000]. In mammals and flowering plants, genomic imprinting is often associated with methylation marks passed from parents to offspring [Reik and Walter, 2001]. There is contradictory evidence for the role of methylation on allele specific expression in social insects. Methylation is associated with allele specific expression in a number of loci in the ants *Camponotus floridanus* and *Harpegnathos saltator* [Bonasio et al., 2012]. Other work on the honeybee *Apis mellifera* found no link between potentially imprinted loci and known methylation sites in that species [Kocher et al., 2015].

Given the diversity within the Hymenoptera, especially in social structure, this research area would benefit from an increase in the breadth of the species studied. The recently sequenced genome of the bumblebee, *Bombus terrestris* displays a full complement of genes involved in the methylation system [Sadd et al., 2015]. Our lab has demonstrated that methylation is important in worker reproduction in this bumblebee [Amarasinghe et al., 2014]. We also recently found allele specific expression in *Bombus terrestris* worker reproduction genes using a candidate gene qPCR based approach [Amarasinghe et al., 2015].

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 [Harris et al., 2010]. Methyl-sensitive restriction enzymes can create libraries that are enriched for non-methylated cytosines (MRE-seq) [Harris et al., 2010]. Genes found in both libraries are monoallelically methylated, with the methylated allele being in the MeDIP-seq data and the unmethylated allele in the MRE-seq data [Harris et al., 2010]. 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 also searched more generally for allele specific expression by analysing thirty published RNA-seq libraries from worker bumblebees [Harrison et al., 2015, Riddell et al., 2014].

Results and discussion

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), see supplementary table 1. Monoallelic expression was confirmed in one of these nineteen (*slit homolog 2 protein-like* (AELG01000623.1)) by allele specific qPCR [Amarasinghe et al., 2015]. The allele with a guanine at the snp position had a mean expression of 6.04 ±8.28 (S.D.) 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 amplification (G mean = 422.70 ±507.36, T mean = 1575.17 ±503.02).

Of the nineteen genes, fourteen had the methylated (MeDIP) allele expressed, while five had the unmethylated (MRE-seq) allele expressed (see supplementary table 1). These 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. We found the MeDIP allele of *yippee-like 1* (AELG01001021.1) was expressed. 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 [Kelley et al., 2010]. The MeDIP allele of *slit homolog 2 protein-like* (AELG01000623.1) was expressed. Slit is produced by midline glia in insects and is involved in cell projection during development [Rothberg et al., 1990]. All three human Slits were found to be hypermethylated in hepatocellular carcinoma cell lines [Zheng et al., 2009]. The MeDIP allele of *methionine aminopeptidase 1-like* (AELG01000544.1) was expressed. Methionine aminopeptidases catalyse N-terminal methionine removal [Leszczyniecka et al., 2006]. MAP1D in humans was found to be potentially oncogenic [Leszczyniecka et al., 2006]. The MRE-seq allele of *calmodulin-lysine N-methyltransferase-like* (AELG01003672.1) was expressed in our study. 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 [Magnani et al., 2010].

Six of the eleven genes with informative hits have functions to do with social organisation in the social insects. We found the MRE-seq allele of *Ecdysone receptor* (AELG01000543.1) was expressed. In *Drosophila melanogaster*, ecdysone receptor interacts with ecdysone to activate a series of ecdysteroid genes [Takeuchi et al., 2007]. In honeybees, *Ecdysone receptor* is expressed in the brain mushroom bodies of both workers and queens and ovaries of queens [Takeuchi et al., 2007]. The MeDIP allele of Shaker (AELG01001021.1) was expressed. It is involved in the operation of potassium ion channel. *Shaker* expression was upregulated in sterile versus reproductive honeybee workers [Cardoen et al., 2011]. The MRE-seq allele of *excitatory amino acid transporter 4-like* (AELG01000969.1) was expressed. Excitatory amino acid transporters are neurotransmitter transporters. *Excitatory amino acid transporter 3* expression was upregulated in sterile honeybee workers [Cardoen et al., 2011]. *Excitatory amino acid transporter 1* expression differences were also associated with worker - queen differentiation in the paper wasp *Polistes metricus* [Toth et al., 2014]. The MeDIP allele of *elongation of very long chain fatty acids protein 6-like* (AELG01004467.1) was expressed. The timing of the upregulation of fatty acid metabolism was found to be different in queen and worker honeybees [Li et al., 2010]. The MeDIP allele of *ras GTPase-activating protein nGAP-like* (AELG01004618.1) was expressed in our sample. *Ras GTPase-activating protein 1* was found to be upregulated in reproductive honeybee workers [Cardoen et al., 2011]. It is involved in oocyte meiosis.

We found the MeDIP allele of *bicaudal D-related protein homolog* (AELG01005399.1) to be expressed in our sample. Bicaudal is involved in embryonic pattern formation in *Drosophila* [Markesich et al., 2000]. It is thought to be involved in the differentiation between soldiers and workers in the termite *Reticulitermes flavipes* [Scharf et al., 2003]. Intriguingly, *bicaudal protein D* has been shown to be methylated more in eggs than sperm in honeybees [Drewell et al., 2014]. To our knowledge this is the first link between gamete specific methylation and adult monoallelic methylation found in insects. This would be a key component of any methylation

based imprinting system.

We then looked at these 19 genes in all thirty RNA-seq libraries. If they are monoallelically expressed in these bees, we would find only one allele in a given RNA-seq library. 15 of these 19 genes were confirmed to show monoallelic expression in all 30 RNA-seq libraries, see supplementary table 2. We would also find only one allele if that bee was homozygote. We can not rule out that these fifteen genes just happen to be homozygote in all thirty bees from five different colonies from multiple sources, but this result at least suggests that the finding in the monoallelic analysis can be generalized. The remaining 4 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 30 RNA-seq libraries. 533 SNPs showed allele-specific expression in $\geq$3 of the 30 RNA-seq libraries (supplementary table 3). Blastn against *Bombus terrestris* returned 275 hits, and blastx against *Drosophila melanogaster* returned 301 hits. 127 of the blastx results had GO annotations.

Of particular interest, allele specific expression was found to be present in the *homeobox protein hox-a3-like* gene (AELG01000285.1). *Hoxa3* is thought to be involved in the regulation of gene expression and embryonic development, and is predicted to be maternally imprinted in humans [Luedi et al., 2007, Wang, 2014]. We also found the *paternally-expressed gene 3* (*PEG3*) (AELGO1005183.1) to show allele specific expression. In mammals *PEG3* is highly expressed in the brain, ovaries, and placenta, where it is involved in the transcriptional control of foetal growth [Hiby et al., 2001, Kim et al., 1997, 2013b, Kohda et al., 2001]. *PEG3* shows allele-specific expression in several organs in humans, including the foetal brain [Murphy et al., 2001]. Furthermore, *PEG3* has previously been demonstrated to be maternally imprinted in humans and mice [Kaneko-Ishino et al., 1995, Kuroiwa et al., 1996].

One hundred and seventy Gene ontology(GO) terms were enriched in the 533 regions showing allele specific expression (Fishers exact test). Figure 2 shows the large number of biological functions associated with these 533 genes. It would be expected that most of these genes are

not involved in epigenetic processes. Allele specific expression is known to be caused by a number of genetic as well as epigenetic processes [Palacios et al., 2009]. The genetic processes usually involve transcription factor binding sites, or less often, involve untranslated regions which alter RNA stability or microRNA binding [Farh et al., 2005]. Several potential epigenetically controlled pathways were also present. The ecdysone receptor-mediated signalling pathway was enriched, as were processes involved in gene transcription (e.g. negative regulation of DNA recombination) (Figure 2).

Conclusions

Our results found that genes that are both monallelically methylated and monoallelically expressed are enriched for functions to do with social organisation in social insects. For example several genes showed roles in ecdysone pathways. In various social insect species ecdysteriods have been shown to be involved in ovary activation and dominance hierarchy in workers and queens [Geva et al., 2005]. Theory has predicted that imprinting should be associated with these types of functions [Queller, 2003]. Of particular note is *bicaudal protein D*, the first evidence in insects of a possible link between gamete specific methylation and adult monoallelic methylation. This link would be a key component of any putative methylation based imprinting system.

Materials and Methods

Samples

Data from thirty RNA-seq libraries were used for the allele specific expression analysis (twenty three from Riddell et al. [2014], six from Harrison et al. [2015] and a RNA-seq library from the bee also used for methylation analysis). 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. The bees used for the MeDIP-seq / MRE-seq / RNA-seq experiment and for the qPCR analysis were from separate Koppert colonies. Samples are outlined in Table 1. 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.

Next generation sequencing

MeDIP-seq, MRE-seq and RNA-seq

RNA and DNA was extracted from a single five day old whole bee (Colony K1). 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 [Harris et al., 2010]. 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 [Harrison et al., 2015, Riddell et al., 2014]. 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 [Kim et al., 2013a] and converted to bam files with Samtools [Li et al., 2009]. 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 [Li and Durbin, 2009]. 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

186 locally re-aligned with GATK [DePristo et al., 2011, McKenna et al., 2010]. PCR duplicates for
187 all bams (mRNA, MeDIP and MRE) were marked with the Picard utility Markduplicates.jar.

Identifying regions of interest and integrating data

188 Coverage of each data type was calculated using GATK DepthofCoverage [McKenna et al., 2010].
189 Only regions with a read depth of at least six in each of the libraries (RNA-seq, MeDIP-seq and
190 MRE-seq) was used. Heterozygotes were identified using Samtools mpileup and bcftools on each
191 data set separately [Li and Durbin, 2009] and results were merged with vcf tools [Danecek et al.,
192 2011]. CpG islands were identified using CpG island searcher [Takai and Jones, 2002]. Regions
193 of mRNA with overlaps of MeDIP, MRE, CpG islands and monoallelic snps were identified with
194 custom perl scripts.

Allele specific expression - Bioinformatic analysis

195 We created a pipeline to search for heterozygous loci that show allele-specific expression and iden-
196 tify the associated enriched gene ontology (GO) terms in thirty previously published RNA-seq
197 libraries [Harrison et al., 2015, Riddell et al., 2014]. Each RNA library was mapped to the *Bom-*
198 *bus terrestris* reference genome (AELG00000000) [Sadd et al., 2015] using the BWA mapper [Li
199 and Durbin, 2009].

200 Bcftools (version 0.1.19-44428cd), bedtools (v2.17.0), and samtools (version 0.1.19-44428cd)
201 were used to prepare the RNA libraries and call the SNPs, before the SNPs were filtered based
202 on mapping quality score [Li and Durbin, 2009, Quinlan and Hall, 2010]. Only SNPs with a
203 mapping quality score of $p < 0.05$ and a read depth of ≥ 6 were included in the analyses. The
204 R package, QuASAR, was then used to identify genotypes (according to the Hardy-Weinberg
205 equilibrium), and locate any allele specific expression at heterozygous sites [Harvey et al., 2014].
206 The loci (the snp position +/- 2900bp) identified as showing ASE in at least three of the thirty
207 libraries, were blasted (Blastx) against *Drosophila melanogaster* proteins (non-redundant (nr)
208 database) [Altschul et al., 1997]. The blast results were annotated using Blast2Go [Gotz et al.,

2008]. Fisher's exact test was implemented to identify enriched GO terms, which were then visualised using REVIGO [Supek et al., 2011]. To identify which bumblebee genes the snps were located in, the snp position +/- 25 bp was blasted (Blastn) against the *Bombus terrestris* genome [Sadd et al., 2015].

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 a sample of the head 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 Gineikiene et al. [2009]. A common reverse primer (CTGGTTCC-CGTCCAATCTAA) was used for all three reactions. A reference forward primer (CGTGTCCA-GAATCGACAATG) 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 [Pfaffl, 2001]. 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 Liu and Saint [2002].

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 [Liu and Saint, 2002, Pfaffl, 2001]. All statistical analysis was carried out using R (3.1.0) [Team, 2015].

Acknowledgements

This work was financially supported by NERC grant no. NE/H010408/1 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.

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). GO-analysis results and lists of differentially expressed transcripts are available as Supporting Information.

Author contributions

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

References

S. F. Altschul, T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25:3389--3402, September 1997. ISSN 0305-1048. URL `://A1997XU79300002`. 17.

Harindra Amarasinghe, B.J. Toghill, D Nathanael, and E. B Mallon. Allele specific expression in worker reproduction genes in the bumblebee Bombus terrestris. *PeerJ*, 3:e1079, 2015. doi: https://dx.doi.org/10.7717/peerj.1079. URL `https://peerj.com/articles/1079/`.

Harindra E. Amarasinghe, Crisenthiya I. Clayton, and Eamonn B. Mallon. Methylation and worker reproduction in the bumble-bee (Bombus terrestris). *Proceedings of the Royal Society B: Biological Sciences*, 281(1780):20132502, April 2014. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2013.2502. URL `http://rspb.royalsocietypublishing.org/content/281/1780/20132502`.

Roberto Bonasio, Qiye Li, Jinmin Lian, Navdeep S. Mutti, Lijun Jin, Hongmei Zhao, Pei Zhang, Ping Wen, Hui Xiang, Yun Ding, Zonghui Jin, Steven S. Shen, Zongji Wang, Wen

276 Wang, Jun Wang, Shelley L. Berger, Jrgen Liebig, Guojie Zhang, and Danny Reinberg.
277 Genome-wide and Caste-Specific DNA Methylomes of the Ants Camponotus floridanus and
278 Harpegnathos saltator. *Current Biology*, 22(19):1755--1764, October 2012. ISSN 0960-9822.
279 doi: 10.1016/j.cub.2012.07.042. URL `http://www.cell.com/current-biology/abstract/`
280 `S0960-9822(12)00867-6`.

281 Dries Cardoen, Tom Wenseleers, Ulrich R. Ernst, Ellen L. Danneels, Dries Laget, Dirk C.
282 DE Graaf, Liliane Schoofs, and Peter Verleyen. Genome-wide analysis of alternative repro-
283 ductive phenotypes in honeybee workers. *Molecular Ecology*, 20(19):4070--4084, October 2011.
284 ISSN 1365-294X. doi: 10.1111/j.1365-294X.2011.05254.x.

285 A. Chittka, Y. Wurm, and L. Chittka. Epigenetics: The making of ant castes. *Current Biology*,
286 22(19):R835--R838, 2012. ISSN 0960-9822. doi: 10.1016/j.cub.2012.07.045.

287 Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. De-
288 Pristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry, Gilean
289 McVean, and Richard Durbin. The variant call format and VCFtools. *Bioinformatics*, 27(15):
290 2156--2158, August 2011. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btr330.
291 URL `http://bioinformatics.oxfordjournals.org/content/27/15/2156`.

292 Mark A. DePristo, Eric Banks, Ryan Poplin, Kiran V. Garimella, Jared R. Maguire, Christo-
293 pher Hartl, Anthony A. Philippakis, Guillermo del Angel, Manuel A. Rivas, Matt Hanna,
294 Aaron McKenna, Tim J. Fennell, Andrew M. Kernytsky, Andrey Y. Sivachenko, Kristian
295 Cibulskis, Stacey B. Gabriel, David Altshuler, and Mark J. Daly. A framework for vari-
296 ation discovery and genotyping using next-generation DNA sequencing data. *Nature Ge-*
297 *netics*, 43(5):491--498, May 2011. ISSN 1061-4036. doi: 10.1038/ng.806. URL `http:`
298 `//www.nature.com/ng/journal/v43/n5/full/ng.806.html`.

299 Robert A. Drewell, Eliot C. Bush, Emily J. Remnant, Garrett T. Wong, Suzannah M. Beeler,
300 Jessica L. Stringham, Julianne Lim, and Benjamin P. Oldroyd. The dynamic DNA methy-
301 lation cycle from egg to sperm in the honey bee Apis mellifera. *Development*, 141(13):

302 2702--2711, July 2014. ISSN 0950-1991, 1477-9129. doi: 10.1242/dev.110163. URL `http:`
303 `//dev.biologists.org/content/141/13/2702`.

304 Kyle Kai-How Farh, Andrew Grimson, Calvin Jan, Benjamin P. Lewis, Wendy K. Johnston,
305 Lee P. Lim, Christopher B. Burge, and David P. Bartel. The widespread impact of mam-
306 malian MicroRNAs on mRNA repression and evolution. *Science (New York, N.Y.)*, 310
307 (5755):1817--1821, December 2005. ISSN 1095-9203. doi: 10.1126/science.1121158.

308 Sharon Geva, Klaus Hartfelder, and Guy Bloch. Reproductive division of labor, dominance, and
309 ecdysteroid levels in hemolymph and ovary of the bumble bee Bombus terrestris. *Journal of*
310 *Insect Physiology*, 51(7):811--823, July 2005. ISSN 0022-1910. doi: 10.1016/j.jinsphys.2005.
311 03.009.

312 Egle Gineikiene, Mindaugas Stoskus, and Laimonas Griskevicius. Single Nucleotide
313 Polymorphism-Based System Improves the Applicability of Quantitative PCR for Chimerism
314 Monitoring. *The Journal of Molecular Diagnostics : JMD*, 11(1):66--74, January 2009. ISSN
315 1525-1578. doi: 10.2353/jmoldx.2009.080039. URL `http://www.ncbi.nlm.nih.gov/pmc/`
316 `articles/PMC2607568/`.

317 K. M. Glastad, B. G. Hunt, S. V. Yi, and M. a. D. Goodisman. DNA methylation in insects:
318 on the brink of the epigenomic era. *Insect Molecular Biology*, 20(5):553--565, 2011. ISSN
319 1365-2583. doi: 10.1111/j.1365-2583.2011.01092.x. URL `http://onlinelibrary.wiley.com/`
320 `doi/10.1111/j.1365-2583.2011.01092.x/abstract`.

321 Stefan Gotz, Juan Miguel Garca-Gmez, Javier Terol, Tim D. Williams, Shivashankar H. Nagaraj,
322 Mara Jos Nueda, Montserrat Robles, Manuel Taln, Joaqun Dopazo, and Ana Conesa. High-
323 -throughput functional annotation and data mining with the Blast2go suite. *Nucleic Acids*
324 *Research*, 36(10):3420--3435, June 2008. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkn176.
325 URL `http://nar.oxfordjournals.org/content/36/10/3420`.

326 D. Haig. The kinship theory of genomic imprinting. *Annual Review of Ecology and Systematics*,

31:9--32, 2000.

R Alan Harris, Ting Wang, Cristian Coarfa, Raman P Nagarajan, Chibo Hong, Sara L Downey, Brett E Johnson, Shaun D Fouse, Allen Delaney, Yongjun Zhao, Adam Olshen, Tracy Ballinger, Xin Zhou, Kevin J Forsberg, Junchen Gu, Lorigail Echipare, Henriette O'Geen, Ryan Lister, Mattia Pelizzola, Yuanxin Xi, Charles B Epstein, Bradley E Bernstein, R David Hawkins, Bing Ren, Wen-Yu Chung, Hongcang Gu, Christoph Bock, Andreas Gnirke, Michael Q Zhang, David Haussler, Joseph R Ecker, Wei Li, Peggy J Farnham, Robert A Waterland, Alexander Meissner, Marco A Marra, Martin Hirst, Aleksandar Milosavljevic, and Joseph F Costello. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature biotechnology*, 28(10):1097--1105, October 2010. ISSN 1546-1696. doi: 10.1038/nbt.1682.

M C Harrison, R L Hammond, and E B Mallon. Reproductive workers show queen-like gene expression in an intermediately eusocial insect, the buff-tailed bumble bee Bombus terrestris. *Molecular Ecology*, 24:121--129, April 2015. ISSN 1365-294X. doi: 10.1111/mec.13215. URL `http://onlinelibrary.wiley.com/doi/10.1111/mec.13215/abstract`.

Chris T. Harvey, Gregory A. Moyerbrailean, Gordon O. Davis, Xiaoquan Wen, Francesca Luca, and Roger Pique-Regi. QuASAR: Quantitative Allele Specific Analysis of Reads. *Bioinformatics*, page btu802, December 2014. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btu802. URL `http://bioinformatics.oxfordjournals.org/content/early/2014/12/04/bioinformatics.btu802`.

Susan E. Hiby, Maria Lough, E. Barry Keverne, M. Azim Surani, Yung Wai Loke, and Ashley King. Paternal monoallelic expression of PEG3 in the human placenta. *Human Molecular Genetics*, 10(10):1093--1100, 2001.

Tomoko Kaneko-Ishino, Yoshimi Kuroiwa, Naoki Miyoshi, Takashi Kohda, Rika Suzuki, Minesuke Yokoyama, Stephane Viville, Sheila C. Barton, Fumitoshi Ishino, and M. Azim Surani.

Peg1/Mest imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. *Nature genetics*, 11:52--59, 1995.

Kevin Kelley, Kelly R. Miller, Amber Todd, Amy Kelley, Rebecca Tuttle, and Steven J. Berberich. YPEL3, a p53-regulated gene that induces cellular senescence. *Cancer research*, 70(9):3566--3575, May 2010. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-09-3219. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2862112/`.

Daehwan Kim, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36, April 2013a. ISSN 1465-6906. doi: 10.1186/gb-2013-14-4-r36. URL `http://genomebiology.com/2013/14/4/R36/abstract`.

Joomyeong Kim, Linda Ashworth, Elbert Branscomb, and Lisa Stubbs. The Human Homolog of a Mouse-Imprinted Gene, Peg3, Maps to a Zinc Finger Gene-Rich Region of Human Chromosome 19q13.4. *Genome Research*, 7:532--540, 1997.

Joomyeong Kim, Wesley D. Frey, Hongzhi He, Hana Kim, Muhammed Ekram, Arundhati Bakshi, Mohammed Faisal, Bambarendage P. U. Perera, An Ye, and Ryoichi Teruyama. Peg3 Mutational Effects on Reproduction and Placenta- Specific Gene Families. *PLOS ONE*, 8(12), 2013b.

Sarah D. Kocher, Jennifer M. Tsuruda, Joshua D. Gibson, Christine M. Emore, Miguel E. Arechavaleta-Velasco, David C. Queller, Joan E. Strassmann, Christina M. Grozinger, Michael R. Gribskov, Phillip San Miguel, Rick Westerman, and Greg J. Hunt. A Search for Parent-of-Origin Effects on Honey Bee Gene Expression. *G3 (Bethesda, Md.)*, June 2015. ISSN 2160-1836. doi: 10.1534/g3.115.017814.

Takashi Kohda, Akio Asai, Yoshimi Kuroiwa, Shin Kobayashi, Kohzoh Aisaka, Goro Nagashima, Michihiro C. Yoshida, Yasumitsu Kondo, Naoto Kagiyama, Takaaki Kirino, Tomoko Kaneko-Ishino, and Fumitoshi Ishino. Tumour suppressor activity of human imprinted gene

iqQ in a glioma cell line. *Genes to Cells*, 6:237--247, 2001.

Yoshimi Kuroiwa, Tomoko Kaneko-Ishino, Fusako Kagitani, Takashi Kohda, Li-Lan Li, Masako Tada, Rika Suzuki, Minesuke Yokoyama, Toshihiko Shiroishi, Sigeharu Wakana, Sheila C. Barton, Fumitoshi Ishino, and M. Azim Surani. Peg3 imprinted gene on proximal chromosome 7 encodes for a zince finger protein. *Nature genetics*, 12:186--190, 1996.

M. Leszczyniecka, U. Bhatia, M. Cueto, N. R. Nirmala, H. Towbin, A. Vattay, B. Wang, S. Zabludoff, and P. E. Phillips. MAP1d, a novel methionine aminopeptidase family member is overexpressed in colon cancer. *Oncogene*, 25(24):3471--3478, June 2006. ISSN 0950-9232. doi: 10.1038/sj.onc.1209383.

Heng Li and Richard Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14):1754--1760, July 2009. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp324.

Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25 (16):2078--2079, August 2009. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp352.

Jianke Li, Jing Wu, Desalegn Begna Rundassa, Feifei Song, Aijuan Zheng, and Yu Fang. Differential Protein Expression in Honeybee (Apis mellifera L.) Larvae: Underlying Caste Differentiation. *PLoS ONE*, 5(10):e13455, October 2010. doi: 10.1371/journal.pone.0013455. URL `http://dx.doi.org/10.1371/journal.pone.0013455`.

Weihong Liu and David A. Saint. A New Quantitative Method of Real Time Reverse Transcription Polymerase Chain Reaction Assay Based on Simulation of Polymerase Chain Reaction Kinetics. *Analytical Biochemistry*, 302(1):52--59, March 2002. ISSN 0003-2697. doi: 10.1006/abio.2001.5530. URL `http://www.sciencedirect.com/science/article/pii/S0003269701955307`.

Philippe P. Luedi, Fred S. Dietrich, Jennifer R. Weidman, Jason M. Bosko, Randy L. Jirtle, and Alexander J. Hartemink. Computational and experimental identification of novel human imprinted genes. *Genome Research*, 17:1723--1730, 2007.

Roberta Magnani, Lynnette M. A. Dirk, Raymond C. Trievel, and Robert L. Houtz. Calmodulin methyltransferase is an evolutionarily conserved enzyme that trimethylates Lys-115 in calmodulin. *Nature Communications*, 1:43, July 2010. doi: 10.1038/ncomms1044. URL `http://www.nature.com/ncomms/journal/v1/n4/full/ncomms1044.html`.

D. C. Markesich, K. M. Gajewski, M. E. Nazimiec, and K. Beckingham. bicaudal encodes the Drosophila beta NAC homolog, a component of the ribosomal translational machinery*. *Development (Cambridge, England)*, 127(3):559--572, February 2000. ISSN 0950-1991.

Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, and Mark A. DePristo. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297--1303, September 2010. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.107524.110. URL `http://genome.cshlp.org/content/20/9/1297`.

Susan K. Murphy, Andrew A. Wylie, and Randy L. Jirtle. Imprinting of PEG3, the Human Homologue of a Mouse Gene Involved in Nurturing Behavior. *Genomics*, 71:110--117, 2001.

Ricardo Palacios, Elodie Gazave, Joaqun Goi, Gabriel Piedrafita, Olga Fernando, Arcadi Navarro, and Pablo Villoslada. Allele-Specific Gene Expression Is Widespread Across the Genome and Biological Processes. *PLoS ONE*, 4(1):e4150, January 2009. doi: 10.1371/journal.pone.0004150. URL `http://dx.doi.org/10.1371/journal.pone.0004150`.

Michael W. Pfaffl. A new mathematical model for relative quantification in real-time RTPCR. *Nucleic Acids Research*, 29(9):e45--e45, May 2001. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/29.9.e45. URL `http://nar.oxfordjournals.org/content/29/9/e45`.

D. C. Queller. Theory of genomic imprinting conflict in social insects. *Bmc Evolutionary Biology*, 3:art. no.--15, 2003. URL ://000188122100015.

Aaron R. Quinlan and Ira M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6):841--842, March 2010. ISSN 1367-4811. doi: 10.1093/bioinformatics/btq033.

W. Reik and J. Walter. Genomic imprinting: Parental influence on the genome. *Nature Reviews Genetics*, 2:21--32, January 2001. ISSN 1471-0056. URL ://000166056000014. 1.

Carolyn E. Riddell, Juan D. Lobaton Garces, Sally Adams, Seth M. Barribeau, David Twell, and Eamonn B. Mallon. Differential gene expression and alternative splicing in insect immune specificity. *BMC Genomics*, 15(1):1031, November 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-1031. URL http://www.biomedcentral.com/1471-2164/15/1031/comments.

J. M. Rothberg, J. R. Jacobs, C. S. Goodman, and S. Artavanis-Tsakonas. slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. *Genes & Development*, 4(12a):2169--2187, December 1990. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.4.12a.2169. URL http://genesdev.cshlp.org/content/4/12a/2169.

Ben M. Sadd, Seth M. Barribeau, Guy Bloch, Dirk C. de Graaf, Peter Dearden, Christine G. Elsik, Jrgen Gadau, Cornelis JP Grimmelikhuijzen, Martin Hasselmann, Jeffrey D. Lozier, Hugh M. Robertson, Guy Smagghe, Eckart Stolle, Matthias Van Vaerenbergh, Robert M. Waterhouse, Erich Bornberg-Bauer, Steffen Klasberg, Anna K. Bennett, Francisco Cmara, Roderic Guig, Katharina Hoff, Marco Mariotti, Monica Munoz-Torres, Terence Murphy, Didac Santesmasses, Gro V. Amdam, Matthew Beckers, Martin Beye, Matthias Biewer, Mrcia MG Bitondi, Mark L. Blaxter, Andrew FG Bourke, Mark JF Brown, Severine D. Buechel, Rossanah Cameron, Kaat Cappelle, James C. Carolan, Olivier Christiaens, Kate L. Ciborowski, David F. Clarke, Thomas J. Colgan, David H. Collins, Andrew G. Cridge, Tamas

452 Dalmay, Stephanie Dreier, Louis du Plessis, Elizabeth Duncan, Silvio Erler, Jay Evans, Tiago
453 Falcon, Kevin Flores, Flvia CP Freitas, Taro Fuchikawa, Tanja Gempe, Klaus Hartfelder,
454 Frank Hauser, Sophie Helbing, Fernanda C. Humann, Frano Irvine, Lars S. Jermiin, Claire E.
455 Johnson, Reed M. Johnson, Andrew K. Jones, Tatsuhiko Kadowaki, Jonathan H. Kidner,
456 Vasco Koch, Arian Khler, F. B. Kraus, H. Mg Lattorff, Megan Leask, Gabrielle A. Lock-
457 ett, Eamonn B. Mallon, David S. Marco Antonio, Monika Marxer, Ivan Meeus, Robin FA
458 Moritz, Ajay Nair, Kathrin Npflin, Inga Nissen, Jinzhi Niu, Francis MF Nunes, John G.
459 Oakeshott, Amy Osborne, Marianne Otte, Daniel G. Pinheiro, Nina Rossi, Olav Rueppell,
460 Carolina G. Santos, Regula Schmid-Hempel, Bjrn D. Schmitt, Christina Schulte, Zil LP
461 Simes, Michelle PM Soares, Luc Swevers, Eva C. Winnebeck, Florian Wolschin, Na Yu,
462 Evgeny M. Zdobnov, Peshtewani K. Aqrawi, Kerstin P. Blankenburg, Marcus Coyle, Liezl
463 Francisco, Alvaro G. Hernandez, Michael Holder, Matthew E. Hudson, LaRonda Jackson,
464 Joy Jayaseelan, Vandita Joshi, Christie Kovar, Sandra L. Lee, Robert Mata, Tittu Mathew,
465 Irene F. Newsham, Robin Ngo, Geoffrey Okwuonu, Christopher Pham, Ling-Ling Pu, Ne-
466 had Saada, Jireh Santibanez, DeNard Simmons, Rebecca Thornton, Aarti Venkat, Kim-
467 berly KO Walden, Yuan-Qing Wu, Griet Debyser, Bart Devreese, Claire Asher, Julie Blom-
468 maert, Ariel D. Chipman, Lars Chittka, Bertrand Fouks, Jisheng Liu, Meaghan P. ONeill,
469 Seirian Sumner, Daniela Puiu, Jiaxin Qu, Steven L. Salzberg, Steven E. Scherer, Donna M.
470 Muzny, Stephen Richards, Gene E. Robinson, Richard A. Gibbs, Paul Schmid-Hempel, and
471 Kim C. Worley. The genomes of two key bumblebee species with primitive eusocial organiza-
472 tion. *Genome Biology*, 16(1):76, April 2015. ISSN 1465-6906. doi: 10.1186/s13059-015-0623-3.
473 URL `http://genomebiology.com/2015/16/1/76/abstract`.

474 Michael E. Scharf, Dancia Wu-Scharf, Barry R. Pittendrigh, and Gary W. Bennett. Caste
475 and development-associated gene expression in a lower termite. *Genome Biology*, 4(10):
476 R62, September 2003. ISSN 1465-6906. doi: 10.1186/gb-2003-4-10-r62. URL `http:`
477 `//genomebiology.com/2003/4/10/R62/abstract`.

478 Fran Supek, Matko Bonjak, Nives kunca, and Tomislav muc. REVIGO Summarizes and Vi-

sualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, 6(7):e21800, July 2011. doi: 10.1371/journal.pone.0021800. URL http://dx.doi.org/10.1371/journal.pone.0021800.

Daiya Takai and Peter A. Jones. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences*, 99(6):3740--3745, March 2002. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.052410099. URL http://www.pnas.org/content/99/6/3740.

Hideaki Takeuchi, Rajib Kumar Paul, Emiko Matsuzaka, and Takeo Kubo. EcR-A expression in the brain and ovary of the honeybee (Apis mellifera L.). *Zoological Science*, 24(6):596--603, June 2007. ISSN 0289-0003. doi: 10.2108/zsj.24.596.

R-Core Team. R: A language and environment for statistical computing., 2015.

Amy L. Toth, John F. Tooker, Srihari Radhakrishnan, Robert Minard, Michael T. Henshaw, and Christina M. Grozinger. Shared genes related to aggression, rather than chemical communication, are associated with reproductive dominance in paper wasps (Polistes metricus). *BMC Genomics*, 15(1):75, January 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-75. URL http://www.biomedcentral.com/1471-2164/15/75/abstract.

Hong-Dan Wang. DNA methylation study of fetus genome through a genome-wide analysis. *BMC Medical Genomics*, 7(18), 2014. URL http://www.biomedcentral.com/1755-8794/7/18.

Susan A Weiner and Amy L Toth. Epigenetics in social insects: a new direction for understanding the evolution of castes. *Genetics research international*, 2012:609810, 2012. ISSN 2090-3162. doi: 10.1155/2012/609810.

Mat Welch and Ryan Lister. Epigenomics and the control of fate, form and function in social insects. *Current Opinion in Insect Science*, 1:31--38, July 2014. ISSN 2214-5745. doi: 10.1016/j.cois.2014.04.005. URL http://www.sciencedirect.com/science/article/pii/S2214574514000066.

504 Hua Yan, Daniel F. Simola, Roberto Bonasio, Jrgen Liebig, Shelley L. Berger, and Danny
505 Reinberg. Eusocial insects as emerging models for behavioural epigenetics. *Nature Reviews*
506 *Genetics*, 15(10):677--688, October 2014. ISSN 1471-0056. doi: 10.1038/nrg3787. URL `http:`
507 `//www.nature.com/nrg/journal/v15/n10/abs/nrg3787.html`.

508 Dan Zheng, Bin-Bin Liu, Yin-Kun Liu, Xiao-Nan Kang, Lu Sun, Kun Guo, Rui-Xia Sun, Jie
509 Chen, and Yan Zhao. Analysis of the expression of Slit/Robo genes and the methylation status
510 of their promoters in the hepatocellular carcinoma cell lines. *Zhonghua Gan Zang Bing Za*
511 *Zhi = Zhonghua Ganzangbing Zazhi = Chinese Journal of Hepatology*, 17(3):198--202, March
512 2009. ISSN 1007-3418.

Figures

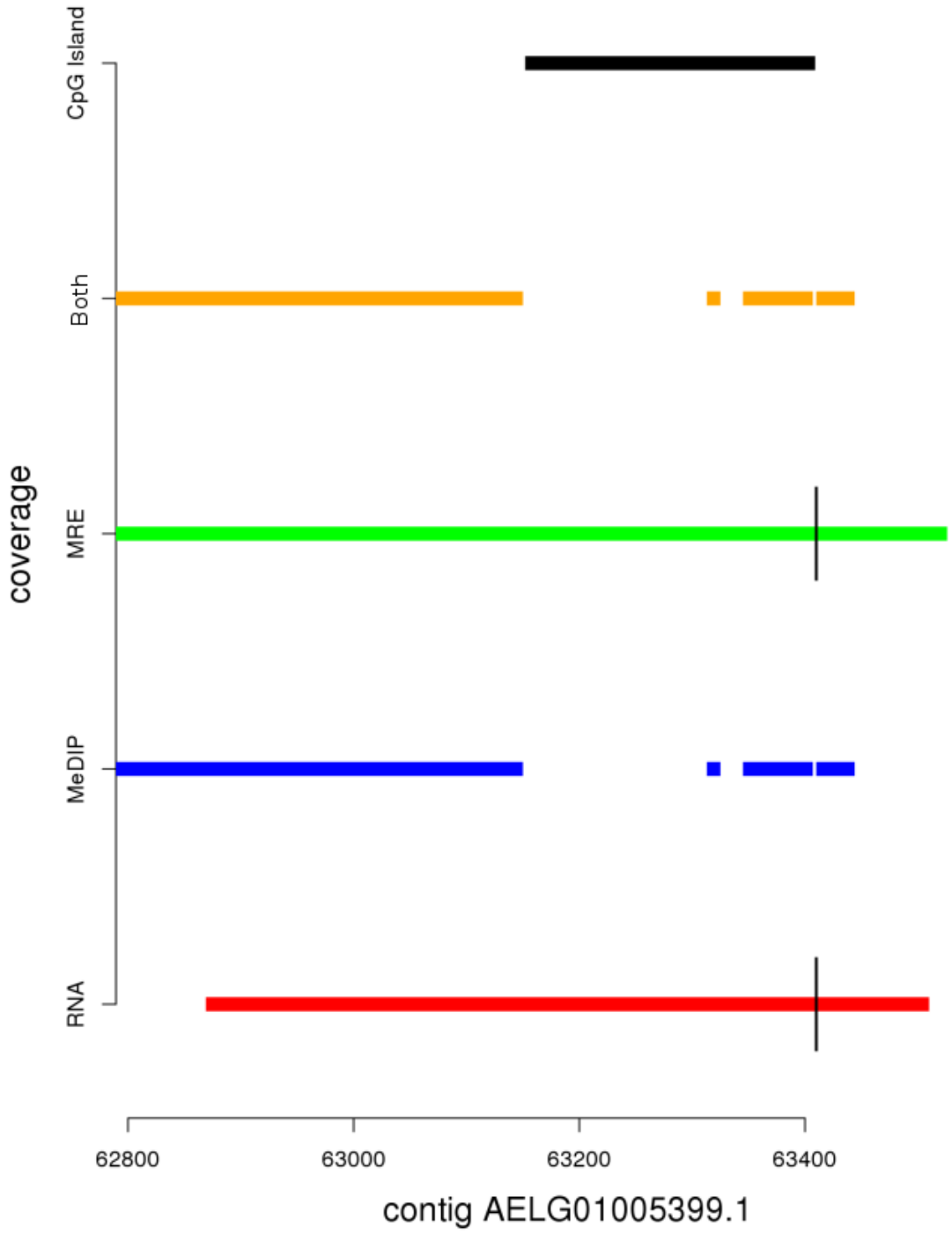

Figure 1. Coverage of three libraries for bicaudal d. Horizontal lines represent available reads for each library over this genomic range (x-axis). The vertical line shows the position of the snp and which genomic library shares the same allele.

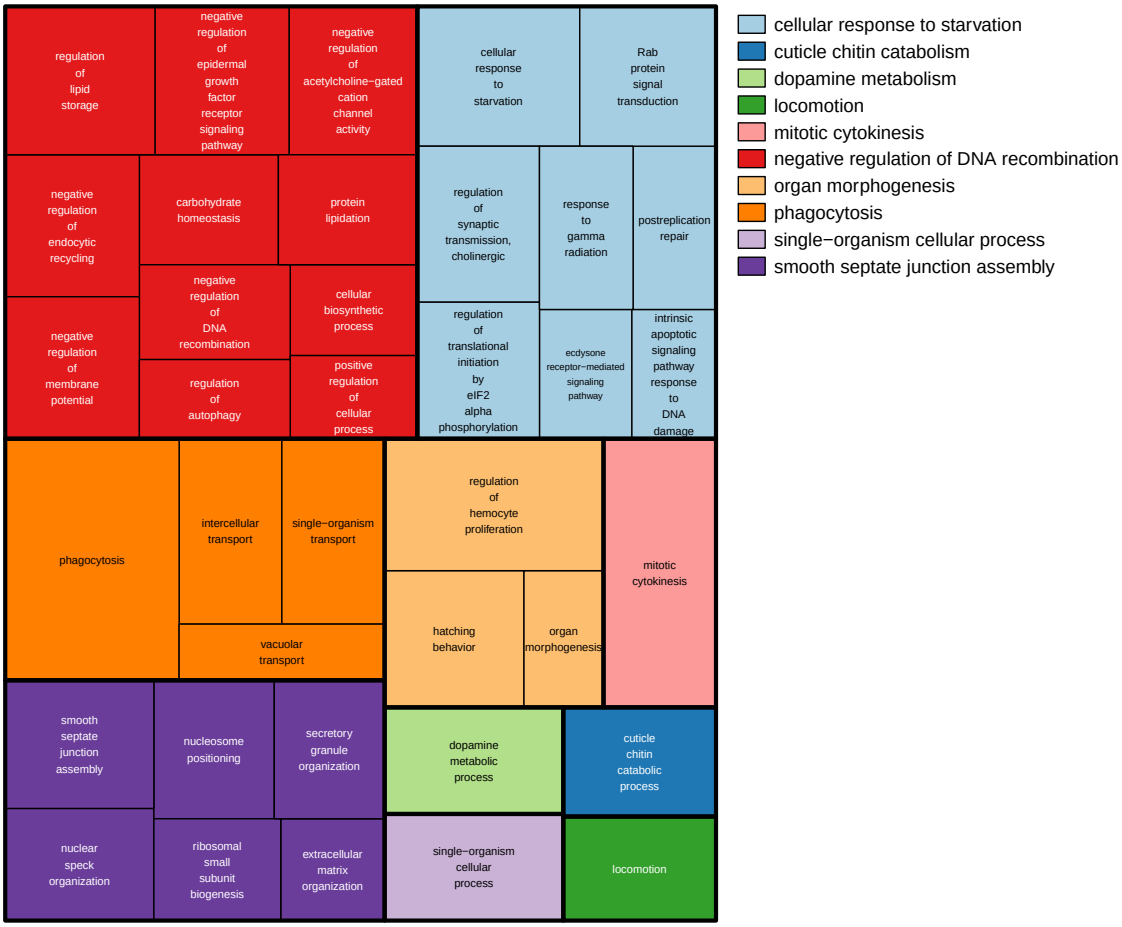

Figure 2. GO terms associated with allele specific expression. A summary of the enriched GO terms ($p < 0.05$, based on Blast2Go annotation) found for genes displaying allele specific expression. This figure was produced using Revigo

Tables

Table 1. 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	2	A1
	2	A2
	2	A3
	1	K1
	14	K2
	9	Q1
MeDip/MRE/RNA-seq	1	K1
qPCR	2	K3
	1	K4
	1	K5

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 thirty previously published transcriptomes. For each of the 19 contigs are the previously published RNA-seq libraries with associated read counts.

Table S3. 533 genes showing allele specific expression in at least three of the 30 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.