

UNIVERSITY OF LEICESTER

DOCTORAL THESIS

---

# Epigenetic mechanisms of insect polyphenisms

---

*Author:*

Zoë Natasha LONSDALE

*Supervisor:*

Associate Professor

Eamonn MALLON

*A thesis submitted in fulfilment of the requirements  
for the degree of Doctor of Philosophy*

*in the*

Department of Genetics and Genome Biology

January 2018

UNIVERSITY OF LEICESTER

# *Abstract*

Department of Genetics and Genome Biology

Doctor of Philosophy

## **Epigenetic mechanisms of insect polyphenisms**

by Zoë Natasha LONSDALE

Insects are emerging as a key lineage for the study of epigenetic phenomena. This is due to the variety of polyphenisms found in insects. In this thesis, the caste polymorphism of the buff-tailed bumblebee *Bombus terrestris* and the phase polymorphism of the desert locust *Schistocerca gregaria* are studied to elucidate the underlying epigenetic mechanisms.

I establish the presence of allele-specific expression and methylation in *B. terrestris*. I used next-generation RNA-sequencing to establish the DNA methylation, alternative splicing, and gene expression patterns of *B. terrestris* worker reproduction. The presence of allele-specific methylation and allele-specific expression were then determined in the same context. Correlations with the aforementioned epigenetic mechanisms were drawn. One major finding was that a higher degree of methylation was witnessed in more highly expressed genes. Higher methylation levels were also associated with more differentially expressed genes and isoforms between workers of a different reproductive state. However, the association between allele-specific expression and allele-specific methylation was weak.

The relationship between alternative splicing and the circadian clock in *S. gregaria* was investigated. The first evidence of genes with differential circadian isoform expression patterns is reported. Finally, I analysed whether genome-wide alternative splicing levels are an important component in ascertaining the varying levels of eusociality found in the Hymenoptera. Fewer splicing events per gene with multiple isoforms was found in more highly eusocial species compared with solitary and more primitively eusocial species. Thus this is the first evidence of an association between level of sociality and alternative splicing.

# *Acknowledgements*

First and foremost, I sincerely thank my supervisor Associate Professor Eamonn Mallon for all his advice and guidance throughout my PhD. I would like to express my gratitude to Dr Rob Hammond for his advice at various stages of my studies. Thanks to Lisa and Ceinwen for their technical support. To Harindra, Hollie, Boris, Alun, Mirko, Jacob, Pshtiwan, and everyone in lab 219, I am grateful for our conversations, your advice, and the enjoyment you've provided in being a part of the Social Epigenetics group.

Thanks to my parents and my brother Sam; your love and support is invaluable. Dad, thank you for always motivating and encouraging me to do the best I can. To my mum, thank you for always believing in me, and Sam your wisdom knows no bounds! I owe my deepest gratitude to the friends near and far that have encouraged and supported me through difficult times and in my research: Liane, Beka, Judy, Kayla, and the cake group.

Finally, I am grateful to the Biotechnology and Biological Sciences Research Council for the financial support and training they have provided via the Midlands Integrative Biosciences Training Partnership.

“We choose to go to the Moon in this decade and do the other things, not because they are easy, but because they are hard!” – John F. Kennedy

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>xiii</b>
<b>Abbreviations</b>	<b>xv</b>
<b>1 General introduction</b>	<b>1</b>
1.1 Polyphenisms . . . . .	1
1.1.1 The evolution of polyphenisms . . . . .	4
1.2 Regulation of polyphenisms . . . . .	6
1.2.1 Epigenetic mechanisms in polyphenisms . . . . .	8
1.2.2 DNA methylation: a regulator of polyphenisms in insects . . .	11
1.2.3 DNA methylation: a mediator of alternative splicing . . . . .	13
1.3 Bumblebee caste polyphenism . . . . .	15
1.4 Phase polyphenism in the desert locust . . . . .	17
1.5 Thesis aims and objectives . . . . .	19
<b>2 Searching for allele-specific expression in <i>Bombus terrestris</i></b>	<b>21</b>
2.1 Introduction . . . . .	21
2.2 Materials and Methods . . . . .	23
2.2.1 Samples . . . . .	24
2.2.2 Next generation sequencing . . . . .	26
2.2.2.1 MeDIP-seq, MRE-seq and RNA-seq . . . . .	26
2.2.2.2 Previously published RNA-seq . . . . .	26
2.2.3 Monoallelic methylation and expression - Bioinformatic analysis	27
2.2.3.1 Alignment to genome . . . . .	27
2.2.3.2 Identifying regions of interest and integrating data .	27
2.2.4 Patrigenic monoallelic expression - Bioinformatic analysis . . .	28



---

2.2.5	Allele-specific expression - Bioinformatic analysis . . . . .	29
2.2.6	Candidate gene allele-specific qPCR . . . . .	30
2.3	Results . . . . .	32
2.3.1	Monoallelic methylation and monoallelic expression . . . . .	32
2.3.2	Confirmation of monoallelic expression by qPCR . . . . .	33
2.3.3	Paternal monoallelic expression . . . . .	34
2.3.4	Allele-specific expression . . . . .	36
2.4	Discussion . . . . .	38
<b>3</b>	<b>Alternative splicing, DNA methylation, and gene expression in <i>Bombus terrestris</i></b>	<b>41</b>
3.1	Introduction . . . . .	41
3.2	Methods . . . . .	44
3.2.1	Bee husbandry and tissue sampling . . . . .	44
3.2.2	RNA and DNA extraction and sequencing . . . . .	44
3.2.3	Identification of differential isoforms and clustering . . . . .	46
3.2.4	Differential expression . . . . .	47
3.2.5	DNA methylation analyses . . . . .	47
3.2.6	Gene ontology annotation . . . . .	48
3.2.7	Comparisons: alternative splicing, methylation, and expression	48
3.3	Results . . . . .	49
3.3.1	RNA-seq and BS-seq mapping efficiencies . . . . .	49
3.3.2	Differential gene expression . . . . .	49
3.3.3	Differential splicing of isoforms . . . . .	55
3.3.4	Differential methylation patterns . . . . .	61
3.3.5	Global expression and methylation patterns . . . . .	64
3.4	Discussion . . . . .	72
<b>4</b>	<b>Allele-specific expression and methylation in reproductive and non-reproductive <i>Bombus terrestris</i> workers.</b>	<b>77</b>
4.1	Introduction . . . . .	77
4.2	Methods . . . . .	79
4.2.1	Samples . . . . .	79
4.2.2	Allele-specific expression analysis . . . . .	79
4.2.3	Allele-specific methylation analysis . . . . .	80
4.2.4	Comparisons: allele-specific methylation and expression . . . .	81
4.3	Results . . . . .	82
4.3.1	Allele-specific expression . . . . .	82
4.3.2	Allele-specific methylation . . . . .	87
4.3.3	Comparisons: allele-specific methylation and expression . . . .	87
4.4	Discussion . . . . .	92

<b>5</b>	<b>Differential circadian isoform expression patterns in the desert locust, <i>Schistocerca gregaria</i></b>	<b>97</b>
5.1	Introduction . . . . .	97
5.2	Methods . . . . .	100
5.2.1	Locust husbandry . . . . .	100
5.2.2	Tissue samples and sequencing . . . . .	101
5.2.3	Alternative splicing analysis . . . . .	102
5.3	Results . . . . .	104
5.4	Discussion . . . . .	114
<b>6</b>	<b>Alternative splicing and eusociality in the Hymenoptera</b>	<b>117</b>
6.1	Introduction . . . . .	117
6.2	Methods . . . . .	119
6.2.1	Transcriptome selection . . . . .	119
6.2.2	Metrics of alternative splicing . . . . .	121
6.2.3	Differential isoform expression analysis . . . . .	123
6.2.4	Splicing event type and gene ontology . . . . .	125
6.3	Results . . . . .	125
6.3.1	Relative splicing proportions . . . . .	125
6.3.1.1	Metric 1: Proportion of multi-isoform genes . . . . .	126
6.3.1.2	Metric 2: Mean isoform number per multi-isoform gene	128
6.3.1.3	Metric 3: Number of splicing events per multi-isoform gene . . . . .	129
6.3.2	Splicing event type . . . . .	131
6.3.3	Gene ontology . . . . .	132
6.4	Discussion . . . . .	141
<b>7</b>	<b>General discussion</b>	<b>145</b>
7.1	A summary of the results . . . . .	145
7.2	Future implications . . . . .	149
7.2.1	Genomic imprinting link . . . . .	149
7.2.2	Methylation: driver, effector, or response? . . . . .	153
7.2.3	Future directions . . . . .	155
<b>8</b>	<b>Conclusions</b>	<b>158</b>
	<b>Bibliography</b>	<b>159</b>

# List of Figures

1.1	The diversity of animal polyphenisms. (A) <i>C. elegans</i> nematode life-history [1]. (B) Second juvenile stage of predator-exposed <i>Daphnia pulex</i> with neckteeth [2]. (C) Snowshoe hare coat colour polyphenism (Cooley Jericho Community Forest 2017). (D) Female and male clownfish (Georgette Douwma, Arkive).	2
1.2	Reaction norm examples. (A) Genotypes A and B both have plastic responses to the environmental variable, but they show opposite phenotypes in response to the same environmental cue [adapted from 3]. The slope of the genotype lines indicates the degree of plasticity. (B) Reaction norm of a polyphenism with Genotype C that has a discrete phenotypic switch at a threshold level of the environmental variable.	5
1.3	<i>Bombus terrestris</i> queen (A) and worker (B) individuals, with silhouettes to show difference in body size.	16
1.4	Morphology of juvenile <i>Schistocerca gregaria</i> desert locusts in the solitary and gregarious phases (Photograph: Compton Tucker, NASA GSFC)	18
2.1	Flow chart of methods for identification of monoallelic methylation and monoallelic expression in <i>Bombus terrestris</i> for a given gene. Blue boxes indicate methylation related methods. Green boxes refer to methods for expression.	24
2.2	Coverage of the RNA-seq, MRE-seq, and MeDIP-seq libraries for <i>ras GTPase-activating protein nGAP-like</i> (LOC100652225). The transcript models come from GCF_000214255.1_Bter_1.0. The y-axis in the coverage plots is $\log(1 + \text{coverage})$ . The red vertical line represents the heterozygote position. The MeDIP-seq allele was expressed in this locus	33
2.3	Base substitutions of the 555 SNPs showing allele-specific expression in at least three of twenty-nine <i>Bombus terrestris</i> transcriptome libraries. The base substitution notation represents the reference base and the alternate base at the polymorphic locus. For example, "G-C" represents a base substitution of guanine to cytosine.	37

2.4	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 [4]. Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment. . . . .	38
3.1	Ovaries of dissected (A) reproductive and (B) non-reproductive <i>Bombus terrestris</i> workers . . . . .	45
3.2	Ovary weight of <i>Bombus terrestris</i> workers immediately after dissection. Red points indicate outliers. . . . .	45
3.3	Largest oocyte length of dissected <i>Bombus terrestris</i> workers . . . . .	46
3.4	Principal component analysis plot based on the 12,072 genes from all 18 <i>Bombus terrestris</i> transcriptome samples, merging replicates of the same colony and reproductive status. Each point represents a gene. R refers to reproductive workers, and N refers to non-reproductive workers. C1, C5, and C8 are the three different <i>B. terrestris</i> colonies. For example, C1R refers to reproductive workers from colony 1. . . .	50
3.5	Principal component analysis plot based on the 12,072 genes from all 18 <i>Bombus terrestris</i> transcriptome samples, with independent replicates. Each point represents a gene. R refers to reproductive workers, and N refers to non-reproductive workers. C1, C5, and C8 are the three different <i>B. terrestris</i> colonies. 0, 1, and 2 refer to the three different replicates for each sample type. For example, C1R_0 refers to the first reproductive worker sample from colony 1. . . . .	51
3.6	MA plot based on the 12,072 genes from all 18 <i>Bombus terrestris</i> transcriptome samples. Red data points show genes which are significantly differentially expressed between reproductive and non-reproductive workers ( $FDR < 0.05$ ) [5, 6]. Red data points with $\log FC > 0$ are up-regulated in reproductive workers compared with non-reproductive workers. Red data points with $\log FC < 0$ are up-regulated in non-reproductive workers compared with reproductive workers. . . . .	52
3.7	Significantly enriched gene ontology terms up-regulated in reproductive workers (REVIGO) [4]. Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment. . . . .	53

3.8	Significantly enriched gene ontology terms up-regulated in non-reproductive workers (REVIGO) [4]. Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment. . . . .	54
3.9	Correlation of expression of multiple-isoform loci based on Jenson-Shannon distance, showing merged replicates. Lower Jenson-Shannon values indicate greater similarity of expression patterns. R denotes reproductive workers, and NR indicates non-reproductive workers. . .	55
3.10	Correlation of expression of multiple-isoform loci based on Jenson-Shannon distance, showing independent replicates. Lower Jenson-Shannon values indicate greater similarity of expression patterns. R denotes reproductive workers, and NR indicates non-reproductive workers. 0, 1, and 2 refer to the three different replicates for each sample type. . . . .	56
3.11	Principal component analysis of the isoform expression of multiple-isoform loci, merging replicates. R refers to reproductive workers, and N refers to non-reproductive workers. C1, C5, and C8 are the three different <i>B. terrestris</i> colonies. For example, C1R refers to reproductive workers from colony 1. . . . .	57
3.12	Significantly differentially expressed isoforms between reproductive and non-reproductive workers within the same colony. Total of 103 significantly differentially expressed isoforms. . . . .	58
3.13	Coverage (total number of reads) of the four multi-isoform genes with differentially expression between reproductive and non-reproductive workers of the three <i>B. terrestris</i> colonies. . . . .	59
3.14	Enriched gene ontology terms of the genes of the 103 differentially expressed isoforms between reproductive and non-reproductive workers within the same colony (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment. . . . .	60
3.15	Enriched gene ontology terms of the differentially methylated genes in reproductive and non-reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment. . . . .	62
3.16	Mean methylation levels (mean methylated reads per cytosine position) of genomic features in reproductive and non-reproductive workers. Error bars are 95% confidence intervals. . . . .	65

3.17	Methylation level (percentage of CpGs methylated) versus expression level (log-transformed FPKM) for (A) reproductive and (B) non-reproductive workers. Each point represents a gene. . . . .	66
3.18	Gene body methylation (percentage of whole gene methylated) versus binned gene expression (Number of total bins = 100, where genes in Bin 0 are least expressed and those in Bin 100 are most highly expressed). Lines were generated using the LOESS local polynomial regression method implemented by <i>ggplot2</i> in R [7]. . . . .	67
3.19	Mean methylation (percentage of CpGs methylated) levels of differentially and non-differentially expressed genes. Error bars are 95% confidence intervals. . . . .	68
3.20	Differential expression (absolute logFC) and binned DNA methylation (percentage of CpGs methylated) (bins=10, where 1 is least methylated, 10 is most highly methylated, and 0 represents unmethylated genes) Error bars are 95% confidence intervals. . . . .	69
3.21	Mean methylation (percentage of CpGs methylated) levels of differentially and non-differentially expressed multi-isoform genes. Error bars are 95% confidence intervals. . . . .	70
3.22	Mean expression (FPKM) levels of differentially and non-differentially expressed multi-isoform genes. Error bars are 95% confidence intervals. . . . .	70
4.1	Enriched gene ontology terms of the genes with allele-specific expression in (A) reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment. . . . .	83
4.2	Enriched gene ontology terms of the genes with allele-specific expression in non-reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment. . . . .	84
4.3	Enriched gene ontology terms of the genes with allele-specific expression <i>only</i> in reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment. . . . .	85

4.4	Enriched gene ontology terms of the genes with allele-specific expression <i>only</i> in non-reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment. . . . .	86
4.5	Enriched gene ontology terms of the genes with allele-specific methylation (A) up-regulated in reproductive workers and (B) down-regulated in reproductive workers compared to non-reproductive workers (created using REVIGO [4]). . . . .	88
5.1	Principal Coordinate Analysis (PCoA) of all genes with differentially expressed isoforms. Notation: “t1a_2” corresponds to time point 1, locust replicate a, and isoform 2. Red represents isoform 1, and blue to isoform 2. . . . .	104
5.2	PCA of the 105 multi-isoform genes with circadian expression patterns and different optimal phase times. Proportion of variance: PC1 = 0.9356, PC2 = 0.0247 . . . . .	105
5.3	Heatmap of 105 genes at eight time points with log-adjusted significantly differentially expressed isoforms. Notation: “t1a” corresponds to timepoint 1, and locust replicate a. . . . .	106
5.4	Congruent expression patterns over 24 hours, based on k-means clustering for the 210 isoforms from the 105 differentially expressed multi-isoform genes. Cl1-9 denote the nine clusters. Different y axis scale bars are used in each sub-figure. . . . .	107
5.5	Genes with differential circadian isoform expression patterns. Different y axis scale bars are used in each sub-figure. Clusters (Cl) determined by k-means clustering of the 105 differentially expressed multi-isoform genes. . . . .	108
5.6	Genes with mapping against the migratory locust <i>L. migratoria</i> with differential circadian isoform expression. Y axis scales are different in each sub-figure. . . . .	109
5.7	Genes with mapping against the German cockroach <i>B. germanica</i> with differential circadian isoform expression. Y axis scales are different in each sub-figure. . . . .	109
5.8	Genes with mapping against the termite <i>Z. nevadensis</i> with differential circadian isoform expression. Y axis scales are different in each sub-figure. . . . .	110
6.1	Origins of eusociality in the Hymenoptera and termites [8]. Red lines indicate evolution of eusocial status. . . . .	120
6.2	Phylogeny of hymenopteran species with diverse social structures used in this study based on NCBI taxonomy (phyloT) [9]. . . . .	121

6.3	Flow chart of methodology to determine the pervasiveness of alternative splicing with regards to social structure in six species of Hymenoptera. <i>ape</i> and <i>cummeRbund</i> are packages in R. . . . .	122
6.4	Phylogeny and percentage of total genes with alternative splicing after controlling for coverage. See Appendix F: Table F.1 for image references.	127
6.5	Phylogenetic Independent Contrast (PIC) adjusted proportion of multi-isoform genes and social index. PIC adjusted multi-isoform genes with phylogenetic distances based on (A) <i>elongation factor 1 alpha F2</i> ( <i>EF1 alpha F2</i> ), and (B) <i>leucine tRNA ligase</i> . Phylogenies generated using (C) <i>EF1 alpha F2</i> (D) <i>leucine tRNA ligase</i> distance matrices with PIC scaled proportion of multi-isoform genes (blue) and PIC scaled social index (red). The axes of scaled index of sociality is calculated from the sociality index and phylogenetic distance matrices using Felsenstein's method [10]. . . . .	128
6.6	Phylogeny and mean number of isoforms per multi-isoform gene. See Appendix F: Table F.1 for image references. . . . .	129
6.7	Phylogeny and mean number of splicing events per multi-isoform gene. See Appendix F: Table F.1 for image references. . . . .	130
6.8	PIC adjusted mean number of splicing events per multi-isoform gene and social index. PIC adjusted multi-isoform genes with phylogenetic distances based on (A) <i>EF1 alpha F2</i> , and (B) <i>leucine tRNA ligase</i> . Phylogenies generated using (C) <i>EF1 alpha F2</i> and (D) <i>leucine tRNA ligase</i> distance matrices with PIC scaled proportion of multi-isoform genes (blue) and PIC scaled social index (red). The axes of scaled index of sociality is calculated from the sociality index and phylogenetic distance matrices using Felsenstein's method [10]. . . . .	131
6.9	Stacked barplot of the proportion of alternative splicing events in six Hymenopteran species. Abbreviations: A3 = alternative 3' splice site, A5 = alternative 5' splice site, ATSS = alternative transcription start site, ATTSS = alternative transcription termination site, ESI = exon skipping/inclusion, ISI = intron skipping/inclusion, MEE = mutually exclusive exons, MESI = multiple exon skipping/inclusion. Amel = <i>Apis mellifera</i> , Bter = <i>Bombus terrestris</i> , Ccin = <i>Cephus cinctus</i> , Hsal = <i>Harpegnathos saltator</i> , Mpha = <i>Monomorium pharaonis</i> , Pbar = <i>Pogonomyrmex barbatus</i> . . . . .	132
6.10	Significantly enriched gene ontology (GO) terms of the multi-isoform genes of <i>A. mellifera</i> in comparison with the reference GO terms of <i>Drosophila melanogaster</i> (Fisher's test: $p < 0.05$ ) (REVIGO)[4]. . . . .	134
6.11	Significantly enriched gene ontology (GO) terms of the multi-isoform genes of <i>B. terrestris</i> in comparison with the reference GO terms of <i>Drosophila melanogaster</i> (Fisher's test: $p < 0.05$ ) (REVIGO)[4]. . . . .	135



6.12	Significantly enriched gene ontology (GO) terms of the multi-isoform genes of <i>C. cinctus</i> in comparison with the reference GO terms of <i>Drosophila melanogaster</i> (Fisher's test: $p < 0.05$ ) (REVIGO)[4]. . . . .	136
6.13	Significantly enriched gene ontology (GO) terms of the multi-isoform genes of <i>H. saltator</i> in comparison with the reference GO terms of <i>Drosophila melanogaster</i> (Fisher's test: $p < 0.05$ ) (REVIGO)[4]. . . . .	137
6.14	Significantly enriched gene ontology (GO) terms of the multi-isoform genes of <i>M. pharaonis</i> in comparison with the reference GO terms of <i>Drosophila melanogaster</i> (Fisher's test: $p < 0.05$ ) (REVIGO)[4]. . . . .	138
6.15	Significantly enriched gene ontology (GO) terms of the multi-isoform genes of <i>P. barbatus</i> in comparison with the reference GO terms of <i>Drosophila melanogaster</i> (Fisher's test: $p < 0.05$ ) (REVIGO)[4]. . . . .	139
6.16	Overlap of enriched GO terms (Fisher test: $p < 0.05$ ) of multi-isoform genes (from Figures 6.10, 6.11, 6.12, 6.13, 6.14, and 6.15) between the six hymenopteran species. The orange bar represents GO terms of alternatively spliced genes that are significantly enriched in all six species. . . . .	140
6.17	Gene ontology terms of multi-isoform genes that are significantly enriched in all six Hymenopteran focal species (Fisher's test: $p < 0.05$ ) (REVIGO) [4]. . . . .	141
7.1	<i>Bombus terrestris</i> pedigree. $r$ = coefficient of relatedness (adapted from Queller 2003 [11]) . . . . .	152

# List of Tables

2.1	Bees used in each experiment. Regarding the different colonies: K refers to Koppert, A to Agralan and Q to the wild caught Leicester queen. . . . .	25
2.2	Number of whole body samples used for paternal monoallelic expression analysis from Harrsion et al. [12]. The colonies were obtained from Agralan Ltd, hence the colony notation of A1, A2, and A3. . . .	28
2.3	The eleven of the nineteen monoallelically methylated and monoallelically expressed genes that returned informative blastx hits compared with the <i>Drosophila melanogaster</i> nr database. . . . .	35
2.4	Genotypes of the eight Harrison et al. [12] RNA-seq libraries at candidate loci identified in Section 2.3.1 to show monoallelic expression and monoallelic methylation. . . . .	36
3.1	Multi-isoform genes that are differentially expressed between reproductive and non-reproductive workers in three <i>B. terrestris</i> colonies. .	58
3.2	The top ten differentially methylated genes between reproductive and non-reproductive bumblebee workers. Gene function was inferred from conserved protein domains identified on NCBI. . . . .	63
3.3	Comparing all known genes (LOC IDs) with differential expression, methylation, and alternative splicing with the hypergeometric test. .	71
3.4	Comparing enriched Gene Ontology (GO) terms with differential expression, methylation, and alternative splicing with the hypergeometric test. . . . .	71
4.1	Comparing all known genes (LOC IDs) with allele-specific expression, allele-specific methylation, and overall differential expression, methylation, and alternative splicing (see Chapter 3) with the hypergeometric test. Genes are referred to as up/down regulated or hyper-/hypo-methylated in the reproductive workers (Rep.), compared to the non-reproductive workers (Non-rep.). . . . .	89

4.2	Comparing enriched Gene Ontology (GO) terms with allele-specific expression, allele-specific methylation, and overall differential expression, methylation, and alternative splicing (see Chapter 3) with the hypergeometric test. Genes are referred to as up/down regulated or hyper-/hypo-methylated in the reproductive workers (Rep.), compared to the non-reproductive workers (Non-rep.). . . . .	90
4.3	Characterised genes in multiple significantly overlapping enriched gene lists with allele-specific expression, allele-specific methylation, differential expression, isoform expression, and methylation (hypergeometric test, $p < 0.05$ ). . . . .	92
4.4	Gene ontology terms in over ten out of twenty-six significantly overlapping enriched GO lists with allele-specific expression, allele-specific methylation, differential expression, isoform expression, and methylation (hypergeometric test, $p < 0.05$ ). . . . .	92
5.1	Gregarious <i>Schistocerca gregaria</i> forskål locust samples (m = male, f = female). . . . .	101
5.2	Genes with differential isoform circadian expression in the gregarious phase of the desert locust <i>Schistocerca gregaria</i> and their gene ontology terms (blastx with the migratory locust <i>Locusta migratoria</i> ). Phase denotes the time at which isoforms have peak expression levels (Isoform 1 : Isoform 2). . . . .	111
5.3	Genes with differential isoform circadian expression in the gregarious phase of the desert locust <i>Schistocerca gregaria</i> and their gene ontology terms (blastx with the cockroach <i>Blattella germanica</i> ). Phase denotes the time at which isoforms have peak expression levels (Isoform 1 : Isoform 2). . . . .	112
5.4	Genes with differential isoform circadian expression in the gregarious phase of the desert locust <i>Schistocerca gregaria</i> and their gene ontology terms (blastx with the termite <i>Zootermopsis nevadensis</i> ). Phase denotes the time at which isoforms have peak expression levels (Isoform 1 : Isoform 2). . . . .	113
6.1	Hymenopteran reference genomes and worker caste transcriptomes accessed via the European Bioinformatics Institute database [13]. . .	123

# Abbreviations

<b>A</b>	<b>A</b> denine
<b>ANCOVA</b>	<b>A</b> nalysis of <b>Co</b> - <b>v</b> ariance
<b>ANOVA</b>	<b>A</b> nalysis of <b>V</b> ariance
<b>ASE</b>	<b>A</b> llele-specific <b>e</b> xpression
<b>ASM</b>	<b>A</b> llele-specific <b>m</b> ethylation
<b>ATP</b>	<b>A</b> denosine <b>t</b> riphosphate
<b>BCV</b>	biological <b>c</b> oefficient of <b>v</b> ariation
<b>BGI</b>	<b>B</b> eijing <b>G</b> enomics <b>I</b> nstitute
<b>BS-seq</b>	<b>B</b> isulphite sequencing
<b>bp</b>	<b>B</b> ase <b>p</b> airs (of nucleic acid)
<b>C</b>	<b>C</b> ytosine
<b>°C</b>	Degrees <b>C</b> entigrade
<b>cDNA</b>	<b>c</b> omplementary <b>D</b> N
<b>CpG</b>	<b>C</b> ytosine <b>p</b> hosphate <b>G</b> uanine sequence within <b>D</b> N
<b>CRISPR</b>	<b>C</b> lustered <b>R</b> egularly <b>I</b> nterspaced <b>S</b> hort <b>P</b> alindromic <b>R</b> epeats
<b>ddH<sub>2</sub>O</b>	<b>D</b> ouble <b>d</b> istilled water
<b>DEG</b>	<b>D</b> ifferentially <b>e</b> xpressed <b>g</b> enes
<b>DEI</b>	<b>D</b> ifferentially <b>e</b> xpressed multi- <b>i</b> soform <b>g</b> enes
<b>df</b>	Degrees of <b>f</b> reedom
<b>DMR</b>	<b>D</b> ifferentially <b>m</b> ethylated <b>r</b> egion
<b>DNA</b>	<b>D</b> eoxyribonucleic <b>a</b> cid
<b>DNMT</b>	<b>D</b> N

---

<b>FDR</b>	<b>F</b> alse <b>d</b> iscovery <b>r</b> ate
<b>FPKM</b>	<b>F</b> ragments <b>P</b> er <b>K</b> ilobase of transcript per <b>M</b> illion mapped reads
<b>g</b>	<b>g</b> rams
<b>G</b>	<b>G</b> uanine
<b>GC-content</b>	<b>G</b> uanine- <b>C</b> ytosine content
<b>GLM</b>	<b>G</b> eneral linear <b>m</b> odel
<b>GO</b>	<b>G</b> ene <b>o</b> ntology
<b>GRIP</b> domain	<b>g</b> olgin-97, <b>R</b> anBP2alpha, <b>I</b> mh1p and <b>p</b> 230/golgin-245 domain
<b>JS</b> distance	<b>J</b> enson- <b>S</b> hannon distance
<b>JTK</b> algorithm	<b>J</b> onckheere- <b>T</b> erpstra- <b>K</b> endall algorithm
<b>logFC</b>	log <b>f</b> old- <b>c</b> hange
<b>MDS</b>	<b>M</b> ultidimensional scaling
<b>MeDIP-seq</b>	<b>M</b> ethylated <b>D</b> N <b>A</b> <b>i</b> mmunoprecipitation sequencing
<b>miRNA</b>	<b>m</b> icro <b>R</b> N <b>A</b>
<b>mm</b>	<b>M</b> illimetres
$\mu$ l	<b>M</b> icrolitres
$\mu$ m	<b>M</b> icrom <b>e</b> tres
$\mu$ M	<b>M</b> icrom <b>o</b> les
<b>MRE-seq</b>	<b>M</b> ethylation-sensitive <b>r</b> estriction <b>e</b> nzyme sequencing
<b>mRNA</b>	<b>m</b> essenger <b>R</b> N <b>A</b>
<b>ncRNA</b>	<b>n</b> on- <b>c</b> oding <b>R</b> N <b>A</b>
<b>ng</b>	<b>N</b> anograms
<b>nr</b>	<b>n</b> on- <b>r</b> edundant genomic (database)
<b>PBS</b>	<b>P</b> hosphate- <b>b</b> uffered <b>S</b> aline
<b>PCA</b>	<b>P</b> rincipal component <b>a</b> nalysis
<b>PCR</b>	<b>P</b> olymerase chain <b>r</b> eaction
<b>PIC</b>	<b>P</b> hylogenetic <b>I</b> ndependent <b>C</b> ontrast
<b>piRNA</b>	<b>P</b> iwi-interacting <b>R</b> N <b>A</b>
<b>qPCR</b>	quantitative <b>P</b> olymerase <b>C</b> hain <b>R</b> eaction

---

<b>QuASAR</b>	<b>Q</b> uantitative <b>A</b> llele <b>S</b> pecific <b>A</b> nalysis of <b>R</b> eads
<b><i>r</i></b>	relatedness
<b>REVIGO</b>	<b>R</b> educe and <b>V</b> isualise <b>G</b> ene <b>O</b> ntology
<b>RNA</b>	<b>R</b> ibonucleic acid
<b>RNAi</b>	<b>R</b> NA interference
<b>RNA-seq</b>	<b>R</b> NA sequencing
<b>SNP</b>	<b>S</b> ingle <b>N</b> ucleotide <b>P</b> olymorphism
<b>T</b>	<b>T</b> hymine
<b>TF</b>	<b>T</b> ranscription <b>F</b> actor
<b>ZT</b>	<b>Z</b> eitgeber time
<b>5mC</b>	<b>5-methylcytosine</b>

## Collaborative work statement

### **Chapter 2: Searching for allele-specific expression in *Bombus terrestris*.**

This chapter includes independent project work by myself (paternal expression analysis), and also details work from Lonsdale et al. [14] which was published in PeerJ on 13th September 2017.

Of this manuscript, Harindra E. Amarasinghe and Despina Nathanael carried out the initial experiments (DNA and RNA extraction, and qPCR). Kate D. Lee aligned the MeDIP-seq, MRE-seq, and RNA-seq datasets. The initial bioinformatic analysis of the 29 libraries was carried out by Maria Kyriakidou. Catherine O'Connor created one of the figures. I analysed the 29 RNA-seq libraries for both allele-specific expression and confirmation of monoallelic expression for 19 candidate genes, and annotated the results. The manuscript was written by Eamonn B. Mallon and myself.

### **Chapter 3: Alternative splicing, DNA methylation, and gene expression in *Bombus terrestris*.**

The RNA extraction, differential expression analysis, and alternative splicing analysis were carried out by myself. The DNA extraction and differential methylation analysis were carried out by Hollie Marshall. The bee husbandry, dissections, and comparative analysis were completed together by Hollie and myself. Vagelis Ladas (Bioinformatics Group, Department of Genetics and Genome Biology, University of Leicester) assisted in the formatting of a data table in preparation for the linear regression comparative analysis.

### **Chapter 4: Allele-specific expression and methylation in reproductive and non-reproductive *Bombus terrestris* workers.**

The same samples were used here as in Chapter 3. The allele-specific expression analysis was carried out by myself. The allele-specific methylation analysis was

executed by Hollie Marshall. Alun Jones created the custom annotation database. Comparative analyses were jointly undertaken by myself and Hollie.

**Chapter 5: Differential circadian isoform expression patterns in the desert locust, *Schistocerca gregaria***

*Schistocerca gregaria* forskål locust samples were previously obtained, extracted, and sequenced by Jonathan Shand. I conducted all the data analysis.

**Chapter 6: Alternative splicing and eusociality in the Hymenoptera.**

All analysis was conducted by myself.



# Chapter 1

## General introduction

### 1.1 Polyphenisms

Ernst Mayr first coined the term “polyphenism” in 1963, to denote non-genetic and genetic variation of phenotype [15]. There is great diversity in polyphenism type and the organisms in which they develop, from nematodes to more complex mammalian species. In a relatively simple animal model, nematodes such as *Caenorhabditis elegans* exhibit a stress-induced polyphenism. In scenarios of high temperature or reduced food availability which is often combined with overcrowding with conspecifics, L1 and L2 larval nematodes will develop into the “dauer” stage which is long-lived, stress-resistant, and non-feeding [16, 17] (Figure 1.1a). This is an alternate state to the L3 larval stage before a nematode continues through the L4 larval stage and finally to become a reproductive adult.

Predator-induced polyphenisms are shown by water flea *Daphnia* species, whereby head crests or neckteeth are produced by the animal when chemoreceptors in the first antennae pair detect kairomones of a predator nearby [18] (Figure 1.1b). The induced development of neckteeth in *Daphnia pulex* tend to cause predation by the phantom midge larvae *Chaoborus* to be difficult [19]. This phenotypic plasticity is

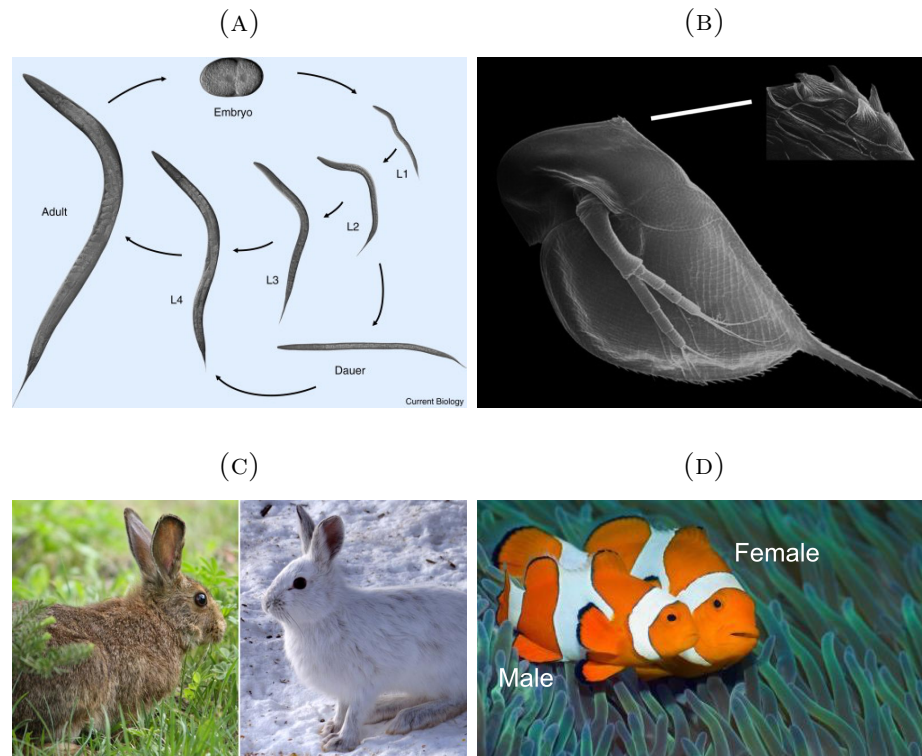


FIGURE 1.1: The diversity of animal polyphenisms. (A) *C. elegans* nematode life-history [1]. (B) Second juvenile stage of predator-exposed *Daphnia pulex* with neckteeth [2]. (C) Snowshoe hare coat colour polyphenism (Cooley Jericho Community Forest 2017). (D) Female and male clownfish (Georggete Douwma, Arkive).

thought to be regulated by the endocrine system, particularly the juvenoid signalling pathway, however the full pathway is as yet unknown [20].

Polyphenisms are also witnessed in higher order Metazoans. Seasonal coat colour polyphenisms occur in at least nine mammalian species [21], such as the Arctic fox *Vulpes lagopus* and the snowshoe hare *Lepus americanus* (Figure 1.1c). In these Arctic species coat colour is white in the winter and brown in the summer, thus assisting the animals to be cryptic to the snow and snow-free woodland/tundra in which they inhabit in different seasons. This polyphenism is key to an individual's survival. The Arctic fox evolved at least 2.6 million years ago (potentially 5 million years ago) [22], but now climate change threatens the survival of these Arctic species if their coat colour polyphenisms are not plastic enough to account for the reduction

in snow fall in Arctic regions. Natural populations of snowshoe hares have limited plasticity in the initiation dates of colour change and the rate of the fall brown-to-white molt [21]. Therefore the correct timing of coat colour polyphenisms are key to survival in these species.

Species' fitness can be maximised by increases in reproduction as well as longevity, and polyphenisms have been witnessed to affect both these fitness aspects. Effects on survival have been described here in the nematode, *Daphnia*, and Arctic mammalian species, and reproduction benefits can be clearly witnessed in sex-determining polyphenisms of several fish species. Functional hermaphroditism is a sex-determining polyphenism that occurs in at least 27 families of *teleost* fish species such as the clownfish [23]. In the monogamous clownfish *Amphiprion percula*, there is a dominance hierarchy determined by size. The female is the dominant fish in a social group also consisting of a large male and smaller immature juveniles (Figure 1.1d). In the event of the death of the female, the largest male will change sex and the largest of the juveniles will mature and become part of the breeding pair [24–26]. This is an example of sequential hermaphroditism and it has been deemed to improve adaptation, increase survival rates, and enhance reproduction [27]. This process is caused by differential gene expression in the brain two weeks after female disappearance, particularly in genes encoding sex steroid hormones. This alteration is followed by expression changes in the gonads from 24 days after female disappearance [23, 28].

The examples I have briefly described consider just a few of the diverse extant polyphenisms of the Metazoa. A great extent of polyphenisms are enclosed in the class Insecta including life history stage polyphenisms (larval/adult stages), seasonal polyphenisms (e.g. morphological changes in lepidoptera), caste polyphenism (worker/queen castes in eusocial insects), and density-dependent polyphenism (phase polyphenism in locusts) [29–31]. There is currently estimated to be around 10 million insect species [32, and references therein], and it has been argued that the great

extent of polyphenisms insects possess hugely contributes to the evolutionary success of the lineage [30].

### 1.1.1 The evolution of polyphenisms

Alternate phenotypes evolve as a response to variation in external selective pressures [33–36]. The concept of a reaction norm is useful in modelling the relationship between an organisms' phenotype and environmental cues in the case of phenotypic plasticity [37]. A reaction norm refers to the range of phenotypes of a particular genotype when exposed to a range of environmental conditions and is typically represented graphically [38]. Figure 1.2a is an example of a representation of a reaction norm showing how a species' phenotype alters according to its genotype and environment. A steeper angle of a genotype line shows a greater degree of plasticity. Thus, for a highly plastic genotype smaller changes in the environmental variable will lead to larger changes in phenotype. Reaction norms are therefore considered to be useful in understanding the selection pressures and evolution of phenotypic plasticity and also the introduction of novel traits [37, 38].

When reaction norms become discrete and phenotypes do not gradually change with small environmental changes (Figure 1.2b), polyphenisms (such as those described in Section 1.1) occur [39]. Consequently, a discrete phenotypic switch will be present whereby an organism's phenotype is fixed until an environmental cue reaches a threshold level that will initiate the development alternative phenotype. In many insects there is a period during larval development when there is a heightened sensitivity to environmental cues such as photoperiod, temperature, crowding, and changes in food quality [39]. This environmental sensitivity then signals regulatory pathways of polyphenisms discussed in Section 1.2 which induces different phenotypes.

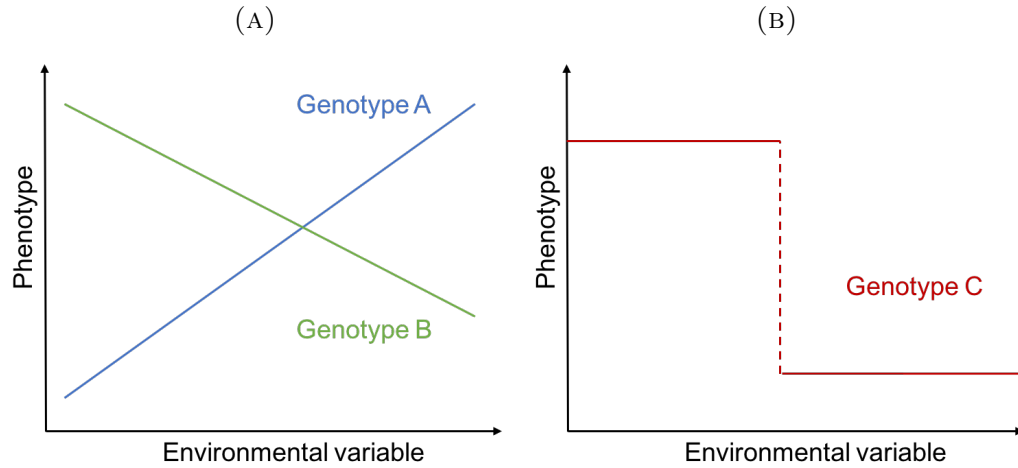


FIGURE 1.2: Reaction norm examples. (A) Genotypes A and B both have plastic responses to the environmental variable, but they show opposite phenotypes in response to the same environmental cue [adapted from 3]. The slope of the genotype lines indicates the degree of plasticity. (B) Reaction norm of a polyphenism with Genotype C that has a discrete phenotypic switch at a threshold level of the environmental variable.

Genetic accommodation has been suggested as a way in which the switch to polyphenisms has evolved [40]. Genetic accommodation is when a “novel phenotype is introduced through a mutation or environmental change is moulded into an adaptive phenotype through quantitative genetic changes” [37, 41]. Thus, it increases the responsiveness of the phenotype to changes in environmental conditions. For example, Suzuki and Nijhout found that a mutation in the control of hormone titer of the tobacco hornworm (*Manduca sexta*) altered the larval colour phenotypic threshold and therefore uncovered a temperature-dependent black larval morph as a second phenotype to the wildtype green larval morph [40]. This study was experimental compared to previous purely observational work, hence causation could be deduced. This temperature-dependent phenotype switch could be represented in a similar format to the reaction norm in Figure 1.2b.

## 1.2 Regulation of polyphenisms

Regulatory genes are often described as key to the evolution of polyphenisms, and it is primarily by this route that polyphenisms are proposed to have evolved in the Hymenoptera [42–45]. Thus, rather than the formation of new genes, it is the differential regulation of the same genes that is a key instigator of polyphenisms. The regulation of genes that determines the initiation and maintenance of polyphenisms can be controlled by genetic factors, the endocrine system, and epigenetic processes [39]. For instance, genetic factors largely determine sexual dimorphism in mammals, and these differential phenotypes are maintained through the endocrine system (with androgens, estrogens, and progestins) [46, 47].

In animals, the endocrine system is important for the control and regulation of genes and therefore different phenotypes [46]. The endocrine system consists of glands that secrete hormones into the circulatory system which then act on target cells [48]. As mentioned above (Section 1.1), *Daphnia* phenotypic plasticity is determined by the juvenoid hormone signalling pathway [20], and the functional hermaphroditism of *teleost* fish (Section 1.1) is indicated to be controlled through differential expression of sex steroid hormones in the brain and gonads [23, 28].

The insect endocrine system (in particular the juvenile hormone, ecdysteroids, and the insulin-signalling pathway) plays an important role in the development and control of different morphologies [46]. In the tobacco hornworm a mutation in the control of hormone titer determines the temperature-dependent threshold between green (wildtype) and black larval morphs [40]. In the desert locust *Schistocerca gregaria* there are two forms (solitary and gregarious) in a phase polyphenism (see Section 1.4). The phase of a locust has been indicated to be transgenerationally inherited involving endocrine regulation. Female locusts raised alone produce a greater number of smaller eggs compared with female locusts raised in crowded conditions [49, 50]. This trend is thought to be associated with ecdysteroid levels in the ovaries. Ovarian ecdysteroid levels are up to four times higher in female

desert locusts raised in crowded conditions with conspecifics compared to that of female locusts raised alone [51]. These ecdysteroids are then found in the developing eggs and have been reported to affect molting during embryonic development [52]. Consequently, this is thought to direct the offsprings' phase-specific behaviour [46, and references therein].

Hormone levels clearly have a significant impact on the initiation and regulation of polyphenisms. Epigenetic mechanisms such as DNA methylation and histone modifications (discussed in Section 1.2.1) can also have large effects in determining distinct morphologies in Metazoan polyphenisms. The term epigenetics was derived from the Greek word “epigenesis” which literally translates as over, outside of, or around genetics. First coined by Waddington in 1942, this terminology originally described how genetic processes affect development [53]. However, the widely accepted modern definition now states epigenetics to be the study of changes of gene expression, and hence phenotype, not due to changes in the underlying DNA sequence [54].

Epigenetics has far reaching effects on polyphenisms, including the regulation of endocrine-related gene expression [55]. Epigenetic mechanisms are responsible for the variety of often reversible polyphenisms involving the endocrine system as a response to environmental stimuli. For instance, DNA methylation is known to regulate key enzymes of hormone biosynthesis. Differential methylation of CpGs near the transcriptional start sites to three steroidogenesis enzymes (CYP11A1, HSD3B1, and CYP19A) was found in bovine follicles [56]. This study also indicated DNA methylation to have a role in the silencing of CYP19A1. Moreover, histone modifications have been indicated to regulate the expression of the steroidogenic acute regulatory (STAR) protein that regulates cholesterol entry to the mitochondria, and influences steroidogenesis [57].

Polyphenism regulation by epigenetic processes have been identified in numerous taxa and can have large effects on an organism's biology. An epigenetic switch

for body weight control has recently been identified in mammals, indicating a polyphenism in energy metabolism. Dalgaard and colleagues found mice with a mutation in the KRAB-zinc-finger transcription factor Trim28 to show a bimodal distribution in body weight [58]. These adult mice could switch between two phenotypes of normal and increased body weight. This was associated with differential expression of paternally imprinted genes between mice with these two phenotypes. These findings indicate the Trim28 gene to affect susceptibility to obesity in humans.

### 1.2.1 Epigenetic mechanisms in polyphenisms

The variety of polyphenisms witnessed in insects (as described in Section 1.1) enables this lineage to be ideal for the study of epigenetics [59]. Epigenetics plays key roles in insect biology including development, aging, disease, and cell differentiation [60]. Epigenetics in insects has been key to our understanding of human diseases, for instance the fruit fly *Drosophila* has been used as a model organism in the study of the underlying mechanisms of the pathogenesis of Huntington's disease [59, 61]. This use of insects to elucidate disease-associated epigenetic patterns is possible due to the underlying molecular mechanisms being evolutionarily conserved across taxa [59]. DNA methylation, histone modification, and microRNAs are the most common epigenetic markers which are highly conserved in insects as well as in mammals and plants [60]. DNA methylation and alternative splicing are the mechanisms which will be the focus of this thesis (described in Sections 1.2.2 and 1.2.3). However, other epigenetic processes can also have a large effect on phenotype.

Acetylation, methylation, phosphorylation, and ubiquitylation of histones have all been reported to affect gene transcription [60]. Post-translational histone modifications affect protein binding properties and chromatin accessibility, which consequently regulates gene expression [62–68]. Furthermore, histone modifications have a close relationship with DNA methylation in mammals [summarised in 69]. Similarly a pattern between DNA methylation and histone modification, primarily on



the third histone of the nucleosome, has been found in the honeybee and a few ant species [70, 71]. However, this relationship is a positive correlation in insects, as opposed to the negative correlation witnessed in mammals [71]. A classic example of insect histone-mediated regulation is of worker/queen development in the honey bee, whereby royal jelly that contains phenyl butyrate (an inhibitor of histone deacetylase) is fed to queen-destined larvae [72]. The phenyl butyrate induces histone modifications which in turn cause transcriptional reprogramming, changes in DNA methylation, and queen development.

Another epigenetic mechanism is that of non-coding RNAs which regulate gene expression post-transcription [73]. MicroRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), and long non-coding RNAs (lncRNAs) are three prominent examples of non-coding RNA. Firstly, miRNAs are short RNA sequences (2123 bp) that can bind to mRNAs thus preventing mRNA translation [74]. Previous studies have identified miRNAs to be partly responsible for caste determination within the Hymenoptera. miRNAs have been demonstrated to be differentially expressed in queen- and worker-destined larvae in the honey bee *Apis mellifera* and bumblebee *Bombus terrestris*, although the associated miRNAs were not conserved between these two species [73, 75–77]. miRNAs are also indicated to be involved in the regulation of an aphid polyphenism. Legeai and colleagues identified differential expression of miRNAs between three morphs: sexual females, parthenogenetic females that produce parthenogenetic female offspring, and parthenogenetic females that produce sexual males and females [78].

Second, piRNAs direct piwi proteins to transposon targets and also affect the DNA methyltransferases (DNMTs) that silence transposons via de novo DNA methylation [79]. PiRNAs are required for spermatogenesis across taxa, and the silencing of transposons appears to be important in embryonic development in mammals [79, 80]. Furthermore, piRNAs have been found to be associated with transgenerational epigenetic inheritance, which is described as when environmental effectors have an impact on the gene expression of the subsequent generation or generations [81]. In

*Drosophila*, piRNAs have been found to be transmitted from mother to offspring in the germ cell cytoplasm and cause gene silencing in the offspring and multiple successive generations [82–84].

Third, being linked to genomic imprinting, lncRNAs are greatly involved in mammalian transcription-regulatory processes [85, 86]. Also, in the honeybee lncRNAs regulate brain development, functional diversification [87], and the transition from nursing to foraging [88].

Transcription factors (TFs) regulate gene expression by changing their binding affinities of target genomic sequences and are highly correlated with epigenetic mechanisms [89]. Histone modifications and DNA methylation have been implicated in regulating TF access and binding [90–97]. For example, promoter hypomethylation causes the overexpression of the TF sonic hedgehog both directly and via nuclear factor-kappa B activation in human tumours [98–100]. Moreover, several TFs are associated with caste fate and behavioural plasticity in eusocial insects, and TF binding sites are more variable among eusocial insects than between solitary and eusocial insects indicating a key role in facilitating the diversity in caste phenotypes witnessed in eusocial insects [42].

Other less widespread mechanisms of epigenetics also exist in insects. For instance, modifications to RNA regulate its stability, localization, transport, splicing, and translation, and it has been proposed that RNA methylation may play a role in the more rapid/plastic expression changes [101]. However one of the most prominent processes in the alteration of gene expression is DNA methylation. DNA methylation is associated with some of the processes already mentioned, and has several existing links with caste polyphenisms in eusocial insects [102].

### 1.2.2 DNA methylation: a regulator of polyphenisms in insects

5mC DNA methylation is an epigenetic mark that occurs when a methyl group is added to the fifth carbon of cytosine of DNA [103]. 5mC methylation usually acts on the cytosine of a CpG (cytosine-phosphate-guanine) site, but it can also occur at CHG and CHH locations where H represents adenine, guanine, or thymine [104]. DNMTs catalyze the methylation of DNA. *Dnmt3* initiates de novo methylation, and *Dnmt1* acts to maintain methylation which assists in copying methylation patterns during DNA replication [105].

In mammals, methylation typically represses expression in repetitive DNA elements and transposons [106]. In the mammalian lineage, it has been established that there is a strong link between genomic imprinting and DNA methylation [107]. Methylation of one parent-of-origin-specific allele occurs on imprinting control regions during gametogenesis, which then initiates the allele-specific down-regulation witnessed in genomic imprinting [108, 109]. This allele-specific methylation is protected from various phases of de-methylation that occur post-fertilisation [110, 111].

Similarly to mammals, social Hymenoptera have a functioning methylation system with a full complement of DNMTs (with the exception of *Polistes* wasps) [44, 112, 113]. DNA methylation is widely known to be present in the eusocial Hymenoptera at CpG sites [114], although levels of methylation are lower than in mammals [106]. Moreover, in contrast to mammalian methylation patterns transcribed exonic regions are the locations in which methylation is found in the honey bee [106].

Methylation has been reported to be involved in insect polyphenisms such as social insect caste differentiation and worker reproduction [102]. For instance, differential methylation has been found between honey bee queen and worker brains [115]. Sparsely methylated genes are associated with caste-specific gene expression in germ lines in honeybees [116]. Moreover, silencing of *Dnmt3* (typically responsible for de

novo DNA methylation) in honey bee larvae caused queen development evocative of a royal-jelly effect [117]. Although another study conducted by Herb and colleagues found no methylation differences between worker and queen castes, they found differences in methylation between honey bee nurse and forager worker subcastes [118]. This study was further able to demonstrate that methylation levels could be reversed when reverting foragers back to nurse behaviours. Also of note, the desert locust *S. gregaria* has displayed particularly high levels of DNA methylation for its placement in the insect lineage which has thought to be indicative of the animal's distinct phase polyphenism [119].

Other studies question the importance of methylation-based epigenetics in the Hymenoptera [120]. Highly methylated genes were found to be uniformly and transcriptionally active in different conditions and representing housekeeping genes expressed in most cell types [121–123]. Libbrecht et al. found that methylated genes typically showed consistently high expression levels and were not associated with different castes in clonal raider ants [124]. Also, Standage and colleagues found a reduced genome-wide methylation system in the primitively eusocial paper wasp, *Polistes dominula*, and the loss of *Dnmt3* [113]. *Dnmt3* was also absent in the closely related *Polistes canadensis* [44]. This points towards a reduced importance of methylation in *Polistes* species.

Although there is some evidence to suggest methylation-dependent social insect caste differentiation and worker reproduction, the relative importance of methylation in the Hymenoptera as a whole so far remains undetermined. Despite methylation appearing of little importance in the wasp [44, 113], this does not necessarily discount its functional relevance in other Hymenopteran species of separate evolutions of eusociality. Therefore, the question remains of if and how much methylation has an effect on insect polyphenisms. Nevertheless, if DNA methylation is to have a meaningful role in the epigenetics of any hymenopteran species it is likely that this methylation is involved in the alternative splicing of genes, which is explored in Section 1.2.3 below.

### 1.2.3 DNA methylation: a mediator of alternative splicing

The process of intron removal and exon ligation of a given gene is considered to be constitutive splicing [125]. However, there can be more than one protein encoded from a gene [126]. Alternative splicing of a gene includes different combinations of exons and introns and subsequently different proteins can be formed. This phenomenon of alternative splicing can have a large effect on the phenotypic profile of an organism [127]. DNA methylation and histone variants are thought to control this process by acting as alternative splicing markers [128–130].

In mammals, methylation has been demonstrated to mediate alternative splicing through complex relationships between different pathways [131], for example DNA methylation has been linked with exon skipping at the *H19/Igf2* locus. Allele-specific methylation of this locus causes exon skipping, and thus the isoform expression of either *H19* (maternal allele) or *Igf2* (paternal allele) [105, 132, 133]. DNA methylation inhibits the binding of CCCTC-binding factor (CTCF) to RNA polymerase II (Pol II). Therefore hypermethylation allows Pol II elongation, and this causes the skipping of exonic regions of DNA in transcription [134]. This type of mechanism has also been suggested to explain why hypomethylation was associated with alternative gene splicing in the gene *ALK* in the honey bee *A. mellifera* and *lipoprotein receptor 2* in the ant *Camponotus floridanus* [62, 128, 131, 135].

In contrast to the CTCF pathway, methylation has been found to be associated with exon retention. When the mammalian methyl-CpG-binding protein 2 (MeCP2) is methylated, MeCP2 pauses Pol II elongation [136, 137]. This is the opposite effect of methylation to Pol II elongation compared with the CTCF pathway, and results in the inclusion of exons at transcription. A pathway similar to this may explain the finding of Bonasio et al. that methylation was involved with exon retention in the ants *C. floridanus* and *Harpegnathos saltator* [122].

In the Hymenoptera, methylated CpGs show a strong pattern of being located near splicing sites and in the regions encoding alternatively spliced exons in the honey bee *Apis mellifera* [115, 138]. Similarly, splicing factors are found at a higher level in methylated genes [128].

Alternative splicing has been shown to be involved in the transition between nurse and foraging worker roles in the honey bee *Apis mellifera* [118]. Lyko et al. and Park et al. first suggested an association between alternative splicing and DNA methylation in the honey bee *Apis mellifera* and a parasitic wasp species, *Nasonia* [115, 139]. Flores and colleagues went on to report this association on a genome-wide scale in *A. mellifera* [140]. Furthermore, in a distinct but closely related lineage (Isoptera) differentially methylated genes between termite (*Zootermopsis nevadensis*) castes were recently found to be associated with alternative splicing [141]. These studies indicate a role for DNA methylation in the regulation of splicing of alternative variants in the Hymenoptera [122], although see also [123].

Furthermore, alternative gene isoforms have been found to be key to locust development. Zhang et al. found both variants of an alternatively spliced gene, chitin synthase 1, to be essential for the growth and development of the oriental migratory locust, *Locusta migratoria manilensis* (Meyen) [142]. One isoform (LmCHS1A) was predominantly expressed in the integument, and the other isoform (LmCHS1B) was mainly expressed in the trachea. This suggested that LmCHS1A may play a major role in chitin biosynthesis for the integument whereas LmCHS1B played a major role in chitin biosynthesis for the trachea. Silencing of either variant led to a high mortality rate of 88% and 51% in the locust nymphs injected with LmCHS1A and LmCHS1B dsRNA, respectively. In a study by Wang et al., 45 genes were found to have differentially expressed isoforms between solitary and gregarious migratory locusts (*Locusta migratoria*) [143]. These included genes associated with cytoskeleton dynamics which are involved in neuronal plasticity, a key mechanism associated with behavioural phase change [144].

DNA methylation-mediation is not the only way in which alternative splicing may be controlled. The histone variant H2A.Z is enriched in hypomethylated loci and has been suggested to have the potential to act as an alternative splicing marker in the plant *Arabidopsis thaliana* and in the green spotted puffer-fish *Tetraodon nigroviridis* [129, 130]. Inhibitors of histone deacetylation increase Pol II elongation rate by making chromatin more open [145]. This implies a role for histone modifications in the regulation of alternative splicing. However no connection between histone modifications and alternative splicing has yet been investigated in social insects. Although histone deacetylases in the ant *C. floridanus* are involved in the transition to foraging/scouting indicating a possible role for histone modifications in determining caste-specific patterns of behaviour [146].

Therefore DNA methylation-mediated (and potentially histone modification-mediated) alternative splicing could be a crucial mechanism in insect polyphenisms, and is a good starting point to elucidating the exact pathways behind these phenomena in the Hymenoptera. Two polyphenisms in the insect lineage that are becoming key subjects for the study of the epigenetic mechanisms of polyphenisms are the caste polyphenism of the buff-tailed bumblebee (*Bombus terrestris*) and the phase polyphenism of the desert locust (*Schistocerca gregaria*). Throughout this thesis these remarkable caste and phase polyphenisms will be studied.

### 1.3 Bumblebee caste polyphenism

The Hymenoptera order consists of the bees, wasps, ants, and sawflies. Eusociality has evolved nine times here as well as in termites, thrips, aphids, one beetle species, and one mammalian species: the naked mole rat [8, 147–155]. Highly eusocial species are recognised by overlapping generations, cooperative brood care,

and reproductive division of labour [156]. Discrete caste systems enable reproductive division of labour whereby the queen typically monopolises reproduction, while daughter workers predominantly look after the brood and forage for food.

While many ant, wasp, and bee species (e.g. the honey bee) are widely regarded as highly eusocial, bumblebees are often thought of as primitively eusocial [157]. Bumblebee colonies are annual and are relatively small with 300-400 individuals in *Bombus terrestris* colonies, compared to tens of thousands in those of the honey bee *Apis mellifera* [158, 159].

The annual buff-tailed bumblebee (*Bombus terrestris*) colony is initiated in early Spring when a queen will emerge from hibernation, find a suitable location to start a nest, forage for food, and begin laying diploid eggs which develop into daughter worker bees. These workers will take over the role of brood care and foraging while the founding queen focuses her energy on ovipositing throughout the Spring and Summer. Later in the Summer, the queen will switch to laying diploid queen eggs and haploid eggs which develop into males. At this point of the year the colony will be reaching its largest size, and worker aggression behaviours of humming, darting, and attacks increase due to competition over reproduction [160]. (Although the founding queen monopolises reproduction, workers may still produce haploid male eggs.) Newly emerged queens (gynes) hibernate and males die soon after mating in late Summer/Autumn.

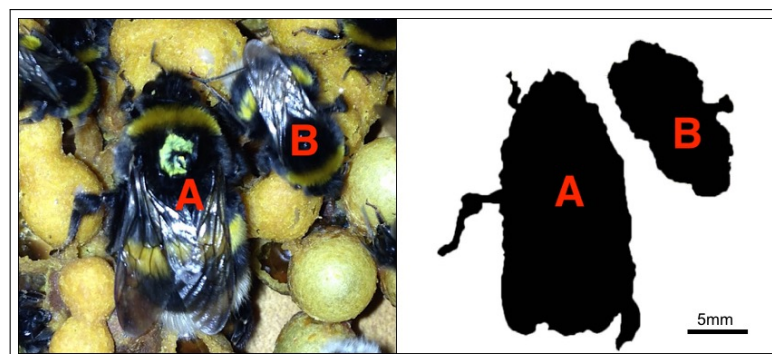


FIGURE 1.3: *Bombus terrestris* queen (A) and worker (B) individuals, with silhouettes to show difference in body size.



The discrete caste system is not as advanced in primitively eusocial species such as *B. terrestris* compared to highly eusocial species. Regarding bumblebee polyphenism, workers and queens are genetically similar, both being diploid as opposed to haploid males. However, the behaviour, longevity, and morphology of these castes is different: they have disparate roles within the colony, queens live longer than workers [161], and queens are two to three times the size of workers (Figure 1.3). These characteristics have sparked interest in bee epigenetics particularly regarding the genetic basis of longevity [103, 162, 163], the evolution of eusociality [164–167], and the evolution of genomic imprinting [11, 168].

## 1.4 Phase polyphenism in the desert locust

The other focal polyphenism of this thesis is the phase polyphenism witnessed in the locust. The desert locust *Schistocerca gregaria* displays phenotypic plasticity according to changes in population density of conspecifics. The two phases are described as solitary and gregarious, and are associated with different behavioural and morphological characteristics (Figure 1.4). Solitarious locusts are typically green in colour which appears cryptic against the foliage in which they inhabit. As the phase name suggests they live alone and given the choice move away from nearby conspecifics [169]. Activity levels are highest during the night [170].

In contrast, gregarious locusts are brighter in colour (yellow/brown) and move in the same direction as other locusts in an “escape-pursuit” behaviour [169]. Gregarious individuals are active during the day, with peak activity previously reported at 1400 hrs [31], and their activity levels drop over night [171].

When there is an increase in population size or a reduction in food availability, solitary locusts can be forced together with conspecifics causing them to become behaviourally gregarious [172]. However it takes several generations for morphological changes to occur [173]. It is in this gregarious phase when locusts swarm, and hence

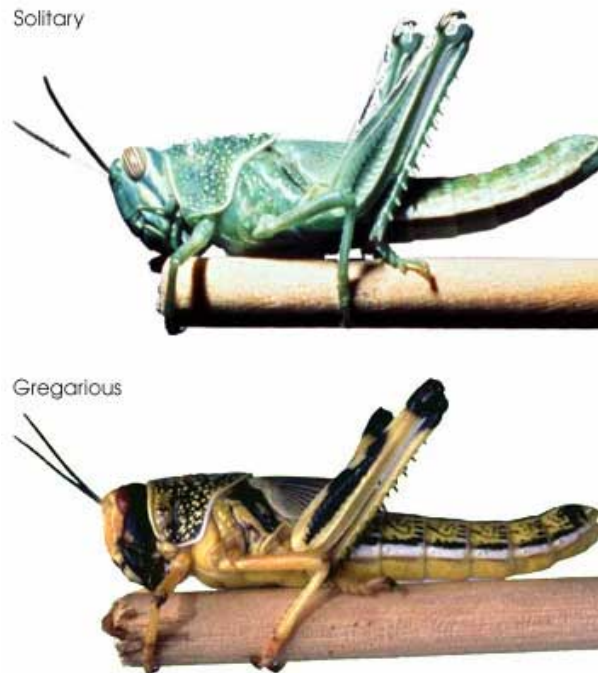


FIGURE 1.4: Morphology of juvenile *Schistocerca gregaria* desert locusts in the solitary and gregarious phases (Photograph: Compton Tucker, NASA GSFC)

are perceived as an agricultural pest species in Africa and the Middle East, where they can cause devastation for crops such as barley, maize, sorghum, and wheat [174].

The evolution of this density-dependent polyphenism has previously been explained by a combination of the risks of predation and cannibalism [169, 175]. A solitary locust uses crypsis to hide from predators and implements an “avoidance” strategy to distance themselves from all conspecific threats. On the other hand, gregarious locusts benefit from individual concealment from predators when part of a group. Gregarious locusts have also developed the escape-pursuit behaviour when at a high population density. Following conspecifics is argued to reduce collisions since the locusts move into areas left vacant by other individuals, which decreases the risk of cannibalism.

## 1.5 Thesis aims and objectives

Through this thesis the following questions are asked in order to elucidate a more defined understanding of the epigenetic mechanisms underlying insect polyphenisms. Chapters 2, 3, and 4 focus on the eusocial buff-tailed bumblebee *Bombus terrestris* as a model organism. Monoallelic methylation is widely known to be associated with allele-specific expression in mammals [106, 107], and due to the complete set of DNMTs in many Hymenoptera the same trend is predicted in social insects (Section 1.2.2). Chapter 2 concentrates on determining the presence of allele-specific expression and monoallelic methylation in *B. terrestris*. Is there an association between allele-specific methylation and expression in social insects?

After establishing the presence of allele-specific effects in *B. terrestris*, chapters 3 and 4 focus on identifying associations between epigenetic mechanisms in reproductive and non-reproductive *B. terrestris* workers. Worker reproduction is a common occurrence in *B. terrestris*, whereby in the later stages of a colony cycle some workers will develop full ovaries and compete over producing haploid male offspring [160] (Section 1.3). This process has previously been implied to be epigenetically regulated [12]. Chapter 3 questions whether DNA methylation, gene expression, and alternative splicing are involved in worker reproduction and determines if there are correlations between these mechanisms. This chapter found general associations between these effects and the reproductive switch in the worker caste.

Chapter 4 then extends this approach with the comparison of allele-specific methylation and allele-specific expression. I identified allele-specific expression in chapter 2, and in chapter 4 I distinguish allele-specific differences that are specifically involved in worker reproduction. This chapter looks to establish the effect that allele-specific methylation and expression can have on a distinct trait in social insects: worker reproduction.

The final two results chapters focus on alternative splicing. As another insect polyphenism example, chapter 5 centres on the gregarious phase of the desert locust *Schistocerca gregaria*. This species is considered a pest in the Middle East and Africa where swarming can cause devastation for crops [174]. Strategies for insect control are often more effective when applied at particular times of the day due to circadian cycles affecting the prevalence of toxicity proteins [176]. Strong circadian rhythms are present in swarming *S. gregaria* [177]. Therefore the relationship between alternative splicing and the circadian clock in the desert locust is investigated. This chapter determines any isoforms with differential circadian rhythms in locusts. This may have the potential instruct when would be the preferential time of day for insecticide application.

After establishing the importance of alternative splicing in another model insect species, chapter 6 takes a broader approach to alternative splicing and caste polymorphisms in eusocial insects. Workers of species with simple social structures have a greater variety of roles to play in a colony compared to workers of species with more complex social structures. Therefore it is of interest to determine the mechanisms that enable the greater variety of behaviours in a particular caste. Alternative splicing can provide a method for multiple proteins to be encoded from a single gene [178]. A greater prevalence of alternative splicing could facilitate the variety of behaviours seen in workers of species with simple social structures. I analyse whether genome-wide alternative splicing levels are an important component in ascertaining the varying levels of eusociality found in the Hymenoptera.

# Chapter 2

## Searching for allele-specific expression in *Bombus terrestris*

### 2.1 Introduction

Allele-specific expression is when there is a difference in expression levels of alleles of the same gene whilst both alleles are both still being expressed at some level. (An extreme form of allele-specific expression is named monoallelic expression, where one allele is completely silenced.) Allele-specific expression is known to be caused by a number of genetic as well as epigenetic processes [179]. The genetic process usually involves cis effects such as transcription factor binding, microRNA binding, and untranslated regions which alter RNA stability. For instance, microRNAs bind to protein-coding genes and cause post-transcriptional gene silencing [180, 181].

Allele-specific expression can also occur through the phenomenon of genomic imprinting, which is when one allele of a gene is down-regulated according to from which parent the allele was inherited. Genomic imprinting is well described in mammals with a high association with DNA methylation (see Section ??). Imprinting has been predicted to have evolved in insects with a eusocial life history by Haig's

theory for the evolution of genomic imprinting [11, 168]. In particular, patrigenes involved in worker reproduction and development are predicted to be imprinted. Recently parent-of-origin allele-specific expression has been found in honey bees [182] and there appears to be a fundamental role in social insect biology for methylation [183].

In mammals and flowering plants, allele-specific expression is often associated with methylation marks passed from parent to offspring [184]. However DNA methylation is involved in numerous other cellular processes [105]. 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* [122]. Recently, evidence was found for allele-specific expression in bumblebee worker reproduction genes [185] and methylation was found to be important in bumblebee worker reproduction [186]. However, other work on the honeybee *Apis mellifera* found no link between genes showing allele-specific expression and known methylation sites in that species [187].

The recently sequenced genome of the bumblebee, *Bombus terrestris*, displays a full complement of genes involved in the methylation system [188]. *B. terrestris* is also a eusocial species with reproductive division of labour. The objective of this study was to investigate allele-specific expression and monoallelic methylation in the bumblebee, *Bombus terrestris*. The identification of a link between allele-specific expression and methylation of worker reproduction and development genes in a eusocial insect would be consistent with the leading theory for the evolution of genomic imprinting.

I then searched specifically for patrigenic monoallelic expression of any candidate loci I identified to show both monoallelic expression and monoallelic methylation. Imprinting of patrigenes is predicted in genes involved in development and worker reproduction in *B. terrestris* colonies, since reproduction of workers is the only way for a patrigen to be continued into the next generation in a queen-right colony

(Section ??). The identification of patrigenic monoallelic expression here would be compatible with Queller's predictions regarding Haig's theory for the evolution of genomic imprinting [11, 168].

## 2.2 Materials and Methods

In this study I 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 (Figure 2.1). MeDIP-seq is an immunoprecipitation technique that creates libraries enriched for methylated cytosines [189]. Methyl-sensitive restriction enzymes can create libraries that are enriched for non-methylated cytosines (MRE-seq) [189]. 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 [189]. Monoallelic expression was identified in these loci from the RNA-seq library. If only one allele was expressed then these loci were both monoallelically methylated and monoallelically expressed in this bee. I confirmed this monoallelic expression in one locus using qPCR. I looked for further confirmation of this monoallelic expression in twenty-nine *Bombus terrestris* RNA-seq libraries.

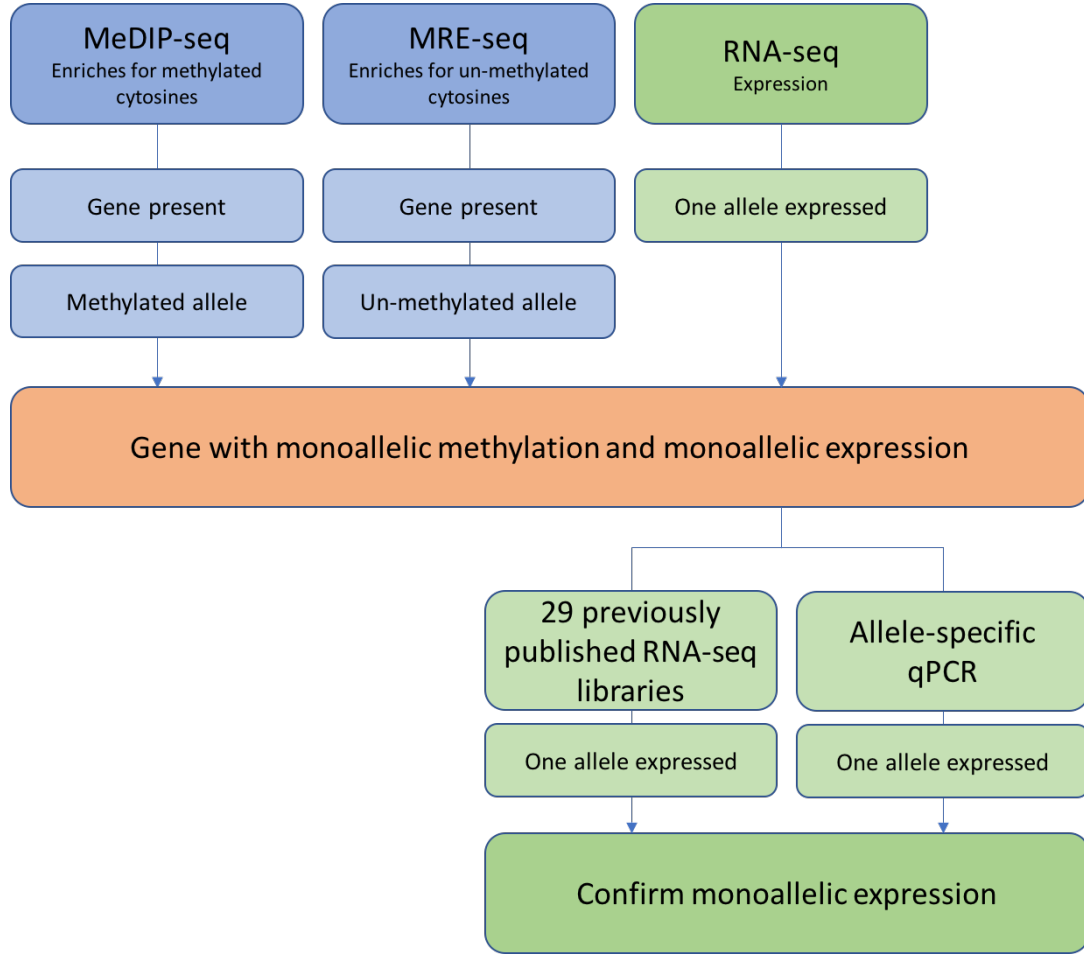


FIGURE 2.1: Flow chart of methods for identification of monoallelic methylation and monoallelic expression in *Bombus terrestris* for a given gene. Blue boxes indicate methylation related methods. Green boxes refer to methods for expression.

I then more generally searched for allele-specific expression by analysing published RNA-seq libraries from worker bumblebees [12, 190]. I identified heterozygotes in the RNA-seq libraries and measured the expression of each allele. I then identified loci that showed significant expression differences between their two alleles.

### 2.2.1 Samples

Data from twenty-nine RNA-seq libraries were used for the allele-specific expression analysis (six from Harrison *et al.* [12], and twenty-three from Riddell *et al.*



[190]). Regarding the paternal monoallelic expression analysis, RNA-seq libraries from three reproductive and two non-reproductive *Bombus terrestris* workers from three colonies were used, along with their three corresponding founding queens (from Harrison et al.) [12] (see Table 2.1).

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 2.1: Bees used in each experiment. Regarding the different colonies: K refers to Koppert, A to Agralan and Q to the wild caught Leicester queen.

Experiment	Number of worker bee samples	Colony	Tissue
Allele-specific expression	1	A1 [12]	Whole body
RNA-seq	2	A2 [12]	Whole body
	2	A3 [12]	Whole body
	1	A4 [12]	Whole body
	14	K1 [190]	Abdomen
	9	Q1 [190]	Abdomen
MeDip/MRE/RNA-seq	1	K2	Whole body
qPCR	2	K3	Head
	1	K4	Head
	1	K5	Head
Paternal monoallelic expression	2	A1[12]	Whole body
RNA-seq	3	A2[12]	Whole body
	3	A3[12]	Whole body

## 2.2.2 Next generation sequencing

### 2.2.2.1 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 [189]. 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-pair reads 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.

### 2.2.2.2 Previously published RNA-seq

Full details of the RNA-seq protocols used have been published previously [12, 190]. 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 here 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.

### **2.2.3 Monoallelic methylation and expression - Bioinformatic analysis**

I 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).

#### **2.2.3.1 Alignment to genome**

RNA-seq reads were aligned to the *Bombus terrestris* genome assembly (AELG000000000) using Tophat [191] and converted to BAM files with Samtools [192]. 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 [193]. 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 [194, 195]. PCR duplicates for all BAMs (mRNA, MeDIP and MRE) were marked with the Picard utility Markduplicates.jar.

#### **2.2.3.2 Identifying regions of interest and integrating data**

Coverage of each data type was calculated using GATK DepthOfCoverage [195]. 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 [193] and results were merged

with vcf tools [196]. CpG islands were identified using CpG island searcher takai comprehensive 2002. Regions of mRNA with overlaps of MeDIP, MRE, CpG islands and monoallelic SNPs were identified with custom Perl scripts.

## 2.2.4 Patrigenic monoallelic expression - Bioinformatic analysis

I searched for patrigenic monoallelic expression in three colonies for the nineteen candidate genes (identified in Section 2.2.3) through comparing the gene expression of the workers and queens (Table 2.2). The BWA mapper [193] was used to align each library to the *Bombus terrestris* reference genome (Bter 1.0, accession AELG00000000.1) [188]. The RNA-seq libraries were prepared, filtered, and any duplicates were removed using samtools (version 0.1.19-44428cd). Bcftools (version 0.1.19-44428cd) was then used to call the SNPs in each library. Each locus of the nineteen candidate contigs previously identified to show monoallelic expression and monoallelic methylation (in Section 2.2.3) was searched for in the eight RNA-seq libraries. Genotypes of these SNPs were then compared between workers and founding queens of the same colonies.

TABLE 2.2: Number of whole body samples used for paternal monoallelic expression analysis from Harrision et al. [12]. The colonies were obtained from Agralan Ltd, hence the colony notation of A1, A2, and A3.

Caste	Colony A1	Colony A2	Colony A3
Founding Queen	1	1	1
Reproductive Worker	1	1	1
Non-reproductive Worker	0	1	1

The problem here was distinguishing homozygotes from parent-specific gene expression. There are three possible scenarios, of which only one would inform us of any parent-specific gene expression that is present. In the event that the worker is homozygote at a locus in question (e.g. AA - where the expressed base is denoted first, and silenced base is second), the queen's genotype, whatever it may be, would not

be able to inform one of any existing parent-specific gene expression, as it would be impossible to determine which base in the worker had been inherited from the queen.

In the second case, if the worker is heterozygote (for instance, AT), and the queen possessed the base expressed in the worker (i.e. the queen's genotype is AT, TA, AC, CA, AG, GA, or AA) then one still would not know whether there is a parental effect or if this locus is merely homozygote in the worker.

However, in the final scenario, if the worker is heterozygote (AT), and the queen does not have the base expressed in the worker (i.e. the queen's genotype is TT, TC, CT, GT, or TG) then one would know that this site is not homozygote in the worker. Instead, this locus would be shown to be monoallelically expressed, with paternal expression and maternal silencing. I would be able to see that the paternally inherited base, A, is expressed, and that the locus is not homozygote in that individual worker since the queen would not have the base, A, at that locus. Searching for this in the same positions in multiple bumblebee colonies would tell me if patrigenic expression of these genes can be generalized among *B. terrestris* colonies.

### 2.2.5 Allele-specific expression - Bioinformatic analysis

I created a pipeline to search for heterozygous loci that show allele-specific expression and identified the associated enriched gene ontology (GO) terms in twenty-nine previously published RNA-seq libraries for those genes that fit this criteria [12, 190].

Each RNA library was mapped to the reference genome (*Bombus terrestris*, Bter 1.0, accession AELG000000000.1) [188] using the BWA mapper [193]. 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 [197, 198]. The success of this control was confirmed by the R package edgeR (version 3.14.0) [5, 6]. The SVA adjustment reduced the edgeR dispersion value from 3.9994 (BCV=2) to 0 (BCV=0.0003) (Appendix A: Figure A.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 [193, 199]. Only SNPs with a mapping quality score of  $p < 0.05$  and a read depth of  $\geq 6$  and present in  $\geq 3$  of the libraries were included in the analyses. The R package, QuASAR, was then used to identify genotypes and locate any allele-specific expression at heterozygous sites [200]. 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 allele-specific expression in at least three of the thirty libraries, were blasted (Blastx) against *Drosophila melanogaster* proteins (non-redundant (nr) database) [201]. The blast results were annotated using Blast2Go [202]. Fisher's exact test was implemented to identify enriched GO terms, which were then visualised using REVIGO [4]. To identify which bumblebee genes the SNPs were located in, the SNP position +/- 25 bp was blasted (Blastn) against the *Bombus terrestris* genome [188].

## 2.2.6 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.

I amplified numerous fragments of the nineteen 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 nineteen loci, which could be amplified with primers of suitable efficiency. I 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 [203]. 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<sub>2</sub>O. 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 [204]. To test for this I repeated all qPCRs with genomic DNA (1μl of diluted DNA (20ng/μl) from the same bees as the template. I would expect equal amounts of each allele in the genomic DNA. I also measured efficiency of each reaction as per [205].

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:

$$ratio_{allele} = \frac{E_{allele}^{-C_{t_{allele}}}}{E_{reference}^{-C_{t_{reference}}}} \quad (2.1)$$

E is the median efficiency of each primer set [204, 205]. All statistical analysis was carried out using R (3.1.0) [7].

## 2.3 Results

### 2.3.1 Monoallelic methylation and monoallelic expression

In total, I 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 *ras GTPase-activating protein nGAP-like* in Figure 2.2. Of the nineteen genes, fourteen had the hypermethylated (MeDIP) allele expressed, while five had the hypomethylated (MRE-seq) allele expressed (see Appendix A: Table A.1).



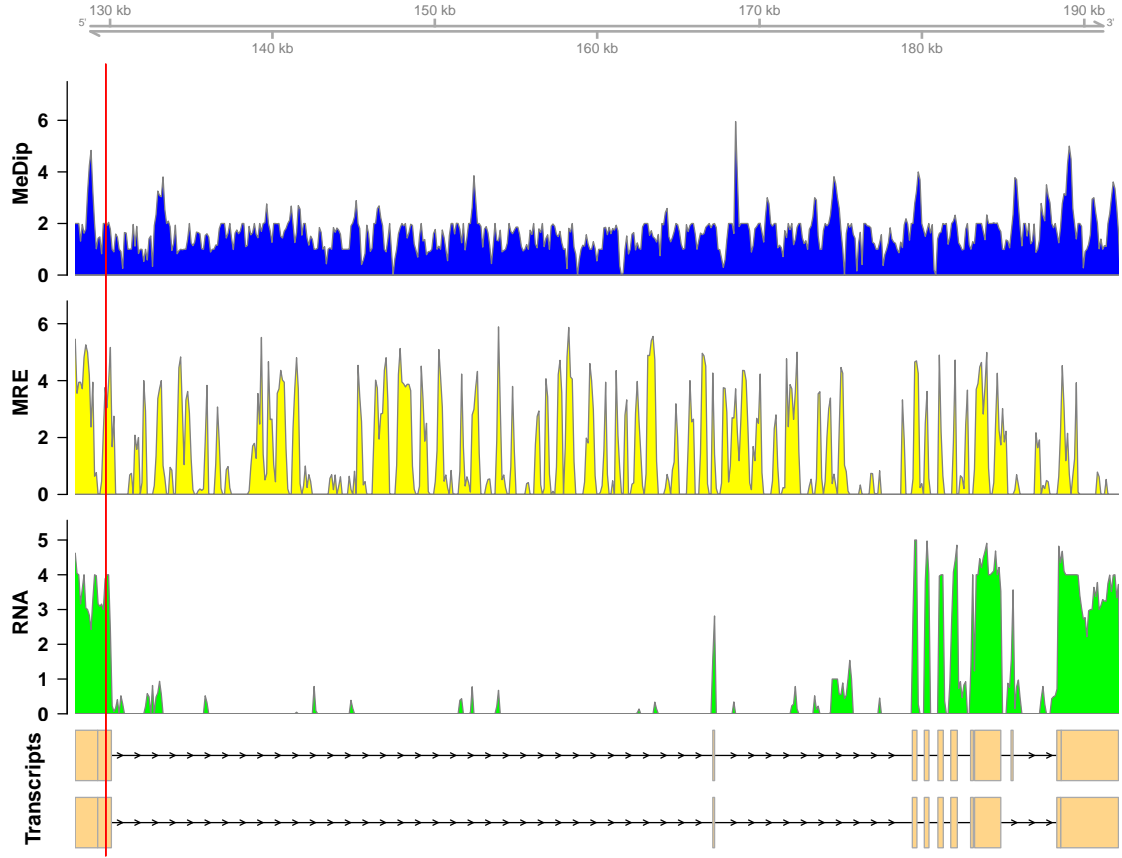


FIGURE 2.2: Coverage of the RNA-seq, MRE-seq, and MeDIP-seq libraries for *ras GTPase-activating protein nGAP-like* (LOC100652225). The transcript models come from GCF\_000214255.1.Bter.1.0. The y-axis in the coverage plots is  $\log(1 + \text{coverage})$ . The red vertical line represents the heterozygote position. The MeDIP-seq allele was expressed in this locus

### 2.3.2 Confirmation of monoallelic expression by qPCR

Monoallelic expression was confirmed in one of these nineteen (*slit homolog 2 protein-like* (AELG01000623.1)) by allele-specific qPCR (amarasinghe allele 2015). The allele with a guanine at the SNP position had a mean expression of  $6.04 \pm 8.28$  (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 =  $422.70 \pm 507.36$ , T mean =  $1575.17 \pm 503.02$ ). In the three other loci tested (*Ras GTPase-activating protein 1*, *Ecdysone receptor*, *methionine aminopeptidase 1-like*)

I found apparent monoallelic expression, but could not dismiss primer efficiency as the cause.

The nineteen genes were blasted against the nr / nt (nucleotide) 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 2.3). Six of the eleven genes with informative hits have functions to do with social organisation in the social insects (Table 2.3).

I 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 Appendix A: Table A.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.

### 2.3.3 Paternal monoallelic expression

I found nine of the nineteen candidate loci to have a corresponding SNP in at least one of the five worker RNA-seq libraries, where only one allele was expressed (thus indicating either homozygosity or monoallelic expression). The same genotypes were identified between workers and their founding queens at all nine locations, in all three colonies (i.e. the second scenario explained above) (Table 2.4). Consequently these results could not confirm any patrigenic monoallelic gene expression which may here be present. This dataset cannot determine whether these loci show paternal-specific expression as opposed to maternal-specific expression or are homozygote.

TABLE 2.3: The eleven of the nineteen monoallelically methylated and monoallelically expressed genes that returned informative blastx hits compared with the *Drosophila melanogaster* nr database.

Gene	Accession	Expressed allele	Function
<i>yippee-like 1</i>	AELG01001021.1	MeDIP	Yippee is an intracellular protein with a zinc-finger like domain. DNA methylation of a CpG island near the <i>yippee-like 3</i> promoter in humans represents a possible epigenetic mechanism leading to decreased gene expression in tumours [206].
<i>slit homolog 2 protein-like</i>	AELG01000623.1	MeDIP	Slit is produced by midline glia in insects and is involved in cell projection during development [207]. All three human Slits were found to be hypermethylated in hepatocellular carcinoma cell lines [208].
<i>methionine aminopeptidase 1-like</i>	AELG01000544.1	MeDIP	Methionine aminopeptidases catalyse N-terminal methionine removal [209]. MAP1D in humans was found to be potentially oncogenic [209].
<i>calmodulin-lysine N-methyltransferase-like</i>	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 [210].
<i>Ecdysone receptor</i>	AELG01000543.1	MRE	In <i>Drosophila melanogaster</i> , ecdysone receptor interacts with ecdysone to activate a series of ecdysteroid genes [211]. In honeybees, <i>Ecdysone receptor</i> is expressed in the brain mushroom bodies of both workers and queens and ovaries of queens [211].
<i>Shaker</i>	AELG01001021.1	MeDIP	Shaker is involved in the operation of potassium ion channel. <i>Shaker</i> expression was upregulated in sterile versus reproductive honeybee workers [212].
<i>excitatory amino acid transporter 3</i>	AELG01000969.1	MRE	Excitatory amino acid transporters are neurotransmitter transporters. <i>Excitatory amino acid transporter 3</i> expression was upregulated in sterile honeybee workers [212]. <i>Excitatory amino acid transporter 1</i> expression differences were also associated with worker - queen differentiation in the paper wasp <i>Polistes metricus</i> [213].
<i>elongation of very long chain fatty acids protein 6-like</i>	AELG01004467.1	MeDIP	The timing of the upregulation of fatty acid metabolism was found to be different in queen and worker honeybees [214].
<i>ras GTPase-activating protein nGAP-like</i>	AELG01004618.1	MeDIP	<i>Ras GTPase-activating protein 1</i> was found to be upregulated in reproductive honeybee workers [212]. It is involved in oocyte meiosis.
<i>bicaudal D-related protein homolog</i>	AELG01005399.1	MeDIP	Bicaudal is involved in embryonic pattern formation in <i>Drosophila</i> [215]. It is thought to be involved in the differentiation between soldiers and workers in the termite <i>Reticulitermes flavipes</i> [216]. <i>Bicaudal protein D</i> has been shown to be methylated more in eggs than sperm in honeybees [217].

TABLE 2.4: Genotypes of the eight Harrison et al. [12] RNA-seq libraries at candidate loci identified in Section 2.3.1 to show monoallelic expression and monoallelic methylation.

Contig	SNP <sup>a</sup>	RNA-seq library <sup>bc</sup>							
		A1 RW	A2 RW	A3 RW	A2 NRW	A3 NRW	A1 Q	A2 Q	A3 Q
AELG01000543.1	3143	-	C/T	-	-	-	C/T	C/T	-
AELG01000623.1	2544	-	A/G	-	A/G	A/G	A/G	-	A/G
AELG01000829.1	3176	-	C/T	C/T	C/T	-	-	C/T	-
AELG01001796.1	10938	-	-	G/A	G/A	-	-	-	G/A
AELG01002621.1	85071	C/T	-	-	C/T	-	C/T	C/T	-
AELG01003672.1	34537	-	T/C	T/C	T/C	-	-	-	-
AELG01004618.1	50955	T/C	T/C	T/C	T/C	T/C	T/C	-	T/C
AELG01005399.1	63410	-	-	-	C/A	-	-	-	-
AELG01006475.1	1492	-	-	-	A/G	-	-	-	-

Genotypes are listed as reference base/alternative base. If “-”, then no SNP is present for this locus in that library. Only the reference allele is expressed.

<sup>a</sup> locus identified to show monoallelic expression and methylation

<sup>b</sup> RW = reproductive worker, NRW = non-reproductive worker, Q = founding queen.

<sup>c</sup> Numbers equate to colony number.

### 2.3.4 Allele-specific expression

I then searched more generally for allele-specific expression in the twenty-nine RNA-seq libraries. 555 loci showed allele-specific expression in  $\geq 3$  of the twenty-nine RNA-seq libraries (Appendix A: Table A.3). No notable difference was seen between the frequency of G/C  $\rightarrow$  A/T and A/T  $\rightarrow$  G/C base substitutions (Figure 2.3), despite the expected mutation bias from G/C to A/T [188, 218]. Searching (blastn) for these 555 loci against *Bombus terrestris* returned 211 hits. To search for gene ontology terms, blastx was used against *Drosophila melanogaster*, which returned 329 hits. 151 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 2.4 shows the large number of biological functions associated with these 555 genes.

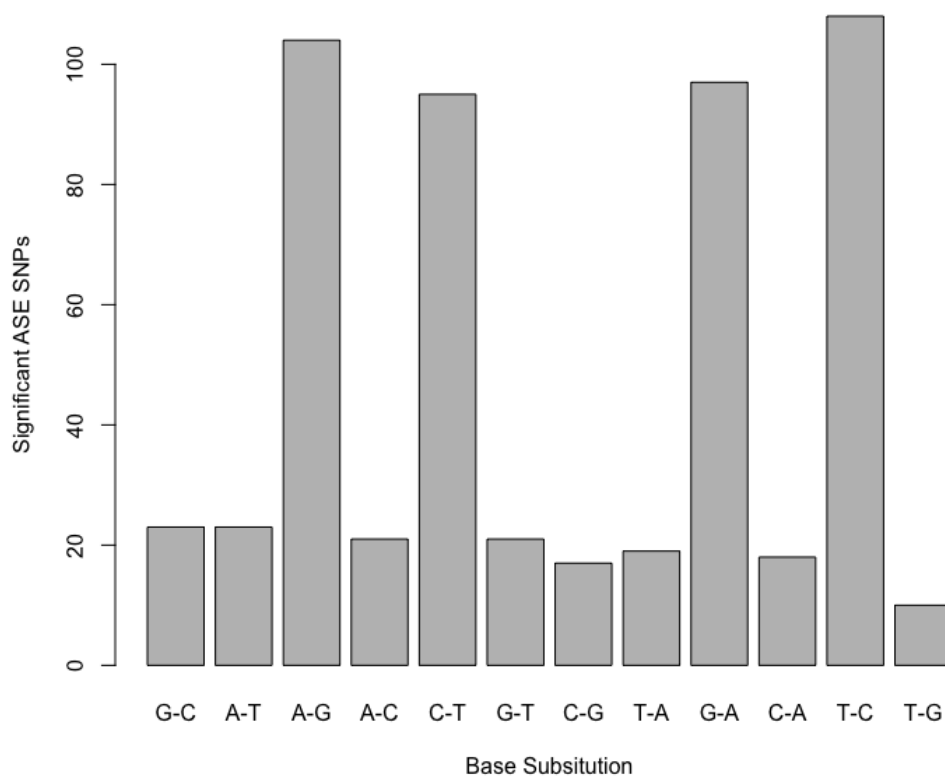


FIGURE 2.3: Base substitutions of the 555 SNPs showing allele-specific expression in at least three of twenty-nine *Bombus terrestris* transcriptome libraries. The base substitution notation represents the reference base and the alternate base at the polymorphic locus. For example, “G-C” represents a base substitution of guanine to cytosine.

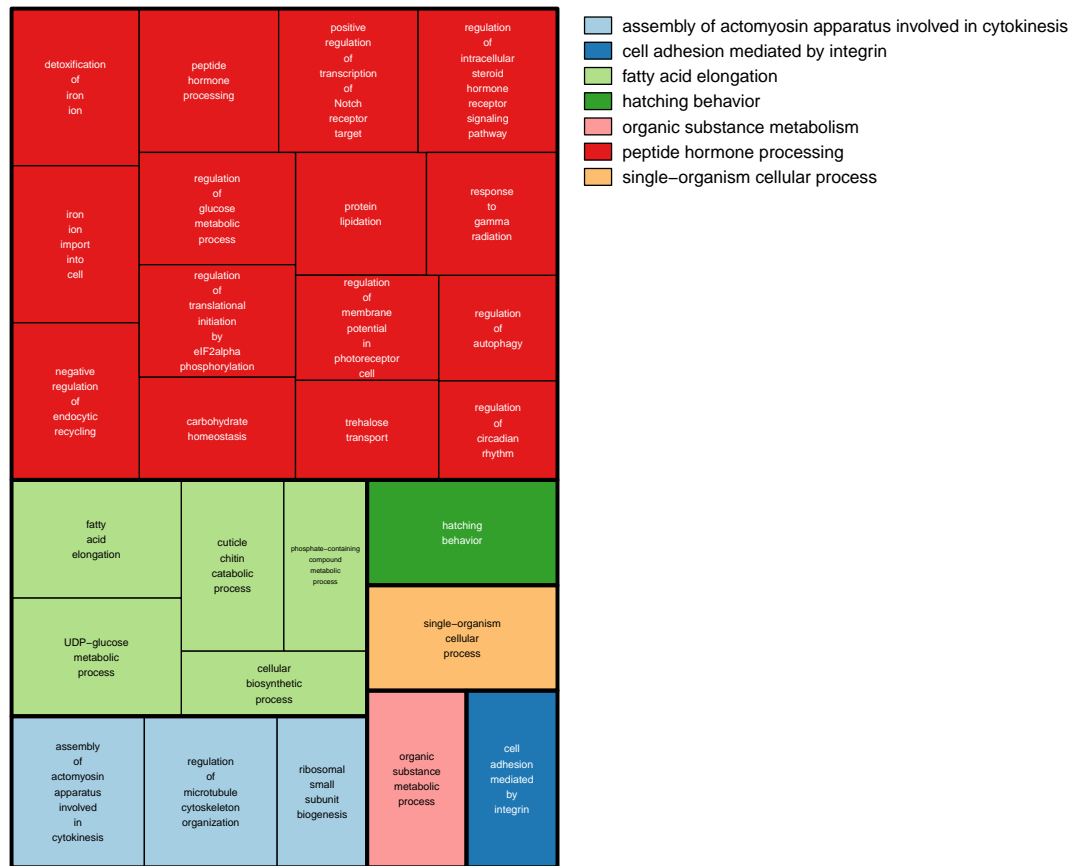


FIGURE 2.4: A summary of the enriched GO terms ( $p < 0.05$ , based on Blast2Go annotation) found for genes displaying allele specific expression. This figure was produced using Revigo [4]. Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment.

## 2.4 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 Appendix A: Table A.1). In ant genes with allele-specific methylation, the hypermethylated allele showed more expression than the hypomethylated allele [122]. This fits with genome wide analysis that shows exonic methylation in insects associated with increased gene expression

[103, 219]. 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 [103]. For example, *EGFR* expression is lower in ant workers that exhibit higher DNA methylation of *EGFR* [220]. Secondly, even in the canonical mammalian methylation system, the “wrong” allele has been shown to be expressed occasionally due to lineage-specific effects [221–225].

I then looked at the expression of these nineteen genes in all twenty-nine RNA-seq libraries. If they are monoallelically expressed in these bees, I 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. I would also find only one allele if that bee was homozygous. I 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.

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 [226]. Another explanation is that these loci are not epigenetically controlled but rather their allele-specific expression is derived from genetic effects [227]. There are three main genetic, as opposed to epigenetic, affectors of allele-specific expression [228]. 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 allele’s 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 2.4).

However, 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 [229]. 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 [230]. This cis-mediated methylation has recently been suggested as being important in social insect biology [227, 231].

I have found that allele-specific expression is widespread in the bumblebee. I have also found that the extreme version of allele-specific expression, monoallelic expression is associated with monoallelic methylation. Genomic imprinting in mammals usually involves monoallelic methylation and expression. It is tempting to associate my results with genomic imprinting, especially as a number of the genes discovered are exactly the type predicted by theory to be imprinted [11]. 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.



## Chapter 3

# Alternative splicing, DNA methylation, and gene expression in *Bombus terrestris*

### 3.1 Introduction

Buff-tailed bumblebee (*Bombus terrestris*) workers generally do not have fully developed ovaries. However, when the bumblebee colony enters the competition phase of its annual cycle some of the workers begin to develop full ovaries and produce haploid sons (Section 1.3). This distinct phenotypic change has previously been linked to divergent gene expression patterns, whereby reproductive workers have gene expression patterns more similar to queens compared to non-reproductive workers [12]. This implies that the epigenetic make-up of a *B. terrestris* individual is key to their reproductive status.

Alternative splicing is a key regulator of gene expression [232]. Different combinations of exons and introns of a gene can be translated and determine the initiation of alternate proteins forms. This can have a large effect on an organism's phenotype.

For instance, human head and body lice occupy different ecological niches. Head lice are a common occurrence especially in children, whereas body lice are usually now restricted to people living in poor sanitary conditions and are vectors of three serious pathogens [233, 234]. Head and body lice have almost identical genomes and transcriptomes, and it is the great degree of differential splicing of their genes which has been indicated to determine their different behaviours and feeding patterns [127]. Therefore it has been demonstrated that alternative splicing can have a large effect on gene expression.

In the Hymenoptera, alternative splicing has previously been found to be involved in worker reproduction. Jarosch and colleagues [235] found alternative splicing of the *geminin* transcription factor to be associated with control of worker sterility in *Apis mellifera* capensis. Knocking out of a specific exon via RNAi resulted in worker ovary activation which is associated with parthenogenetically producing diploid female offspring [235]. Hence, a worker would be able to produce offspring from an ovum without fertilization.

The initiation of worker reproduction and caste differentiation between queens and workers have been indicated to rely on the same gene regulatory networks in *A. mellifera* since Formesyn and colleagues found the knockdown of an epidermal growth factor receptor to be involved in caste differentiation and induction of reproduction in workers [236]. In light of this, caste differentiation has been found to be associated with alternative splicing. In *Apis mellifera* the transition from nursing to foraging is linked with alternative splicing events [118]. Thus, *A. mellifera* regulatory genes are differentially expressed or differentially spliced resulting in different behavioural phenotypes in nurse and forager workers. This adds to the concept that alternative splicing may be key to the initiation of worker reproduction.

Alternative splicing has been associated with DNA methylation in insects. In the hymenopteran lineage, CpG methylation has been previously found to be linked with controlling splicing in *A. mellifera* [115, 118, 237], *Nasonia vitripennis* [139],

*Camponotus floridanus*, and *Harpegnathos saltator* [122]. This is described in detail in Section 1.2.3 with possible antagonistic (MeCP2-like, CTCF-like) methylation-mediated splicing processes. Hypomethylation was associated with alternative gene splicing in *A. mellifera* and *C. floridanus* [62, 135], and hypermethylation was associated with exon retention in *C. floridanus* and *H. saltator* [122]. DNA methylation is not thought to be a “universal driver” of insect eusociality [238], but it is probable that DNA methylation is acting in concert with other epigenetic modifiers (e.g. ncRNAs, histone modifications) to regulate alternative splicing [62, 239, 240]. Furthermore, methylation has been indicated to play a role in *B. terrestris* worker reproduction [186]. Therefore in this chapter and the next I explore epigenetic differences in worker reproduction.

Here I implement a combination of RNA-sequencing and bisulfite-sequencing techniques in the buff-tailed bumblebee *B. terrestris* workers. I identify differential patterns of expression, methylation, and alternative splicing in reproductive and non-reproductive workers, and also determine associations here between these epigenetic mechanisms. Subsequently, gene ontology analysis is conducted to determine any functions that are enriched in genes up-regulated in reproductive and non-reproductive workers. The process of a *B. terrestris* worker developing ovaries has been reported to require epigenetic signalling and changes in methylation [12, 118, 186]. Thus, genes with functions associated with ovary development, epigenetic regulation, and signalling are expected to be differentially expressed and differentially methylated between reproductive and non-reproductive workers.

## 3.2 Methods

### 3.2.1 Bee husbandry and tissue sampling

Three *B. terrestris* colonies (referred to as colonies 1, 5, and 8) from Agralan were kept at a temperature of 28°C and at 60% humidity. They were kept in red light and fed pollen and a solution of 50% water and 50% honey *ad libitum*. Callow workers, less than 24 hours old, were taken from each colony and placed in boxes of five workers.

The worker bees were sacrificed at six days old. For each bee, the head was snap frozen in liquid nitrogen immediately after it was sacrificed. Through dissection in 1% phosphate-buffered saline (PBS) solution, the reproductive status of the bees was determined and classed as either reproductive (Figure 3.1a), non-reproductive (Figure 3.1b), or intermediate. Workers were classed as having developed ovaries, and therefore reproductive, if the largest oocyte was larger than the trophocyte follicle [241]. Immediately after dissection the ovaries of each worker were weighed (Figure 3.2), and the length of the largest oocyte of each ovary of the reproductive and intermediate workers was measured (Figure 3.3). In order to sample the reproductive workers with the most developed ovaries, the bees that had the longest oocyte length were selected as this measurement is tightly correlated with reproductive status [242, 243].

### 3.2.2 RNA and DNA extraction and sequencing

The heads of three reproductive and three non-reproductive worker bees from each of the three colonies were each sampled. For each head, an incision was made in the sagittal plane dividing it in two. Left and right sides of the head were randomly distributed to RNA and DNA extractions. RNA was extracted using the Sigma-Aldrich GenElute Mammalian Total RNA Miniprep kit from half of the heads of

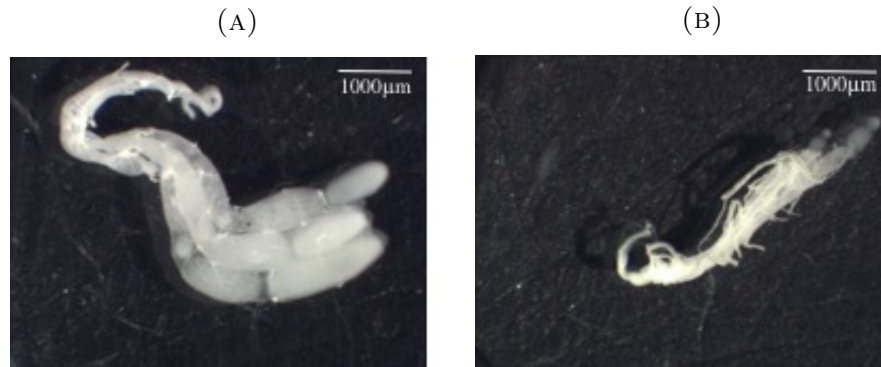


FIGURE 3.1: Ovaries of dissected (A) reproductive and (B) non-reproductive *Bombus terrestris* workers

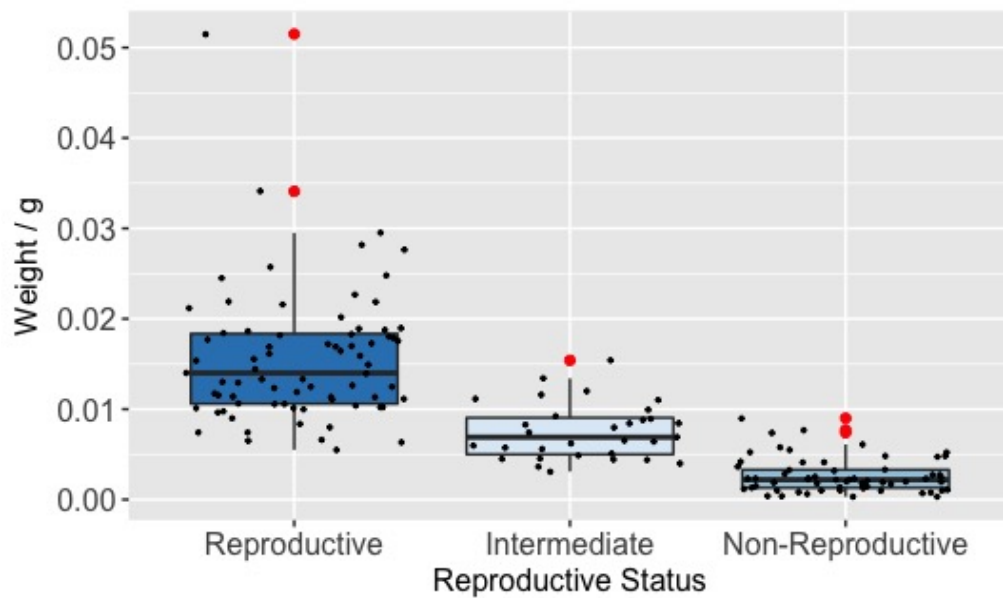


FIGURE 3.2: Ovary weight of *Bombus terrestris* workers immediately after dissection. Red points indicate outliers.

the worker bees. The other half of the heads of the same bees were used for DNA extraction for bisulphite sequencing. The extracted RNA was treated with DNase (Sigma-Aldrich DNase I treatment kit), and RNA quality was then determined by Nanodrop and Agilent 2100 Bioanalyzer. A total of 18 RNA samples were sequenced by BGI Tech Solution Co., Ltd. (Hong Kong).

DNA samples from the same eighteen worker bees were taken from half heads (Qiagen kit). Samples from the same colony and with the same reproductive status

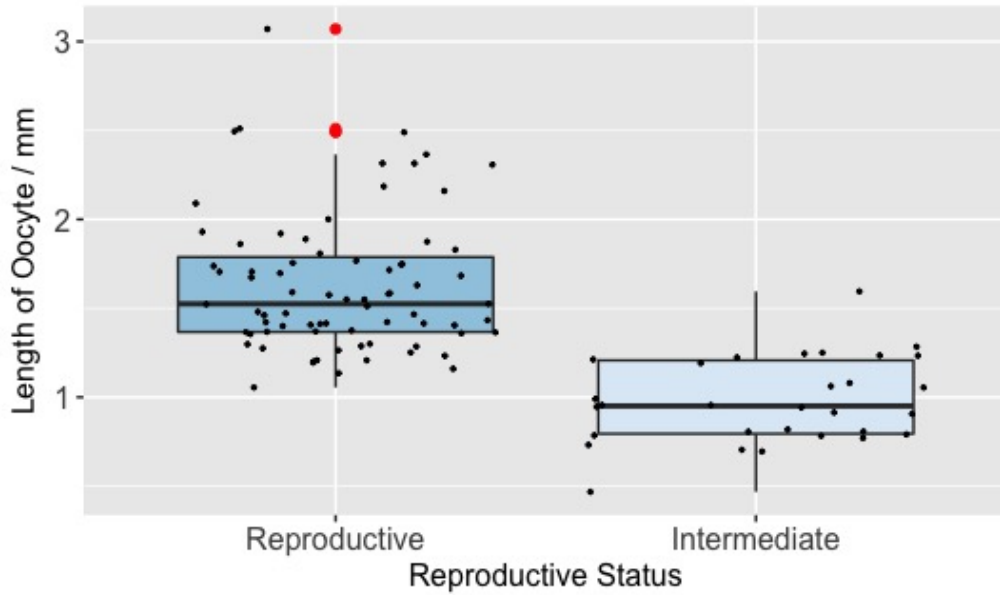


FIGURE 3.3: Largest oocyte length of dissected *Bombus terrestris* workers

(reproductive or non-reproductive) were pooled producing six pooled samples in total, and were sent for 100bp paired-end bisulfite sequencing (BS-seq) (BGI Tech Solution Co., Ltd. (Hong Kong)).

### 3.2.3 Identification of differential isoforms and clustering

The RNA-seq libraries were aligned to the *Bombus terrestris* Bter1.0 reference genome (Refseq accession number GCF\_000214255.1) [188] using Tophat2 [244]. The aligned transcripts were assembled using cufflinks (version 2.2.1) as described by Trapnell et al. [245]. Cuffmerge was used to create a merged transcriptome annotation from the transcripts of all eighteen libraries. Then cuffdiff identified differentially expressed transcripts between samples from different colonies and with a different reproductive status (reproductive versus non-reproductive). The *cummeRbund* [246] package in R (version 3.3.0) was used for further downstream processing [7].

### 3.2.4 Differential expression

After Tophat2 alignment (see Section 3.2.3), samtools and bcftools were used for filtering and identifying variants. A general linear model (GLM) was conducted in the R (version 3.3.0) [7] package *edgeR* (version 3.14.0) [5, 6] to identify differentially expressed genes between reproductive and non-reproductive workers.

### 3.2.5 DNA methylation analyses

BS-seq libraries were aligned to same reference genome (Bter.1.0) using Bismark [247] v.0.16.1 and bowtie2 [248] v.2.2.6 with standard parameters. Bismark was also used to extract methylation calls and generate M-bias plots. After examination of the M-bias plots up to 8bp were removed from the 5' end of each read and up to 5bp were removed from the 3' end of each read. The anti-sense strand of the reproductive Colony 5 sample was the exception, with 25bp being removed from the 3' end. Alignment and methylation calling was then re-run with the new trimmed reads.

Two methods were used to identify differentially methylated regions (DMRs), SeqMonk and the *bsseq* R package (described below). Subsequently the DMRs found in both analyses were compared, and the regions found in both lists were deemed to make up the more conservative final list of DMRs used in the downstream analysis.

First, SeqMonk [249] was used to identify differentially methylated regions between reproductive and non-reproductive worker samples. Coverage outliers and duplicates were removed and sample replicate sets were defined along with windows of 20 CpGs. Average window length was assessed and outliers were removed, leaving windows between 20-3000bp. Methylation levels were quantified using the bisulfite methylation quantitation pipeline available in SeqMonk, with a minimum of 4 reads in every sample per CpG. Logistic regression analysis was then used to determine

windows which significantly differed between the two sample sets, with a minimum percentage difference in methylation levels of 10%.

Coverage files generated from Bismark were also used to carry out differential CpG methylation analysis using the *bsseq* [250] (v.1.11.0) package in R (version 3.3.0) [7]. Data were smoothed and paired t-tests were carried out on each CpG with correction for multiple testing. Differentially methylated regions were then calculated.

The DMRs generated from *bsseq* were read into SeqMonk and the logistic regression results were filtered based on overlap with the *bsseq* DMRs. These common regions were then further filtered producing a list of regions also overlapping a gene.

### 3.2.6 Gene ontology annotation

The loci (the SNP position +/- 2900bp) identified as showing differential expression, differential methylation, and alternative splicing, were blasted (Blastx) against *Drosophila melanogaster* proteins (non-redundant, nr, database) [201]. The blast results were annotated using Blast2Go [202]. Fisher's exact test was implemented to identify enriched GO terms compared with the list of GO terms from the *B. terrestris* reference transcriptome. Enriched GO terms were then visualised using REVIGO [4]. REVIGO is a web server that summarizes lists of GO terms in graphical formats by using a similarity clustering algorithm.

### 3.2.7 Comparisons: alternative splicing, methylation, and expression

In R (version 3.3.0) [7], a multiple logistic regression was applied to methylation percentages of each gene versus expression counts, and methylation percentages were also measured against FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values which are a measure of relative gene expression.



A hypergeometric test [251] was applied to identify if more genes and gene ontology terms overlapped between regions with differential expression, differential methylation, and alternative splicing than would be expected by chance.

## 3.3 Results

### 3.3.1 RNA-seq and BS-seq mapping efficiencies

RNA-seq reads were mapped with a mean efficiency of 92.2% (91.0-92.7%). The mean mapping efficiency was 63.6% (63.0-65.5%) and the mean coverage per gene was 23.2 reads (20.2-24.2) for the BS-seq. The lower mapping efficiency for the BS-seq is expected for bisulphite-treated DNA. Usually only 40-80% of reads from bisulphite-treated DNA can be mapped to a reference genome [252].

### 3.3.2 Differential gene expression

A total of 12,072 genes were identified from all eighteen *Bombus terrestris* transcriptome samples. After implementing a principal component analysis, non-reproductive workers were higher on the PC2 axis compared to reproductive workers of the corresponding colony (Figures 3.4 and 3.5). This indicates a reproductive status effect. There was also a colony effect as reproductive and non-reproductive colony 8 workers clustered more closely together than with any other samples. This implies more similar gene expression patterns between colonies 1 and 5 compared to colony 8.

After applying a two-way general linear model (GLM) with colony and reproductive status as factors, 849 genes were up-regulated in reproductive workers compared to non-reproductive workers, and 344 genes were found to be significantly up-regulated in non-reproductive workers compared to reproductive worker bees (GLM [5], FDR

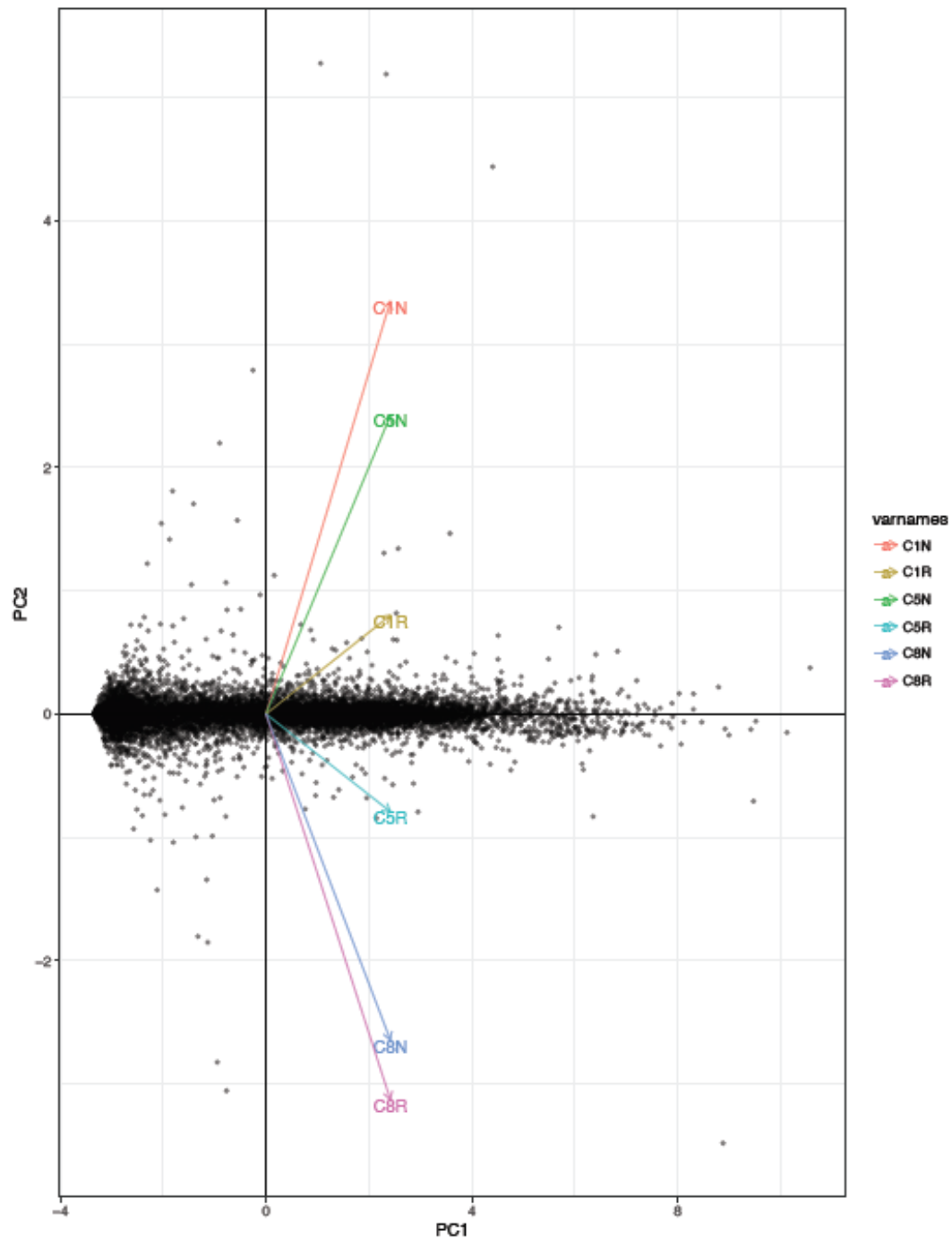


FIGURE 3.4: Principal component analysis plot based on the 12,072 genes from all 18 *Bombus terrestris* transcriptome samples, merging replicates of the same colony and reproductive status. Each point represents a gene. R refers to reproductive workers, and N refers to non-reproductive workers. C1, C5, and C8 are the three different *B. terrestris* colonies. For example, C1R refers to reproductive workers from colony 1.

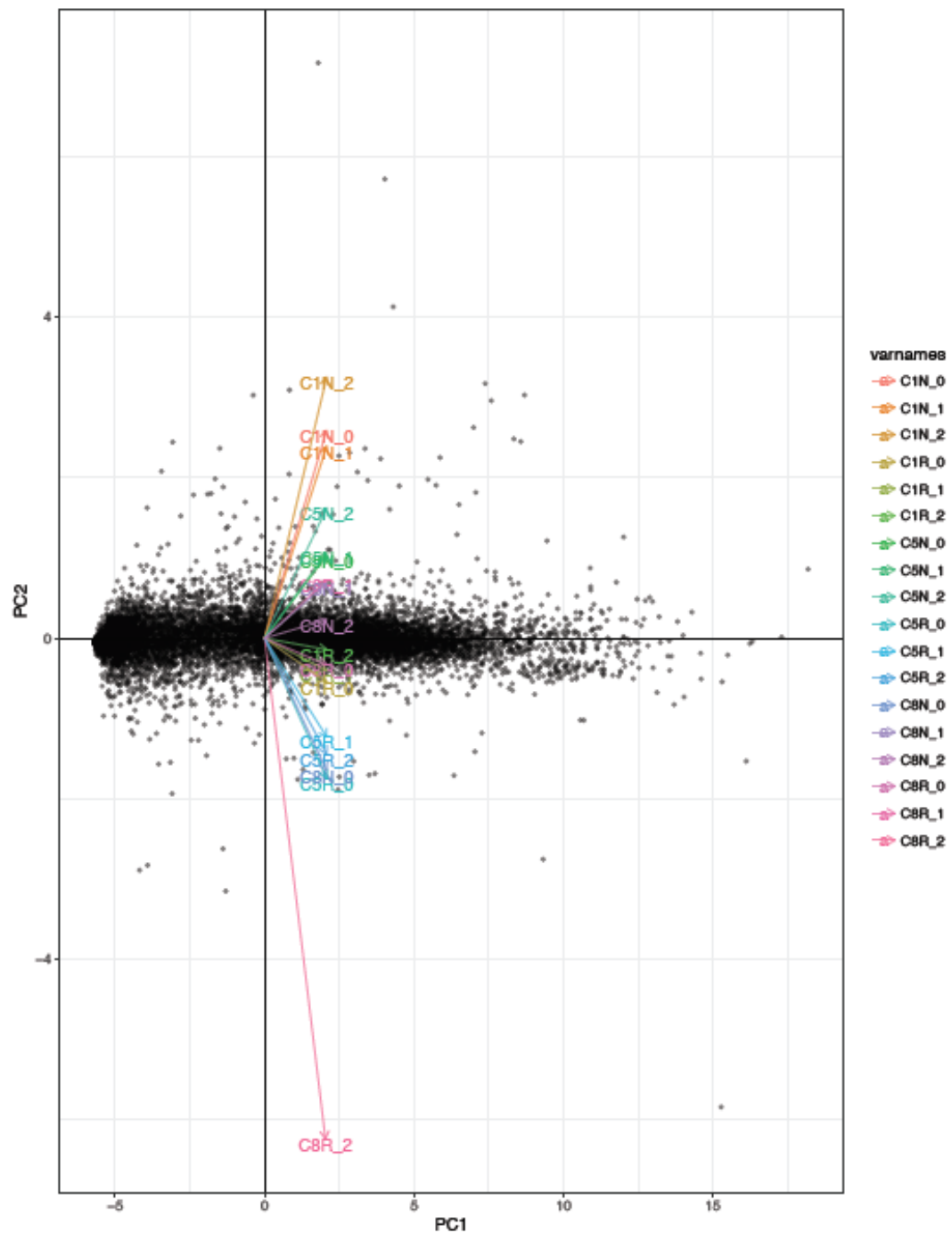


FIGURE 3.5: Principal component analysis plot based on the 12,072 genes from all 18 *Bombus terrestris* transcriptome samples, with independent replicates. Each point represents a gene. R refers to reproductive workers, and N refers to non-reproductive workers. C1, C5, and C8 are the three different *B. terrestris* colonies. 0, 1, and 2 refer to the three different replicates for each sample type. For example, C1R.0 refers to the first reproductive worker sample from colony 1.

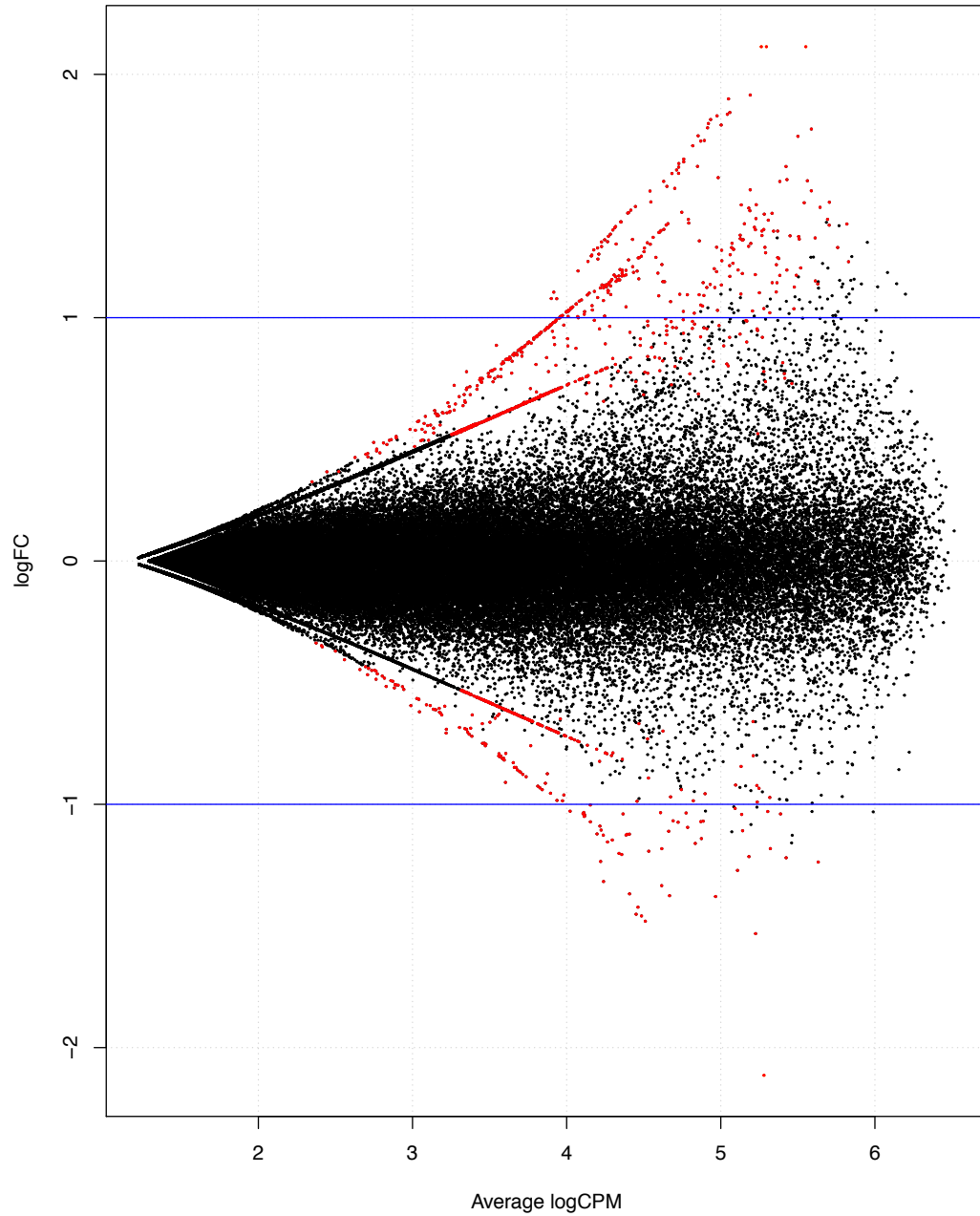


FIGURE 3.6: MA plot based on the 12,072 genes from all 18 *Bombus terrestris* transcriptome samples. Red data points show genes which are significantly differentially expressed between reproductive and non-reproductive workers ( $\text{FDR} < 0.05$ ) [5, 6]. Red data points with  $\text{logFC} > 0$  are up-regulated in reproductive workers compared with non-reproductive workers. Red data points with  $\text{logFC} < 0$  are up-regulated in non-reproductive workers compared with reproductive workers.

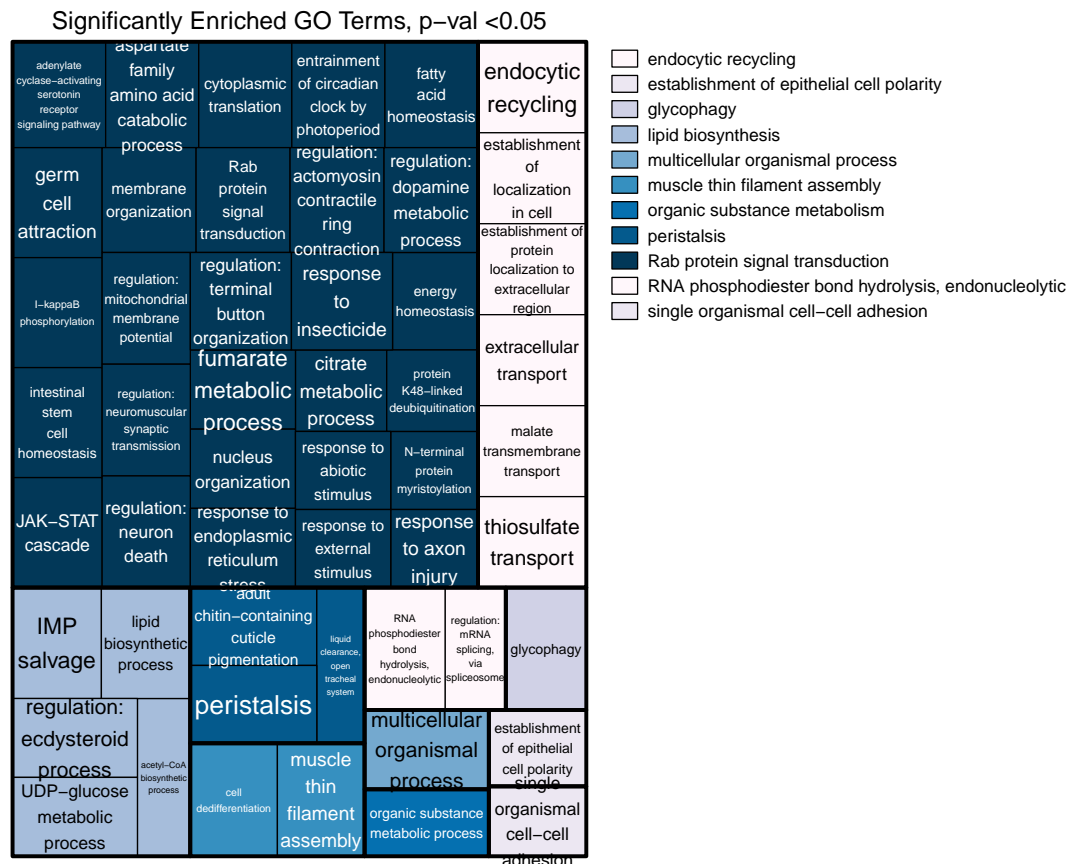


FIGURE 3.7: Significantly enriched gene ontology terms up-regulated in reproductive workers (REVIGO) [4]. Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment.

< 0.05). The eighteen transcriptome libraries had a dispersion value of 2.30891 and a biological coefficient of variation (BCV) of 1.5195 (Figure 3.6).

These significantly differentially expressed genes were compared against the *Drosophila melanogaster* protein non-redundant (nr) database (blastx). 610 out of the 849 loci up-regulated in reproductive workers had blast hits, with 339 of these loci also showing annotation. Of particular interest, the gene *ecdysone* was found to be up-regulated in reproductive workers. 247 out of the 344 loci up-regulated in non-reproductive workers had blast hits, with 145 of these also showing annotation.

320 gene ontology (GO) terms were enriched in regions up-regulated in reproductive

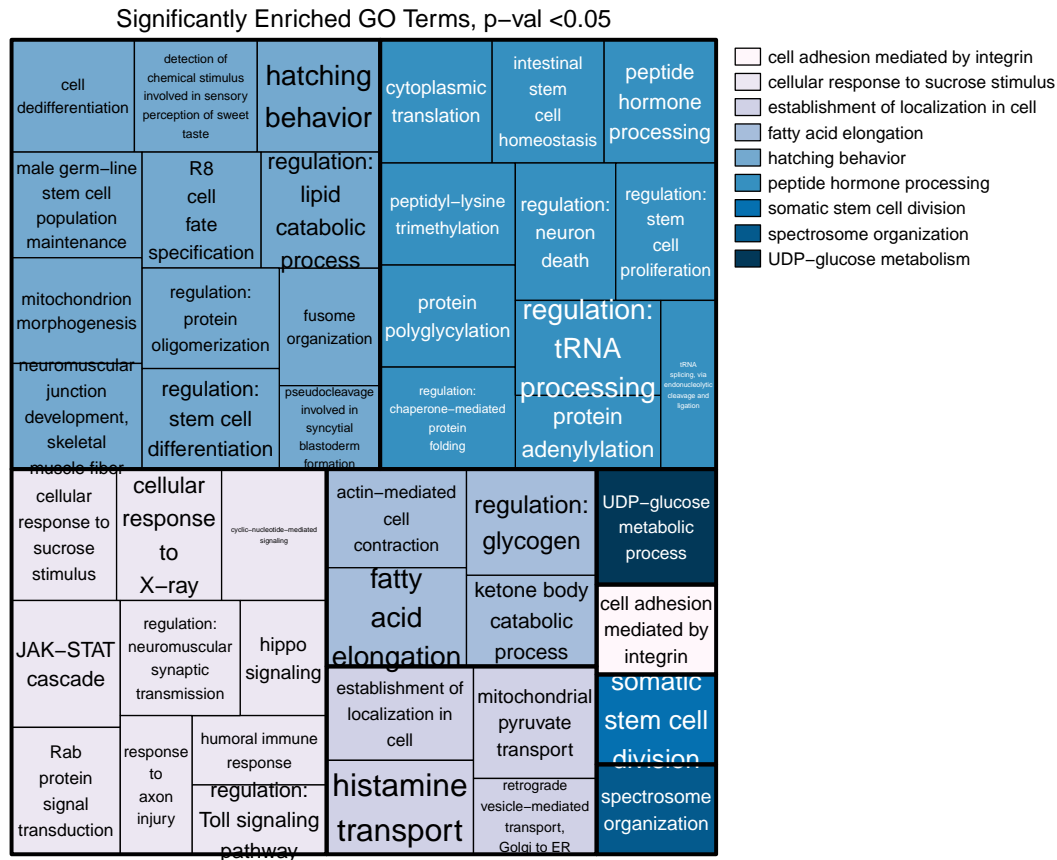


FIGURE 3.8: Significantly enriched gene ontology terms up-regulated in non-reproductive workers (REVIGO) [4]. Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment.

workers (Figure 3.7), and 264 GO terms were enriched in the regions up-regulated in non-reproductive workers (Figure 3.8) (Fisher’s exact test:  $p < 0.05$ ). Functions associated with epigenetic regulation and signalling were present in both groups. For instance, genes involved in the production of miRNAs involved in gene silencing and genes related to the regulation of mRNA splicing via the spliceosome were found to be enriched in regions up-regulated in reproductive workers. Genes involved in peptide hormone processing (including histone H3-K27 acetylation and RNA splicing) were significantly enriched in regions up-regulated in non-reproductive workers.

### 3.3.3 Differential splicing of isoforms

A total of 51,153 isoforms were found to be expressed from the 12,072 genes found in the eighteen *B. terrestris* transcriptome libraries. After filtering for an FPKM value greater than 0.5, 6,901 of the 12,072 genes showed more than one isoform. These 6,901 genes had a total of 45,452 corresponding isoforms and will hence be known as multi-isoform loci.

The expression patterns of non-reproductive worker multi-isoform loci from colonies 1 and 5 were more similar to each other (Jenson-Shannon distance = 0.0421) than compared to reproductive workers of the same colonies (Jenson-Shannon distance = 0.0582, and 0.0482) (Figures 3.9). Reproductive workers in colonies 1 and 5 had expression patterns more similar to each other (Jenson-Shannon distance = 0.0406) than compared to non-reproductive workers of the same colony. However, in colony 8 there is a colony effect. Reproductive and non-reproductive worker gene expression is more similar in colony 8 than compared to workers of other colonies. Isoforms of replicate 2 of the reproductive workers from colony 8 exhibited a greater difference in expression compared to all other libraries (Figure 3.10).



FIGURE 3.9: Correlation of expression of multiple-isoform loci based on Jenson-Shannon distance, showing merged replicates. Lower Jenson-Shannon values indicate greater similarity of expression patterns. R denotes reproductive workers, and NR indicates non-reproductive workers.

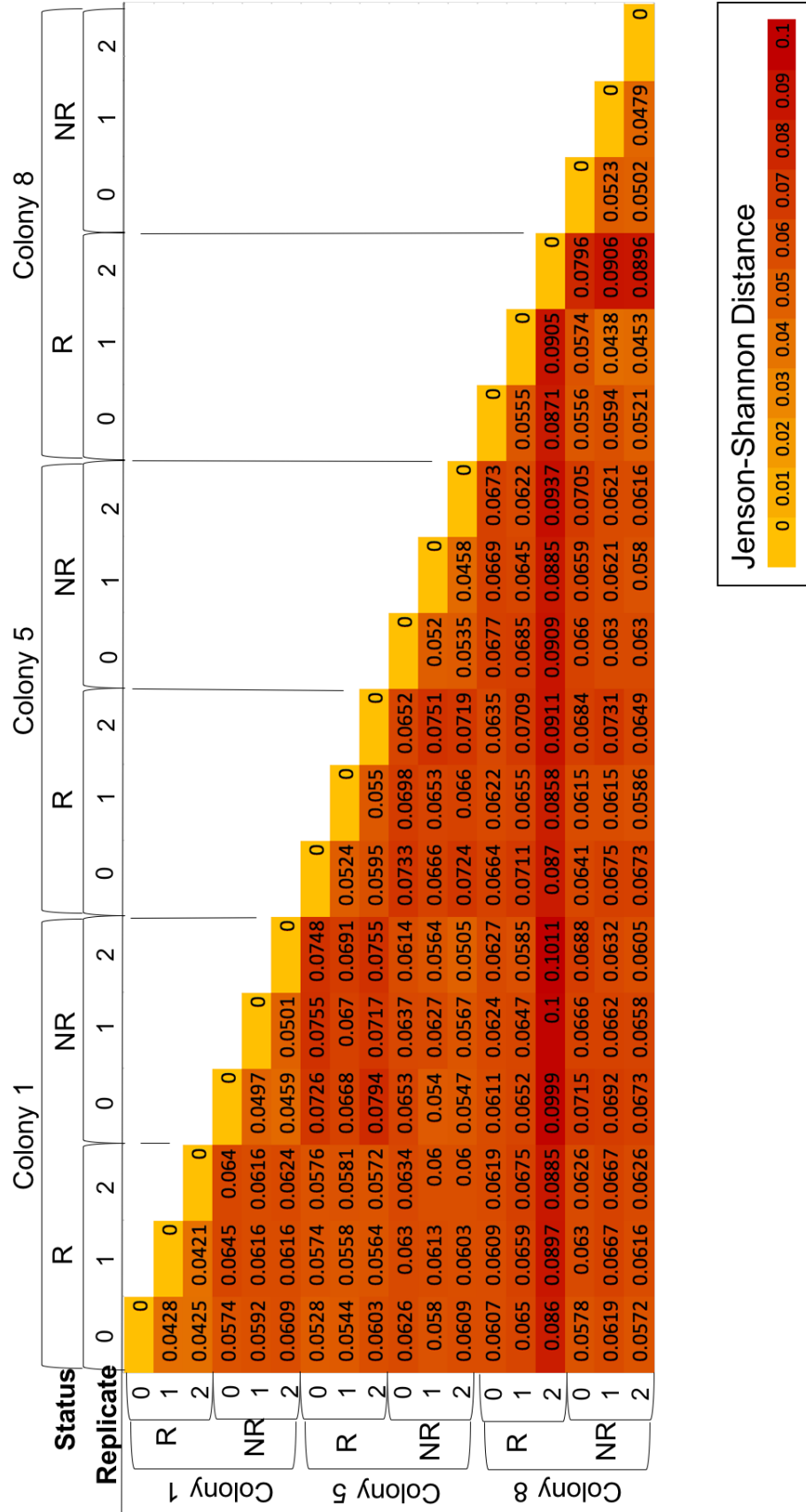


FIGURE 3.10: Correlation of expression of multiple-isoform loci based on Jensen-Shannon distance, showing independent replicates. Lower Jensen-Shannon values indicate greater similarity of expression patterns. R denotes reproductive workers, and NR indicates non-reproductive workers. 0, 1, and 2 refer to the three different replicates for each sample type.



After applying a principal component analysis a colony effect is seen with samples corresponding more closely within colonies compared to reproductive status (Figure 3.11). However, the expression patterns of the isoforms of reproductive and non-reproductive worker multi-isoform loci still separate in the same direction. The reproductive worker samples are positioned higher on the PC2 axis than the corresponding non-reproductive workers of the same colony.

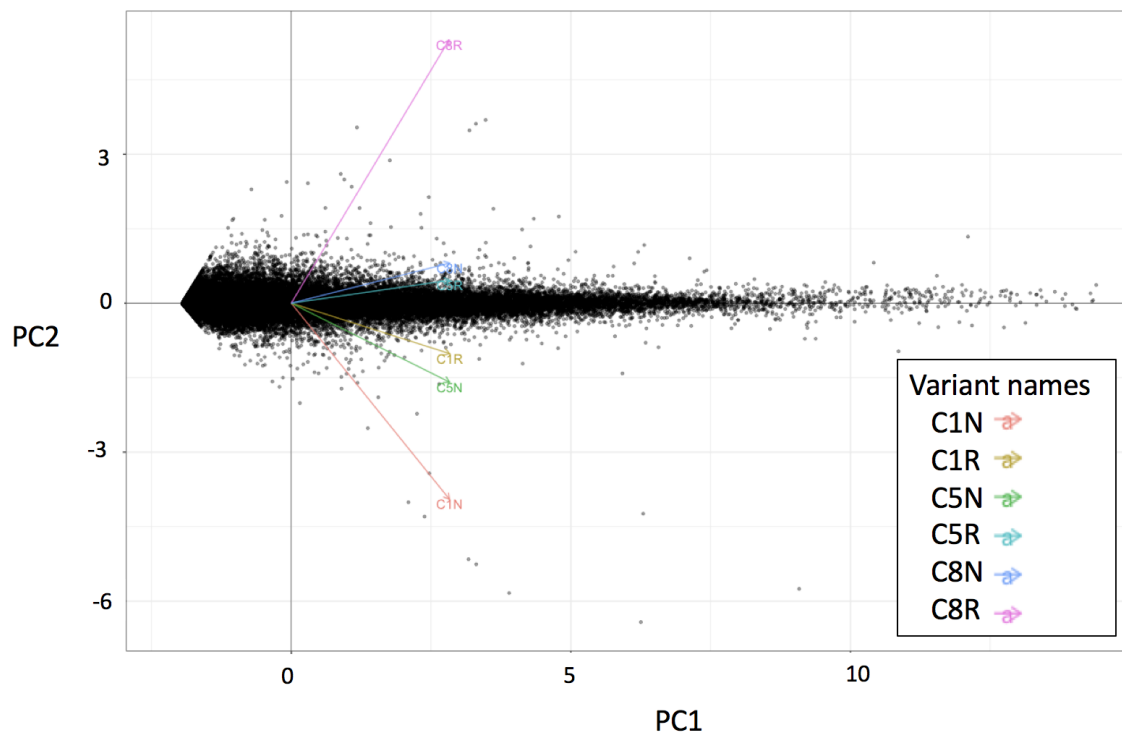


FIGURE 3.11: Principal component analysis of the isoform expression of multiple-isoform loci, merging replicates. R refers to reproductive workers, and N refers to non-reproductive workers. C1, C5, and C8 are the three different *B. terrestris* colonies. For example, C1R refers to reproductive workers from colony 1.

From the multi-isoform loci, 584 significantly differentially expressed isoforms were identified between the eighteen transcriptome libraries (FDR adjusted  $p < 0.05$ ). Of these, 61 isoforms were significantly differentially expressed between reproductive and non-reproductive workers in colony 1, 62 isoforms were differentially expressed between reproductive and non-reproductive workers in colony 5, and nine isoforms were differentially expressed between reproductive and non-reproductive workers in

colony 8 (FDR adjusted  $p < 0.05$ ). Thus a total of 103 isoforms were significantly differentially expressed between reproductive and non-reproductive workers of the same colony. 23 of these isoforms were differentially expressed in both colony 1 and colony 5, five isoforms were differentially expressed in colonies 1 and 8, and five isoforms were differentially expressed in colonies 5 and 8 (Figure 3.12). Four isoforms were significantly differentially expressed between reproductive and non-reproductive workers in all three colonies (Table 3.1, Figure 3.13). Of these four isoforms, two have been previously found to be components of Hymenoptera venom and this is discussed further in Section 3.4.

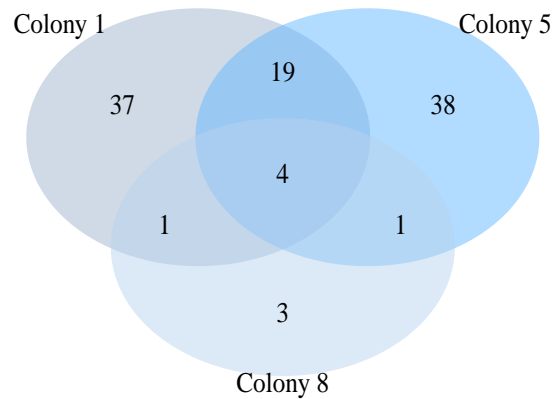


FIGURE 3.12: Significantly differentially expressed isoforms between reproductive and non-reproductive workers within the same colony. Total of 103 significantly differentially expressed isoforms.

TABLE 3.1: Multi-isoform genes that are differentially expressed between reproductive and non-reproductive workers in three *B. terrestris* colonies.

Transcript ID	Gene ID	Gene Name
TCONS_00048635	LOC100651307	Serine protease inhibitor 3-like
TCONS_00006883	LOC100647178	Venom acid phosphatase Acph-1-like
TCONS_00018880	LOC100644966	Uncharacterized
TCONS_00035686	LOC100643625	Putative pyridoxamine 5'-phosphate oxidase

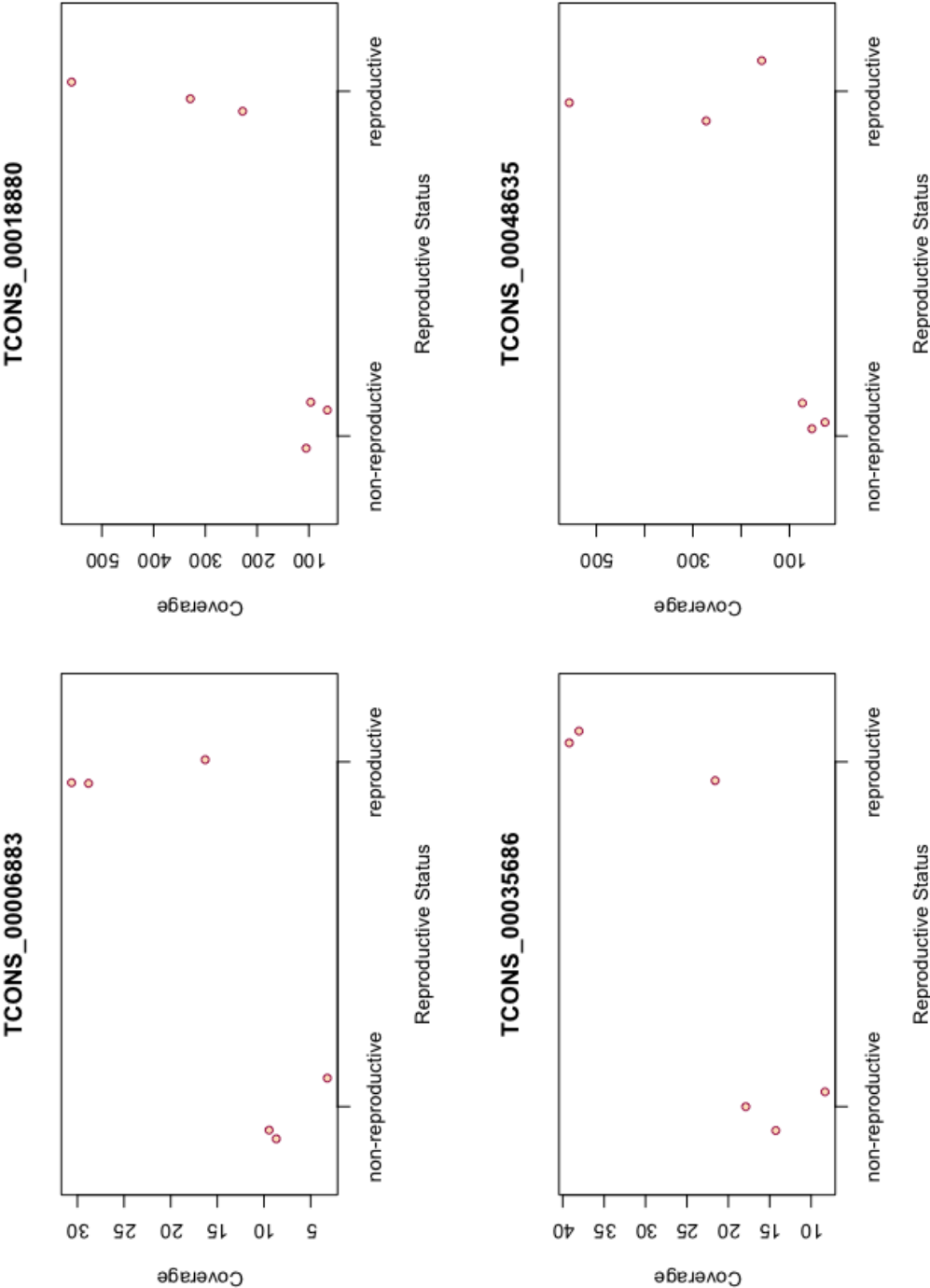


FIGURE 3.13: Coverage (total number of reads) of the four multi-isoform genes with differentially expression between reproductive and non-reproductive workers of the three *B. terrestris* colonies.

87 enriched gene ontology (GO) terms were found for the 103 genes of the isoforms significantly differentially expressed between reproductive and non-reproductive workers of the same colony (Figure 3.14). Of particular interest, the regulation of histone methylation was enriched in this dataset, which indicates histone methylation is differentially regulated between reproductive and non-reproductive workers of the same colony.

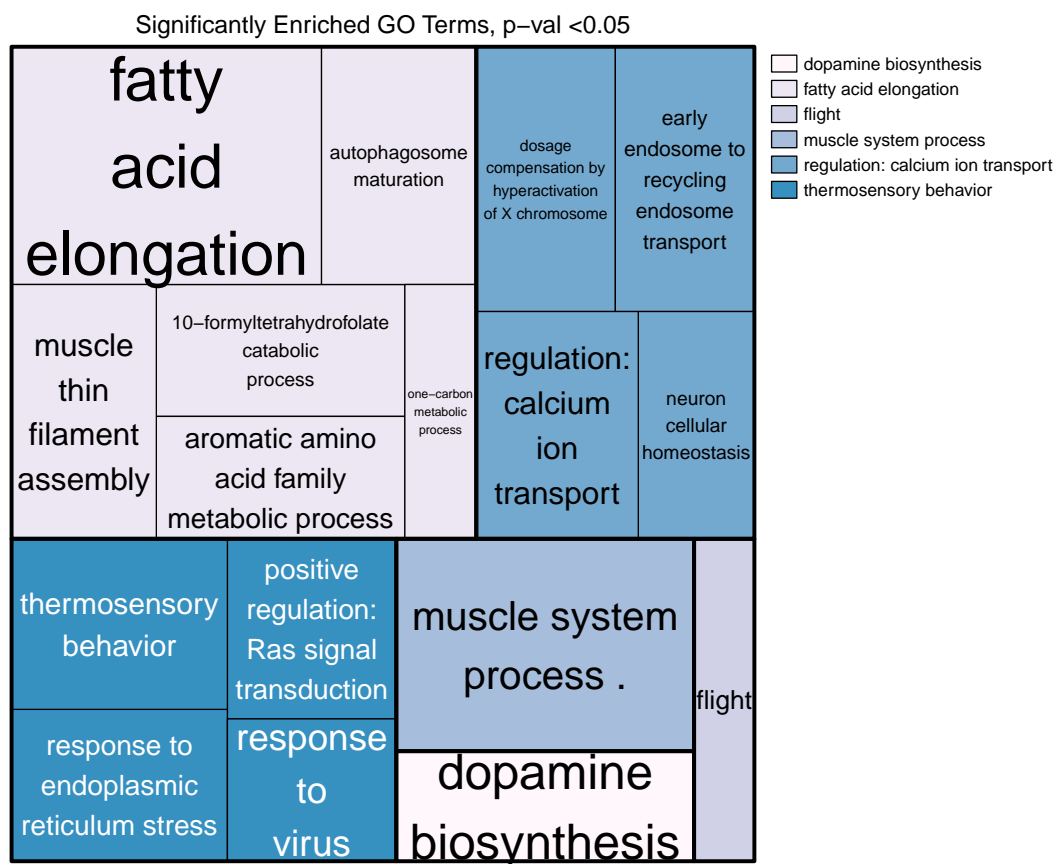


FIGURE 3.14: Enriched gene ontology terms of the genes of the 103 differentially expressed isoforms between reproductive and non-reproductive workers within the same colony (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment.

### 3.3.4 Differential methylation patterns

203 genes were differentially methylated between reproductive and non-reproductive worker bees. Of these, 79 were hypermethylated in reproductive workers compared to non-reproductive workers. 94 showed hypomethylation in reproductive workers compared with non-reproductive workers. 35 genes displayed instances of both hypermethylation and hypomethylation, indicating that in the same gene some CpGs showed higher levels of methylation and other CpGs showed reduced levels of methylation in reproductive compared to non-reproductive workers. This may be resultant of differential splicing of isoforms of the same gene, or simply due to cis effects [253]. The ten most differentially methylated genes are listed in Table 3.2 and are largely associated with epigenetic mechanisms including chromatin organisation, mRNA splicing, and histone deacetylation.

245 GO terms were found to be enriched in the genes that showed differential methylation between reproductive and non-reproductive worker bees (Fisher's exact test,  $p < 0.05$ ) (Figure 3.15). 260 GO terms were found to be enriched in the genes that showed hypermethylation in reproductive worker bees compared to non-reproductive workers, and 320 GO terms were found to be enriched in the genes that showed hypomethylation in reproductive workers. Of notable interest, the enriched GO terms included positive regulation of the Ecdysteroid process (GO:005998), regulation of alternative mRNA splicing (GO:0000381), gene silencing by RNA (GO:0060966), epigenetic gene regulation (GO:0040029), positive epigenetic gene regulation (GO:0045815), and inter-male aggressive behaviour (GO:0002121).

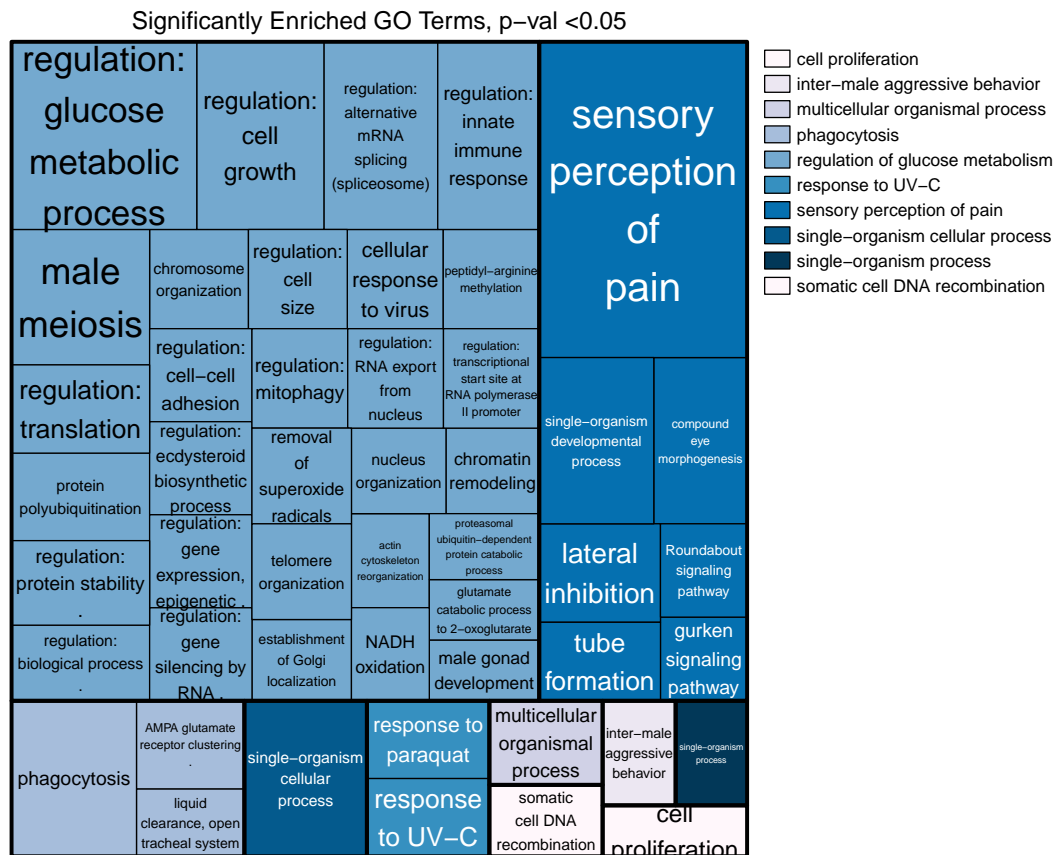


FIGURE 3.15: Enriched gene ontology terms of the differentially methylated genes in reproductive and non-reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the  $p$ -value associated with that clusters enrichment.

TABLE 3.2: The top ten differentially methylated genes between reproductive and non-reproductive bumblebee workers. Gene function was inferred from conserved protein domains identified on NCBI.

Gene ID	Description	Function	Non-Repro Methylation (%)	Repro Methylation (%)	FDR
LOC100644478	NAD-dependent protein deacetylase Sirt6	Gene silencing via chromatin, life span	37.5	76.6	<0.01
LOC100645572	threonylcarbamoyladenosine tRNA methylthiotransferase	Protein synthesis, tRNA and rRNA base modification	36.9	72.3	<0.01
LOC100645370	protein phosphatase 1B	Regulating cellular responses to stress	29.6	64.1	<0.04
LOC100643120	RILP-like protein homolog	Membrane trafficking	49.3	71.4	<0.01
LOC100650146	retinol dehydrogenase 12	Retinol metabolism	0	21.4	<0.01
LOC100651986	chromodomain-helicase-DNA-binding protein 1	Chromatin organisation, histone deacetylation	43.8	23.1	<0.04
LOC100646999	Kv channel-interacting protein 2	Calcium sensing, protein inactivation/activation	74.7	53.5	<0.01
LOC100646653	protein phosphatase 1 regulatory subunit 7	Protein-protein interactions	36.0	9.3	<0.001
LOC100650375	integrator complex subunit 2	snRNA transcription and processing	45.8	17.4	<0.04
LOC100651995	cleavage and polyadenylation specificity factor subunit 6	mRNA splicing	33.0	2.8	<0.001

### 3.3.5 Global expression and methylation patterns

There was a higher degree of methylation in exons and CDS regions, compared to that of introns (Figure 3.16). Similar patterns were witnessed in reproductive and non-reproductive workers. This is consistent with the methylation patterns found in other studies of Hymenoptera [122, 238].

Relationships between methylation and expression levels showed only a slight correlation at the gene level (Figures 3.17a and 3.17b), which is consistent with previous studies [44, 122]. Yet there is a clear association between genome-wide methylation and expression, with higher levels of methylation witnessed in more highly expressed genes (Figure 3.18). After applying a linear model, there was a significant correlation between expression rank and methylation level (ANCOVA:  $F = 1588.044$ ,  $df = 1$ ,  $p < 0.0001$ ), but no significant relationship between expression rank and reproductive status (ANOVA:  $F = 1.965$ ,  $df = 1$ ,  $p = 0.161$ ). This pattern is similar to that observed in the ants *Camponotus floridanus* and *Harpegnathos saltator* by Bonasio and colleagues [122]. The methylation level was slightly lower for those genes with the highest expression levels in Bonasio et al. [122], but no such drop in methylation at high expression levels was seen in the present study.



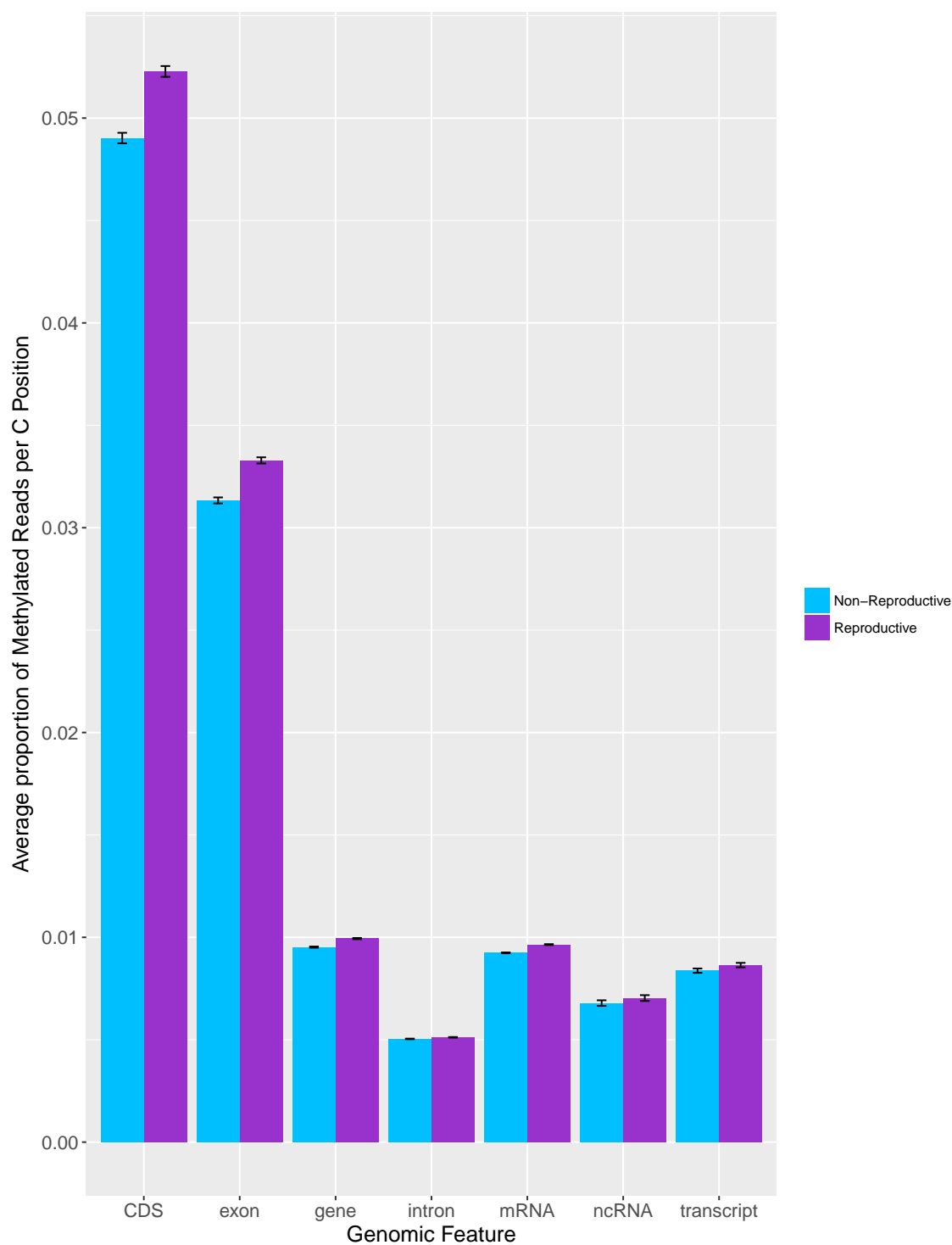


FIGURE 3.16: Mean methylation levels (mean methylated reads per cytosine position) of genomic features in reproductive and non-reproductive workers. Error bars are 95% confidence intervals.

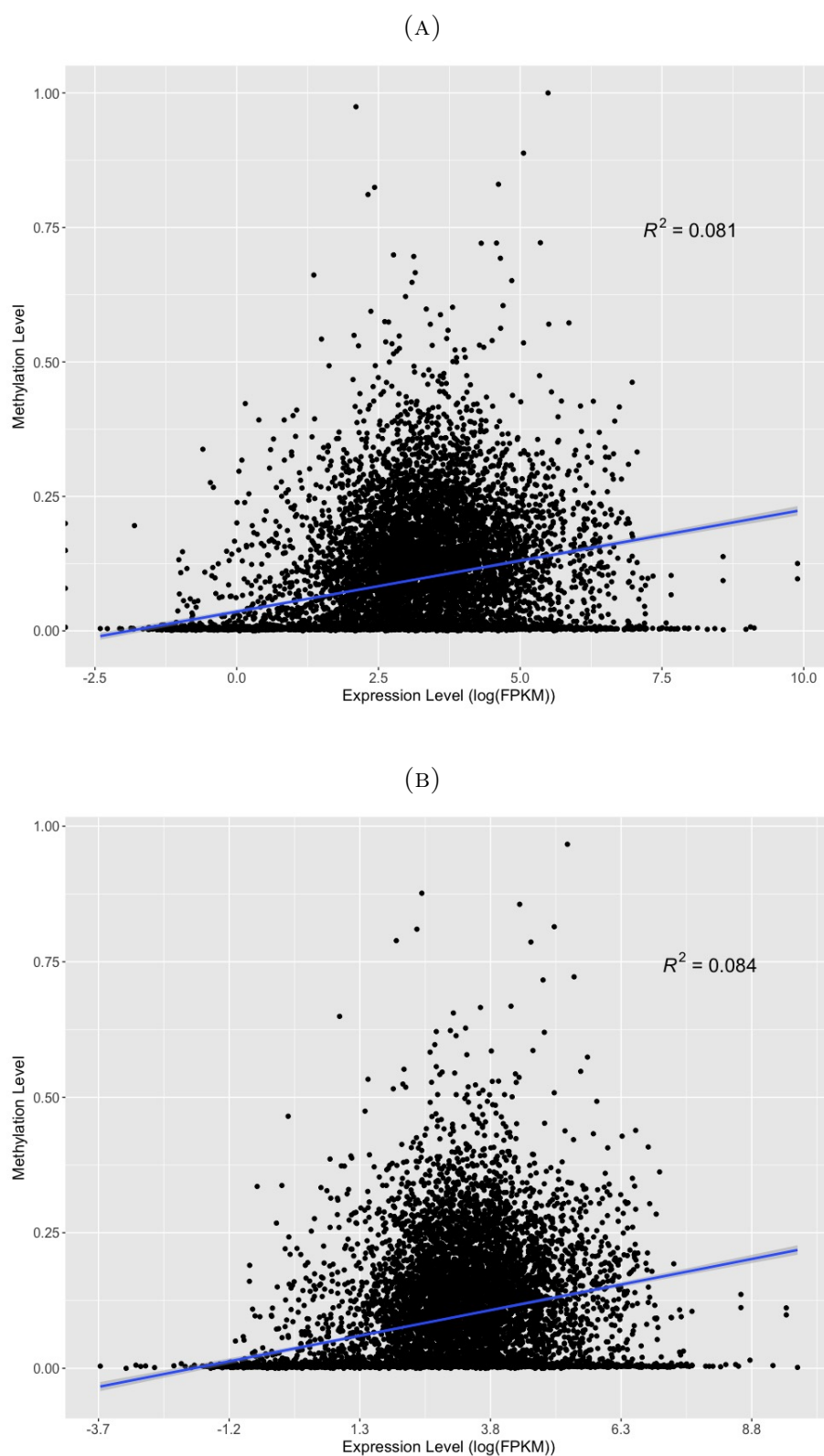


FIGURE 3.17: Methylation level (percentage of CpGs methylated) versus expression level (log-transformed FPKM) for (A) reproductive and (B) non-reproductive workers. Each point represents a gene.

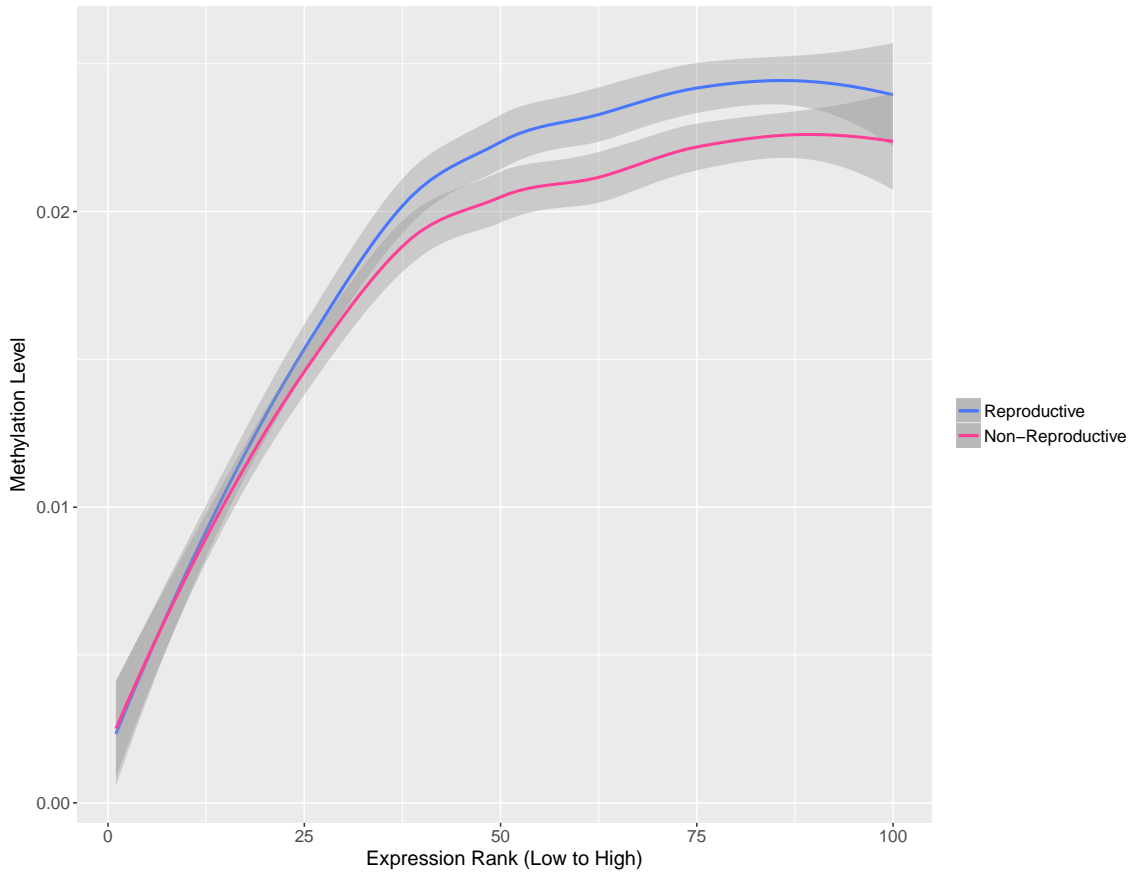


FIGURE 3.18: Gene body methylation (percentage of whole gene methylated) versus binned gene expression (Number of total bins = 100, where genes in Bin 0 are least expressed and those in Bin 100 are most highly expressed). Lines were generated using the LOESS local polynomial regression method implemented by *ggplot2* in R [7].

There was a significant difference in methylation per CpG between differentially expressed genes (DEG) and non-DEG (t-test:  $t = -7.2494$ ,  $df = 130.11$ ,  $p < 0.0001$ ) with higher methylation levels in non-DEG (Figure 3.19). When only focusing on the DEG, there was a trend of higher levels of methylation in genes with a greater degree of differential expression between reproductive and non-reproductive workers after applying a linear model (ANOVA:  $F = 792.78$ ,  $df = 1$ ,  $p < 0.0001$ ) (Figure 3.20). Differentially expressed multi-isoform (DEI) genes showed lower gene body methylation compared to that of non-DEI genes (t-test:  $t = -5.516$ ,  $df = 100.49$ ,  $p < 0.0001$ ) (Figure 3.21). Interestingly, the same trend was significant for *Dinoponera*

*quadriceps* in Patalano et al. [44]. There was also a significant difference in mean expression levels between DEI genes and non-DEI genes (t-test:  $t = 5.9856$ ,  $df = 101.64$ ,  $p < 0.0001$ ) with a higher average expression percentage in DEI genes (Figure 3.22).

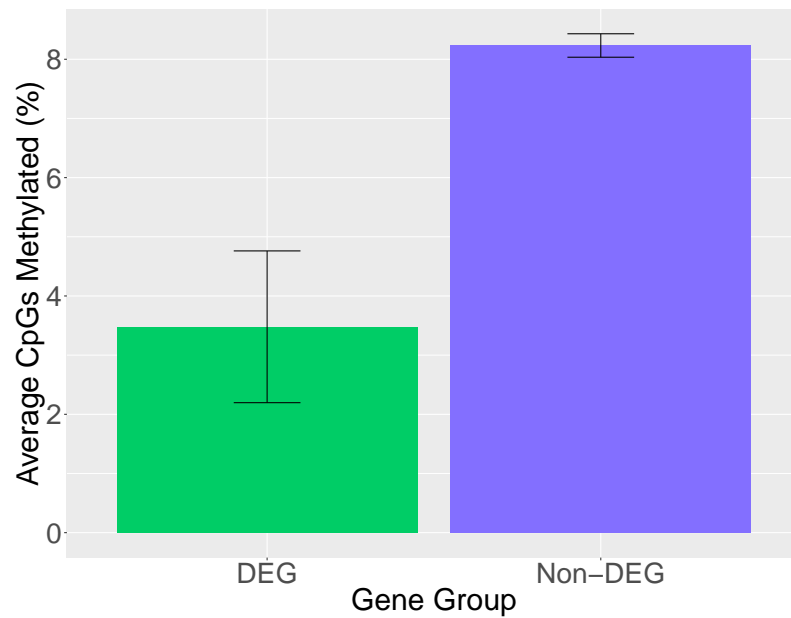


FIGURE 3.19: Mean methylation (percentage of CpGs methylated) levels of differentially and non-differentially expressed genes. Error bars are 95% confidence intervals.

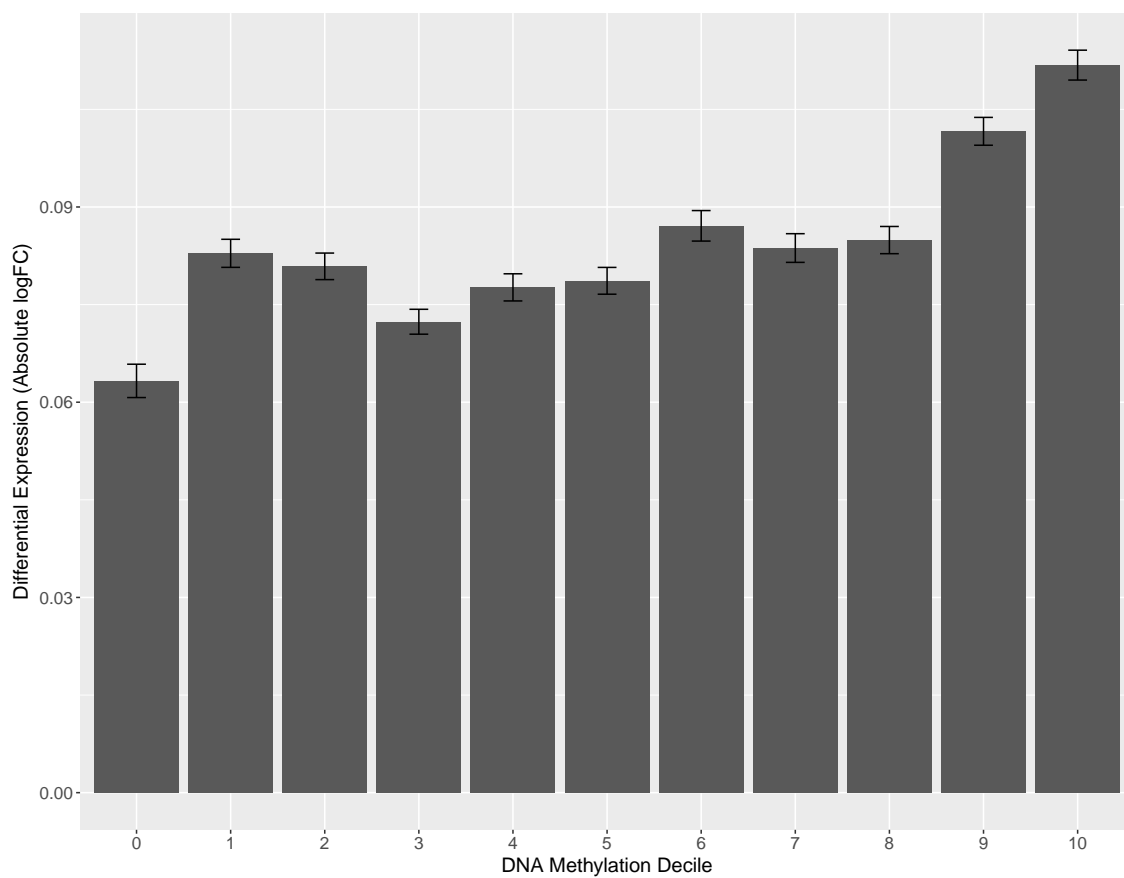


FIGURE 3.20: Differential expression (absolute logFC) and binned DNA methylation (percentage of CpGs methylated) (bins=10, where 1 is least methylated, 10 is most highly methylated, and 0 represents unmethylated genes) Error bars are 95% confidence intervals.

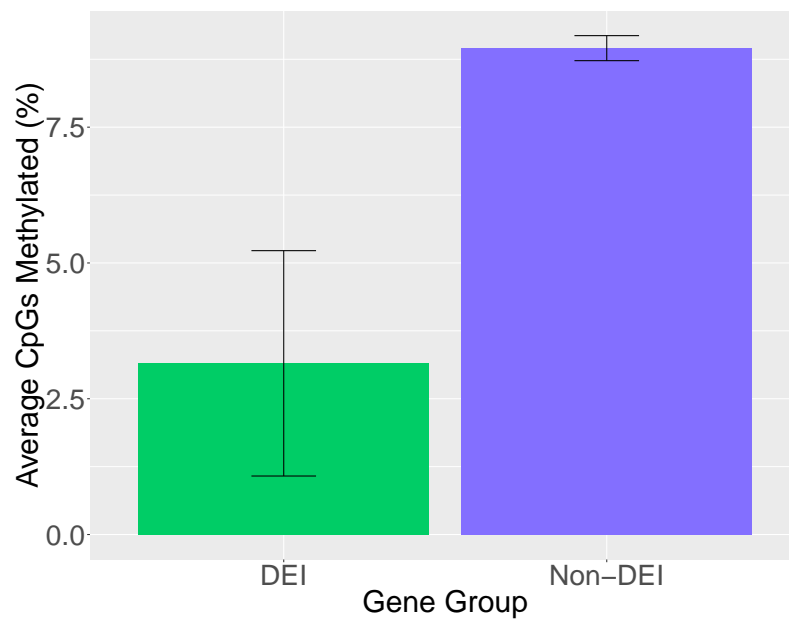


FIGURE 3.21: Mean methylation (percentage of CpGs methylated) levels of differentially and non-differentially expressed multi-isoform genes. Error bars are 95% confidence intervals.

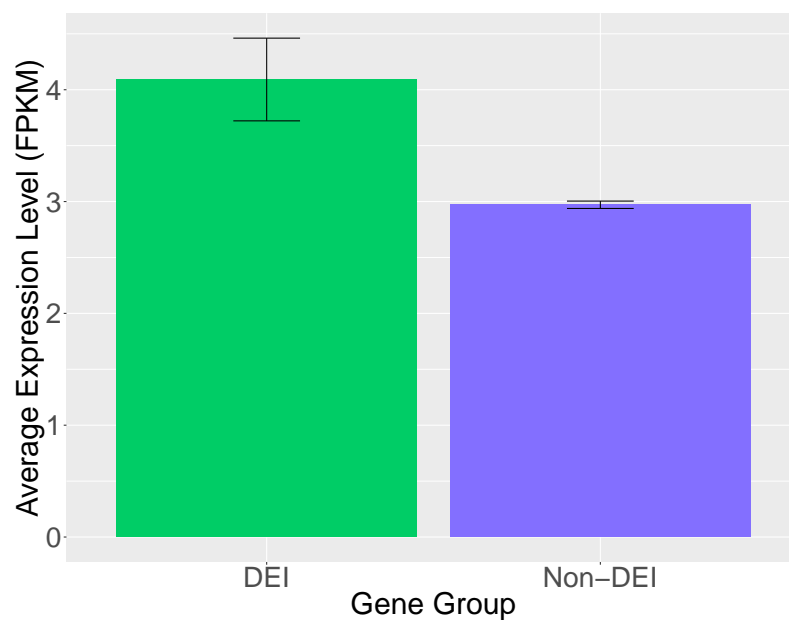


FIGURE 3.22: Mean expression (FPKM) levels of differentially and non-differentially expressed multi-isoform genes. Error bars are 95% confidence intervals.

TABLE 3.3: Comparing all known genes (LOC IDs) with differential expression, methylation, and alternative splicing with the hypergeometric test.

Gene set 1	Gene set 2	Genes overlapped	p	Significant overlap
Up regulated <sup>a</sup>	Down regulated <sup>b</sup>	11	<0.0001	yes
Up regulated <sup>a</sup>	Diff. Isoforms	0	-	-
Up regulated <sup>a</sup>	Hypermethylation	0	-	-
Up regulated <sup>a</sup>	Hypomethylation	1	0.65	no
Down regulated <sup>b</sup>	Diff. Isoforms	0	-	-
Down regulated <sup>b</sup>	Hypermethylation	0	-	-
Down regulated <sup>b</sup>	Hypomethylation	0	-	-
Diff. Isoforms	Hypermethylation	0	-	-
Diff. Isoforms	Hypomethylation	0	-	-

<sup>a</sup> Gene expression up regulated in reproductive workers compared to non-reproductive workers.<sup>b</sup> Gene expression down regulated in reproductive workers compared to non-reproductive workers.

TABLE 3.4: Comparing enriched Gene Ontology (GO) terms with differential expression, methylation, and alternative splicing with the hypergeometric test.

GO set 1	GO set 2	GO terms overlapped	p	Significant overlap
Up regulated <sup>a</sup>	Down regulated <sup>b</sup>	36	<0.0001	yes
Up regulated <sup>a</sup>	Diff. Isoforms	0	-	-
Up regulated <sup>a</sup>	Hypermethylation	0	-	-
Up regulated <sup>a</sup>	Hypomethylation	0	-	-
Down regulated <sup>b</sup>	Diff. Isoforms	4	0.00023	yes
Down regulated <sup>b</sup>	Hypermethylation	0	-	-
Down regulated <sup>b</sup>	Hypomethylation	0	-	-
Diff. Isoforms	Hypermethylation	4	0.00022	yes
Diff. Isoforms	Hypomethylation	2	0.18	no

<sup>a</sup> Gene expression up regulated in reproductive workers compared to non-reproductive workers.<sup>b</sup> Gene expression down regulated in reproductive workers compared to non-reproductive workers.

Genes and GO terms present in multiple analyses out of the expression, methylation, and alternative splicing analyses were identified. Hypergeometric tests were carried out to determine whether more genes and GO terms overlapped in these analyses more than expected (hypergeometric test,  $p < 0.05$ ) (Tables 3.3 and 3.4). Eleven genes overlapped between up- and down-regulated regions (Appendix B), and 36 gene ontology terms were also found to overlap between these categories. The gene ontology terms which overlapped between up-regulated and down-regulated

regions included genes functions associated with splicing (GO:0000974 - generation of catalytic spliceosome for first transesterification step) and the female germline (GO:0007301 - female germline ring canal formation, GO:0007312 - oocyte nucleus migration involved in oocyte dorsal/ventral axis specification) (Appendix C). Down regulated GO terms in reproductive workers also significantly overlapped with those with differential isoform expression, as did GO terms associated with hypermethylation and differential isoform expression (Appendix C).

### 3.4 Discussion

This is the first instance in which differential expression, methylation, and alternative splicing status has been directly compared between reproductive and non-reproductive *Bombus terrestris* workers. Here, there are higher levels of gene body methylation for genes with a higher degree of expression. The average methylation level was higher for those genes and isoforms consistently expressed to the same level between reproductive and non-reproductive genes compared to differentially expressed genes and isoforms. This is reminiscent of a study of *Nasonia* that reported that housekeeping genes had higher levels of methylation than genes with other functions [123]. One would expect housekeeping genes to have uniform levels of expression between workers of a different reproductive status. Therefore it follows that genes that show consistent expression levels in bumblebees have higher levels of methylation on average.

Despite methylation levels being higher in consistently expressed genes compared to differentially expressed genes, a different pattern is seen when focusing on only the subset of genes with differential expression. Of those genes with differential expression between reproductive and non-reproductive workers, there is a higher degree of differential expression in those genes with more methylation. This suggests methylation to be involved in regulating gene expression levels in bees of alternate



reproductive status. This is in line with the assertions of Wang et al. who conclude that DNA methylation could play a role in maintaining differential expression [230].

Other epigenetic mechanisms potentially underlying these expression patterns are here indicated; gene ontology terms involved in RNA splicing and histone H3-K27 acetylation were enriched in regions up-regulated in non-reproductive workers (Figure 3.8). Production of gene silencing miRNAs and regulation of mRNA via the spliceosome were enriched in regions up-regulated in reproductive workers (Figure 3.7). In addition, histone methylation was associated with differentially regulated isoforms (Figure 3.14). Genes associated with the regulation of the ecdysteroid process and RNA splicing were also differentially methylated (Figure 3.15). Thus alternative splicing, miRNAs, ecdysteroids, and histone modifications via methylation and acetylation are all here implicated to be associated with bumblebee worker reproduction. These correlations are consistent with current literature. For instance, ecdysteroids were previously found to be associated with reproductive division of labour, being present at higher levels in dominant reproductive workers in *B. terrestris* queenless colonies [242].

Genes involved in histone acetylation were up-regulated in non-reproductive workers compared to reproductive workers. Histone methylation acts in preventing transcription of proteins through the tightening of the DNA strands around the histone core, whereas acetylation loosens the DNA strands thus allowing transcription [254]. Therefore these processes act antagonistically and the histone acetyltransferases contribute to histone demethylation [255]. Here, the up-regulation of genes involved in histone acetylation in non-reproductive workers suggests initiation of gene expression.

Four multi-isoform genes were found to be differentially expressed between reproductive and non-reproductive workers in all three *B. terrestris* colonies in this study. One of these genes was uncharacterised, yet the three characterised genes had gene functions of interest. Serine protease inhibitor 3-like was one of these four genes

that were differentially expressed. Serine protease and serine protease inhibitors are typically found in bumblebee venom [256, 257]. Serine proteases are fibrinolytic agents that facilitate the enzyme plasmin to breakdown fibrin inside blood clots. This facilitates the spread of bee venom throughout the victim's bloodstream. In contrast, serine protease inhibitors have been found to inhibit plasmin and exhibit antifibrinolytic activity [256, 257]. Antifibrinolytic agents such as serine protease inhibitors prevent the breakdown of fibrin in blood clots. Therefore serine protease inhibitors reduce bleeding at a sting site [257].

Another differentially expressed multi-isoform gene, venom acid phosphatase Acph-1-like, has also previously been identified to be present in venom. Acid phosphatase is considered to be a common component in Hymenoptera venom, being present in the venom of the honey bees *Apis mellifera* [258] and *Apis cerana* [259], as well as endoparasitoid and ectoparasitoid wasps [260–263]. Acid phosphatase is the allergen in bee venom which releases histamine from sensitized human basophils [258].

A putative pyridoxamine 5'-phosphate oxidase was found to be a multi-isoform gene that was differentially expressed between reproductive and non-reproductive workers in all three *B. terrestris* colonies. Pyridoxamine 5'-phosphate oxidase is an enzyme that is involved in the de novo synthesis of pyridoxine (vitamin B6) which is required for amino acid, glucose, and lipid metabolism [264]. Interestingly, in the honey bee pyridoxamine 5'-phosphate oxidase was previously found to be differentially expressed between bees trained to associate a floral odour with a sugar reward compared with control bees exposed only to air [265]. This implied its importance for learning and memory formation from olfactory signals.

Moreover, it is of interest to note the number of genes with differential methylation between reproductive and non-reproductive animals with regards to degree of eusociality in comparison with previous studies. Patalano et al. found there to be no differentially methylated regions between queen or “gamergate” reproductive castes and non-reproductive workers in the paper wasp (*Polistes canadensis*) and

the dinosaur ant (*Dinoponera quadriceps*), which are both very primitively eusocial species [44]. In the present study 203 differentially methylated genes were identified between primitively eusocial reproductive and non-reproductive *B. terrestris* workers. Whereas Lyko et al. found 561 differentially methylated regions between queen and worker honeybees (*Apis mellifera*) [115], a highly eusocial species. Therefore an increasing prevalence of differential methylation may assist with morphological and behavioural caste differentiation, that is associated with a higher degree of eusociality. A pattern such as this would be similar to the observations made by Simola and colleagues [42]. They found that the evolution of transcription factors is more divergent among eusocial insects than between solitary and eusocial insects.

In contrast, Bewick et al. [238] previously found no association between overall methylation levels and degree of sociality in insects. This led to their conclusions that methylation levels are highly variable across insects and methylation to not be a “universal driver” of social behaviour [238]. However, differential methylation patterns were not taken into account. It appears that overall methylation is not a determinant of sociality, yet the degree of differential methylation between castes could associate with insect sociality. The facilitation of different expression patterns is key to the distinct caste polyphenisms witnessed in more highly eusocial species and this is explored in regards to alternative splicing in Chapter 6.

In the future, it would be interesting to use these RNA-seq and BS-seq datasets to explore the exact correlations between alternative splicing and DNA methylation levels in a similar manner to the expression-methylation genome-wide comparisons (Figures 3.18, 3.19, and 3.20). Also, a future study could compare the genes found here to have differential gene expression to those previously found by Harrison et al. [12]. Harrison et al. explored gene expression patterns in various *B. terrestris* life history stages including reproductive and non-reproductive workers [12]. The comparison of these studies would more firmly establish any gene expression patterns of worker reproduction that are a common trend in this species, as opposed to colony-specific effects.

In conclusion, higher methylation levels have been found in genes with a higher degree of expression in the bumblebee *B. terrestris*. Higher methylation levels were also found in genes with a greater degree of differential expression between workers with alternate reproductive states, this is indicative of methylation playing a role in the control or maintenance of differential expression.

## Chapter 4

# Allele-specific expression and methylation in reproductive and non-reproductive *Bombus terrestris* workers.

### 4.1 Introduction

Eusocial insects are becoming key model organisms for the study of epigenetics [102, 266–268]. This is partly due to the presence of allele-specific expression and methylation mentioned in Chapter 2. However, the exact relationship between methylation and allele-specific expression in this lineage is not clear [122, 185–187, 230].

In the last couple of years, two studies in particular have focused on searching for parent-specific gene expression in the honey bee [182, 187]. These papers aimed to uncouple parent-specific and lineage-specific gene expression through the implementation of reciprocal cross experiments. The few differences between these studies in

design (e.g. tissue sampled) corresponds to divergent results [see 69]. Kocher and colleagues [187] used different honey bee subspecies and reported predominantly maternally biased parent-specific gene expression in the brain and full body. Galbraith et al. [182] used European and African honey bees of the same subspecies and identified strong paternally biased parent-specific gene expression in the worker fat body/ovaries. The genes found in Kocher et al. were compared with a list of known methylated genes, but no methylation analysis was carried out on the same bee samples to search for any methylation of the genes found to show parent-specific gene expression. Methylation of genes was not studied in Galbraith et al.

Both allele-specific expression and methylation have been implicated to be associated with worker reproduction. Allele-specific expression of the worker reproduction genes *Ecdysone 20 monooxygenase* and *IMP-L2-like* was found in the bumblebee [185]. Moreover, in Chapter 2, I found genes with allele-specific expression in the bumblebee, and several of these had caste differentiation functions. For example, *bicaudal-D* showed allele-specific expression in the bumblebee, and this gene is involved in the differentiation between soldiers and workers in the termite *Reticulitermes flavipes* [216] and it is methylated more in eggs than sperm in the honeybee [217]. Regarding methylation patterns, Amarasinghe et al. suggested that this mechanism also plays an important role in bumblebee worker reproduction [186] and I have established this in Chapter 3.

In this study I identify allele-specific gene expression and allele-specific methylation in the same samples of reproductive and non-reproductive *Bombus terrestris* workers in multiple colonies. Comparisons of these factors are drawn in combination with the differentially expressed, differentially methylated and alternatively spliced genes found in Chapter 3. Using these data I aim to distinguish any genome-wide allele-specific effects of both expression and methylation in worker reproduction.

## 4.2 Methods

### 4.2.1 Samples

The same worker bee samples with RNA-sequencing and bisulphite-sequencing were selected as described in Chapter 3.

Alternative splicing can introduce biological variation in allele-specific expression in different exons [269]. Therefore, it is important that phasing is taken into account in this study to ensure good sequencing accuracy. During the process of sequencing, phasing is when a nucleotide is not correctly removed after signal detection [270]. This would mean that no new nucleotide would be able to bind on this DNA fragment so the old nucleotide would be incorrectly detected as the next base in the sequence. Phasing is the main cause of decreasing per base sequence quality towards the end of a read. Thus, per base sequence quality was calculated for all libraries in order to account for any phasing that may be present. A per base sequence quality score is an estimate of the probability of that base being incorrectly called. Sequence quality checks were conducted with FastQC [271]. Per base sequence quality scores (q) range from 0 to 40, where higher numbers indicate better quality. After applying FastQC, per base sequence quality was found to be good (q score between 28 and 40) across all read positions for all samples.

### 4.2.2 Allele-specific expression analysis

The eighteen RNA-seq libraries were aligned to the *B. terrestris* Bter\_1.0 reference genome (Refseq accession number GCF\_000214255.1) [188] using Tophat [244]. Reads were mapped with a mean efficiency of 92.2% (91.0-92.7%). Bcftools (version 0.1.19-44428cd), bedtools (version 2.17.0), and samtools (version 0.1.19-44428cd) were then used to prepare the RNA libraries and identify variants, before the variants were filtered based on mapping quality score [193, 199]. The R package, QuASAR,

identified genotypes and located any allele-specific expression at heterozygous sites [200]. QuASAR removes SNPs with extreme differential allele expression from the analyses, thus controlling for any base-calling errors. Genomic regions ( $\pm 2900$ bp) of the significant loci were blastx against the reference non-redundant (nr) *Drosophila melanogaster* genome to give the associated gene ontology (GO) terms. Fisher's exact test was then conducted to determine enriched GO terms.

### 4.2.3 Allele-specific methylation analysis

Bam files of aligned reads from Bismark (as described in Chapter 3) were used as input files for *armfinder* within the Methpipe software (v.2.4.3) [272] to identify allele-specific methylation. *Armfinder* implements a sliding window and admixture model to identify allele-specific methylation in whole genome sequencing datasets [272, 273]. For each “window” of the genome two models are implemented. The first model is a “single allele” model which assumes the same degree of methylation for both alleles. The second model is a “two allele” model that assumes the alleles have different methylation levels. The model which fits the window with the highest likelihood is accepted, then the programme moves onto the next genomic window until the whole genome is analysed.

All reproductive and non-reproductive worker data were pooled separately, giving a mean coverage of 69 reads per gene for each condition. Duplicates were removed and standard parameters were used. Genes with allele-specific methylation were compared against the nr *D. melanogaster* reference genome and enriched GO terms found using Fisher's exact test.



#### 4.2.4 Comparisons: allele-specific methylation and expression

The hypergeometric test [251] was applied to identify if more genes and gene ontology terms overlapped between regions with allele-specific expression and allele-specific methylation than would be expected by chance.

Recently (September 2017), a colleague (Alun Jones) produced custom annotations of the *B. terrestris* reference transcriptome as part of his doctoral research. Due to time constraints this could not be implemented to find annotations for all of the genes of interest found here. Instead, a candidate gene approach was taken for only the genes that showed both allele-specific expression and allele-specific methylation. Annotations of the transcriptome were generated using Trinotate which integrates a variety of methods for functional annotation. Within Trinotate, BLAST and SwissProt were used to conduct homology searches in the transcript and protein databases [274]. HMMER and PFAM were used for protein domain identification [275, 276], and protein signal peptide and transmembrane domain prediction was carried out by signalP and tmHMM [277, 278]. Existing annotation databases were searched using the eggNOG, GO, and Kegg databases [279–281]. This generated a more comprehensive annotation database of the *B. terrestris* genome compared to the current reference transcriptome. The genes I identified to show both allele-specific expression and allele-specific methylation were compared with this custom database.

## 4.3 Results

### 4.3.1 Allele-specific expression

447 SNPs were identified to show allele-specific expression in reproductive worker bees in at least three of the nine transcriptome libraries. When searching for the genes with blastx against the *Bombus terrestris* reference genome 321 of the 447 genes had blast hits, but only three of these had gene ontology (GO) annotation due to the *B. terrestris* genome not currently being as well annotated compared to model organisms such as *Drosophila melanogaster*. 116 genes had GO annotation when conducting blastx with *D. melanogaster*. Thus the subsequent downstream analyses were carried out using the *D. melanogaster* blastx results.

In the non-reproductive worker libraries, 497 SNPs were found to show allele-specific expression in at least three libraries. 355 of the 497 had blastx hits for *B. terrestris*, but again only three possessed GO annotation. However, 126 genes showed GO annotation when blastx against *D. melanogaster*.

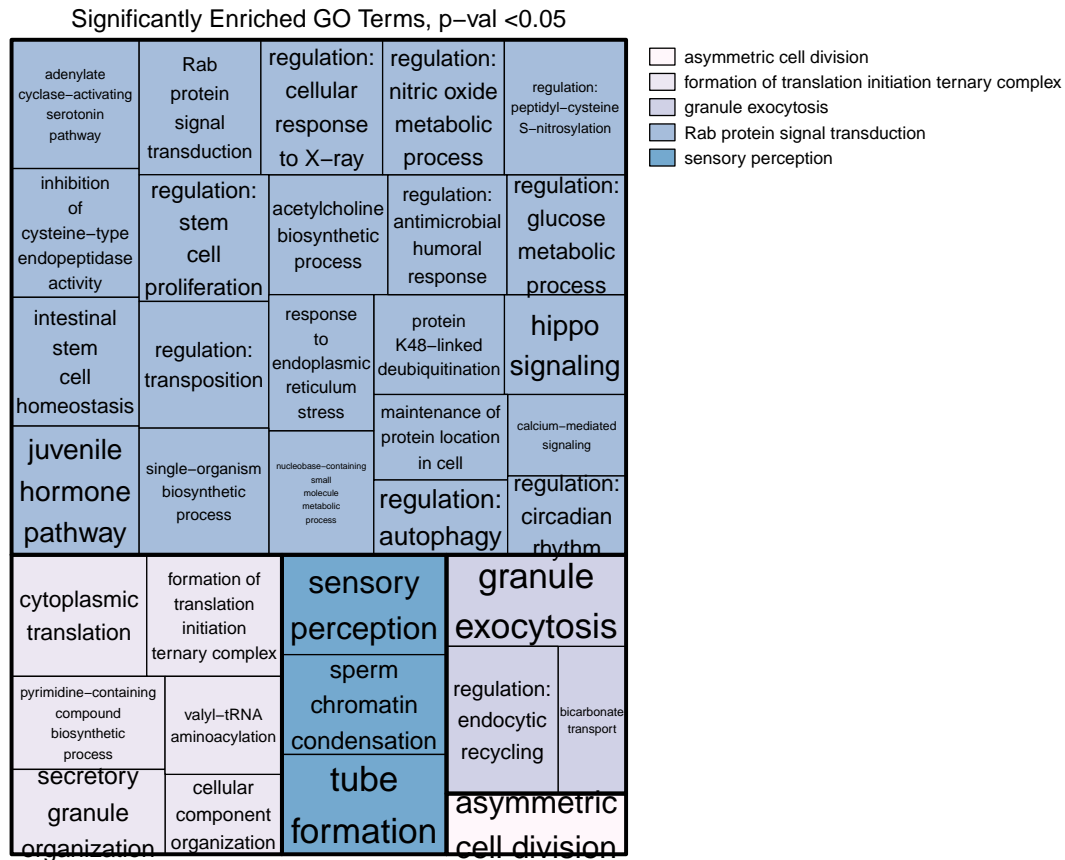


FIGURE 4.1: Enriched gene ontology terms of the genes with allele-specific expression in (A) reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment.

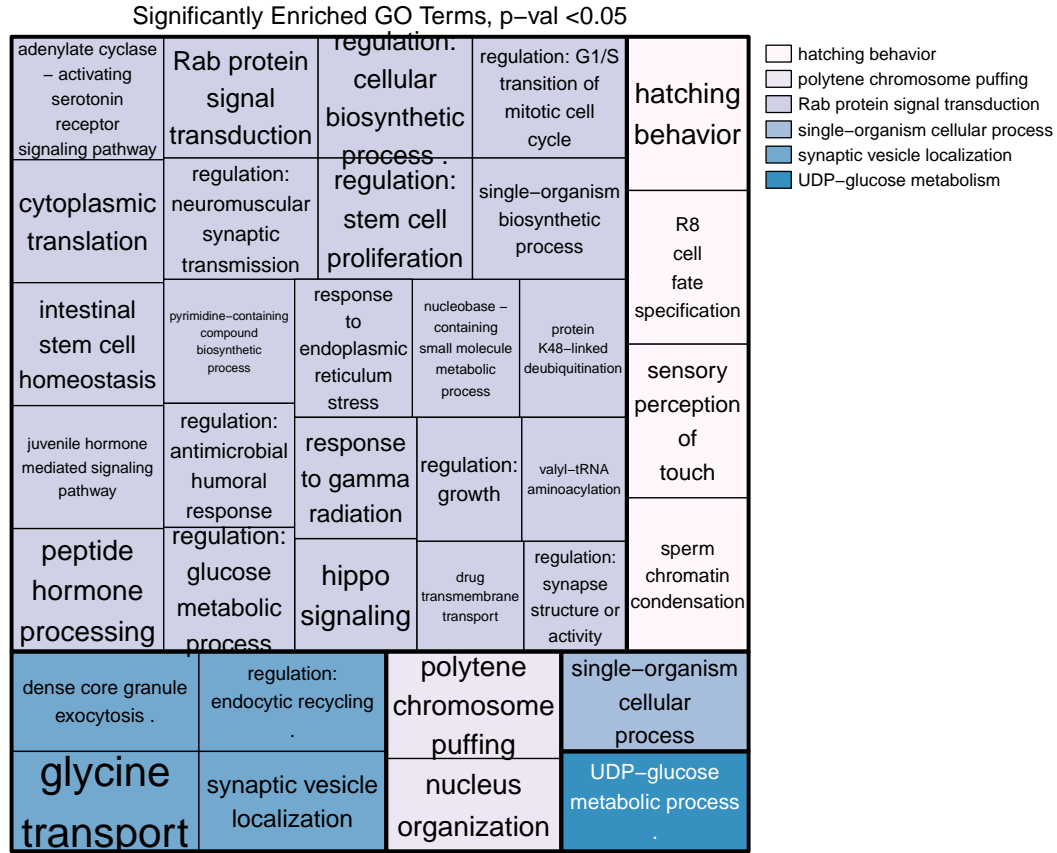


FIGURE 4.2: Enriched gene ontology terms of the genes with allele-specific expression in non-reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment.

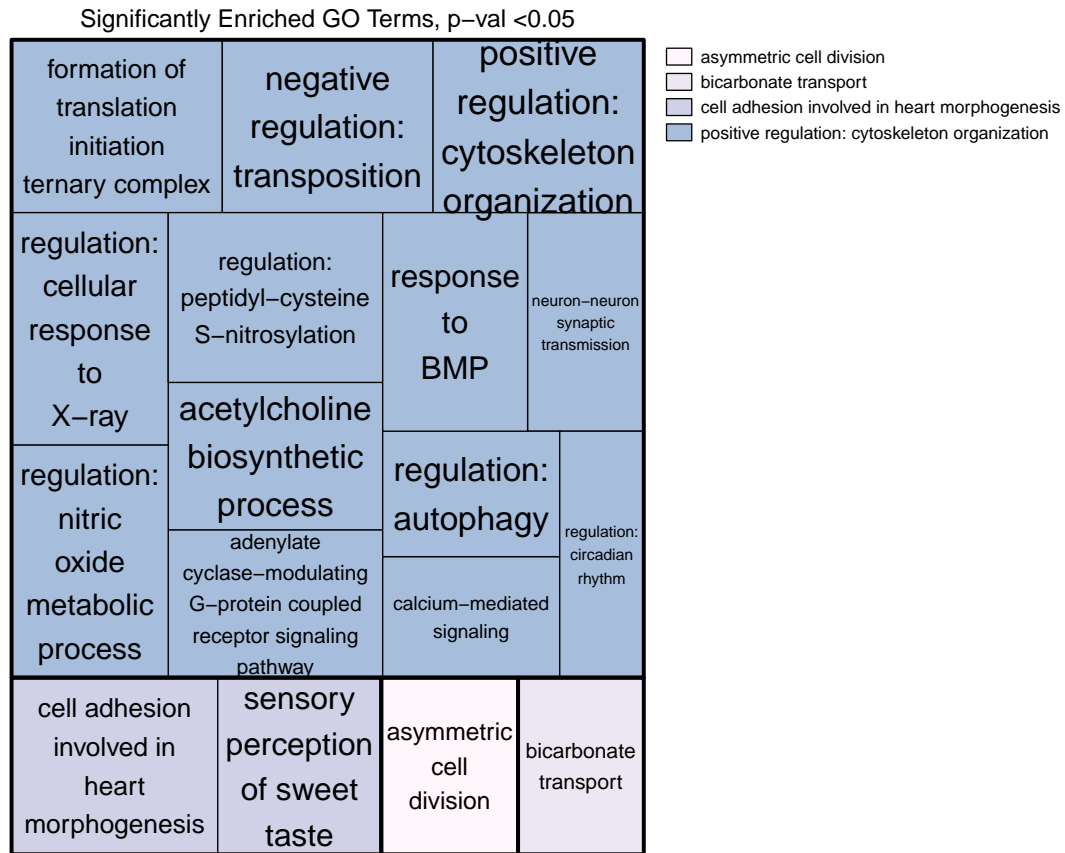


FIGURE 4.3: Enriched gene ontology terms of the genes with allele-specific expression *only* in reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment.

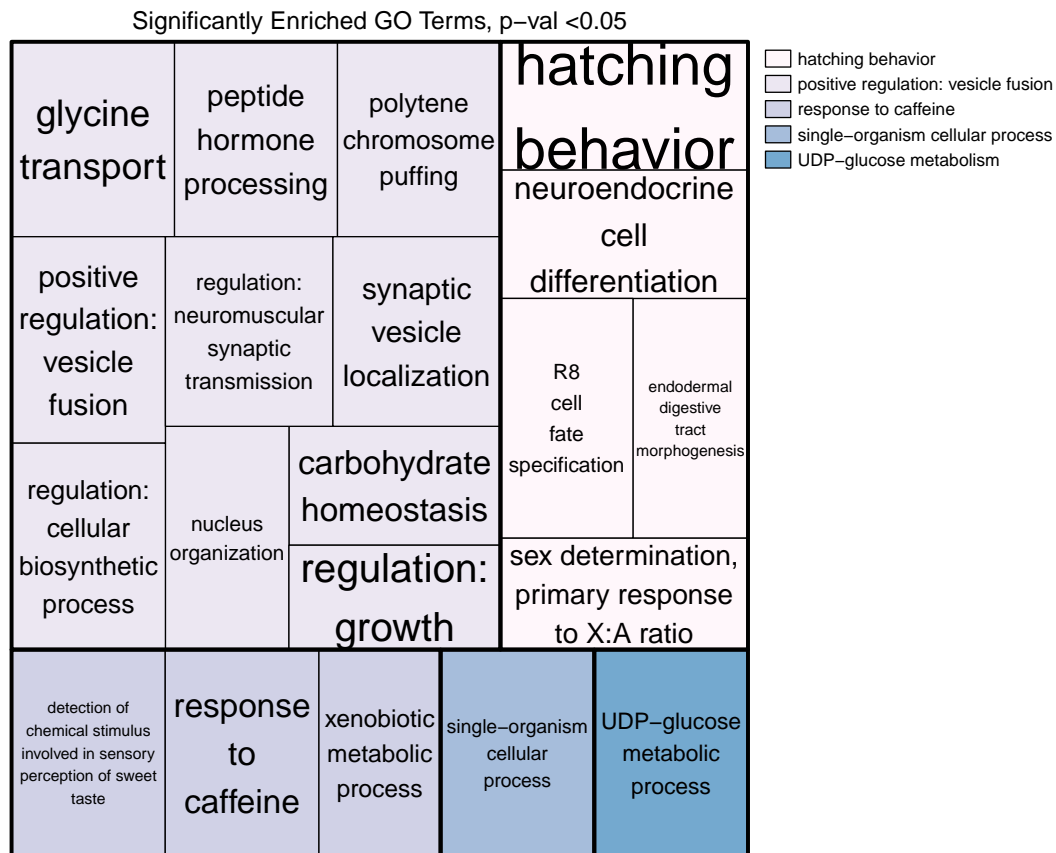


FIGURE 4.4: Enriched gene ontology terms of the genes with allele-specific expression *only* in non-reproductive workers (created using REVIGO [4]). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured “superclusters” of loosely related terms. The area of the rectangles represents the p-value associated with that clusters enrichment.

Of the 447 and 497 SNPs showing allele-specific expression in reproductive and non-reproductive workers respectively, 247 SNPs showed allele-specific expression in both worker types. 200 SNPs only showed allele-specific expression in reproductive workers, and 250 SNPs showed allele-specific expression in non-reproductive workers only.

195 GO terms were found to be significantly enriched compared with the whole *B. terrestris* transcriptome (Fisher’s exact test  $p < 0.05$ ) for genes with allele-specific expression in reproductive workers (Figure 4.1). 189 significantly enriched GO terms

(Fisher's exact test  $p < 0.05$ ) were found for genes with allele-specific expression in non-reproductive workers (Figure 4.2). 123 of these were found to be enriched GO terms of genes with allele-specific expression in both reproductive and non-reproductive workers. 72 and 65 enriched GO terms were associated with genes that showed allele-specific expression in reproductive workers only and non-reproductive workers only (Figures 4.3 and 4.4).

Several GO terms associated with epigenetic mechanisms were enriched in genes with allele-specific expression in both reproductive and non-reproductive workers. These GO terms include histone *H3-K27* acetylation and chromatin-mediated maintenance of transcription, as well as GO terms associated with canonical reproduction pathways in social insects.

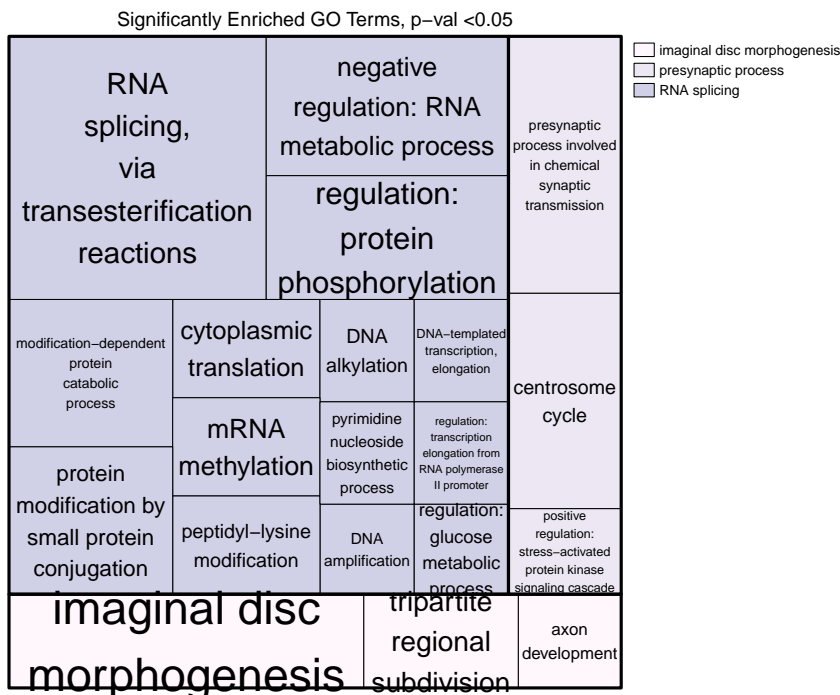
### 4.3.2 Allele-specific methylation

117 genes showed allele-specific methylation in non-reproductive workers, and 152 genes showed allele-specific methylation in reproductive workers. 77 of these genes showed allele-specific methylation in both reproductive and non-reproductive workers. Gene ontology terms enriched in the genes showing allele-specific methylation in reproductive and non-reproductive workers are displayed in Figures 4.5a and 4.5b. Of note, genes associated with RNA splicing, mRNA methylation, and histone *H3-K9* modification were enriched in genes with allele-specific methylation.

### 4.3.3 Comparisons: allele-specific methylation and expression

Genes and enriched GO terms that exhibited allele-specific expression or allele-specific methylation were then compared with the genes and enriched GO terms that were found to show differential expression, differential methylation, and alternative splicing from Chapter 3. Numbers of genes and enriched GO terms that were

(A)



(B)

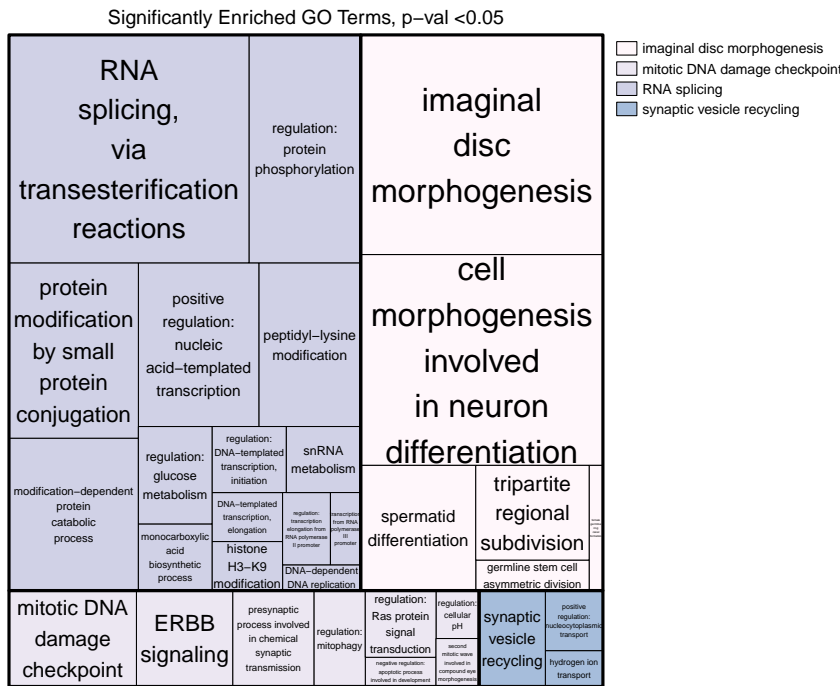


FIGURE 4.5: Enriched gene ontology terms of the genes with allele-specific methylation (A) up-regulated in reproductive workers and (B) down-regulated in reproductive workers compared to non-reproductive workers (created using REVIGO [4]).



TABLE 4.1: Comparing all known genes (LOC IDs) with allele-specific expression, allele-specific methylation, and overall differential expression, methylation, and alternative splicing (see Chapter 3) with the hypergeometric test. Genes are referred to as up/down regulated or hyper-/hypo-methylated in the reproductive workers (Rep.), compared to the non-reproductive workers (Non-rep.).

Gene set 1	Gene set 2	Genes overlapped	p value	Significant overlap (p < 0.05)
Rep. ASE <sup>a</sup>	Non-rep. ASE <sup>a</sup>	31	<0.0001	yes
Rep. ASE <sup>a</sup>	Rep. ASM <sup>b</sup>	3	0.022	yes
Rep. ASE <sup>a</sup>	Non-rep. ASM <sup>b</sup>	3	0.011	yes
Rep. ASE <sup>a</sup>	Diff. Isoforms	0	-	-
Rep. ASE <sup>a</sup>	Up regulation	18	<0.0001	yes
Rep. ASE <sup>a</sup>	Down regulation	6	<0.0001	yes
Rep. ASE <sup>a</sup>	Hypermethylation	0	-	-
Rep. ASE <sup>a</sup>	Hypomethylation	0	-	-
Non-rep. ASE <sup>a</sup>	Rep. ASM <sup>b</sup>	2	0.14	no
Non-rep. ASE <sup>a</sup>	Non-rep. ASM <sup>b</sup>	4	0.0016	yes
Non-rep. ASE <sup>a</sup>	Diff. Isoforms	0	-	-
Non-rep. ASE <sup>a</sup>	Up regulation	17	<0.0001	yes
Non-rep. ASE <sup>a</sup>	Down regulation	6	<0.0001	yes
Non-rep. ASE <sup>a</sup>	Hypermethylation	1	0.4	no
Non-rep. ASE <sup>a</sup>	Hypomethylation	0	-	-
Rep. ASM <sup>b</sup>	Non-rep. ASM <sup>b</sup>	77	<0.0001	yes
Rep. ASM <sup>b</sup>	Diff. Isoforms	0	-	-
Rep. ASM <sup>b</sup>	Up regulation	5	0.0066	yes
Rep. ASM <sup>b</sup>	Down regulation	2	0.12	no
Rep. ASM <sup>b</sup>	Hypermethylation	4	0.058	no
Rep. ASM <sup>b</sup>	Hypomethylation	7	0.0013	yes
Non-rep. ASM <sup>b</sup>	Diff. Isoforms	0	-	-
Non-rep. ASM <sup>b</sup>	Up regulation	5	0.0022	yes
Non-rep. ASM <sup>b</sup>	Down regulation	1	0.375	no
Non-rep. ASM <sup>b</sup>	Hypermethylation	2	0.311	no
Non-rep. ASM <sup>b</sup>	Hypomethylation	3	0.14	no

<sup>a</sup> Allele-specific expression

<sup>b</sup> Allele-specific methylation

observed to overlap more than expected are shown in Tables 4.1 and 4.2. See Appendices D and E for all significantly overlapping gene names, GO terms, and all genes with both allele-specific expression and allele-specific methylation.

TABLE 4.2: Comparing enriched Gene Ontology (GO) terms with allele-specific expression, allele-specific methylation, and overall differential expression, methylation, and alternative splicing (see Chapter 3) with the hypergeometric test. Genes are referred to as up/down regulated or hyper-/hypo-methylated in the reproductive workers (Rep.), compared to the non-reproductive workers (Non-rep.).

GO set 1	GO set 2	GO terms overlapped	p value	Significant overlap (p < 0.05)
Rep. ASE <sup>a</sup>	Non-rep. ASE <sup>a</sup>	123	<0.0001	yes
Rep. ASE <sup>a</sup>	Rep. ASM <sup>b</sup>	12	<0.0001	yes
Rep. ASE <sup>a</sup>	Non-rep. ASM <sup>b</sup>	28	<0.0001	yes
Rep. ASE <sup>a</sup>	Diff. Isoforms	10	<0.0001	yes
Rep. ASE <sup>a</sup>	Up regulation	0	-	-
Rep. ASE <sup>a</sup>	Down regulation	10	<0.0001	yes
Rep. ASE <sup>a</sup>	Hypermethylation	7	<0.0001	yes
Rep. ASE <sup>a</sup>	Hypomethylation	9	<0.0001	yes
Non-rep. ASE <sup>a</sup>	Rep. ASM <sup>b</sup>	12	<0.0001	yes
Non-rep. ASE <sup>a</sup>	Non-rep. ASM <sup>b</sup>	31	<0.0001	yes
Non-rep. ASE <sup>a</sup>	Diff. Isoforms	8	<0.0001	yes
Non-rep. ASE <sup>a</sup>	Up regulation	0	-	-
Non-rep. ASE <sup>a</sup>	Down regulation	27	<0.0001	yes
Non-rep. ASE <sup>a</sup>	Hypermethylation	8	<0.0001	yes
Non-rep. ASE <sup>a</sup>	Hypomethylation	8	<0.0001	yes
Rep. ASM <sup>b</sup>	Non-rep. ASM <sup>b</sup>	277	<0.0001	yes
Rep. ASM <sup>b</sup>	Diff. Isoforms	6	<0.0001	yes
Rep. ASM <sup>b</sup>	Up regulation	0	-	-
Rep. ASM <sup>b</sup>	Down regulation	5	<0.0001	yes
Rep. ASM <sup>b</sup>	Hypermethylation	127	<0.0001	yes
Rep. ASM <sup>b</sup>	Hypomethylation	149	<0.0001	yes
Non-rep. ASM <sup>b</sup>	Diff. Isoforms	14	<0.0001	yes
Non-rep. ASM <sup>b</sup>	Up regulation	0	-	-
Non-rep. ASM <sup>b</sup>	Down regulation	12	0.0002	yes
Non-rep. ASM <sup>b</sup>	Hypermethylation	220	<0.0001	yes
Non-rep. ASM <sup>b</sup>	Hypomethylation	261	<0.0001	yes

<sup>a</sup> Allele-specific expression

<sup>b</sup> Allele-specific methylation

A small number of genes significantly overlapped between genes showing allele-specific expression in reproductive workers and allele-specific methylation in reproductive and non-reproductive workers (Table 4.1). Four of these six genes with allele-specific expression and allele-specific methylation were present in more than one comparison. LOC105666144 was found to show allele-specific expression

and allele-specific methylation in both reproductive and non-reproductive workers. Second, LOC100644154 showed allele-specific expression in reproductive and non-reproductive workers and allele-specific methylation in non-reproductive workers. LOC100646154 had allele-specific expression in non-reproductive workers and allele-specific methylation in reproductive and non-reproductive workers. Finally, LOC100647771 showed allele-specific expression in reproductive workers and allele-specific methylation in reproductive and non-reproductive workers. When conducting blastx searches against the reference nr *D. melanogaster* genome these six genes with allele-specific expression and allele-specific methylation were uncharacterised. Yet, after comparing these genes with the custom Trinotate annotation (Section 4.2.4), the six genes were found to have functions associated with the GRIP (golgin-97, RanBP2alpha, Imh1p and p230/golgin-245) protein domain (LOC100642556), nucleic acid binding (LOC105666144), protein phosphorylation (LOC100646154), transcription factors (LOC100647771), transposase (LOC105666784), chromatin modification, and the regulation of DNA binding (LOC100644154) (Appendix D: Table D12).

Allele-specific expression in reproductive workers was also associated with overall differential expression levels between workers of a different reproductive status. Similarly, allele-specific expression in non-reproductive workers was associated with overall differential expression levels. Genes with allele-specific methylation in reproductive and non-reproductive workers significantly overlapped with differentially expressed and differentially methylated genes. However genes with differential splicing between reproductive and non-reproductive workers did not overlap with any other gene group. 87 known genes appeared in multiple significantly overlapping gene lists, of these all were uncharacterised except four genes (Table 4.3).

Regarding gene ontology, a significant number of GO terms overlapped between those GO terms with allele-specific expression and allele-specific methylation in reproductive and non-reproductive workers, overall down regulation in reproductive workers, overall differential methylation, and differential splicing (Table 4.2). Only

TABLE 4.3: Characterised genes in multiple significantly overlapping enriched gene lists with allele-specific expression, allele-specific methylation, differential expression, isoform expression, and methylation (hypergeometric test,  $p < 0.05$ ).

LOC ID	Gene name	Number of overlapping gene lists	GO term
LOC100644972	glycine-rich protein DOT1-like	2	-
LOC100645332	Golgi resident protein GCP60	3	fatty-acyl-CoA-binding, integral component of membrane transport
LOC100643807	protein Jumonji	2	polytene chromosome, regulation of histone H3-K27 methylation, ESC/E(Z) complex, nuclear euchromatin
LOC100647906	junctionophilin-1	3	-

enriched GO terms that were up regulated in reproductive workers did not overlap with any other GO list.

271 GO terms appeared in multiple significantly overlapping GO lists. Four GO terms appeared in over ten of the lists of significantly overlapping GO terms (Table 4.4). This included negative regulation of phosphorylation, an epigenetic modifier of histones.

TABLE 4.4: Gene ontology terms in over ten out of twenty-six significantly overlapping enriched GO lists with allele-specific expression, allele-specific methylation, differential expression, isoform expression, and methylation (hypergeometric test,  $p < 0.05$ ).

GO ID	GO term	Number of overlapping GO lists
GO:0002181	cytoplasmic translation	17
GO:0046579	positive regulation: Ras protein signal transduction	14
GO:0010906	regulation: glucose metabolic process	15
GO:0042326	negative regulation: phosphorylation	13

## 4.4 Discussion

Close associations between the genes and gene ontology terms with allele-specific expression and allele-specific methylation have been identified in the bumblebee

*Bombus terrestris*. This reflects the potential for a relationship between these phenomena. Only six genes showed both allele-specific expression and allele-specific methylation (Table 4.1). This possibly demonstrates the previous assertion that most allele-specific expression is driven by factors other than epigenetics [see 282, and Chapter 2]. Moreover, the software available for allele-specific methylation detection is limited. *Armfinder* in the Methpipe software was implemented and a false discovery rate (FDR) was used [272]. The use of an FDR has the possibility of producing a greater number of false negative results and hence cause more genes with allele-specific methylation to be excluded from the results compared with other tests of determining significance [283]. Therefore the FDR has recently been reported to be too rigorous a method for bisulphite-sequencing datasets [284]. In the future, the Benjamini & Hochberg method could be used to determine genes showing significant allele-specific methylation because this test has the greatest sensitivity, specificity, and FDRs closest to 0.05 in comparison to other methods [283, 285].

The fact still remains that out of the genes that were identified to have allele-specific methylation, a relatively small number of these genes also showed allele-specific expression. Thus, perhaps there is no direct link between allele-specific expression and allele-specific methylation, and other mechanisms have greater roles to play in worker reproduction (Section 1.2.1). For instance, transcription factors (TF) bind to enhancer or promoter regions upstream of their target gene. This affects the transcription of the target gene to either be up- or down-regulated [286]. Therefore TFs could be involved in the differential gene expression changes between reproductive and non-reproductive workers.

Even though there were not many genes with both allele-specific expression and allele-specific methylation, the gene annotations that were identified using the custom annotation database were associated with a variety of epigenetic mechanisms. Allele-specific methylation is indicated to affect genes involved in chromatin modification, protein phosphorylation, and TFs which regulate gene expression and protein

activation. Here, a gene coding for a TF (LOC100647771) was found to have allele-specific methylation and allele-specific expression in reproductive workers. This implies that this TF may be partly epigenetically controlled. The allele-specific methylation of LOC100647771 could affect the expression of the TF and this may change its pervasiveness in the cell to regulate other genes that could be involved in worker reproduction. Therefore, allele-specific methylation may affect genes that have the potential to regulate the expression of worker reproduction genes downstream. The allele-specific expression of the six genes (with allele-specific expression and allele-specific methylation) may be a result of the allele-specific methylation. Alternatively they could be linked through other downstream mechanisms, or there is still the possibility that the methylation status is a by-product of the allele-specific expression (explored in Section 7.2.2). As well as the causation of this relationship being unclear, other mechanisms are shown to be involved.

Several genes identified in this study show links with DNA methylation, alternative splicing, and other epigenetic mechanisms. For instance, the *jumonji* gene was one of four named genes that occurred in more than one of the significantly overlapping methylation/expression gene lists. It was found to show allele-specific methylation in both reproductive and non-reproductive workers, and showed overall hypermethylation in non-reproductive workers. Therefore, it is interesting that the *jumonji* group of proteins have previously been found to be involved in regulation of histone demethylation [287–289]. This gene family has also been linked with cardiac disease, obesity, neurological disorders and some human cancers [see review 289, and references therein].

Here it was found that genes with histone modification associations show allele-specific methylation and expression. This is consistent with previous studies that indicate histone modifications to be highly associated with insect methylation and allele-specific expression. Hunt et al. reported that insect DNA methylation patterns are consistent with those of histone modifications [71]. Also, Galbraith et al. [182]

found several genes that were associated with histone modification to show allele-specific expression with a paternal bias in reproductive workers.

Negative regulation of phosphorylation was here found to be highly involved in bumblebee allele-specific expression and methylation processes. Phosphorylation describes the addition of a phosphoryl group to a protein and is one of the most abundant post-translational modifications [290]. Histone phosphorylation changes the availability of DNA for transcription. Of note, histone acetylation can suppress an inhibitory phosphorylation and therefore initiate transcription [291]. Also, phosphorylation of histone 3 prevents heterochromatin spreading by antagonizing H3K9 methylation in *Drosophila* [292]. Hence phosphorylation plays key roles in association with other mechanisms of epigenetic regulation.

Interestingly, a transposase which regulates the movement of transposons was found here to have allele-specific methylation and expression. In mammals and plants methylation acts to silence transposable elements [293]. Also, methylation was previously found to control transposase activity in maize [294]. In insects transposons are typically unmethylated [129, 295, 296]. Therefore it is surprising that a transposase is methylated in the bumblebee.

In this study only the genes showing both allele-specific expression and allele-specific methylation were compared with the custom annotation database generated, but in the future all the genes of interest could be compared with the custom database to give a better understanding of the underlying expression patterns of worker reproduction. It would also be interesting to use these same RNA- and bisulphite-sequencing libraries to determine whether the same allele is methylated and expressed. This would help indicate whether allele-specific methylation and allele-specific expression are directly related or if they are potentially linked through other downstream mechanisms.

In conclusion, allele-specific expression and allele-specific methylation have a weak

correlation in reproductive and non-reproductive workers. However the insect epigenetic process is implicated to be complex, with the involvement and interaction of DNA methylation, alternative splicing, hormone signalling pathways, and various forms of histone modification.



# Chapter 5

## Differential circadian isoform expression patterns in the desert locust, *Schistocerca gregaria*

### 5.1 Introduction

The first three results chapters of this thesis (Chapters 2, 3, and 4) focused on elucidating the underlying epigenetic mechanisms of the caste polyphenism found in the Hymenoptera lineage. Here, I move onto studying another insect which displays a polyphenism, and I explore the isoform expression patterns of the desert locust *Schistocerca gregaria*.

The desert locust displays phenotypic plasticity according to changes in population density of conspecifics (Section 1.4). The two phases are described as solitary and gregarious, which are associated with different behavioural and morphological characteristics. When there is an increase in population size or a reduction in food availability, solitary locusts can be forced together with conspecifics causing them to become behaviourally gregarious [172]. However it takes several generations for

morphological changes to occur [173]. It is in this gregarious phase when desert locusts swarm, and hence are perceived as an agricultural pest species in Africa and the Middle East, where they can cause devastation for crops such as barley, maize, sorghum, and wheat [174].

Several insecticides have been developed with varying degrees of efficacy to counteract this pest. Populations of the oriental migratory locust, *Locusta migratoria manilensis*, are extremely destructive in China. Organophosphate insecticides have been used to control this species for a few decades [38]. Organophosphates inhibit acetylcholinesterase which is a catalyst for the breakdown of the neurotransmitter acetylcholine and other choline esters [297]. However, the use of organophosphates has now led to the resistance of this insecticide through increased detoxification by esterases and glutathione S-transferases, and increased activity and reduced sensitivity of acetylcholinesterase to organophosphate inhibition [298].

New strategies for locust control are being developed. RNAi silencing of chitin synthase 1 led to high mortality rates in *L. migratoria manilensis* [142]. Chitin synthase 1 is essential for locust development and growth of the trachea and integument in particular. Therefore this gene has been identified as a potential target for the control of destructive locust swarms.

Moreover, the time of day at which an insecticide is applied has been observed to affect its efficacy in mosquitoes [176]. Mosquito detoxification processes and therefore pesticide susceptibility have been demonstrated to be affected by the circadian cycle [176]. By applying insecticides at the time of day when detoxification processes are at their most down-regulated the efficacy level of the insecticide may be maximised. If there is a better understanding of the circadian cycle of genes involved in locust detoxification the efficacy of locust insecticides could also be maximised. This would be beneficial with whichever insecticide is used. Therefore, this chapter focuses on the identification of genes with a circadian cycle in the desert locust *S. gregaria*.

Gregarious *S. gregaria* individuals are active during the day, with peak activity reported at 1400 hours [31, 171]. In contrast, solitary locusts are considered to be nocturnal. Ould Ely and colleagues [170] observed solitary locust activity to peak 1-2 hours after dusk, and during the day solitary locusts exhibit hiding behaviours and only tend to fly if they are disturbed [299]. These distinctive observed activity patterns of solitary and gregarious desert locusts thus indicate strong circadian rhythms.

Since the discovery of the *period* (*per*) locus in *Drosophila melanogaster* [300], circadian rhythms have been studied for over four decades to elucidate the genetic basis of how organisms respond to different selection pressures that vary over a 24 hour time period. The clock genes *per*, *timeless* (*tim*), and *clock* (*clk*), are known to be important in the *D. melanogaster* circadian rhythm [301] and have been reported to also have a rhythmic expression pattern in *S. gregaria* [177]. However, no difference in expression patterns of these genes has yet been found between solitary and gregarious locusts. This could be due to differences in photosensitivity between laboratory and wild locust populations [171], or the circadian rhythm may be regulated further downstream of the *per*, *tim*, and *clk* genes in the locust.

Little is yet known about interactions between circadian rhythms and post-transcriptional processes. The circadian clock has been found to regulate alternative splicing in mice [302] and microRNAs have been demonstrated to be necessary for the maintenance of the circadian rhythm in *D. melanogaster* [303]. Furthermore, alternative splicing has been demonstrated to be important in the locust. In a recent study, 45 genes were found to have differentially expressed isoforms between solitary and gregarious migratory locusts (*Locusta migratoria*) [143]. These included genes associated with cytoskeleton dynamics which are involved in neuronal plasticity, a key mechanism associated with behavioural phase change [144]. However, there are biological differences between *S. gregaria* and *L. migratoria*. *S. gregaria* and *L. migratoria* are both members of the Acrididae family, but are from different subfamilies of Cyrtacanthacridinae and Oedipodinae, respectively. Phase

polyphenism is not observed in all species of the Acrididae. Therefore it has been suggested that the phase polyphenisms witnessed in *S. gregaria* and *L. migratoria* evolved independently [304]. The underlying mechanisms behind gregarization of these two subfamilies is indicated to be different. Tactile stimulation of the antennae triggers gregarization in an Oedipodinae species (*Chortoicetes terminifera*) [305], whereas tactile stimulation of the hind legs as well as visual and olfactory cues lead to gregarious behaviours in the Cyrtacanthacridinae subfamily (*S. gregaria*) [306, 307]. Hence isoform expression differences may not be the same in these two subfamilies.

It is currently unknown whether there are alternative isoform circadian expression patterns in desert locusts (*S. gregaria*). Here I use whole genome RNA-sequencing data to study how alternative splicing varies in gregarious locusts (*Schistocerca gregaria* forskål) over a 24 hour period. This will facilitate the identification of detoxification isoforms with circadian patterns of expression.

## 5.2 Methods

### 5.2.1 Locust husbandry

Gregarious phase *Schistocerca gregaria* forskål locusts were reared at the University of Leicester under a 12:12 hour light:dark cycle. They were kept in groups of 100-300 individuals together in 50 x 50 x 50 cm cages. During the photophase (light) light levels were 750-1000 lx with a temperature of 36°C, whereas in scotophase (dark), locusts were kept in a light level of 0 lx and at 25°C. All locusts were fed wheat seedlings and bran flakes *ad libitum*.

### 5.2.2 Tissue samples and sequencing

Samples were taken at eight time points over a period of 24 hours (i.e. every 3 hours), starting at 1 hour after lights on (ZT 1). At each time point three gregarious locusts were sampled, giving a total of 24 locust samples (Table 5.1). Locust age was controlled for as all locusts were sampled 2 weeks ( $\pm 3$  days) after the adult molt.

TABLE 5.1: Gregarious *Schistocerca gregaria* *forskål* locust samples (m = male, f = female).

Time point (ZT)	Replicate			Circadian phase
	A	B	C	
1	m	f	m	Photophase
4	f	f	m	
7	f	f	m	
10	m	f	f	
13	m	f	f	Scotophase
16	m	f	f	
19	f	f	m	
22	m	f	f	

At least one male and at least one female were sampled at each time point in order to discern a general pattern for differential circadian isoform expression patterns across *S. gregaria* genders. No sex differences in the circadian rhythm of *S. gregaria* locusts have been reported in previous literature. *S. gregaria* sexual dimorphism is limited to morphological differences in size, reproductive organs, courtship behaviour, and cuticle colour [308, 309]. Females do not display any courtship behaviours [309, and references therein]. However, male *S. gregaria* locusts demonstrate courtship behaviours of orientation towards the female, a slow approach, and mounting a female locust [309, and references therein]. *S. gregaria* females are a brown-beige to yellowish colour and tend to be larger in size compared to males [309]. Gregarious *S. gregaria* males display a yellow cuticle colour 10 days after the adult molt, which is partly due to juvenile hormone synthesis [310].

Dissections were carried out at 25°C under low intensity red light. The brain and optic lobe were extracted and the tissues were pooled together for each individual

before RNA extraction. The optic lobe was chosen for RNA extraction because it has been indicated as the location of the central circadian clock in hemimetabolous insects such as locusts. The compound eye is the primary photo-entrainment site for the central circadian clock, and severing of the optic nerves or removal of the optic lobes has been demonstrated to remove the activity rhythm [311, and references therein].

The Trizol reaction was used to extract the RNA. Subsequently, TURBO DNase treatment (Life technologies) and the Qiagen RNeasy mini kit were used to remove any DNA and phenol contamination. RNA quality was checked using a bioanalyzer (Agilent, G2939A). The 24 RNA samples were then sequenced with the NextSeq platform (Illumina, Glasgow Polyomics, UK). Paired-end reads at 2x75bp with 25M reads were sequenced. RNA sequencing resulted in approximately  $1.3 \times 10^9$  reads, 25-75 bp in length.

### 5.2.3 Alternative splicing analysis

Transcriptome quality checks were carried out using fastQC (version 0.11.5) [271]. Subsequently 10 base pairs were removed from the 5' ends of all RNA-seq libraries (cutadapt version 1.11 [312]). In the previous three results chapters (Chapters 2, 3, and 4) I began my analyses by aligning the *Bombus terrestris* RNA-seq libraries with the reference genome of that species. However, the locust *S. gregaria* does not yet have a published genome, so a different pipeline was used. KisSplice (version 2.4.0-p1) [313] was used to identify alternative splicing events directly from the trimmed transcriptome libraries. Pairwise comparisons with the R (version 3.4.1) package *kissDE* (version 1.4.0) was carried out to test if variants were significantly differentially expressed at the eight timepoints (coverage: at least 10 reads per variant) [7, 314]. *kissDE* is based on count regression with a negative binomial distribution. A general linear model framework is used with isoforms, experimental conditions,

and the interaction between them as the variables. This pipeline separates all transcripts into two groups: isoforms that include the exon of interest and isoforms that exclude the exon of interest [314].

Subsequently, the genes identified by *kissDE* were further filtered using the Jonckheere-Terpstra-Kendall (JTK) algorithm (*JTK\_cycle* in R) to uncover genes with a period of 20-28 hours in both isoforms, and different isoform phase times (i.e. time point of peak expression) [315]. The JTK algorithm is reported to detect cycling transcripts more accurately and efficiently than the other commonly used methods of COSOPT and Fishers G test [315]. The JTK algorithm was compiled from the Jonckheere-Terpstra (JT) test and Kendalls tau. The JT test takes ordered independent groups and identifies monotonic orders of data across the groups [316, 317]. By contrast, Kendall's tau is used to measure the association between two measured quantities [318]. Combining these measures in the JTK algorithm allows the JT test to be applied to hypothesized group orderings, while keeping the group sizes fixed. Kendall's tau is calculated between the hypothesized time series and each cyclical ordering. The period and phase with the minimum Kendall's tau p value is then selected to correlate to the expression pattern with the highest likelihood. Hence, this algorithm identified differential isoform circadian expression patterns. Further downstream analyses of k-means clustering and principal component analysis were carried out in R [7].

Significantly differentially expressed isoforms with circadian patterns were searched for in the migratory locust *L. migratoria*, the termite *Zootermopsis nevadensis*, and German cockroach *Blattella germanica* nr databases (blastx) to find corresponding gene IDs. The *L. migratoria* reference genome was selected on account of its phylogenetically close placement to *S. gregaria*, and the other two reference genomes were chosen due to their slightly better annotation and close relation to the locust lineage. Blast2GO was used to obtain associated gene ontology (GO) terms for the genes identified [319].

### 5.3 Results

7,580 genes with significantly differentially expressed isoforms over the eight time points were identified in the gregarious locust (*kissDE*: FDR adjusted  $p < 0.05$ ) (Figure 5.1). Of these genes, 1,707 isoforms were identified to show a circadian pattern of expression with a period of 20-28 hours (JTK:  $p < 0.05$ ), and 111 isoforms showed a circadian expression pattern with Benjamini—Hochberg  $q < 0.05$ . 105 genes had both isoforms (210 isoforms) showing a significant circadian pattern (period of 20-28 hours) (JTK:  $p < 0.05$ ) and different phase times. No genes were identified to have both isoforms showing a significant circadian pattern and different phase times with  $q < 0.05$ . Hence, the 105 genes (210 isoforms) with  $p < 0.05$  were used for the downstream analysis.

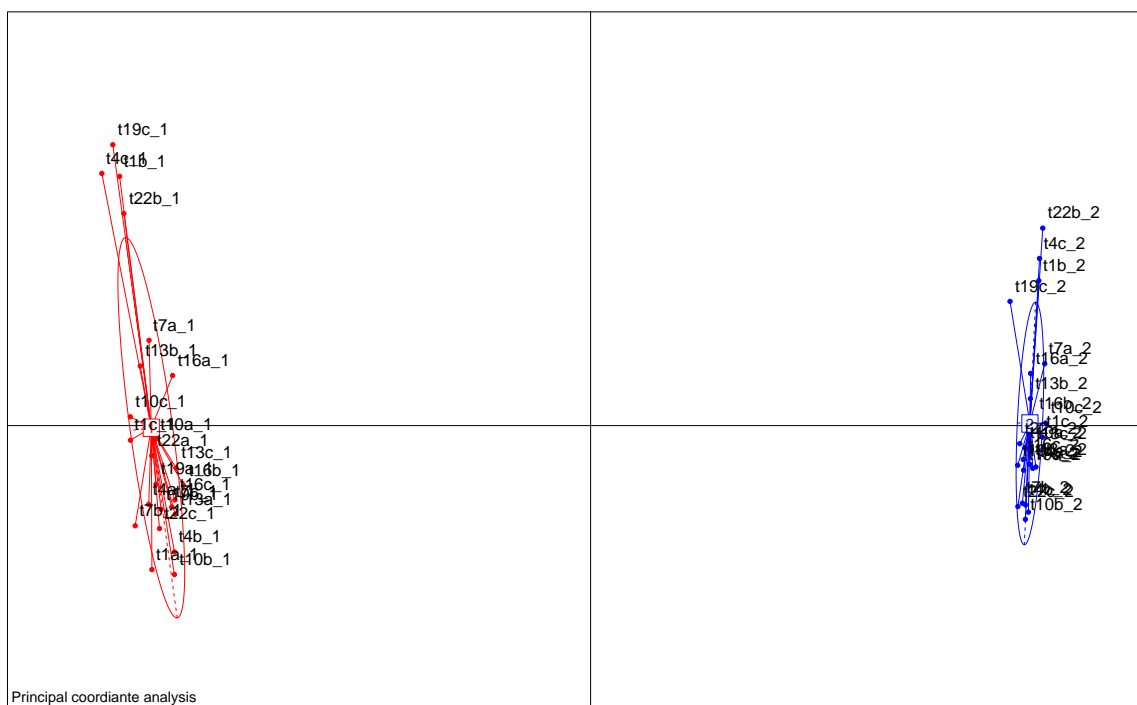


FIGURE 5.1: Principal Coordinate Analysis (PCoA) of all genes with differentially expressed isoforms. Notation: “t1a.2” corresponds to time point 1, locust replicate a, and isoform 2. Red represents isoform 1, and blue to isoform 2.

93.56% of the variance in expression of the 210 isoforms was explained by the difference in isoforms of the same genes (Figure 5.2), this reflects that these genes have



differentially expressed isoforms. Several isoforms showed marked differences in expression between photophase (light) and scotophase (dark) (Figure 5.3). K-means clustering (Figure 5.4) indicated that the majority of differences between isoforms were due to distinct magnitudes of expression rather than temporal differences in expression, even though in all 105 genes the two isoforms have disparate phase times to each other ( $p < 0.05$ ).

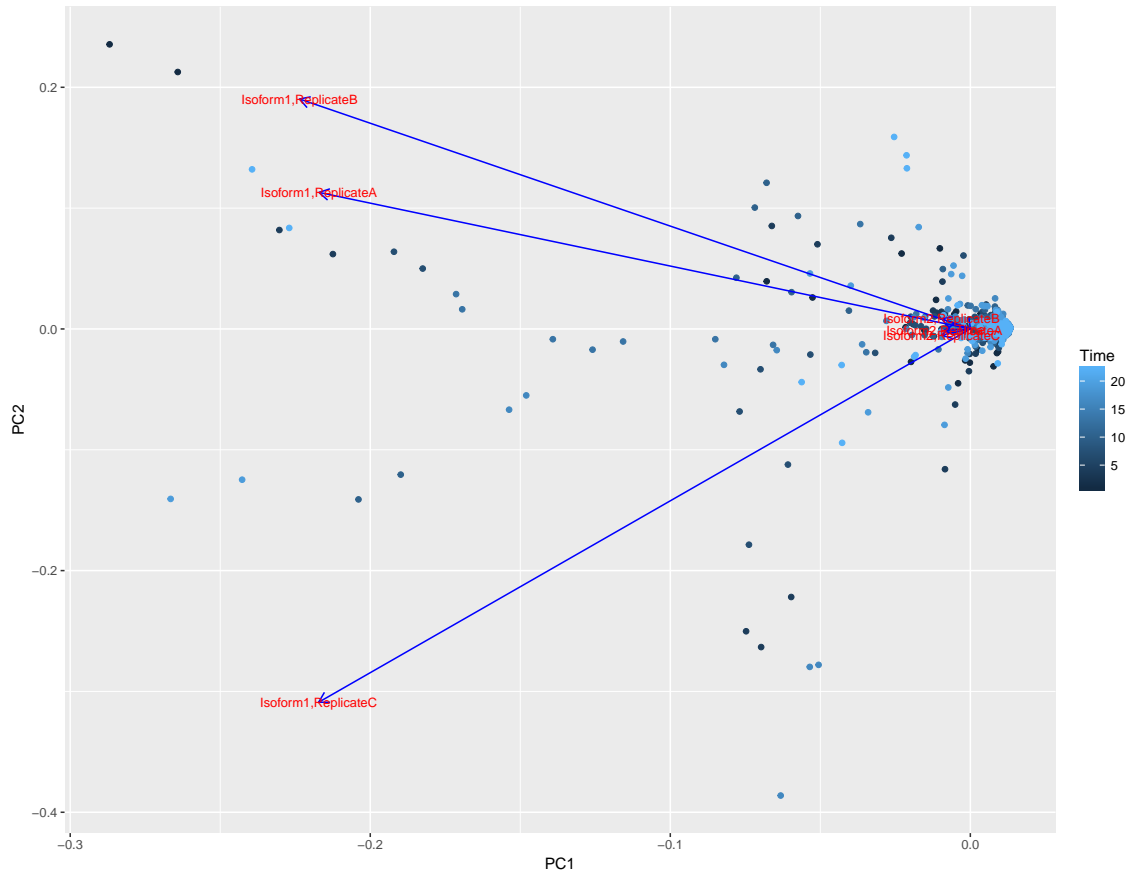


FIGURE 5.2: PCA of the 105 multi-isoform genes with circadian expression patterns and different optimal phase times. Proportion of variance: PC1 = 0.9356, PC2 = 0.0247

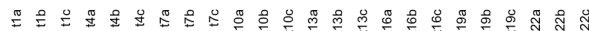


FIGURE 5.3: Heatmap of 105 genes at eight time points with log-adjusted significantly differentially expressed isoforms. Notation: “t1a” corresponds to timepoint 1, and locust replicate a.

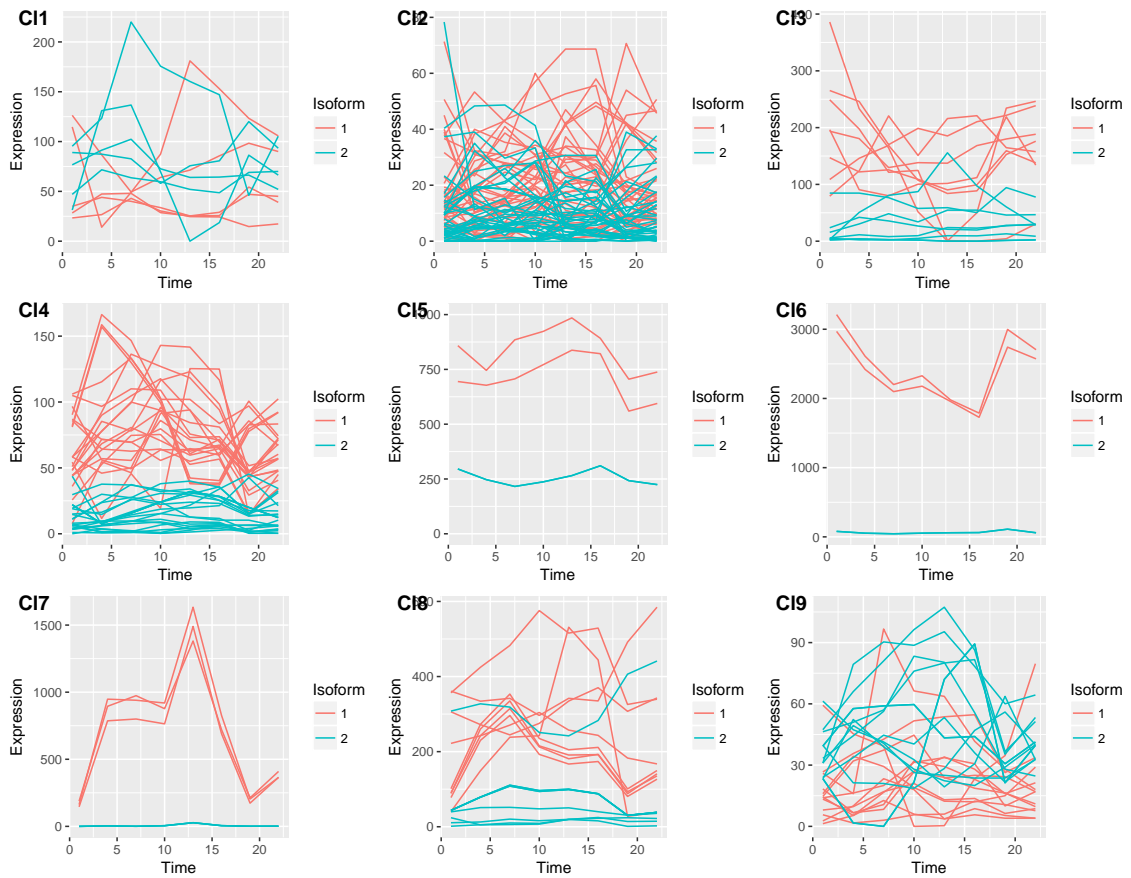


FIGURE 5.4: Congruent expression patterns over 24 hours, based on k-means clustering for the 210 isoforms from the 105 differentially expressed multi-isoform genes. CI1-9 denote the nine clusters. Different y axis scale bars are used in each sub-figure.

Figure 5.5 depicts some examples of the genes found with predominant temporal differences in their circadian isoform expression patterns. In gene 5376 the first isoform has peak expression at 0 ZT which is when lights turn on (Figure 5.5a). Whereas isoform 2 has a phase of 4.5 ZT which is just before the time when the expression of isoform 1 is at its lowest. Similarly, maximum isoform expression levels are opposed at 4.5 ZT and 13.5 ZT for gene 731 (Figure 5.5b), and for gene 1426 the two isoform phases are exact opposites at 0 ZT and 12 ZT (Figure 5.5c). Interestingly, after searching in the nr database (blastn) gene 731 had the most similarity with DNA polymerase beta-like (LOC108772560) in the ant *Cyphomyrmex costatus* (e-Value = 0.012). The beta subunit of DNA polymerase III was previously

found to show a circadian rhythm of expression in a *Synechococcus* species strain [320]. Thus the circadian expression pattern of a DNA polymerase beta-like gene in the locust is supportive of previous assertions that the cell cycle and circadian rhythms are closely linked and regulate each other [321].

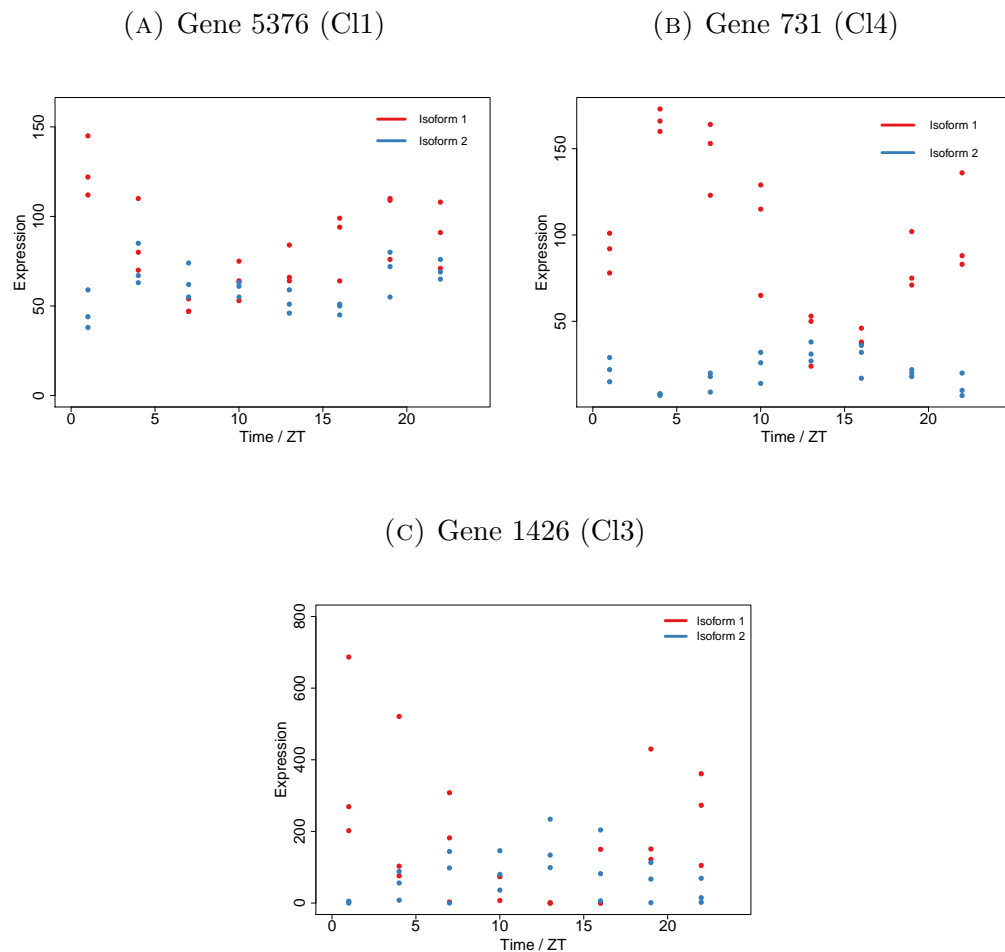


FIGURE 5.5: Genes with differential circadian isoform expression patterns. Different y axis scale bars are used in each sub-figure. Clusters (Cl) determined by k-means clustering of the 105 differentially expressed multi-isoform genes.

After blastx searches, two of the 105 genes with differential circadian isoform expression patterns could be mapped to the *L. migratoria* genome and had associated gene ontology (GO) terms (Table 5.2). Six of the 105 genes could be mapped to the *Z. nevadensis* genome, and two other genes of the 105 could be mapped to the

*B. germanica* genome (Tables 5.4 and 5.3). This low mapping rate is a reflection of the current limited annotation of insect species outside the Dipteran lineage. Although the circadian patterns are clear, the isoform differences in these ten genes are predominantly due to differences in magnitude of expression despite the isoforms having different phases (Figures 5.6, 5.7, and 5.8). The GO terms of these ten genes were generally associated with mitochondrion organization, transmembrane transport, the apoptotic process, and the oxidation-reduction process.

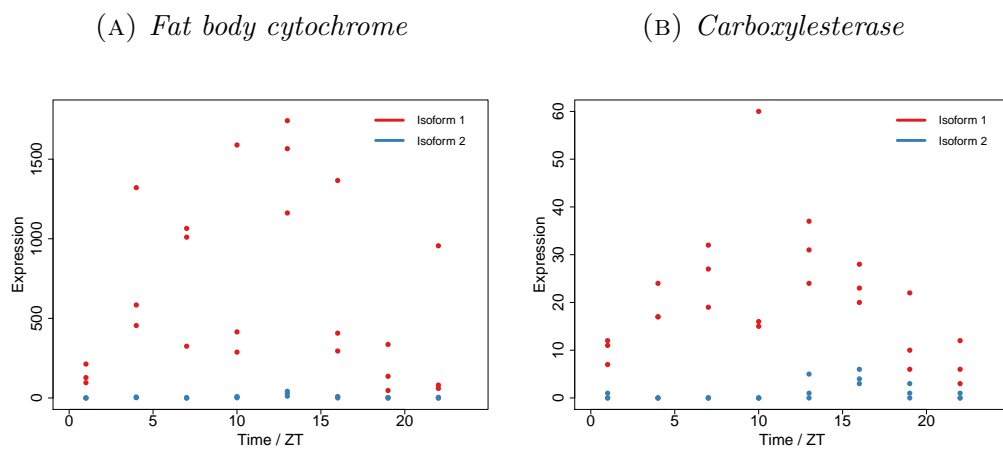


FIGURE 5.6: Genes with mapping against the migratory locust *L. migratoria* with differential circadian isoform expression. Y axis scales are different in each sub-figure.

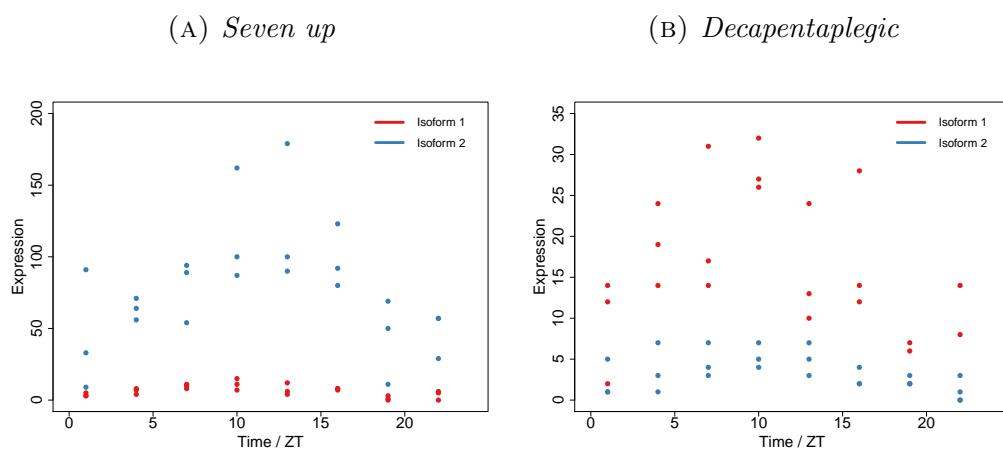


FIGURE 5.7: Genes with mapping against the German cockroach *B. germanica* with differential circadian isoform expression. Y axis scales are different in each sub-figure.

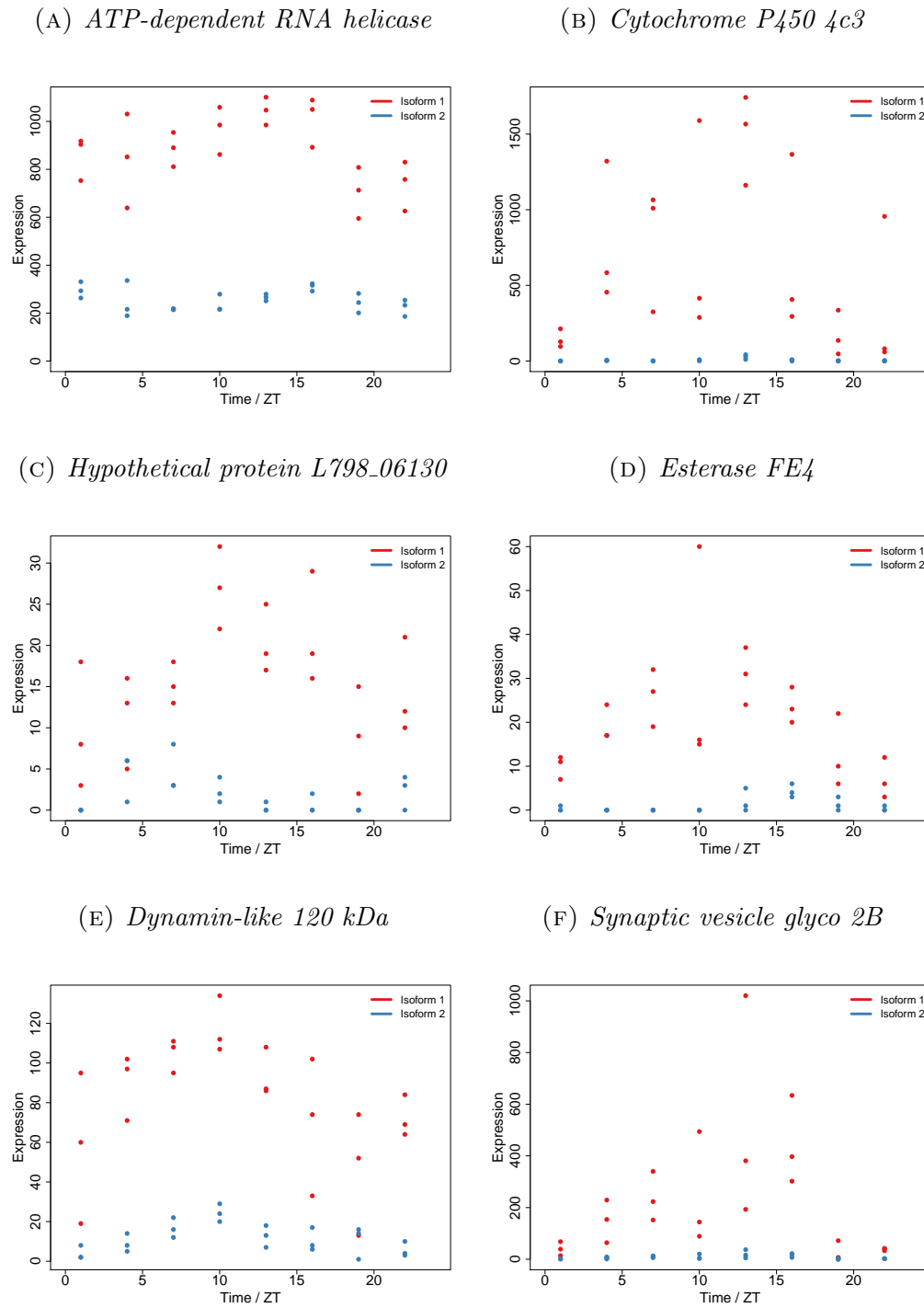


FIGURE 5.8: Genes with mapping against the termite *Z. nevadensis* with differential circadian isoform expression. Y axis scales are different in each sub-figure.

TABLE 5.2: Genes with differential isoform circadian expression in the gregarious phase of the desert locust *Schistocerca gregaria* and their gene ontology terms (blastx with the migratory locust *Locusta migratoria*). Phase denotes the time at which isoforms have peak expression levels (Isoform 1 : Isoform 2).

Gene	Phase (ZT)	Length (bp)	e-Value	GO IDs	GO Names
<i>Fat body cytochrome</i>	10.5 : 12	311	5.94076E-7	GO:0005506; GO:0016705; GO:0055114; GO:0020037	iron ion binding; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; oxidation- reduction process; heme binding
<i>Carboxylesterase</i>	10.5 : 16.5	229	1.30523E-24	GO:0016787	hydrolase activity

TABLE 5.3: Genes with differential isoform circadian expression in the gregarious phase of the desert locust *Schistocerca gregaria* and their gene ontology terms (blastx with the cockroach *Blattella germanica*). Phase denotes the time at which isoforms have peak expression levels (Isoform 1 : Isoform 2).

Gene	Phase (ZT)	Length (bp)	e-Value	GO IDs	GO Names
<i>Seven up</i>	9 : 12	81	2.45222E-7	GO:0003677; GO:0005634; GO:0003700; GO:0008270; GO:0004879; GO:0046872; GO:0006351; GO:0003707; GO:0006355; GO:0043565; GO:0043401; GO:0030522	DNA binding; nucleus; transcription factor activity, sequence-specific DNA binding; zinc ion binding; RNA polymerase II transcription factor activity, ligand-activated sequence-specific DNA binding; metal ion binding; transcription, DNA-templated; steroid hormone receptor activity; regulation of transcription, DNA-templated; sequence-specific DNA binding; steroid hormone mediated signalling pathway; intracellular receptor signalling pathway
<i>decapentaplegic, partial</i>	10.5 : 9	123	3.19787E-6	GO:0008083; GO:0005576; GO:0040007	growth factor activity; extracellular region; growth



TABLE 5.4: Genes with differential isoform circadian expression in the gregarious phase of the desert locust *Schistocerca gregaria* and their gene ontology terms (blastx with the termite *Zootermopsis nevadensis*). Phase denotes the time at which isoforms have peak expression levels (Isoform 1 : Isoform 2).

Gene	Phase (ZT)	Length (bp)	e-Value	GO IDs	GO Names
<i>ATP-dependent RNA helicase DDX17</i>	12 : 16.5	285	9.02E-28	GO:0000166; GO:0003676; GO:0005524; GO:0004386; GO:0016787	nucleotide binding; nucleic acid binding; ATP binding; helicase activity; hydrolase activity
<i>Cytochrome P450 4c3</i>	10.5 : 12	311	5.47E-11	GO:0005506; GO:0046872; GO:0016491; GO:0016705; GO:0016020; GO:0016021; GO:0055114; GO:0004497; GO:0020037	iron ion binding; metal ion binding; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; membrane; integral component of membrane; oxidation- reduction process; monooxygenase activity; heme binding
<i>Hypothetical protein L798_06130</i>	10.5 : 6	156	6.38E-05	GO:0003676; GO:0046872	nucleic acid binding; metal ion binding
<i>Esterase FE4</i>	10.5 : 16.5	229	1.18E-12	GO:0016787	hydrolase activity
<i>Dynamin-like 120 kDa</i>	9 : 10.5	190	2.41E-10	GO:0000166; GO:0005525; GO:0003924; GO:0005739; GO:0006915; GO:0007005	nucleotide binding; GTP binding; GTPase activity; mitochondrion; apoptotic process; mitochondrion organization
<i>Synaptic vesicle glyco 2B</i>	12 : 10.5	202	6.76E-17	GO:0055085; GO:0022857; GO:0016020; GO:0016021	transmembrane transport; transmembrane transporter activity; membrane; integral component of membrane

## 5.4 Discussion

This is the first account of differential circadian isoform expression patterns in the desert locust *Schistocerca gregaria*. Alternate isoforms have been identified to be expressed at differential levels throughout the day, and previous literature shows that several of these genes have already been associated with circadian rhythmic patterns.

*ATP-dependent RNA helicase DDX17* is an ortholog of the *period-1* (*prd-1*) gene. *prd-1* has a long (25 hour) period length in the circadian rhythm of asexual spore formation (conidiation) in the fungus *Neurospora* [322], and it was previously found that *prd-1* encodes an ATP-dependent RNA helicase that influences nutritional compensation of the *Neurospora* circadian clock [323]. Therefore the finding here that *ATP-dependent RNA helicase DDX17* has a circadian pattern of expression in the locust (Table 5.4, Figure 5.8a) reflects the ancestry of this conserved gene [324].

The expression of cytochrome P450 gene family members is regulated by the circadian clock in mammals [325, 326]. Consistent with previous studies, *cytochrome P450 4c3* and the *fat body cytochrome* have now been found here to have a circadian pattern of expression in the desert locust (Tables 5.4 and 5.2, Figures 5.8b and 5.6a). It is also of note that *cytochrome P450 4C1* (LOCMI16195) and *cytochrome C* (LOCMI10443) were previously found to be differentially spliced genes between gregarious and solitary *L. migratoria* brain samples [143].

Interestingly, cytochrome P450 genes have been found to be associated with detoxification in insects (in the house fly, *Musca domestica*) [327]. Similarly, up-regulation of *esterase FE4* (which was also found here to have a circadian pattern in the locust (Table 5.4, Figure 5.8d)) was previously found to be associated with causing insecticide resistance in the aphid *Myzus persicae* [328]. This echoes the notion made by Balmert et al. that mosquito detoxification processes and therefore pesticide susceptibility are affected by the circadian cycle [176]. Hence, perhaps a similar

trend is true for the desert locust. This knowledge could potentially be used in the application of insecticides to combat crop-threatening locust swarms. By applying the insecticide at the time of day when the *cytochrome*- and *esterase FE4*-associated pathways are at their most down-regulated (at dawn) the efficacy level of the insecticide may be maximised.

The gene *dynammin* (*dnm*) that is involved in apoptosis [329] was found here to have a circadian pattern of expression in the locust (Table 5.4, Figure 5.8e). In previous literature, the circadian gene *clock* (*clk*) was found to play an important role in apoptosis [330]. Silencing *clk* expression in human cell lines affected the expression of apoptosis-related genes and caused an increase in the rate of apoptosis. Comparably, in insects the *Drosophila* homolog of *dnm* (*shibire<sup>ts1</sup>*) is involved in membrane vesicle scission and has been proposed to be linked to setting the circadian period [331]. The observed pattern of expression in the locust indicates a similar tight link between apoptosis and circadian rhythms in this lineage.

The gene *seven up* (*svp*) controls photoreceptor cell fates in *Drosophila* [332] and the knockdown of *svp* has an effect on *Drosophila* circadian rhythms [333]. Hence, it is interesting that this gene that is involved in light detection shows a circadian rhythm of expression here in the locust (Table 5.3, Figure 5.7a).

Finally, the *decapentaplegic* (*dpp*) gene encodes a bone morphogenetic protein and in *Drosophila* *dpp* is positioned on the left arm of the second chromosome next to *timeless* (*tim*) which is a key gene involved in circadian clock regulation [334]. Therefore, if the chromosomal positioning is similar in the locust there is a possibility that there is linkage between *dpp* and *tim* causing *dpp* to have the observed circadian pattern of expression (Table 5.3, Figure 5.7b).

Despite the low mapping rates to the current annotations of the reference genomes available, several of the genes here identified to show circadian patterns of expression in gregarious locusts have been previously described to be associated with the circadian clock in other organisms. This reflects the high rate of conservation of

these functions across the Metozoa, and the number and breadth of processes which are affected by the circadian clock. Furthermore, circadian alternative splicing has previously been found to be tissue dependent [302], therefore focusing solely on the brain and optic lobe in this study may have excluded notable circadian rhythms in other organs. Therefore, in the future it would be interesting to conduct a similar study comparatively with other tissues and when there are improved annotations of closely related reference genomes available. Also, since solitary and gregarious locusts have opposite activity levels throughout the day it would be of interest to look into the circadian differences between the different phases of this polyphenism.

# Chapter 6

## Alternative splicing and eusociality in the Hymenoptera

### 6.1 Introduction

Across the hymenopteran lineage there is a remarkable variety of life histories from parasitic wasps and solitary bees to highly eusocial species. In species with a high degree of eusociality, there are different castes that are highly specialised for different roles in the running of the colony. In the honey bee, female bees are categorized into queens that monopolise reproduction, and predominantly sterile workers that are responsible for brood care and foraging. These different castes are behaviourally and morphologically distinct, yet they are formed from the same genome. The presence of these multiple phenotypes from the same genome must be due to differential gene usage.

Alternative splicing permits the same gene to encode several different proteins through different transcription start and termination sites and the inclusion or exclusion of different exons and introns [178]; thus it is one mechanism through which different phenotypes can form. Chen et al. found a strong relationship between

alternative splicing and species complexity across 47 eukaryotic species [335], where species complexity was measured by the number of cell types. Chordates were observed to have the highest levels of alternative splicing than any other taxonomic group [335]. Alternative splicing has also been demonstrated to be important in recent lineage-specific radiations of phenotype. For instance, alternative splicing has been established to play a key role in the adaptive radiation of cichlids, and the adaptation of lice to different ecological niches [127, 336].

In the Hymenoptera, alternative splicing is known to play important roles in the division of labour and caste differentiation in eusocial insects. Alternative splicing has been shown to be involved in the transition between nurse and foraging worker roles in the honey bee *Apis mellifera* [118]. In Jarosch et al. 2011 [235], alternative splicing of a transcription factor was associated with control of worker sterility in *Apis mellifera capensis*. Knocking out of a specific exon via RNAi resulted in worker ovary activation which is associated with parthenogenetically producing diploid female offspring in this subspecies [235]. However, it is unknown whether the degree of alternative splicing of a worker varies with species social complexity in the Hymenoptera.

An individual worker in a highly eusocial species has less varied behaviours to perform in comparison to primitively eusocial workers and solitary animals that have a greater number of different roles. For example, the highly eusocial ant species *M. pharaonis* and *P. barbatus* have sterile workers that do not need to exhibit mating behaviours or develop reproductive organs [337, 338]. On the other hand, primitively eusocial *H. saltator* colonies often contain “gamergate” workers that can mate and lay fertilised eggs as well as assisting with raising conspecifics [163, 339]. Therefore, it is anticipated that workers in highly eusocial species will have lower levels of alternative splicing than workers in less socially complex species. I also predict any genes with alternative splicing to be primarily associated with regulation and signalling as these pathways are important for the initiation and maintenance of different phenotypes [340–342]. Here, I assess the relative levels of alternative splicing

between workers of several hymenopteran species with different social structures.

## 6.2 Methods

### 6.2.1 Transcriptome selection

Eusociality has evolved in nine separate events in the Hymenoptera [8, 147–153] (Figure 6.1). In this study, publicly available worker transcriptomes obtained from six insect species were selected with varying degrees of sociality and phylogenetic relatedness (Figure 6.2, Table 6.1) and cover three of these evolutionary events: Apini (*Apis mellifera*), Bombini (*Bombus terrestris*), and Formicidae (*Harpegnathos saltator*, *Pogonomyrmex barbatus*, *Monomorium pharaonis*).

Three highly eusocial Hymenopteran species were chosen: the honey bee (*A. mellifera*), the red harvester ant (*P. barbatus*), and the pharaoh ant (*M. pharaonis*). In these three species there are distinct sterile worker castes and the queen monopolises reproduction. The honey bee (*A. mellifera*) has perennial colonies that have an average number of 40,000 - 80,000 bees in Summer. *M. pharaonis* is a highly invasive ant species [343, 344]. *M. pharaonis* colonies have an average of 170 queens that make up around 5% of the colony [337]. They display uniclonality (i.e. there is no aggression between colonies) which gives the impression of large “super” colonies when *M. pharaonis* nests are at a high density. *P. barbatus* reproduction is monopolised by a mating “winged alate” caste that fly away to mate with males who exhibit a mammal-like lekking behaviour [345]. Although *P. barbatus* again has sterile workers, worker roles are not constant and appear relatively fluid [338] compared with *A. mellifera* and *M. pharaonis* indicating that this species may not be as highly eusocial as *A. mellifera* and *M. pharaonis*.

The two primitively eusocial Hymenopteran species selected are the buff-tailed bumblebee (*B. terrestris*) and the indian jumping ant (*H. saltator*) [346]. These species

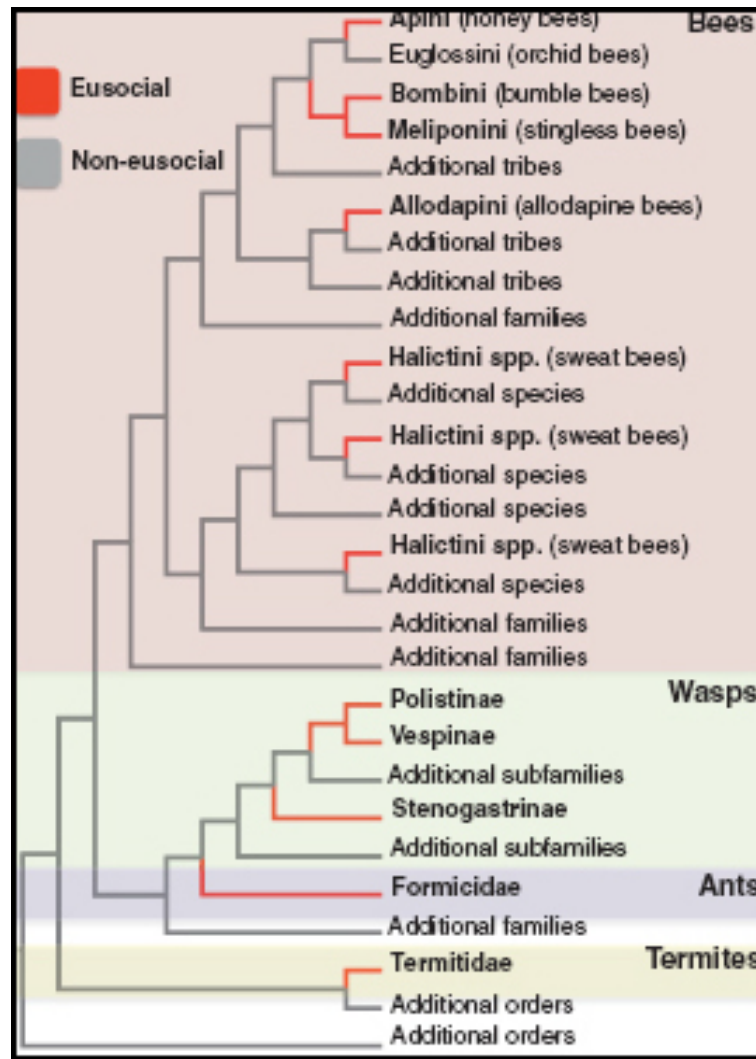


FIGURE 6.1: Origins of eusociality in the Hymenoptera and termites [8]. Red lines indicate evolution of eusocial status.

exhibit signs of eusociality, but are classified as primitively eusocial due to the limited morphological differences between their castes [347]. *B. terrestris* bees have relatively small annual colonies with a monandrous queen that largely monopolises reproduction (see Section 1.3), although workers often compete over reproduction to produce haploid males [348]. *H. saltator* has small colonies and “gamergate” workers often mate and reproduce [163, 339]. Therefore, the castes are less distinguished in primitively eusocial species compared to highly eusocial species [349].

Sawflies are solitary, yet are still part of the Hymenopteran order. Therefore, the



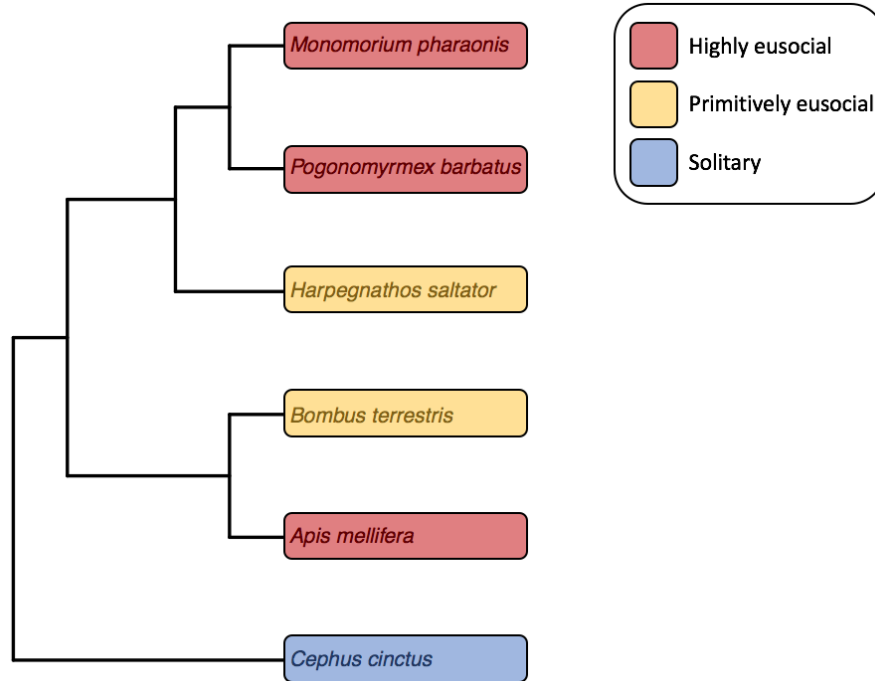


FIGURE 6.2: Phylogeny of hymenopteran species with diverse social structures used in this study based on NCBI taxonomy (phyloT) [9].

wheat stem sawfly, *Cephus cinctus*, was selected as an outgroup. Sawfly larvae reside in stems of wheat for around one month before they hibernate over Winter in stem “stubs” and pupate in early Spring [350]. They emerge as adults from May to early June, when they mate and lay eggs. Generally one egg is laid per stem, but in the event of multiple eggs being laid in a single stem only one larva will survive.

### 6.2.2 Metrics of alternative splicing

Three metrics were used as measures of the pervasiveness of alternative splicing. First, the proportion of genes with multiple isoforms was compared between the six species with varying social structures. This is referred to as Metric 1. A worker in a highly eusocial species has less varied behaviours to perform in comparison to species with less complex social structures. Hence, workers in highly eusocial species are anticipated to have lower levels of alternative splicing than workers in

less socially complex species (Section 6.1). Therefore, for Metric 1 it is predicted that workers of species with more complex social structures will have fewer genes with multiple isoforms. Second, the mean number of isoforms per multi-isoform gene was measured (Metric 2). I predict workers of species with more complex social structures will have fewer isoforms per multi-isoform gene. Third, the number of splicing events per multi-isoform gene was identified (Metric 3). It is anticipated that workers of species with more complex social structures will have fewer splicing events per multi-isoform gene. Metrics 1 and 3 were also used in the eukaryote-wide study by Chen et al. that found a strong relationship between organism complexity and alternative splicing [335]. The pipeline for determining the three metrics of alternative splicing used in this study is summarised in Figure 6.3.

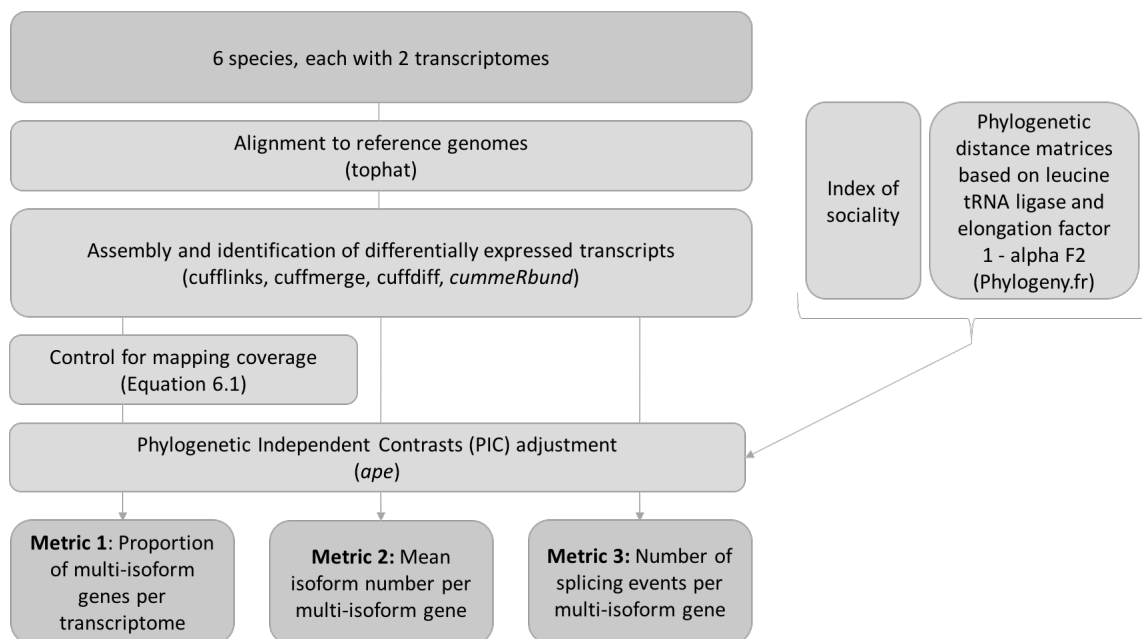


FIGURE 6.3: Flow chart of methodology to determine the pervasiveness of alternative splicing with regards to social structure in six species of Hymenoptera. *ape* and *cummeRbund* are packages in R.

### 6.2.3 Differential isoform expression analysis

Tophat (version 3.3.6) [244] was used to align the transcriptomes to the appropriate reference genomes (Table 6.1). The aligned transcripts were assembled using cufflinks (version 2.2.1) as described by Trapnell et al. [245]. Cufflinks accounts for positional biases in coverage, thus resulting in the mapping coverage being representative of the associated reference transcriptome [351, 352]. Cuffmerge was used to create merged transcriptome annotations from the transcripts. Then cuffdiff identified differentially expressed transcripts between different species. The *cummeRbund* (version 2.14.0) [246] package in R (version 3.3.0) was used for further downstream processing [7].

To control for mapping coverage the following equation (6.1) was implemented;

$$NMG = \frac{MG}{C} \times 100$$

where:  $MG$  = multi-isoform genes;  $C$  = mean coverage, and (6.1)

$NMG$  = new multi-isoform genes that are controlled for coverage.

TABLE 6.1: Hymenopteran reference genomes and worker caste transcriptomes accessed via the European Bioinformatics Institute database [13].

Index of sociality <sup>1</sup>	Common name	Latin binomial	Worker transcriptomes	Sample	Coverage (%)	Reference genome assembly
3	Buff-tailed bumblebee	<i>Bombus terrestris</i>	C1R_1 <sup>2</sup> , C1N_3 <sup>2</sup>	Head	91.8	Bter_1.0
				Head	92.2	[188]
5	Honeybee	<i>Apis mellifera</i>	SRR1028783, SRR1028784	Body	95.9	Amel_4.5
				Body	95.5	[353]
4	Red harvester ant	<i>Pogonomyrmex barbatus</i>	SRR1910415 SRR1814129	Head	71.9	Pbar_UMD_V03
				Head	79.4	[346]
3	Indian jumping ant	<i>Harpegnathos saltator</i>	SRR330972 SRR330973	Body	60.5	HarSal_1.0
				Body	81.4	[354]
5	Pharoah ant	<i>Monomorium pharaonis</i>	DRR029612 DRR029613	Body	80.0	M.pharaonis_V2.0
				Body	83.2	scaffold (EBI)
1	Wheat stem sawfly	<i>Cephus cinctus</i>	SRR3048750, SRR3052013	Body	84.7	Ccin1 scaffold
				Body	90.7	(EBI)

<sup>1</sup> 1 = solitary, 2 = subsocial/parasocial, 3 = primitively eusocial, 4 = eusocial, and 5 = highly eusocial

<sup>2</sup> see Chapter 3

An index of sociality was created in order to be incorporated into a Phylogenetic Independent Contrasts (PIC) adjustment which controlled for varying degrees of

phylogenetic relatedness between the six species. Each species was given a rating on a scale from 1 to 5, where 1 = solitary, 2 = subsocial/parasocial, 3 = primitively eusocial, 4 = eusocial, and 5 = highly eusocial [based on 347, and references therein]. Therefore: *C. cinctus* = 1, *B. terrestris* and *H. saltator* = 3, *P. barbatus* = 4, and *A. mellifera* and *M. pharaonis* = 5. The *leucine tRNA ligase* and *elongation factor 1 - alpha F2* genes were used to create phylogenetic distance matrices. These two genes have previously been demonstrated to give an accurate representation of phylogenetic relationships in *Atta fabrizius* ants [355]. The online platform *Phylogeny.fr* was used to create phylogenies with MUSCLE as an aligner, then PhyML and TreeDyn were used to render the phylogenies and trees [356–362].

Using the generated distance matrices, a PIC adjustment in the R (version 3.4.0) package *ape* (version 4.1) was then applied to the three metrics of alternative splicing identified and the social index. This is based on Felsenstein’s method of PIC which takes into account differences in evolutionary rates and phylogeny between focal species and therefore it controls for the inherent non-independence of taxa [10]. Correct topology and branch lengths are assumed, as is a Brownian motion (BM) model of character evolution [363]. (The BM model perceives evolution of a continuous variable as an independent “random walk” in each species that causes a normal distribution of the variable across a phylogeny.) Contrasts in phenotype (X; e.g. number of multi-isoform genes) are calculated between each species and divided by its standard deviation [as detailed in 10]. Contrasts in a known variable (Y; e.g. social index) are also calculated. Therefore the X contrasts will be independent of each other but not of the Y contrasts. Thus the PIC adjustment has enabled sociality and the three metrics of alternative splicing to be comparable between the six focal species here without phylogenetic relationships contributing as a confounding factor [10, 364–367].

### 6.2.4 Splicing event type and gene ontology

The R package *spliceR* (version 1.14.0) was used to identify different types of alternative splicing event [368]. Regarding the gene ontology (GO) analysis, the DNA sequences of the multi-isoform gene transcripts were retrieved from the reference transcripts of the same species (GCF\_000002195.4\_Amel\_4.5\_rna\_from\_genomic.fna, GCF\_000214255.1\_Bter\_1.0\_rna\_from\_genomic.fna, GCF\_000341935.1\_Ccin1\_rna\_from\_genomic.fna, GCF\_000147195.1\_HarSal\_1.0\_rna\_from\_genomic.fna, GCF\_000980195.1\_M.pharaonis\_V2.0\_rna\_from\_genomic.fna, GCF\_000187915.1\_Pbar\_UMD\_V03\_rna\_from\_genomic.fna). The first 2000 bases of the isoforms of the multi-isoform genes were selected for efficiency to find the corresponding gene names by comparison with the *Drosophila melanogaster* nr database (blastx), and to give GO terms. Fisher's exact test was implemented to identify GO terms which were enriched in the multi-isoform genes compared with the reference transcriptome of *D. melanogaster* (FlyBase [369]), a solitary insect.

## 6.3 Results

### 6.3.1 Relative splicing proportions

For the insects classified as highly eusocial, initially 4,750 genes out of a total 14,339 genes had more than one isoform (FPKM > 0.5) in *Apis mellifera*. This corresponded to 21,373 isoforms, with 9,351 splicing events. In *M. pharaonis*, 6,621 out of 15,155 genes have more than one isoform (27,488 isoforms and 13,167 splicing events). *P. barbatus* had 6,039 out of 13,959 genes with more than one isoform (24,329 isoforms, 11,912 splicing events).

By contrast, in the primitively eusocial species, *B. terrestris* showed 6,034 out of 12,254 genes to have more than one isoform (26,502 isoforms, 12,800 splicing events),

and in *H. saltator* 6,800 out of 16,610 genes had more than one isoform (29,109 isoforms, 14,670 splicing events). The sawfly *C. cinctus* outgroup had 6,471 genes out of 12,287 with more than one isoform (33,379 isoforms with 15,533 splicing events). The relative splicing proportions were then identified with regards to sociality index in three different formats.

#### 6.3.1.1 Metric 1: Proportion of multi-isoform genes

Firstly, the degree of sociality and proportion of multi-isoform genes was compared. After controlling for coverage, the solitary sawfly (*C. cinctus*) had the highest proportion of multi-isoform genes. The primitively eusocial bee (*B. terrestris*) showed a greater proportion of genes with multiple isoforms than the highly eusocial bee (*A. mellifera*) (Figure 6.4). The same trend was also witnessed in the ant lineage with the primitively eusocial ant *H. saltator* having a greater proportion of multi-isoform genes compared to the highly eusocial ants (*M. pharaonis* and *P. barbatus*). However, after applying a PIC this trend was found to not be significant (Linear regression. *EF1 alpha F2* distance matrix: Adjusted R-squared = 0.4305, F-statistic = 4.779, df = 1,4, p = 0.0941; *leucine tRNA ligase* distance matrix: Adjusted R-squared = 0.4624, F-statistic = 5.3, df = 1,4, p = 0.0827) (Figures 6.5a and 6.5b). In other words, there was no correlation found here between degree of sociality and proportion of multi-isoform genes.

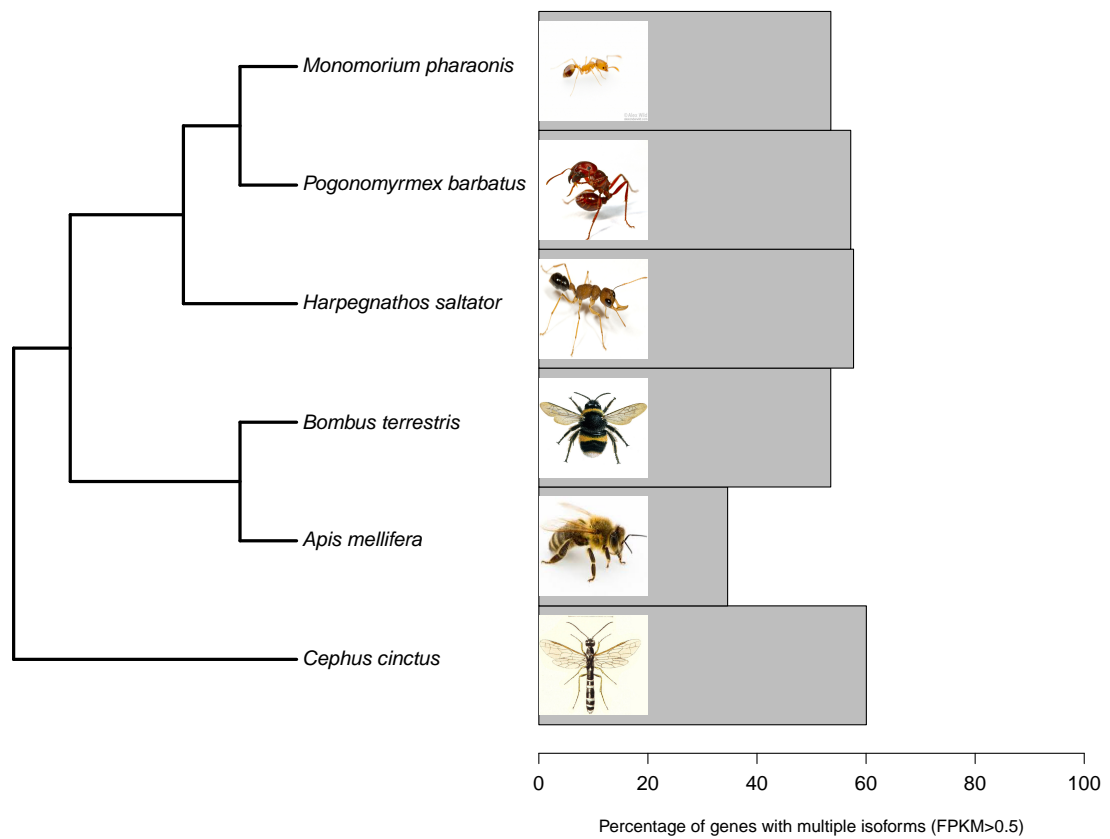


FIGURE 6.4: Phylogeny and percentage of total genes with alternative splicing after controlling for coverage. See Appendix F: Table F.1 for image references.

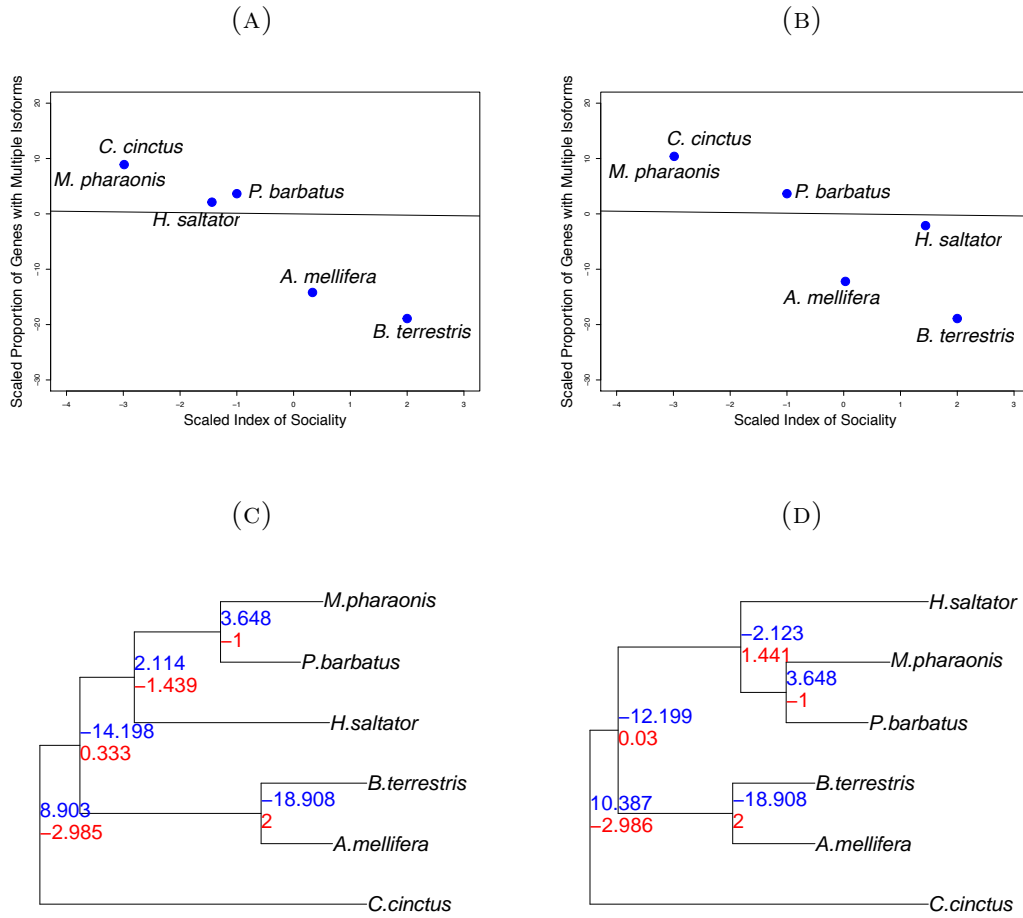


FIGURE 6.5: Phylogenetic Independent Contrast (PIC) adjusted proportion of multi-isoform genes and social index. PIC adjusted multi-isoform genes with phylogenetic distances based on (A) *elongation factor 1 alpha F2* (*EF1 alpha F2*), and (B) *leucine tRNA ligase*. Phylogenies generated using (C) *EF1 alpha F2* (D) *leucine tRNA ligase* distance matrices with PIC scaled proportion of multi-isoform genes (blue) and PIC scaled social index (red). The axes of scaled index of sociality is calculated from the sociality index and phylogenetic distance matrices using Felsenstein's method [10].

### 6.3.1.2 Metric 2: Mean isoform number per multi-isoform gene

Second, the mean isoform number per multi-isoform gene was identified in the six focal species. *C. cinctus* displayed the highest mean isoform number per multi-isoform gene (Figure 6.6). The highly eusocial honeybee had a higher mean isoform



count than the primitively eusocial bumblebee. Conversely, the primitively eusocial ant had a higher mean isoform count than the two highly eusocial ant species. After applying the PIC, there was no significant correlation between social status and number of isoforms per multi-isoform gene (Linear regression. *EF1 alpha F2* distance matrix: Adjusted R-squared = 0.266, F-statistic = 2.807, df = 1,4, p = 0.1692; *leucine tRNA ligase* distance matrix: Adjusted R-squared = 0.296, F-statistic = 3.103, df = 1,4, p = 0.153).

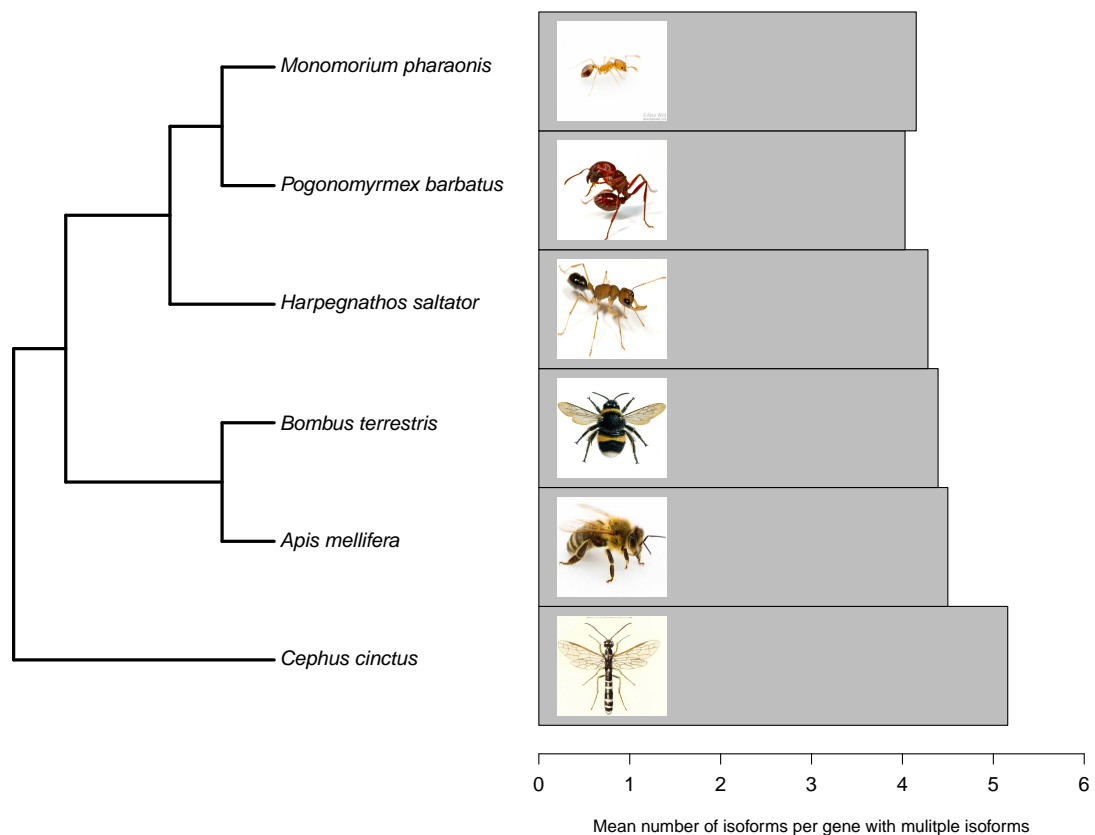


FIGURE 6.6: Phylogeny and mean number of isoforms per multi-isoform gene. See Appendix F: Table F.1 for image references.

### 6.3.1.3 Metric 3: Number of splicing events per multi-isoform gene

Third, there was a negative correlation between number of splicing events per multi-isoform gene and degree of eusociality (Linear regression. *EF1 alpha F2* distance

matrix: Adjusted R-squared = 0.860, F-statistic = 31.78, df = 1,4, p = 0.00487; *leucine tRNA ligase* distance matrix: Adjusted R-squared = 0.860, F-statistic = 31.77, df = 1,4, p = 0.00488) (Figures 6.8a and 6.8b), with *B. terrestris* and *H. saltator* having a greater mean number of splicing events per multi-isoform gene than *A. mellifera*, *M. pharaonis*, and *P. barbatus* (Figure 6.7). The sawfly outgroup again had the highest mean number of splicing events per multi-isoform gene.

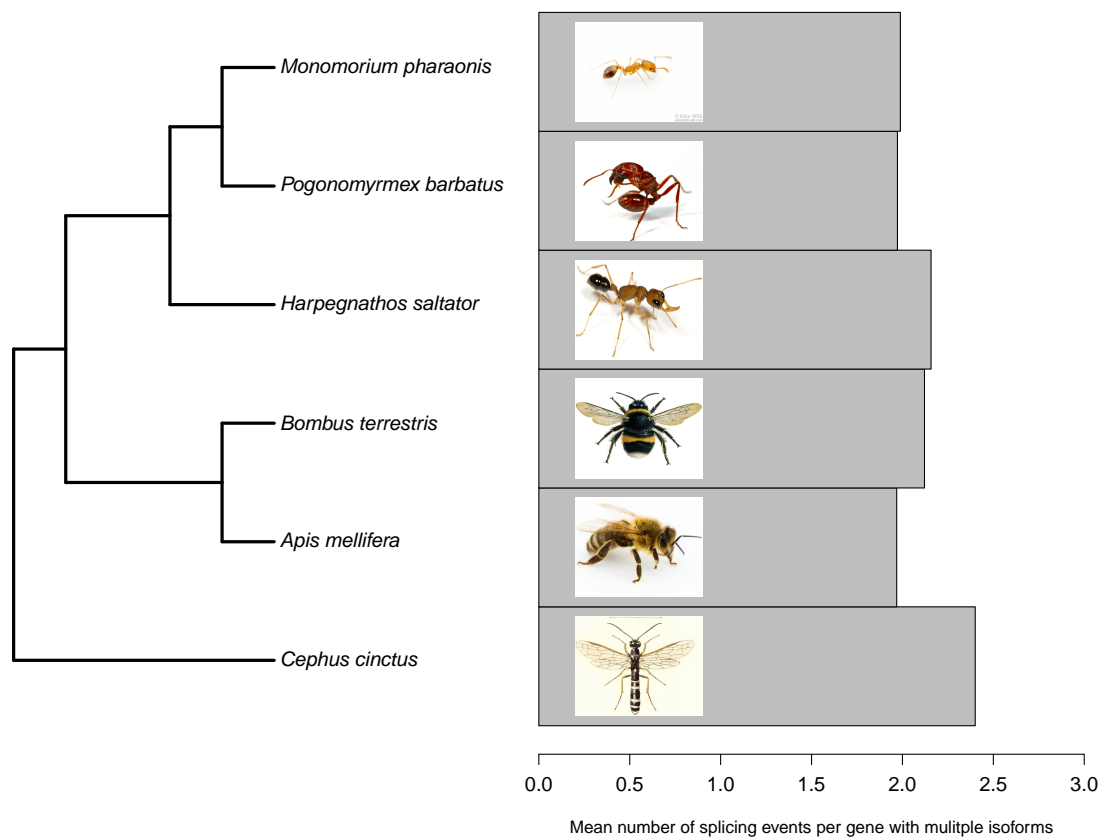


FIGURE 6.7: Phylogeny and mean number of splicing events per multi-isoform gene. See Appendix F: Table F.1 for image references.

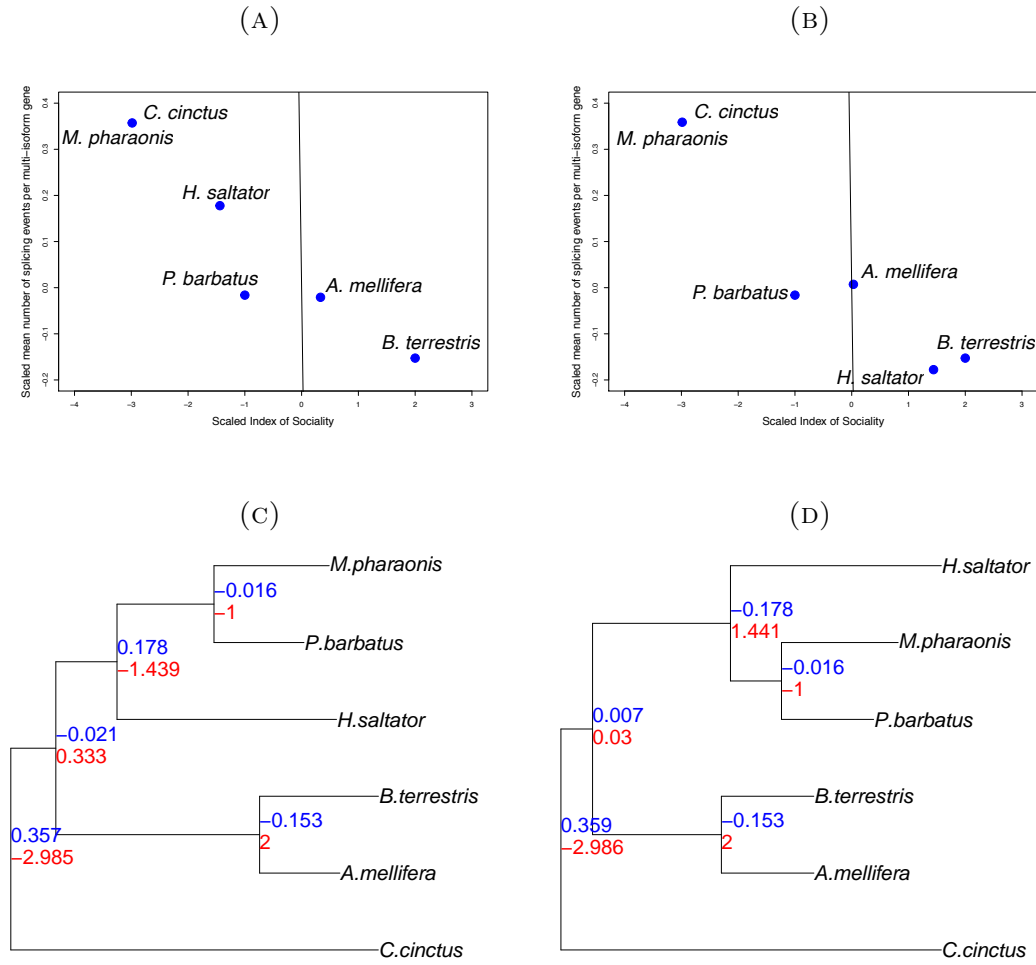


FIGURE 6.8: PIC adjusted mean number of splicing events per multi-isoform gene and social index. PIC adjusted multi-isoform genes with phylogenetic distances based on (A) *EF1 alpha F2*, and (B) *leucine tRNA ligase*. Phylogenies generated using (C) *EF1 alpha F2* and (D) *leucine tRNA ligase* distance matrices with PIC scaled proportion of multi-isoform genes (blue) and PIC scaled social index (red). The axes of scaled index of sociality is calculated from the sociality index and phylogenetic distance matrices using Felsenstein's method [10].

### 6.3.2 Splicing event type

A variety of splicing events are witnessed in these six species (Figure 6.9). The majority of these splicing events tend to be at alternative transcription start sites, alternative transcription termination sites, and exon skipping/inclusion events. Very

few alternative splicing events showed mutually exclusive exons in their isoforms. However, no significant difference in splicing event type between the six species was found (chi-squared contingency table: chi-square = 21.0, df = 35,  $p = 0.970$ ).

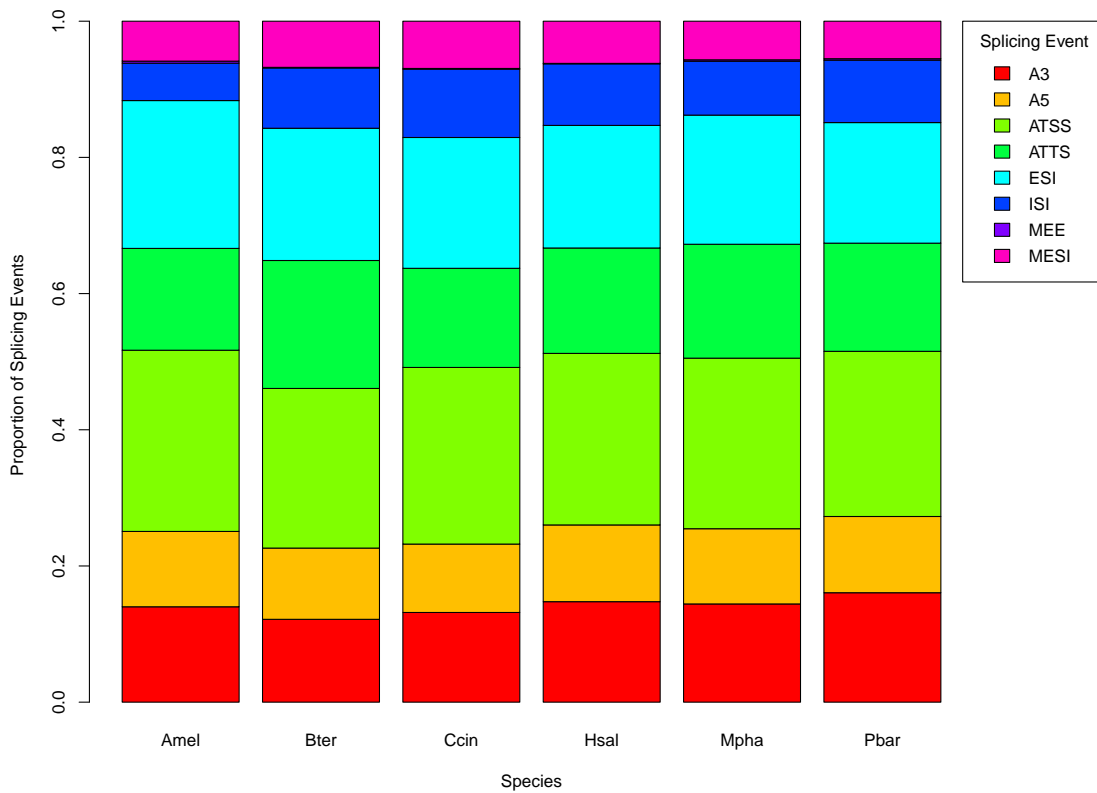


FIGURE 6.9: Stacked barplot of the proportion of alternative splicing events in six Hymenopteran species. Abbreviations: A3 = alternative 3' splice site, A5 = alternative 5' splice site, ATSS = alternative transcription start site, ATTS = alternative transcription termination site, ESI = exon skipping/inclusion, ISI = intron skipping/inclusion, MEE = mutually exclusive exons, MESI = multiple exon skipping/inclusion. Amel = *Apis mellifera*, Bter = *Bombus terrestris*, Ccin = *Cephus cinctus*, Hsal = *Harpegnathos saltator*, Mpha = *Monomorium pharaonis*, Pbar = *Pogonomyrmex barbatus*.

### 6.3.3 Gene ontology

The alternatively spliced genes of the six hymenopteran species studied were significantly enriched for the functions of neurogenesis, signalling pathways, and

various housekeeping functions in comparison with the solitary insect *Drosophila melanogaster* (Diptera) (Figures 6.10, 6.11, 6.12, 6.13, 6.14, and 6.15).

There was significant overlap between the enriched GO terms of the multi-isoform genes for each of the six focal species (Hypergeometric test:  $p < 0.00001$ ). This is likely to be a reflection of the divergence of the Dipteran and hymenopteran lineages resulting in differential gene ontology patterns. The gene ontology library of *Drosophila melanogaster* is likely to be different to that of a hymenopteran species. Therefore this gene ontology analysis indicates the gene functions which are particularly prone to alternative splicing in a largely eusocial lineage.

The majority of GO terms were identified in only a single species, with a surprisingly high number of enriched GO terms found only in *A. mellifera*. Phylogeny was conveyed in the number of enriched GO terms in species intersections (Figure 6.16) with a greater number of GO terms solely shared between closely related *P. barbatus* and *M. pharaonis* compared with any other two species. Moreover, seventeen GO terms were enriched in all six species studied (Figure 6.16: bar depicted in orange). These seventeen GO terms were found to be associated with cellular housekeeping functions indicating the probable importance of this epigenetic conservation (Figure 6.17).

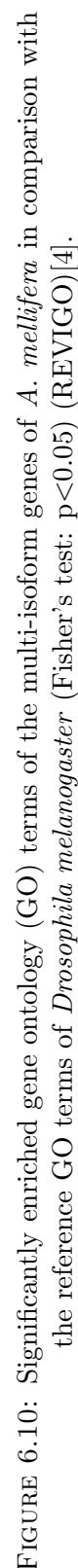


FIGURE 6.10: Significantly enriched gene ontology (GO) terms of the multi-isoform genes of *A. mellifera* in comparison with the reference GO terms of *Drosophila melanogaster* (Fisher's test:  $p < 0.05$ ) (REVIGO)[4].

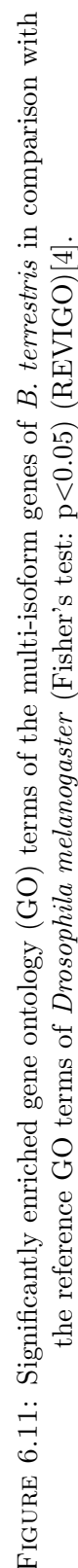


FIGURE 6.11: Significantly enriched gene ontology (GO) terms of the multi-isoform genes of *B. terrestris* in comparison with the reference GO terms of *Drosophila melanogaster* (Fisher's test:  $p < 0.05$ ) (REVIGO)[4].

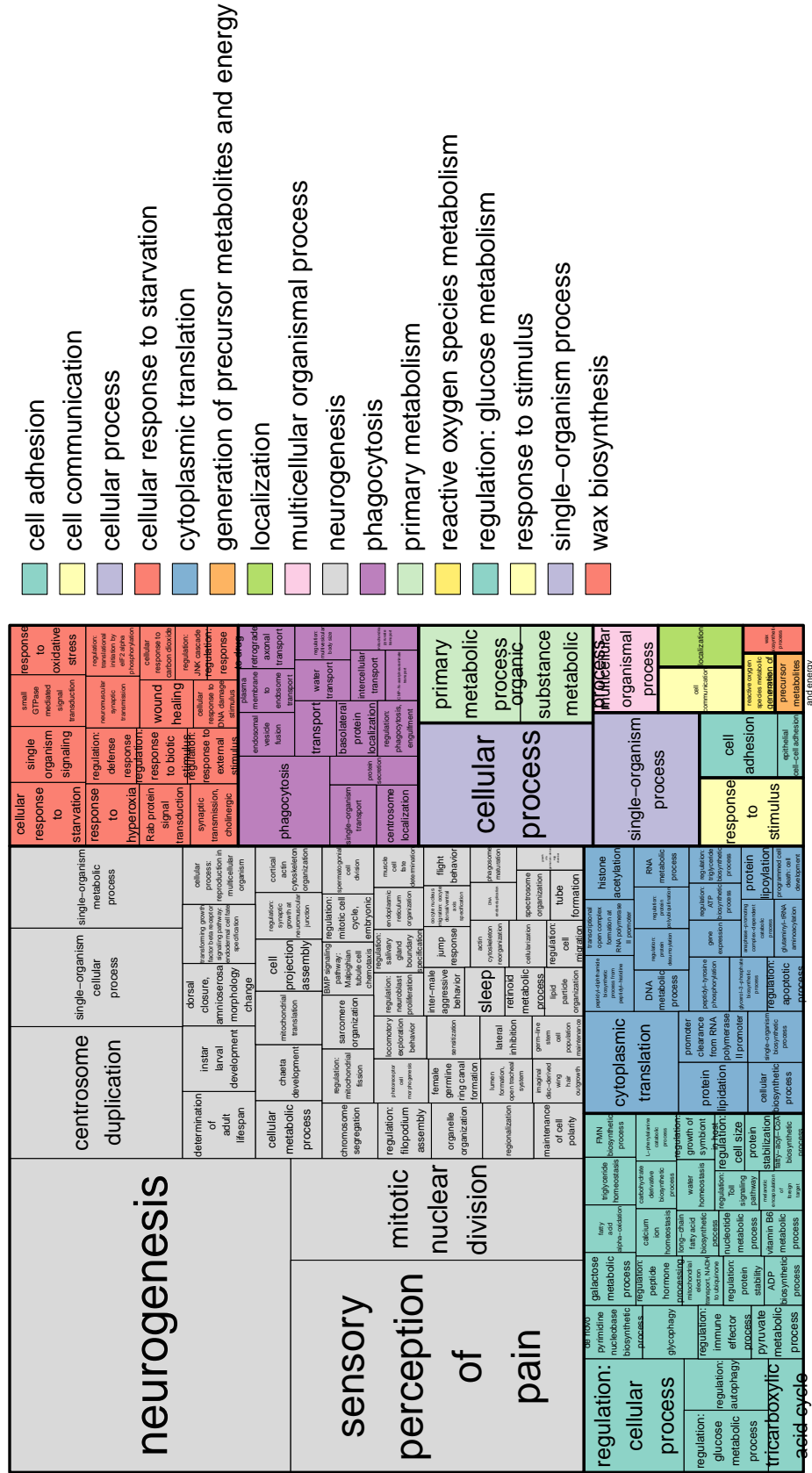


FIGURE 6.12: Significantly enriched gene ontology (GO) terms of the multi-isoform genes of *C. cinctus* in comparison with the reference GO terms of *Drosophila melanogaster* (Fisher's test:  $p < 0.05$ ) (REVIGO)[4].



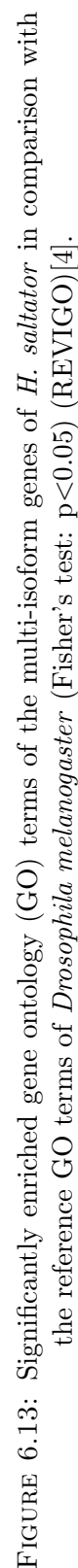


FIGURE 6.13: Significantly enriched gene ontology (GO) terms of the multi-isoform genes of *H. saltator* in comparison with the reference GO terms of *Drosophila melanogaster* (Fisher's test:  $p < 0.05$ ) (REVIGO)[4].

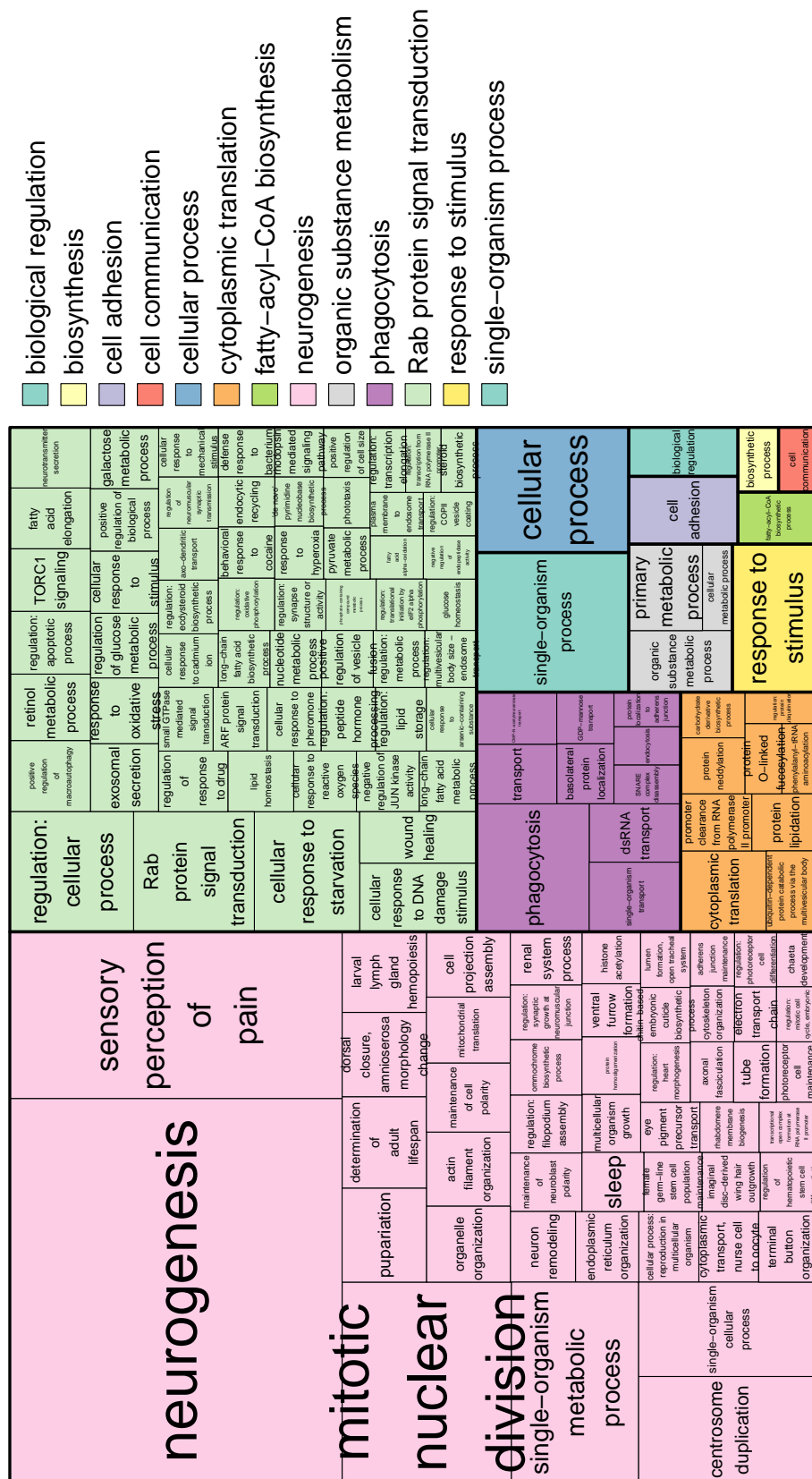
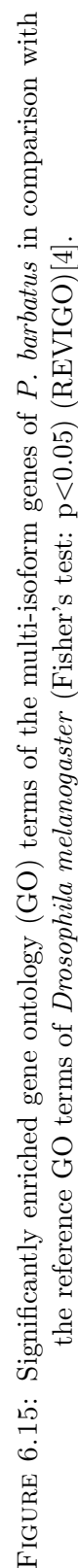


FIGURE 6.14: Significantly enriched gene ontology (GO) terms of the multi-isoform genes of *M. pharaonis* in comparison with the reference GO terms of *Drosophila melanogaster* (Fisher's test:  $p < 0.05$ ) (REVIGO)[4].



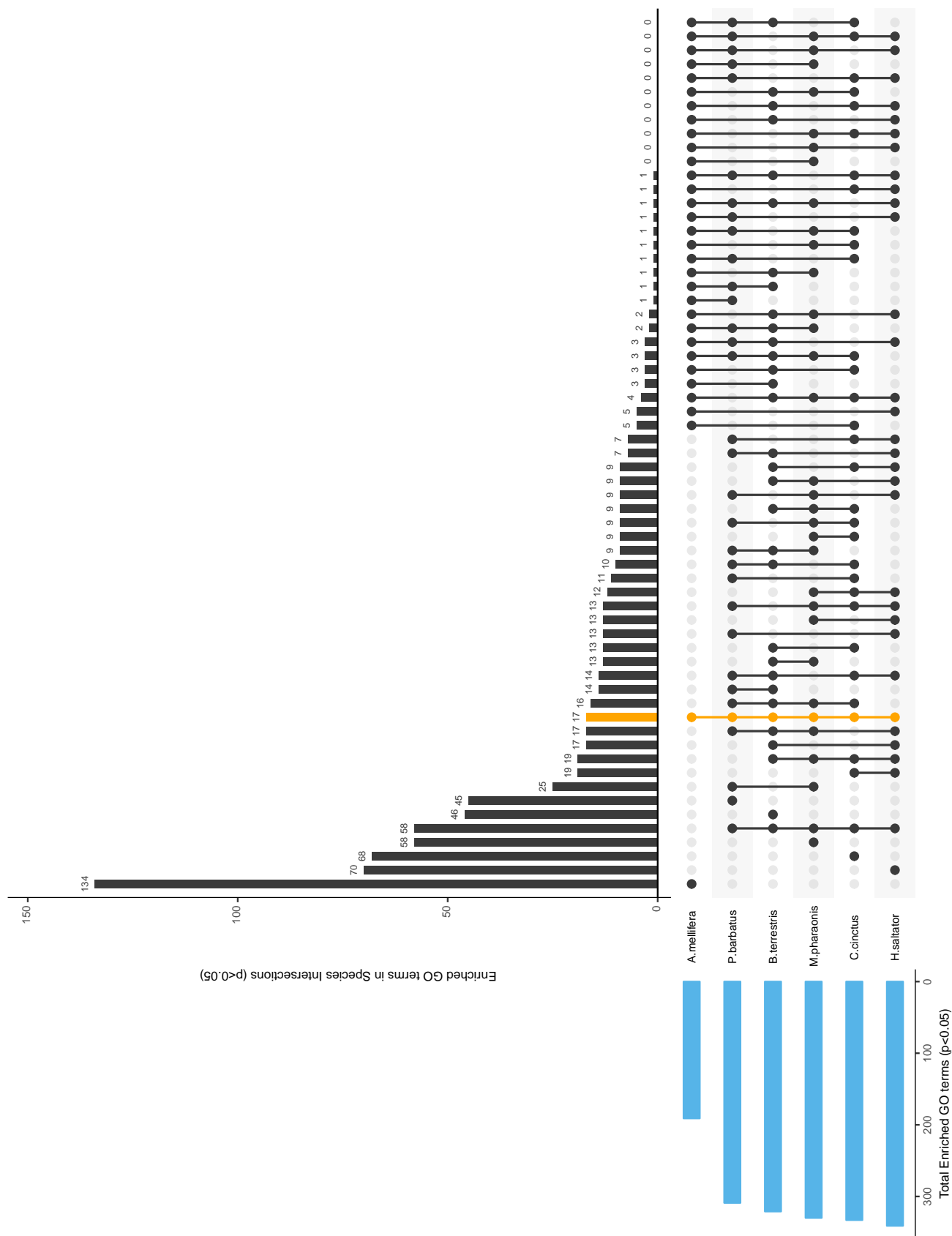


FIGURE 6.16: Overlap of enriched GO terms (Fisher test:  $p < 0.05$ ) of multi-isoform genes (from Figures 6.10, 6.11, 6.12, 6.13, 6.14, and 6.15) between the six hymenopteran species. The orange bar represents GO terms of alternatively spliced genes that are significantly enriched in all six species.

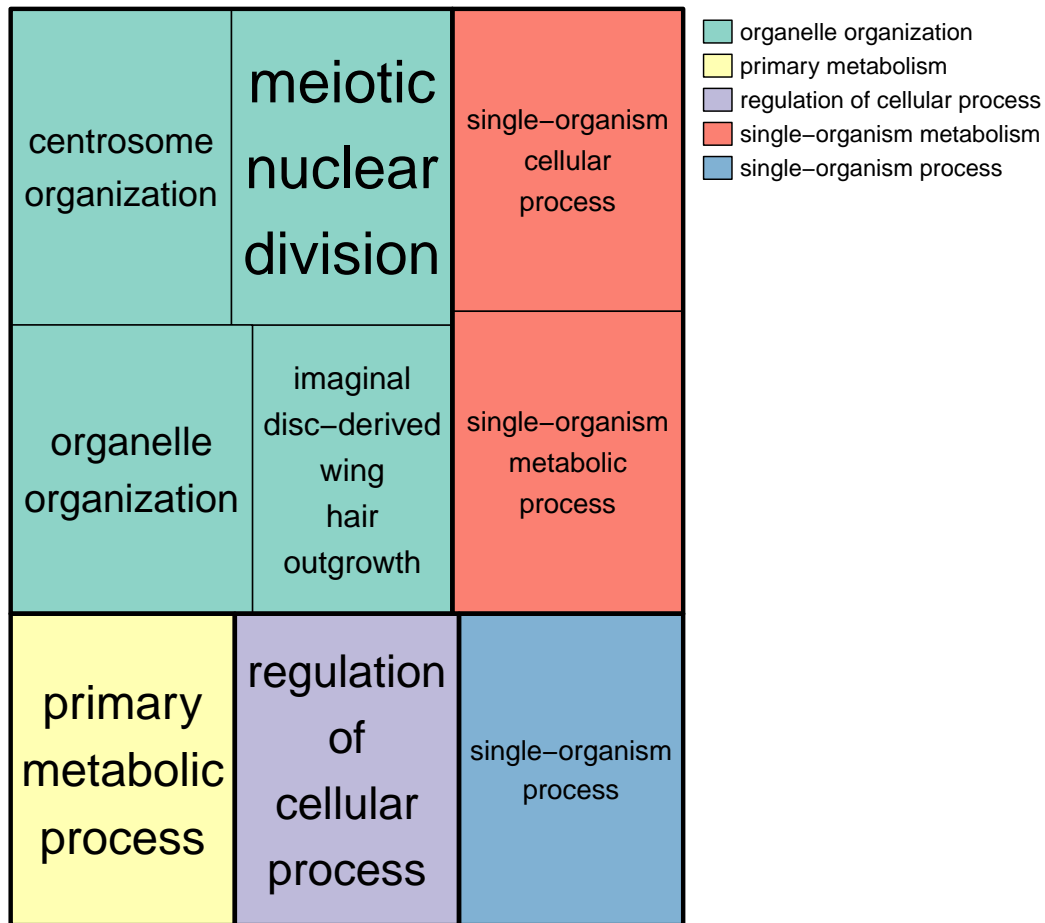


FIGURE 6.17: Gene ontology terms of multi-isoform genes that are significantly enriched in all six Hymenopteran focal species (Fisher's test:  $p < 0.05$ ) (REVIGO) [4].

## 6.4 Discussion

The prevalence of alternative splicing events here observed in the Hymenoptera (34-60%) is comparable to the pervasiveness of splicing events previously recorded in other insect species. 41.66% of all genes had alternative splicing events in the fruit fly, *Drosophila melanogaster*, 29.74% in the head louse, *Pediculus humanus capitis*, and 30.89% in the body louse, *Pediculus humanus corporis* [127].

A significantly fewer number of splicing events in multi-isoform genes were found to occur in more highly eusocial species. This pattern is consistent with the loss of

splicing sites in more highly eusocial caste individuals as hypothesised (see section 6.1). However, in contrast with the predictions no association between degree of eusociality and proportion of multi-isoform genes in workers is here witnessed. There are two possible reasons why this is the case. Firstly, alternative splicing may have a negligible effect on allowing fewer proteins to be expressed in more highly eusocial castes and instead other epigenetic mechanisms such as DNA methylation and histone modifications may have greater roles to play. Secondly, in this study six Hymenopteran species were used with two transcriptomes each, thus there may not have been sufficient power here to detect a correlation of sociality with alternative splicing. There was a 9% and 8% probability that the proportion of multi-isoform genes here witnessed was not due to the insects' social status, with phylogenies applied from *leucine tRNA ligase* and *EF1 alpha F2* respectively (Section 6.3.1: PIC). Therefore, it could be of interest in the future for repetition with a more extensive representation of the Hymenopteran lineage.

Moreover, one must be mindful of the limitations of this study. Due to the number of transcriptomes currently available different body samples were used. Transcriptomes sequenced from whole body samples were used for *M. pharaonis*, *H. saltator*, *A. mellifera*, and *C. cinctus*, and transcriptomes sequenced from head samples were used for *B. terrestris* and *P. barbatus*. If this would have any effect on the degree of alternative splicing one would expect a higher number of genes showing alternative splicing using whole body samples compared to head samples. If this was the case it could cause highly eusocial *P. barbatus* to have a greater number of multi-isoform genes than the primitively eusocial *H. saltator*. This would also cause an even greater negative correlation between degree of eusociality and pervasiveness of alternative splicing in the bee lineage. Furthermore, it would be of interest in the future to apply the eusociality index developed by Keller and Perrin [370]. This would require accurate information for each study species of the proportion of total energy spent on offspring production and the proportion of the number of genetic copies made by the  $i$ th individual from the parental generation, but it could potentially produce a

more accurate depiction of the social status of the Hymenopteran species.

Previously exon skipping was found to be the most common type of alternative splicing in mammals (38.4%) [371, 372], and alternative 3' splice site events were found to be most prevalent in *Drosophila* [372]. Keren and colleagues [178] describe the prevalence of exon skipping to increase with greater complexity in the Eukaryota. Therefore, it has been suggested that exon skipping demonstrates the greatest contribution to phenotypic complexity, whereas the predominance of alternative 5' and 3' splice sites are seen as an “intermediate evolutionary stage” [178]. In contrast, alternative transcription start sites are here observed to be the most common splicing event type amongst the Hymenoptera. Few insect species have yet been studied regarding their splicing event range. Thus, in the future it would be of interest to investigate whether the prevalence of splicing types in other insects is similar to that of Hymenoptera or more consistent with the well studied model species *Drosophila melanogaster*.

Alternatively spliced genes were here found to be particularly associated with neurogenesis, signalling, and housekeeping gene ontology terms. This is consistent with previous studies that showed alternative splicing to be prevalent in genes associated with regulation and signalling [340–342].

In conclusion, the finding that on average fewer splicing events are present per multi-isoform gene in more highly eusocial species is indicative of a reduced degree of splicing in caste individuals of more highly eusocial species. This points towards alternative splicing having a role to play in the diversity of phenotypes and behaviour witnessed in the Hymenoptera. In the future, it would be of interest to discover if the multi-isoform genes identified have orthologs with multiple isoforms in other species. It would also be interesting to see if this trend in workers is continued in male and queen members of eusocial species. In addition, it was previously found that the evolution of transcription factors - another facilitator of phenotypic diversity (Section 1.2.1) - is more divergent among eusocial insects than between solitary and

---

eusocial insects [42]. Therefore, future research could investigate whether there are similarly higher overall levels of alternative splicing in species of high eusociality compared to that of solitary or primitively eusocial species.



# Chapter 7

## General discussion

### 7.1 A summary of the results

This thesis has focused on the epigenetic mechanisms underlying two main polyphenisms in insects: the eusocial insect caste polyphenism and the locust phase polyphenism. I summarise my findings below.

### Chapter 2: Searching for allele-specific expression in *Bombus terrestris*.

Eusocial insects in the Hymenoptera lineage are emerging as models for epigenetics. DNA methylation is known to affect allele-specific expression in the epigenetics of mammals and flowering plants. However, there is contradictory evidence on whether this pattern is conserved in eusocial insects.

Here, I investigated allele-specific expression and monoallelic methylation in the bumblebee, *Bombus terrestris*. I found nineteen genes that were both monoallelically methylated and monoallelically expressed in a single bee. A number of these genes are involved in reproduction. Fourteen of these genes expressed the hypermethylated

allele, while the other five expressed the hypomethylated allele. I also searched for allele specific expression in twenty-nine published RNA-seq libraries. I found 555 loci with allele-specific expression. The presence of allele-specific expression was here established, however the underlying cis or epigenetic mechanisms are not clear. Therefore epigenetic routes of biased gene expression were explored in Chapters 3 and 4.

### **Chapter 3: Alternative splicing, DNA methylation, and gene expression in *Bombus terrestris*.**

In the buff-tailed bumblebee *Bombus terrestris* workers generally do not have fully developed ovaries. However, when the colony enters the competition phase of its annual cycle some of the workers begin to develop full ovaries and produce haploid sons. This distinct phenotypic change has previously been linked to DNA methylation [12] but the exact pathway is largely unknown. DNA methylation is known to affect alternative splicing and therefore gene expression in mammals. Here, I consider whether this association between these epigenetic mechanisms continues in *B. terrestris*, and if this correlates with worker reproduction.

In this study I found higher methylation levels in genes with a higher degree of expression in a similar manner to that found by Bonasio et al. [122]. Higher methylation levels were also found to be associated with differential expression between reproductive and non-reproductive workers. A similar trend was found for differential isoform expression. This indicates methylation to play a role in the maintenance or control of differential expression. This study was then extended to explore allele-specific effects in Chapter 4.

## **Chapter 4: Allele-specific expression and methylation in reproductive and non-reproductive *Bombus terrestris* workers.**

Both allele-specific expression and methylation have been implicated to be associated with worker reproduction. Allele-specific expression of worker reproduction genes were previously found in the bumblebee [185], and bumblebee reproductive workers have gene expression patterns more similar to queens compared to non-reproductive workers [12]. This chapter builds on work in Chapter 2 where I found genes with allele-specific expression in the bumblebee, and several of these had caste differentiation functions. It has also been suggested that methylation patterns play an important role in bumblebee worker reproduction [186].

Here I identified allele-specific gene expression and allele-specific methylation of reproductive and non-reproductive *Bombus terrestris* workers in multiple colonies. Comparisons of these factors were drawn in combination with the differentially expressed, differentially methylated and alternatively spliced genes found in Chapter 3. Allele-specific expression and allele-specific methylation show a weak correlation in reproductive and non-reproductive workers, so they may not be directly related. Thus the insect epigenetic process is implicated to be complex, with the involvement and interaction of DNA methylation, alternative splicing, hormone signalling pathways, and various forms of histone modification. For example, negative regulation of phosphorylation was here found to be highly involved in the bumblebee allele-specific expression and methylation processes.

## **Chapter 5: Differential circadian isoform expression patterns in the desert locust, *Schistocerca gregaria*.**

A different polyphenism system was considered in Chapter 5, the phase polyphenism of the desert locust *Schistocerca gregaria*. The gregarious form of *S. gregaria* is often considered a significant pest of crops [174], and it possesses a distinct circadian

pattern with activity peaking at 1400 hours and at it's lowest levels during the night [31, 171]. The circadian clock has been found to regulate alternative splicing in mice [302], and alternative splicing has been previously demonstrated to be important in alternate phases of a different locust species [143]. However it was unknown whether any isoforms showed expression patterns to reflect the strong circadian cycle of the locust.

Through this chapter I identified genes with differential circadian isoform expression patterns. Several of these genes had previously been described to be associated with the circadian clock in other organisms. This reflects the high rate of conservation of these functions across the Metazoa, and the number and breadth of processes that are affected by the circadian clock.

## **Chapter 6: Alternative splicing and eusociality in the Hymenoptera.**

Highly eusocial insects are distinguished from solitary and more primitively eusocial animals by the caste system which creates reproductive and sterile individuals in the same colony [156]. Due to the division of labour a single caste will carry out a smaller variety of tasks compared to an individual belonging to a less highly eusocial species. Alternative splicing facilitates multiple proteins to be expressed from the same gene [178]. Thus it was hypothesised that there would be lower caste-specific splicing levels in more highly eusocial insects.

In contrast to predictions, no difference was deduced in the proportion of genes with multiple isoforms or the mean number of isoforms per multi-isoform gene compared with level of eusociality. However, I found that there were fewer splicing events per gene with multiple isoforms in more highly eusocial hymenopteran species compared with primitively eusocial and solitary species. This is consistent with the proposed hypothesis. The reduction in the degree of splicing in caste individuals of more

highly eusocial species points towards alternative splicing having a role to play in the diversity of phenotypes and behaviour witnessed in the Hymenoptera. In addition, consistent with previous literature [340–342], alternatively spliced genes were here found to be particularly associated with neurogenesis, signalling, and housekeeping gene ontology terms.

## 7.2 Future implications

Throughout the first three results chapters (Chapters 2, 3, and 4) focusing on the caste polyphenism of the bumblebee, I concentrated on methylation and expression profiles in the context of worker reproduction in particular. The relationship between these factors and other epigenetic mechanisms has become more apparent. Furthermore the types of gene functions (e.g. caste differentiation and worker reproduction) that are associated with these patterns has been revealed. These observations along with the prevalence of allele-specific effects are reminiscent of the phenomenon of genomic imprinting and the predictions made by Queller [11] (detailed in Section 7.2.1 below).

### 7.2.1 Genomic imprinting link

Genomic imprinting is described as parent-specific gene expression, whereby the allele expressed is dependent on from which parent the allele was inherited [373]. One example of genomic imprinting is the *Zdbf2* locus that is imprinted in mice, causing the long isoform of *Zdbf2* (*Liz*) transcript to be expressed briefly in early embryos and embryonic stem cells [374]. Notably a study recently found that this transient early embryonic expression programs the adult epigenetic state [375]. Genomic imprinting has been identified in mammals and angiosperms, and chromosomal genomic imprinting has been reported in various other organisms [376–380]. In mammals genomic imprinting is predominantly present in the reproductive tissues

(e.g. placenta), but it has also been reported in the brain [107, 381]. However, it has been questioned since the 1990's why the established phenomenon of genomic imprinting first evolved [382].

Genomic imprinting is counter-intuitive. In diploid organisms, natural selection is expected to favour expression of both alleles of a gene to ensure normal expression levels even when a non-functional or deleterious allele arises [383]. Therefore it is unexpected for a mechanism such as imprinting to evolve that silences one allele [384].

Several explanations have been put forward as to why genomic imprinting evolved in the first place. The three main theories comprise of Day and Bonduriansky's sexual antagonism theory [385], Wolf and Hager's maternal-offspring co-adaptation theory [386], and Haig's kinship theory [384]. Firstly, the sexual antagonism theory relies on sex-specific selection pressures acting on a gene [385]. Genomic imprinting is said to allow the more adaptive allele for the offspring's gender to be expressed. However this theory does not account for imprinted genes expressing the opposite sex-of-parent allele, which has repeatedly been demonstrated through reciprocal crosses [384].

Secondly, the maternal-offspring co-adaptation theory claims the fitness of offspring can largely depend on interactions between the offspring and their mother [386]. Genomic imprinting is proposed to evolve when the two alleles of a gene in the offspring relate to different fitness levels due to their effect on interactions with the mother. One major drawback of this approach is that there is no prediction of genomic imprinting in adults, unless erasure of an imprint is costly [384]. Yet adult genomic imprinting has been reported in the literature in several scenarios [387, 388].

The leading explanation for the evolution of imprinting is Haig's kinship theory [168]. This proposes that genomic imprinting arose due to maternal- and paternal-derived alleles having different selection pressures in relation to kin resource allocation [168].

Haig's theory is the principal line of thought, because it is consistent with the currently found imprinted genes in mammals and angiosperms [384, 389]. Unlike the other two theories, the kinship theory is compatible with both expression of the opposite sex-of-parent allele and its presence in adults.

Haig's theory goes on to predict imprinting to occur in organisms with different degrees of relatedness in close social groups, such as eusocial insects [11]. Haplodiploid social insects are prime candidates for tests of kinship theory as there are many contexts within a colony in which an individual interacts with conspecifics of varying relatedness, for instance worker reproduction, queen competition, and colony fission. Queller developed this theory to make predictions of the types of genes functions, and which alleles (matrigene/patrigene) should be imprinted in the social Hymenoptera in various contexts [11]. To note, matrigenes are alleles in self which have been inherited from the mother, whereas maternal alleles are genes within the mother. The same concept applies for patrigenes and paternal alleles.

In haplodiploid species females hatch from fertilized (diploid) eggs, whilst males hatch from unfertilised (haploid) eggs, so the degree of relatedness between offspring is different for matrigenes and patrigenes. The coefficient of relatedness ( $r$ ) between a worker (Figure 7.1) and other sister workers is 0.75, which is higher than their relatedness with their own offspring ( $r = 0.5$ ). Therefore, according to the classic Hamiltonian inclusive fitness theory, in the decision of whether to help care for workers laid by the queen or produce their own offspring, a worker is predicted to choose to help raise their sister workers in a bumblebee colony [11, 390].

However, in a bumblebee colony where there is genomic imprinting different conclusions are made. Queller's predictions are laid out in the decision of a worker bee to either help care for the sons laid by the colony queen (i.e. brothers), help care for the offspring of workers (i.e. nephews), or to reproduce herself to produce sons [11, 69].

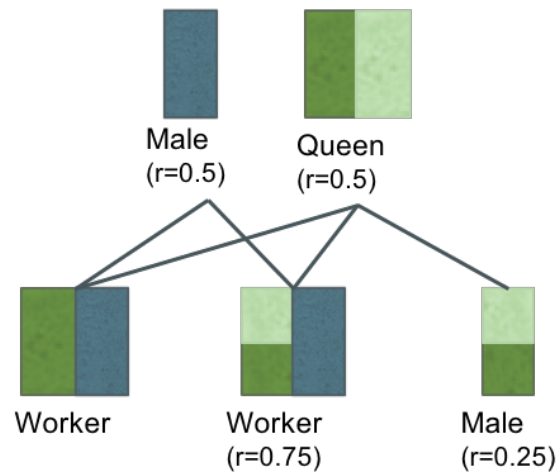


FIGURE 7.1: *Bombus terrestris* pedigree.  $r$  = coefficient of relatedness (adapted from Queller 2003 [11])

In bumblebee colonies the queen is monogynous (i.e. singly mated) causing all worker offspring to have the same father. Compared to other workers  $r = 1$  for a patriline. Yet  $r = 0$  with sons produced by the queen (brothers) for the patriline. Therefore in a queen-right colony reproduction of sisters or the focal worker is the only way for the same patriline as in the focal worker to be continued into the next generation. Conversely, for matriline  $r = 0.5$  with other workers and  $r = 0.5$  with brothers. So matriline is expected to regulate the focal worker's reproduction when it is associated with high costs to brothers, because matriline is less likely to occur in nephews ( $r = 0.25$ ) than they are in brothers.

In a scenario when the founding bumblebee queen dies (i.e. queenless), the only remaining options are to either help care for the offspring of workers, or to reproduce herself to produce sons. There now should not be any differential selective pressure on the patriline since it has the same degree of relatedness to the offspring of the focal worker and the offspring of her sisters ( $r = 0.5$ ). Whereas the matriline would favour reproduction by the focal worker since the degree of relatedness is higher with her sons ( $r = 0.5$ ) compared to with nephews ( $r = 0.25$ ).

Identifying genomic imprinting in the social Hymenoptera (ants, bees, and wasps)



would provide an opportunity to test this evolutionary theory [11]. Recently Galbraith et al. [182] and Kocher et al. [187] carried out reciprocal cross experiments which tested Haig's predictions in the honey bee, and they each provide evidence for parent-of-origin effects (described in Section 4.1). However these two studies were somewhat contradictory, leaving this debate still remaining unresolved. Galbraith and colleagues found an association between patrigene expression and worker reproduction. Whereas, Kocher and colleagues reported a small group of significant parent-specific gene expression which was primarily maternally biased. These studies were carried out with different tissues. Moreover, European and African *Apis mellifera ligustica* bees were sampled in Galbraith et al., but Kocher et al. used two different honeybee lineages from these continents, namely *A. mellifera carnica* and *A. mellifera scutellata*. Therefore Pegoraro and colleagues [69] suggested that the parent-of-origin effects Kocher et al. [187] and Galbraith et al. [391] witnessed could be conditional to the environmental and tissue-specific context in which the difference in expression was tested.

A method of determining indicators of the presence of genomic imprinting in social insects is the identification of the underlying mechanisms. DNA methylation is key to imprinting in mammals. The addition of a methyl group to the fifth carbon of cytosine (5-methylcytosine - 5mC) in an imprinting control region on only one of the two alleles of a gene is associated with the allele-specific expression of that gene [392]. This association is well described in mammals [69]. However, any relationship between methylation and parent-specific gene expression in social insects is not so clear, as is discussed below (Section 7.2.2).

### 7.2.2 Methylation: driver, effector, or response?

In mammals, methylation acts as a marker on imprinting control regions which initiates the silencing of the associated allele [69, and references therein]. The presence of a complete set of DNMTs (apart from *Polistes* wasps) in social insects [44, 103, 113]

and the prediction of imprinting in this lineage [11] leads us to question whether methylation initiates or “drives” expression changes in the Hymenoptera in a process similar to that of mammals.

In mammals more than 70% of CpGs are methylated and DNA methylation acts to silence specific genes via imprinting control regions [104]. Whereas hymenopteran methylation levels are much lower with less than 2% of CpGs methylated, exons are targeted (Chapter 3 and [122]), and this epigenetic marker is generally believed to have an opposite effect with being associated with increased gene expression [106]. Therefore if methylation is a driver of parent-specific gene expression the exact process is predicted to differ to that of the mammalian lineage. In several previous studies differential methylation has been found to be associated with processes predicted to be imprinted in social insects such as caste differentiation and worker reproduction (see Section 1.2.2). Moreover, my studies found higher methylation levels to be associated with differential expression between reproductive and non-reproductive workers (Chapter 3). This indicates methylation to play a role in the maintenance or control of differential expression. However, the absence of *Dnmt3* and reduced methylation system in a *Polistes* wasp [113] implies that either DNA methylation has a variable effect across the social Hymenoptera or that methylation is not the key driver to allele-specific expression in this lineage.

Another more probable scenario is that methylation could be an effector of allele-specific expression in combination with other mechanisms such as histone modifications and ncRNAs. My research demonstrates that there may not be a direct link between allele-specific expression and allele-specific methylation (Chapter 4) and other factors are implied to be involved, specifically alternative splicing, hormone signalling pathways, and histone modifications (Chapters 3 and 4). Moreover, methylation-independent genomic imprinting has been recently reported in mammals, showing that methylation is not necessary for parent-specific gene expression [393].

Alternatively, there is still the possibility that methylation is not a driver or effector of allele-specific expression in the Hymenoptera. Methyl groups could simply be opportunistic in their binding. Highly methylated genes are typically associated with housekeeping functions and are uniformly expressed in the Hymenoptera [121–123], which is supported by my findings in Chapter 3. Unravelling of chromatin which facilitates the transcription of genes would also help enable methylation to act on those regions. This process would explain any correlations witnessed between methylation and expression in the Hymenoptera. In this field correlations have only been able to be inferred thus far, yet novel methods like CRISPR will be able to investigate causation of methylation related effects. This is explored below in Section 7.2.3, along with a method to resolve the discrepancies between the studies of Kocher et al. and Galbraith et al. [187, 391].

### 7.2.3 Future directions

In the future, a reciprocal cross could be conducted with the same sub-species of bee to resolve the discrepancy between the studies by Kocher et al. and Galbraith et al. [182, 187] (Section 7.2.1). Galbraith sampled the ovaries and the fat body in combination [182]. Hence, one cannot determine whether the expression patterns seen were resultant of primarily the reproductive tissue, the fat body (which acts as a liver and resource provision for the oocytes), or if expression patterns were similar in both tissues. Ovaries largely consist of the oocytes. However, methylation and imprinting are erased in primordial germ cells and new imprints are established in mature gametes (assuming a methylation based method of imprinting in social insects). I would presume the ovaries would show expression patterns representative of the potential offspring of the worker bee, rather than the worker itself. Therefore, I would expect the fat body could be the tissue determining the parent-specific gene expression witnessed in Galbraith et al. [182]. Thus it would be of interest to

separately sample the ovaries, fat body, and brain in any future reciprocal cross. This would test the predicted tissue-specific nature of parent-specific gene expression.

Experiments such as these reciprocal crosses are facilitated by the inception of next generation sequencing. The breadth and depth of knowledge of hymenopteran epigenetics has been significantly advanced due to this methodology over the past few decades. Now, a new gene editing technology, CRISPR, has emerged which greatly builds on the capabilities of RNAi [394]. As described in Pegoraro et al. [69], CRISPR gene editing technology could be beneficial in searching for the underlying mechanisms of imprinting-linked expression patterns. Recently, CRISPR has been tested in the hymenopteran species of *Nasonia vitripennis*, *Apis mellifera*, and the clonal raider ant *Ooceraea biroi* [395–397]. This demonstrates that this method could be used in hymenopteran species to induce knock-downs or mutations of imprinting-related genes. Knock-down of genes involved in DNA methylation (e.g. DNMTs) and the measurement of the expression status of key worker reproduction or caste differentiation genes could elucidate any direct relationship between these factors. Studies so far (including my own) have established correlations, but this approach has the advantage of being able to determine causation.

Targeting of specific protein domains is also a benefit of CRISPR [394]. Therefore, this approach could be applied to histone modification proteins. In mammals, histone modifications are the epigenetic markers that are most frequently associated with imprinting [398]. Notably, histone modifications were often indicated throughout Chapters 3 and 4 to be involved in worker reproduction and linked to allele-specific expression and methylation patterns. Therefore, it would be of interest to target specific modifications, particularly of histone 3 (H3K9, H3K27, H3K4) and 4 (H4K20) [69], to elucidate the exact pathways involved.

The adaptability of CRISPR allows its application in many areas. Methylation tags and histone modifications are two common epigenetic markers, therefore it is thought that these factors are more likely to be involved in hymenopteran caste

differentiation and imprinting. However, other markers such as modifications on cytosines have been described in mammals and have been suggested to have a distinct role to play in the epigenome [see 69, and references therein]. For instance, 5-hydroxymethylationcytosine (5hmC) has been indicated to be an epigenetic tag and may regulate DNA methylation [399]. Hence the underlying pathways are likely to be complex and involve numerous tissue- and cell-specific marker types. CRISPR provides an exciting new method to establish these complex pathways.

# Chapter 8

## Conclusions

I have made the following contributions to scientific knowledge through my doctoral research:

1. The identification of genes with allele-specific expression in *Bombus terrestris*.
2. Genome-wide associations between gene expression and DNA methylation in *Bombus terrestris* were found for the first time, with higher exonic methylation levels compared to that of introns.
3. A weak association between allele-specific expression and allele-specific methylation in *Bombus terrestris* was elucidated.
4. For the first time, the presence of genes with differential circadian isoform expression patterns in the gregarious phase of the desert locust *S. gregaria* was identified.
5. The first evidence of fewer splicing events per gene with multiple isoforms in more highly eusocial hymenopteran species compared with primitively eusocial and solitary species.

# Bibliography

- [1] Ralf J. Sommer and Akira Ogawa. Hormone signaling and phenotypic plasticity in nematode development and evolution. *Current Biology*, 21(18):R758–R766, September 2011. ISSN 09609822. doi: 10.1016/j.cub.2011.06.034. URL <http://linkinghub.elsevier.com/retrieve/pii/S096098221100707X>.
- [2] C. Laforsch, W. Ngwa, W. Grill, and R. Tollrian. An acoustic microscopy technique reveals hidden morphological defenses in *Daphnia*. *Proceedings of the National Academy of Sciences*, 101(45):15911–15914, November 2004. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0404860101. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0404860101>.
- [3] M. Pigliucci, Courtney J. Murren, and Carl D. Schlichting. Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*, 209(12):2362–2367, June 2006. ISSN 0022-0949, 1477-9145. doi: 10.1242/jeb.02070. URL <http://jeb.biologists.org/cgi/doi/10.1242/jeb.02070>.
- [4] Fran Supek, Matko Bonjak, Nives kunca, and Tomislav muc. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, 6(7):e21800, July 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0021800. URL <http://dx.plos.org/10.1371/journal.pone.0021800>.
- [5] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression

- data. *Bioinformatics*, 26(1):139–140, January 2010. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btp616. URL <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btp616>.
- [6] D. J. McCarthy, Y. Chen, and G. K. Smyth. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10):4288–4297, May 2012. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gks042. URL <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gks042>.
- [7] R Development Core Team. R: A language and environment for statistical computing. 2015.
- [8] Joan E. Strassmann, David Queller, John Avise, and Francisco Ayala. *In the Light of Evolution: Volume V: Cooperation and Conflict*. National Academies Press, Washington, D.C., February 2012. ISBN 978-0-309-21836-8. URL <http://www.nap.edu/catalog/13223>. DOI: 10.17226/13223.
- [9] Ivica Letunic and Peer Bork. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1):W242–W245, July 2016. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkw290. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw290>.
- [10] Joseph Felsenstein. Phylogenies and the comparative method. *The American Naturalist*, 125(1):1–15, January 1985. ISSN 0003-0147, 1537-5323. doi: 10.1086/284325. URL <http://www.journals.uchicago.edu/doi/10.1086/284325>.
- [11] David C Queller. Theory of genomic imprinting conflict in social insects. *BMC Evol Biol.*, 3(15), 2003.



- [12] Mark C. Harrison, Robert L. Hammond, and Eamonn B. Mallon. Reproductive workers show queenlike gene expression in an intermediately eusocial insect, the buff-tailed bumble bee *Bombus terrestris*. *Molecular ecology*, 24(12):3043–3063, 2015. URL <http://onlinelibrary.wiley.com/doi/10.1111/mec.13215/full>.
- [13] Silvano Squizzato, Young Mi Park, Nicola Buso, Tamer Gur, Andrew Cowley, Weizhong Li, Mahmut Uludag, Sangya Pundir, Jennifer A. Cham, Hamish McWilliam, and Rodrigo Lopez. The EBI Search engine: providing search and retrieval functionality for biological data from EMBL-EBI. *Nucleic Acids Research*, 43(W1):W585–W588, July 2015. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkv316. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv316>.
- [14] Zoe N Lonsdale, Kate Lee, Maria Kyriakidou, Despina Nathanael, Harindra E Amarasinghe, Catherine O'Connor, and Eamonn B. Mallon. Monoallelic methylation and allele specific expression in a social insect. *PeerJ*, 2017. doi: 10.1101/022657.
- [15] E. Mayr. *Animal Species and Evolution*. Cambridge: Belknap Press of Harvard University Press, 1963.
- [16] Patrick J. Hu. Dauer. *WormBook*, 2007. ISSN 1551-8507. doi: 10.1895/wormbook.1.144.1. URL [http://www.wormbook.org/chapters/www\\_dauer/dauer.html](http://www.wormbook.org/chapters/www_dauer/dauer.html).
- [17] Donald L. Riddle, Margaret M. Swanson, and Patrice S. Albert. Interacting genes in nematode dauer larva formation. *Nature*, 290(5808):668–671, April 1981. ISSN 0028-0836, 1476-4687. doi: 10.1038/290668a0. URL <http://www.nature.com/doifinder/10.1038/290668a0>.
- [18] L. C. Weiss, J. Leimann, and R. Tollrian. Predator-induced defences in *Daphnia longicephala*: location of kairomone receptors and timeline of sensitive

- phases to trait formation. *Journal of Experimental Biology*, 218(18):2918–2926, September 2015. ISSN 0022-0949, 1477-9145. doi: 10.1242/jeb.124552. URL <http://jeb.biologists.org/cgi/doi/10.1242/jeb.124552>.
- [19] Jonathan M. Jeschke, Michael Kopp, and Ralph Tollrian. Predator functional responses: discriminating between handling and digesting prey. *Ecological Monographs*, 72(1):95–112, February 2002. ISSN 0012-9615. doi: 10.1890/0012-9615(2002)072[0095:PFRDBH]2.0.CO;2. URL [http://doi.wiley.com/10.1890/0012-9615\(2002\)072\[0095:PFRDBH\]2.0.CO;2](http://doi.wiley.com/10.1890/0012-9615(2002)072[0095:PFRDBH]2.0.CO;2).
- [20] Stuart R. Dennis, Gerald A. LeBlanc, and Andrew P. Beckerman. Endocrine regulation of predator-induced phenotypic plasticity. *Oecologia*, 176(3):625–635, November 2014. ISSN 0029-8549, 1432-1939. doi: 10.1007/s00442-014-3102-8. URL <http://link.springer.com/10.1007/s00442-014-3102-8>.
- [21] L. S. Mills, M. Zimova, J. Oyler, S. Running, J. T. Abatzoglou, and P. M. Lukacs. Camouflage mismatch in seasonal coat color due to decreased snow duration. *Proceedings of the National Academy of Sciences*, 110(18):7360–7365, April 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1222724110. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1222724110>.
- [22] X. Wang, Z. J. Tseng, Q. Li, G. T. Takeuchi, and G. Xie. From ‘third pole’ to north pole: a Himalayan origin for the arctic fox. *Proceedings of the Royal Society B: Biological Sciences*, 281(1787):20140893–20140893, June 2014. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2014.0893. URL <http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2014.0893>.
- [23] Laura Casas, Fran Saborido-Rey, Taewoo Ryu, Craig Michell, Timothy Ravasi, and Xabier Irigoien. Sex change in clownfish: molecular insights from transcriptome analysis. *Scientific Reports*, 6(1), December 2016. ISSN 2045-2322. doi: 10.1038/srep35461. URL <http://www.nature.com/articles/srep35461>.

- [24] Peter Buston. Social hierarchies: Size and growth modification in clownfish. *Nature*, 424(6945):145–146, July 2003. ISSN 00280836, 14764679. doi: 10.1038/424145a. URL <http://www.nature.com/doifinder/10.1038/424145a>.
- [25] Haruki Ochi. Mating behavior and sex change of the anemonefish, *Amphiprion clarkii*, in the temperate waters of southern Japan. *Environmental Biology of Fishes*, 26(4):257–275, December 1989. ISSN 0378-1909, 1573-5133. doi: 10.1007/BF00002463. URL <http://link.springer.com/10.1007/BF00002463>.
- [26] H. Fricke and S. Fricke. Monogamy and sex change by aggressive dominance in coral reef fish. *Nature*, 266(5605):830–832, April 1977. ISSN 0028-0836.
- [27] Yvonne Sadovy de Mitcheson and Min Liu. Functional hermaphroditism in teleosts. *Fish and Fisheries*, 9(1):1–43, March 2008. ISSN 1467-2960, 1467-2979. doi: 10.1111/j.1467-2979.2007.00266.x. URL <http://doi.wiley.com/10.1111/j.1467-2979.2007.00266.x>.
- [28] Yasuhisa Kobayashi, Ryo Horiguchi, Saori Miura, and Masaru Nakamura. Sex- and tissue-specific expression of P450 aromatase (cyp19a1a) in the yellowtail clownfish, *Amphiprion clarkii*. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 155(2):237–244, February 2010. ISSN 10956433. doi: 10.1016/j.cbpa.2009.11.004. URL <http://linkinghub.elsevier.com/retrieve/pii/S1095643309011684>.
- [29] Paul M. Brakefield. Seasonal polyphenism in butterflies and natural selection. *Trends in Ecology & Evolution*, 11(7):275–277, July 1996. ISSN 01695347. doi: 10.1016/0169-5347(96)30025-6. URL <http://linkinghub.elsevier.com/retrieve/pii/0169534796300256>.
- [30] Stephen J. Simpson, Gregory A. Sword, and Nathan Lo. Polyphenism in insects. *Current Biology*, 21(18):R738–R749, September 2011. ISSN 09609822.

- doi: 10.1016/j.cub.2011.06.006. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982211006518>.
- [31] E. Gaten, S. J. Huston, H. B. Dowse, and T. Matheson. Solitary and gregarious locusts differ in circadian rhythmicity of a visual output neuron. *Journal of Biological Rhythms*, 27(3):196–205, June 2012. ISSN 0748-7304, 1552-4531. doi: 10.1177/0748730412440860. URL <http://jbr.sagepub.com/cgi/doi/10.1177/0748730412440860>.
- [32] S. Kosakovsky Pond, S. Wadhawan, F. Chiaromonte, G. Ananda, W.-Y. Chung, J. Taylor, A. Nekrutenko, and The Galaxy Team. Windshield splatter analysis with the Galaxy metagenomic pipeline. *Genome Research*, 19(11): 2144–2153, November 2009. ISSN 1088-9051. doi: 10.1101/gr.094508.109. URL <http://genome.cshlp.org/cgi/doi/10.1101/gr.094508.109>.
- [33] Nancy A. Moran. The evolutionary maintenance of alternative phenotypes. *The American Naturalist*, 139(5):971–989, May 1992. ISSN 0003-0147, 1537-5323. doi: 10.1086/285369. URL <http://www.journals.uchicago.edu/doi/10.1086/285369>.
- [34] D. A. Roff. The evolution of dimorphic traits: predicting the genetic correlation between environments. *Genetics*, 136(1):395–401, January 1994. ISSN 0016-6731.
- [35] Douglas J. Emlen. Artificial selection on horn length-body size allometry in the horned beetle *Onthophagus acuminatus* (Coleoptera: Scarabaeidae). *Evolution*, 50(3):1219, June 1996. ISSN 00143820. doi: 10.2307/2410662. URL <http://www.jstor.org/stable/2410662?origin=crossref>.
- [36] Armin P. Moczek and H. Frederik Nijhout. Developmental mechanisms of threshold evolution in a polyphenic beetle. *Evolution and Development*, 4(4): 252–264, July 2002. ISSN 1520-541X, 1525-142X. doi: 10.1046/j.1525-142X.

- 2002.02014.x. URL <http://doi.wiley.com/10.1046/j.1525-142X.2002.02014.x>.
- [37] Mary Jane West-Eberhard. *Developmental Plasticity and Evolution*. Oxford University Press, Oxford ; New York, 2003. ISBN 978-0-19-512234-3 978-0-19-512235-0.
- [38] Hitoshi Miyakawa, Naoki Sugimoto, Tetsuo I. Kohyama, Taisen Iguchi, and Toru Miura. Intra-specific variations in reaction norms of predator-induced polyphenism in the water flea *Daphnia pulex*. *Ecological Research*, 30(4):705–713, July 2015. ISSN 0912-3814, 1440-1703. doi: 10.1007/s11284-015-1272-4. URL <http://link.springer.com/10.1007/s11284-015-1272-4>.
- [39] H. Frederik Nijhout. Control Mechanisms of Polyphenic Development in Insects. *BioScience*, 49(3):181–192, March 1999. ISSN 1525-3244, 0006-3568. doi: 10.2307/1313508. URL <https://academic.oup.com/bioscience/article-lookup/doi/10.2307/1313508>.
- [40] Y. Suzuki. Evolution of a polyphenism by genetic accommodation. *Science*, 311(5761):650–652, February 2006. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1118888. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1118888>.
- [41] John Gerhart and Marc Kirschner. The theory of facilitated variation. *Proceedings of the National Academy of Sciences of the United States of America*, 104 Suppl 1:8582–8589, May 2007. ISSN 0027-8424. doi: 10.1073/pnas.0701035104.
- [42] Daniel F. Simola, Lothar Wissler, Greg Donahue, Robert M. Waterhouse, Martin Helmkampf, Julien Roux, Sanne Nygaard, Karl M. Glastad, Darren E. Hagen, Lumi Viljakainen, Justin T. Reese, Brendan G. Hunt, Dan Graur, Eran Elhaik, Evgenia V. Kriventseva, Jiayu Wen, Brian J. Parker, Elizabeth

- Cash, Eyal Privman, Christopher P. Childers, Monica C. Muoz-Torres, Jacobus J. Boomsma, Erich Bornberg-Bauer, Cameron R. Currie, Christine G. Elsik, Garret Suen, Michael A. D. Goodisman, Laurent Keller, Jrgen Liebig, Alan Rawls, Danny Reinberg, Chris D. Smith, Chris R. Smith, Neil Tsutsui, Yannick Wurm, Evgeny M. Zdobnov, Shelley L. Berger, and Jrgen Gadau. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Research*, 23(8):1235–1247, August 2013. ISSN 1549-5469. doi: 10.1101/gr.155408.113.
- [43] K. M. Kapheim, H. Pan, C. Li, S. L. Salzberg, D. Puiu, T. Magoc, H. M. Robertson, M. E. Hudson, A. Venkat, B. J. Fischman, A. Hernandez, M. Yandell, D. Ence, C. Holt, G. D. Yocum, W. P. Kemp, J. Bosch, R. M. Waterhouse, E. M. Zdobnov, E. Stolle, F. B. Kraus, S. Helbing, R. F. A. Moritz, K. M. Glastad, B. G. Hunt, M. A. D. Goodisman, F. Hauser, C. J. P. Grimmelikhuijzen, D. G. Pinheiro, F. M. F. Nunes, M. P. M. Soares, E. D. Tanaka, Z. L. P. Simoes, K. Hartfelder, J. D. Evans, S. M. Barribeau, R. M. Johnson, J. H. Massey, B. R. Southey, M. Hasselmann, D. Hamacher, M. Biewer, C. F. Kent, A. Zayed, C. Blatti, S. Sinha, J. S. Johnston, S. J. Hanrahan, S. D. Kocher, J. Wang, G. E. Robinson, and G. Zhang. Genomic signatures of evolutionary transitions from solitary to group living. *Science*, 348(6239):1139–1143, June 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa4788. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.aaa4788>.
- [44] Solenn Patalano, Anna Vlasova, Chris Wyatt, Philip Ewels, Francisco Camara, Pedro G. Ferreira, Claire L. Asher, Tomasz P. Jurkowski, Anne Segonds-Pichon, Martin Bachman, Irene Gonzlez-Navarrete, Andr E. Minoche, Felix Krueger, Ernesto Lowy, Marina Marcet-Houben, Jose Luis Rodriguez-Ales, Fabio S. Nascimento, Shankar Balasubramanian, Toni Gabaldon, James E. Tarver, Simon Andrews, Heinz Himmelbauer, William O. H. Hughes, Roderic Guig, Wolf Reik, and Seirian Sumner. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proceedings*

- of the National Academy of Sciences*, 112(45):13970–13975, November 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1515937112. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1515937112>.
- [45] Antonia Klein, Eva Schultner, Helena Lowak, Lukas Schrader, Jrgen Heinze, Luke Holman, and Jan Oettler. Evolution of social insect polyphenism facilitated by the sex differentiation cascade. *PLOS Genetics*, 12(3):e1005952, March 2016. ISSN 1553-7404. doi: 10.1371/journal.pgen.1005952. URL <http://dx.plos.org/10.1371/journal.pgen.1005952>.
- [46] K. Hartfelder and D.J. Emlen. Endocrine Control of Insect Polyphenism. In *Insect Endocrinology*, pages 464–522. Elsevier, 2012. ISBN 978-0-12-384749-2. URL <http://linkinghub.elsevier.com/retrieve/pii/B9780123847492100111>. DOI: 10.1016/B978-0-12-384749-2.10011-1.
- [47] R. M. Cox, D. S. Stenquist, and R. Calsbeek. Testosterone, growth and the evolution of sexual size dimorphism. *Journal of Evolutionary Biology*, 22(8):1586–1598, August 2009. ISSN 1010061X, 14209101. doi: 10.1111/j.1420-9101.2009.01772.x. URL <http://doi.wiley.com/10.1111/j.1420-9101.2009.01772.x>.
- [48] V. Hartenstein. The neuroendocrine system of invertebrates: a developmental and evolutionary perspective. *Journal of Endocrinology*, 190(3):555–570, September 2006. ISSN 0022-0795, 1479-6805. doi: 10.1677/joe.1.06964. URL <http://joe.endocrinology-journals.org/cgi/doi/10.1677/joe.1.06964>.
- [49] Seiji Tanaka and Koutaro Maeno. A review of maternal and embryonic control of phase-dependent progeny characteristics in the desert locust. *Journal of Insect Physiology*, 56(8):911–918, August 2010. ISSN 00221910. doi: 10.1016/j.jinsphys.2010.05.013. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022191010001629>.

- [50] Koutaro Maeno and Seiji Tanaka. Maternal effects on progeny size, number and body color in the desert locust, *Schistocerca gregaria*: Density- and reproductive cycle-dependent variation. *Journal of Insect Physiology*, 54(6):1072–1080, June 2008. ISSN 00221910. doi: 10.1016/j.jinsphys.2008.04.010. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022191008000784>.
- [51] Amer I. Tawfik and Frantisek Sehnal. A role for ecdysteroids in the phase polymorphism of the desert locust. *Physiological Entomology*, 28(1):19–24, March 2003. ISSN 0307-6962, 1365-3032. doi: 10.1046/j.1365-3032.2003.00316.x. URL <http://doi.wiley.com/10.1046/j.1365-3032.2003.00316.x>.
- [52] J. W. Truman and L. M. Riddiford. The origins of insect metamorphosis. *Nature*, 401(6752):447–452, September 1999. ISSN 0028-0836. doi: 10.1038/46737.
- [53] C.H. Waddington. The epigenotype. *Endeavour*, (1):18–20, 1942.
- [54] AD Riggs, VE Russo, and RA Martienssen. *Epigenetic Mechanisms of Gene Regulation*. Plainview, N.Y.: Cold Spring Harbor Laboratory Press, 1996.
- [55] Xiang Zhang and Shuk-Mei Ho. Epigenetics meets endocrinology. *Journal of Molecular Endocrinology*, 46(1):R11–32, February 2011. ISSN 1479-6813.
- [56] Jens Vanselow, Marion Spitschak, Marika Nimz, and Rainer Frbass. DNA Methylation Is Not Involved in Preovulatory Down-Regulation of CYP11a1, HSD3b1, and CYP19a1 in Bovine Follicles but May Have a Role in Permanent Silencing of CYP19a1 in Large Granulosa Lutein Cells1. *Biology of Reproduction*, 82(2):289–298, February 2010. ISSN 0006-3363, 1529-7268. doi: 10.1095/biolreprod.109.079251. URL <https://academic.oup.com/biolreprod/article-lookup/doi/10.1095/biolreprod.109.079251>.



- [57] Hisahiko Hiroi, Lane K. Christenson, and Jerome F. Strauss. Regulation of transcription of the steroidogenic acute regulatory protein (StAR) gene: temporal and spatial changes in transcription factor binding and histone modification. *Molecular and Cellular Endocrinology*, 215(1-2):119–126, February 2004. ISSN 0303-7207. doi: 10.1016/j.mce.2003.11.014.
- [58] Kevin Dalgaard, Kathrin Landgraf, Steffen Heyne, Adelheid Lempradl, John Longinotto, Klaus Gossens, Marius Ruf, Michael Orthofer, Ruslan Strogantsev, Madhan Selvaraj, Tess Tsai-Hsiu Lu, Eduard Casas, Raffaele Teperino, M. Azim Surani, Ilona Zvetkova, Debra Rimmington, Y.C. Loraine Tung, Brian Lam, Rachel Larder, Giles S.H. Yeo, Stephen ORahilly, Tanya Vavouri, Emma Whitelaw, Josef M. Penninger, Thomas Jenuwein, Ching-Lung Cheung, Anne C. Ferguson-Smith, Anthony P. Coll, Antje Krner, and J. Andrew Pospisilik. Trim28 Haploinsufficiency Triggers Bi-stable Epigenetic Obesity. *Cell*, 164(3):353–364, January 2016. ISSN 00928674. doi: 10.1016/j.cell.2015.12.025. URL <http://linkinghub.elsevier.com/retrieve/pii/S009286741501689X>.
- [59] Krishnendu Mukherjee, Richard M. Twyman, and Andreas Vilcinskas. Insects as models to study the epigenetic basis of disease. *Progress in Biophysics and Molecular Biology*, 118(1-2):69–78, July 2015. ISSN 00796107. doi: 10.1016/j.pbiomolbio.2015.02.009. URL <http://linkinghub.elsevier.com/retrieve/pii/S0079610715000346>.
- [60] Warren W. Burggren. Epigenetics in Insects: Mechanisms, Phenotypes and Ecological and Evolutionary Implications. In *Advances in Insect Physiology*. Elsevier, 2017. URL <http://linkinghub.elsevier.com/retrieve/pii/S0065280617300073>. DOI: 10.1016/bs.aiip.2017.04.001.
- [61] Kevin N. Dietz, Luisa Di Stefano, Robert C. Maher, Hui Zhu, Marcy E. MacDonald, James F. Gusella, and James A. Walker. The *Drosophila* Huntington’s disease gene ortholog *dhtt* influences chromatin regulation during development.

- Human Molecular Genetics*, 24(2):330–345, January 2015. ISSN 0964-6906, 1460-2083. doi: 10.1093/hmg/ddu446. URL <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddu446>.
- [62] Reini F. Luco, Mariano Allo, Ignacio E. Schor, Alberto R. Kornblihtt, and Tom Misteli. Epigenetics in alternative pre-mRNA splicing. *Cell*, 144(1): 16–26, January 2011. ISSN 00928674. doi: 10.1016/j.cell.2010.11.056. URL <http://linkinghub.elsevier.com/retrieve/pii/S0092867410013784>.
- [63] R. F. Luco, Q. Pan, K. Tominaga, B. J. Blencowe, O. M. Pereira-Smith, and T. Misteli. Regulation of alternative splicing by histone modifications. *Science*, 327(5968):996–1000, February 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1184208. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1184208>.
- [64] Nicolas Negre, Christopher D. Brown, Lijia Ma, Christopher Aaron Bristow, Steven W. Miller, Ulrich Wagner, Pouya Kheradpour, Matthew L. Eaton, Paul Loriaux, Rachel Sealfon, Zirong Li, Haruhiko Ishii, Rebecca F. Spokony, Jia Chen, Lindsay Hwang, Chao Cheng, Richard P. Auburn, Melissa B. Davis, Marc Domanus, Parantu K. Shah, Carolyn A. Morrison, Jennifer Zieba, Sarah Suchy, Lionel Senderowicz, Alec Victorsen, Nicholas A. Bild, A. Jason Grundstad, David Hanley, David M. MacAlpine, Mattias Mannervik, Koen Venken, Hugo Bellen, Robert White, Mark Gerstein, Steven Russell, Robert L. Grossman, Bing Ren, James W. Posakony, Manolis Kellis, and Kevin P. White. A cis-regulatory map of the *Drosophila* genome. *Nature*, 471(7339):527–531, March 2011. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature09990. URL <http://www.nature.com/doi/10.1038/nature09990>.
- [65] Paulina Kolasinska-Zwierz, Thomas Down, Isabel Latorre, Tao Liu, X Shirley Liu, and Julie Ahringer. Differential chromatin marking of introns and expressed exons by H3k36me3. *Nature Genetics*, 41(3):376–381, March 2009.

- ISSN 1061-4036. doi: 10.1038/ng.322. URL <http://www.nature.com/doifinder/10.1038/ng.322>.
- [66] Steven Henikoff. Nucleosome destabilization in the epigenetic regulation of gene expression. *Nature Reviews. Genetics*, 9(1):15–26, January 2008. ISSN 1471-0064. doi: 10.1038/nrg2206.
- [67] Swaminathan Venkatesh, Michaela Smolle, Hua Li, Madelaine M. Gogol, Malika Saint, Shambhu Kumar, Krishnamurthy Natarajan, and Jerry L. Workman. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature*, 489(7416):452–455, August 2012. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11326. URL <http://www.nature.com/doifinder/10.1038/nature11326>.
- [68] Gabriel E Zentner and Steven Henikoff. Regulation of nucleosome dynamics by histone modifications. *Nature Structural & Molecular Biology*, 20(3):259–266, March 2013. ISSN 1545-9993, 1545-9985. doi: 10.1038/nsmb.2470. URL <http://www.nature.com/doifinder/10.1038/nsmb.2470>.
- [69] Mirko Pegoraro, Hollie Marshall, Zoe N Lonsdale, and Eamonn B Mallon. Do social insects support Haig’s kin theory for the evolution of genomic imprinting? *Epigenetics*, pages 00–00, July 2017. ISSN 1559-2294, 1559-2308. doi: 10.1080/15592294.2017.1348445. URL <https://www.tandfonline.com/doi/full/10.1080/15592294.2017.1348445>.
- [70] K. M. Glastad, B. G. Hunt, and M. A. D. Goodisman. DNA methylation and chromatin organization in insects: insights from the ant *Camponotus floridanus*. *Genome Biology and Evolution*, 7(4):931–942, April 2015. ISSN 1759-6653. doi: 10.1093/gbe/evv039. URL <http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evv039>.
- [71] Brendan G. Hunt, Karl M. Glastad, Soojin V. Yi, and Michael A.D. Goodisman. Patterning and regulatory associations of DNA methylation are mirrored

- by histone modifications in insects. *Genome Biology and Evolution*, 5(3):591–598, March 2013. ISSN 1759-6653. doi: 10.1093/gbe/evt030. URL <https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evt030>.
- [72] Alexandra Chittka and Lars Chittka. Epigenetics of royalty. *PLoS Biology*, 8(11):e1000532, November 2010. ISSN 1545-7885. doi: 10.1371/journal.pbio.1000532. URL <http://dx.plos.org/10.1371/journal.pbio.1000532>.
- [73] Daniel B Weaver, Juan M Anzola, Jay D Evans, Jeffrey G Reid, Justin T Reese, Kevin L Childs, Evgeny M Zdobnov, Manoj P Samanta, Jonathan Miller, and Christine G Elsik. Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome Biology*, 8(6):R97, 2007. ISSN 14656906. doi: 10.1186/gb-2007-8-6-r97. URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2007-8-6-r97>.
- [74] Victor Ambros. The functions of animal microRNAs. *Nature*, 431(7006):350–355, September 2004. ISSN 0028-0836, 1476-4679. doi: 10.1038/nature02871. URL <http://www.nature.com/doifinder/10.1038/nature02871>.
- [75] Yuan-Yuan Shi, Hua-Jun Zheng, Qi-Zhong Pan, Zi-Long Wang, and Zhi-Jiang Zeng. Differentially expressed microRNAs between queen and worker larvae of the honey bee (*Apis mellifera*). *Apidologie*, 46(1):35–45, January 2015. ISSN 0044-8435, 1297-9678. doi: 10.1007/s13592-014-0299-9. URL <http://link.springer.com/10.1007/s13592-014-0299-9>.
- [76] Regan Ashby, Sylvain Fort, Iain Searle, and Ryszard Maleszka. MicroRNAs in honey bee caste determination. *Scientific Reports*, 6(1), May 2016. ISSN 2045-2322. doi: 10.1038/srep18794. URL <http://www.nature.com/articles/srep18794>.

- [77] David H. Collins, Irina Mohorianu, Matthew Beckers, Vincent Moulton, Tamas Dalmay, and Andrew F. G. Bourke. MicroRNAs associated with caste determination and differentiation in a primitively eusocial insect. *Scientific Reports*, 7:45674, March 2017. ISSN 2045-2322. doi: 10.1038/srep45674. URL <http://www.nature.com/articles/srep45674>.
- [78] Fabrice Legeai, Guillaume Rizk, Thomas Walsh, Owain Edwards, Karl Gordon, Dominique Lavenier, Nathalie Leterme, Agns Mreau, Jacques Nicolas, Denis Tagu, and Stphanie Jaubert-Possamai. Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrtosiphon pisum*. *BMC Genomics*, 11(1):281, 2010. ISSN 1471-2164. doi: 10.1186/1471-2164-11-281. URL <http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-281>.
- [79] Alexei A. Aravin, Ravi Sachidanandam, Deborah Bourc’his, Christopher Schaefer, Dubravka Pezic, Katalin Fejes Toth, Timothy Bestor, and Gregory J. Hannon. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular Cell*, 31(6):785–799, September 2008. ISSN 10972765. doi: 10.1016/j.molcel.2008.09.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S1097276508006199>.
- [80] Guilin Wang and Valerie Reinke. A *C. elegans* Piwi, PRG-1, regulates 21u-RNAs during spermatogenesis. *Current Biology*, 18(12):861–867, June 2008. ISSN 09609822. doi: 10.1016/j.cub.2008.05.009. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982208006064>.
- [81] Edith Heard and Robert A. Martienssen. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell*, 157(1):95–109, March 2014. ISSN 00928674. doi: 10.1016/j.cell.2014.02.045. URL <http://linkinghub.elsevier.com/retrieve/pii/S0092867414002864>.

- [82] J. Brennecke, C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark, and G. J. Hannon. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science*, 322(5906):1387–1392, November 2008. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1165171. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1165171>.
- [83] T. Greutzinger, C. Armenise, C. Brun, B. Mugat, V. Serrano, A. Pelisson, and S. Chambeyron. piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Research*, 22(10):1877–1888, October 2012. ISSN 1088-9051. doi: 10.1101/gr.136614.111. URL <http://genome.cshlp.org/cgi/doi/10.1101/gr.136614.111>.
- [84] Augustin de Vanssay, Anne-Laure Boug, Antoine Boivin, Catherine Hermant, Laure Teyssset, Valrie Delmarre, Christophe Antoniewski, and Stphane Ronsseray. piRNAs and epigenetic conversion in *Drosophila*. *Fly*, 7(4):237–241, October 2013. ISSN 1933-6934, 1933-6942. doi: 10.4161/fly.26522. URL <http://www.tandfonline.com/doi/abs/10.4161/fly.26522>.
- [85] Ruifang Wu, Yuwen Su, Haijing Wu, Yong Dai, Ming Zhao, and Qianjin Lu. Characters, functions and clinical perspectives of long non-coding RNAs. *Molecular Genetics and Genomics*, 291(3):1013–1033, June 2016. ISSN 1617-4615, 1617-4623. doi: 10.1007/s00438-016-1179-y. URL <http://link.springer.com/10.1007/s00438-016-1179-y>.
- [86] Chandrasekhar Kanduri. Long noncoding RNAs: Lessons from genomic imprinting. *Biochimica Et Biophysica Acta*, 1859(1):102–111, January 2016. ISSN 0006-3002. doi: 10.1016/j.bbagr.2015.05.006.
- [87] Taketoshi Kiya, Atsushi Ugajin, Takekazu Kunieda, and Takeo Kubo. Identification of kakusei, a nuclear non-coding RNA, as an immediate early gene from the honeybee, and its application for neuroethological study. *International Journal of Molecular Sciences*, 13(12):15496–15509, November 2012.

- ISSN 1422-0067. doi: 10.3390/ijms131215496. URL <http://www.mdpi.com/1422-0067/13/12/15496/>.
- [88] H. Tadano, Y. Yamazaki, H. Takeuchi, and T. Kubo. Age- and division-of-labour-dependent differential expression of a novel non-coding RNA, *Nb-1*, in the brain of worker honeybees, *Apis mellifera* L. *Insect Molecular Biology*, 18(6):715–726, December 2009. ISSN 09621075, 13652583. doi: 10.1111/j.1365-2583.2009.00911.x. URL <http://doi.wiley.com/10.1111/j.1365-2583.2009.00911.x>.
- [89] L. Liu, G. Jin, and X. Zhou. Modeling the relationship of epigenetic modifications to transcription factor binding. *Nucleic Acids Research*, 43(8):3873–3885, April 2015. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkv255. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv255>.
- [90] R. D. Kornberg and Y. Lorch. Interplay between chromatin structure and transcription. *Current Opinion in Cell Biology*, 7(3):371–375, June 1995. ISSN 0955-0674.
- [91] Michael Shogren-Knaak, Haruhiko Ishii, Jian-Min Sun, Michael J. Pazin, James R. Davie, and Craig L. Peterson. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science (New York, N.Y.)*, 311(5762):844–847, February 2006. ISSN 1095-9203. doi: 10.1126/science.1124000.
- [92] F. Watt and P. L. Molloy. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & Development*, 2(9):1136–1143, September 1988. ISSN 0890-9369.

- [93] Z. Siegfried, S. Eden, M. Mendelsohn, X. Feng, B. Z. Tsuberi, and H. Cedar. DNA methylation represses transcription in vivo. *Nature Genetics*, 22(2):203–206, June 1999. ISSN 1061-4036. doi: 10.1038/9727.
- [94] Y. Zhang and D. Reinberg. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & Development*, 15(18):2343–2360, September 2001. ISSN 0890-9369. doi: 10.1101/gad.927301.
- [95] Yulia A. Medvedeva, Abdullah M. Khamis, Ivan V. Kulakovskiy, Wail Ba-Alawi, Md Shariful I. Bhuyan, Hideya Kawaji, Timo Lassmann, Matthias Harbers, Alistair R. R. Forrest, Vladimir B. Bajic, and FANTOM consortium. Effects of cytosine methylation on transcription factor binding sites. *BMC genomics*, 15:119, March 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-119.
- [96] Weixing Feng, Zengchao Dong, Bo He, and Kejun Wang. Analysis method of epigenetic DNA methylation to dynamically investigate the functional activity of transcription factors in gene expression. *BMC genomics*, 13:532, October 2012. ISSN 1471-2164. doi: 10.1186/1471-2164-13-532.
- [97] Aime M. Deaton and Adrian Bird. CpG islands and the regulation of transcription. *Genes & Development*, 25(10):1010–1022, May 2011. ISSN 1549-5477. doi: 10.1101/gad.2037511.
- [98] Li-Hui Wang, Yoon-La Choi, Xuan-Yan Hua, Young-Kee Shin, Young-Jin Song, Sei-Jin Youn, Hyo-Young Yun, Seon-Mee Park, Wun-Jae Kim, Hee-Jin Kim, Jong-Sun Choi, and Seok-Hyung Kim. Increased expression of sonic hedgehog and altered methylation of its promoter region in gastric cancer and its related lesions. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, 19(5):675–683, May 2006. ISSN 0893-3952. doi: 10.1038/modpathol.3800573.



- [99] Wei Cui, Li-Hui Wang, Yuan-Yuan Wen, Min Song, Bai-Lin Li, Xiao-Ling Chen, Man Xu, Su-Xiang An, Jing Zhao, Yi-Yan Lu, Xiao-Yi Mi, and En-Hua Wang. Expression and regulation mechanisms of sonic hedgehog in breast cancer. *Cancer Science*, 101(4):927–933, April 2010. ISSN 1349-7006. doi: 10.1111/j.1349-7006.2010.01495.x.
- [100] Zhao-Heng Duan, Hao-Chuan Wang, Dong-Mei Zhao, Xiao-Xin Ji, Min Song, Xiao-Jun Yang, and Wei Cui. Cooperatively transcriptional and epigenetic regulation of sonic hedgehog overexpression drives malignant potential of breast cancer. *Cancer Science*, 106(8):1084–1091, August 2015. ISSN 1349-7006. doi: 10.1111/cas.12697.
- [101] Ye Fu, Dan Dominissini, Gideon Rechavi, and Chuan He. Gene expression regulation mediated through reversible m6a RNA methylation. *Nature Reviews Genetics*, 15(5):293–306, March 2014. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg3724. URL <http://www.nature.com/doifinder/10.1038/nrg3724>.
- [102] Hua Yan, Daniel F. Simola, Roberto Bonasio, Jrgen Liebig, Shelley L. Berger, and Danny Reinberg. Eusocial insects as emerging models for behavioural epigenetics. *Nature Reviews Genetics*, 15(10):677–688, September 2014. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg3787. URL <http://www.nature.com/doifinder/10.1038/nrg3787>.
- [103] Hua Yan, Roberto Bonasio, Daniel F. Simola, Jrgen Liebig, Shelley L. Berger, and Danny Reinberg. DNA methylation in social insects: how epigenetics can control behavior and longevity. *Annual Review of Entomology*, 60(1):435–452, January 2015. ISSN 0066-4170, 1545-4487. doi: 10.1146/annurev-ento-010814-020803. URL <http://www.annualreviews.org/doi/abs/10.1146/annurev-ento-010814-020803>.
- [104] S. Feng, S. J. Cokus, X. Zhang, P.-Y. Chen, M. Bostick, M. G. Goll, J. Hetzel, J. Jain, S. H. Strauss, M. E. Halpern, C. Ukomadu, K. C. Sadler, S. Pradhan,

- M. Pellegrini, and S. E. Jacobsen. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences*, 107(19):8689–8694, May 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1002720107. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1002720107>.
- [105] A. Bird. DNA methylation patterns and epigenetic memory. *Genes & Development*, 16(1):6–21, January 2002. ISSN 08909369. doi: 10.1101/gad.947102. URL <http://www.genesdev.org/cgi/doi/10.1101/gad.947102>.
- [106] Y. Wang, M. Jorda, P. L. Jones, R. Maleszka, X. Ling, H. M. Robertson, C. A. Mizzen, M. A. Peinado, and G. E. Robinson. Functional CpG methylation system in a social insect. *Science*, 314(5799):645–647, October 2006. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1135213. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1135213>.
- [107] Andrew J. Wood and Rebecca J. Oakey. Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genetics*, 2(11):e147, 2006. ISSN 1553-7390, 1553-7404. doi: 10.1371/journal.pgen.0020147. URL <http://dx.plos.org/10.1371/journal.pgen.0020147>.
- [108] Anne C. Ferguson-Smith. Genomic imprinting: the emergence of an epigenetic paradigm. *Nature Reviews. Genetics*, 12(8):565–575, 2011. ISSN 1471-0064. doi: 10.1038/nrg3032.
- [109] Marisa S. Bartolomei. Genomic imprinting: employing and avoiding epigenetic processes. *Genes & Development*, 23(18):2124–2133, September 2009. ISSN 1549-5477. doi: 10.1101/gad.1841409.
- [110] Stefanie Seisenberger, Simon Andrews, Felix Krueger, Julia Arand, Jörn Walter, Ftima Santos, Christian Popp, Bernard Thienpont, Wendy Dean, and Wolf Reik. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Molecular Cell*, 48(6):849–862, December 2012.

- ISSN 10972765. doi: 10.1016/j.molcel.2012.11.001. URL <http://linkinghub.elsevier.com/retrieve/pii/S109727651200932X>.
- [111] Meelad M. Dawlaty, Achim Breiling, Thuc Le, Gnter Raddatz, M. Inmaculada Barrasa, Albert W. Cheng, Qing Gao, Benjamin E. Powell, Zhe Li, Mingjiang Xu, Kym F. Faull, Frank Lyko, and Rudolf Jaenisch. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Developmental Cell*, 24(3):310–323, February 2013. ISSN 1878-1551. doi: 10.1016/j.devcel.2012.12.015.
- [112] Hua Yan, Roberto Bonasio, Daniel F. Simola, Jrgen Liebig, Shelley L. Berger, and Danny Reinberg. DNA methylation in social insects: how epigenetics can control behavior and longevity. *Annual Review of Entomology*, 60(1):435–452, January 2015. ISSN 0066-4170, 1545-4487. doi: 10.1146/annurev-ento-010814-020803. URL <http://www.annualreviews.org/doi/10.1146/annurev-ento-010814-020803>.
- [113] Daniel S. Standage, Ali J. Berens, Karl M. Glastad, Andrew J. Severin, Volker P. Brendel, and Amy L. Toth. Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Molecular Ecology*, 25(8):1769–1784, April 2016. ISSN 09621083. doi: 10.1111/mec.13578. URL <http://doi.wiley.com/10.1111/mec.13578>.
- [114] Marcus R. Kronforst, David C. Gilley, Joan E. Strassmann, and David C. Queller. DNA methylation is widespread across social Hymenoptera. *Current Biology*, 18(7):R287–R288, April 2008. ISSN 09609822. doi: 10.1016/j.cub.2008.02.015. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982208001607>.
- [115] Frank Lyko, Sylvain Foret, Robert Kucharski, Stephan Wolf, Cassandra Falckenhayn, and Ryszard Maleszka. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biology*, 8(11):e1000506,

- November 2010. ISSN 1545-7885. doi: 10.1371/journal.pbio.1000506. URL <http://dx.plos.org/10.1371/journal.pbio.1000506>.
- [116] N. Elango, B. G. Hunt, M. A. D. Goodisman, and S. V. Yi. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proceedings of the National Academy of Sciences*, 106(27):11206–11211, July 2009. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0900301106. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0900301106>.
- [117] R. Kucharski, J. Maleszka, S. Foret, and R. Maleszka. Nutritional control of reproductive status in honeybees via DNA methylation. *Science*, 319(5871):1827–1830, March 2008. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1153069. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1153069>.
- [118] Brian R Herb, Florian Wolschin, Kasper D Hansen, Martin J Aryee, Ben Langmead, Rafael Irizarry, Gro V Amdam, and Andrew P Feinberg. Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature Neuroscience*, 15(10):1371–1373, September 2012. ISSN 1097-6256, 1546-1726. doi: 10.1038/nn.3218. URL <http://www.nature.com/doi/10.1038/nn.3218>.
- [119] C. Falckenhayn, B. Boerjan, G. Raddatz, M. Frohme, L. Schoofs, and F. Lyko. Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*. *Journal of Experimental Biology*, 216(8):1423–1429, April 2013. ISSN 0022-0949, 1477-9145. doi: 10.1242/jeb.080754. URL <http://jeb.biologists.org/cgi/doi/10.1242/jeb.080754>.
- [120] Manlika Kilaso, Nadine C Chapman, Emily J Remnant, Benjamin P Oldroyd, and Chanpen Chanchao. No evidence that DNA methylation is associated with the regulation of fertility in the adult honey bee *Apis mellifera* (Hymenoptera: Apidae) worker ovary: DNA methylation and worker sterility.

- Austral Entomology*, 2016. ISSN 2052174X. doi: 10.1111/aen.12218. URL <http://doi.wiley.com/10.1111/aen.12218>.
- [121] B. G. Hunt, J. A. Brisson, S. V. Yi, and M. A. D. Goodisman. Functional Conservation of DNA Methylation in the Pea Aphid and the Honeybee. *Genome Biology and Evolution*, September 2010. ISSN 1759-6653. doi: 10.1093/gbe/evq057. URL <https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evq057>.
- [122] 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 Wang, Jun Wang, Shelley L. Berger, Jurgen Liebig, Guojie Zhang, and Danny Reinberg. Genome-wide and Caste-Specific DNA Methylomes of the Ants *Camponotus floridanus* and *Harpegnathos saltator*. *Current Biology*, 22(19):1755–1764, October 2012. ISSN 09609822. doi: 10.1016/j.cub.2012.07.042. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982212008676>.
- [123] Xu Wang, David Wheeler, Amanda Avery, Alfredo Rago, Jeong-Hyeon Choi, John K. Colbourne, Andrew G. Clark, and John H. Werren. Function and evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genetics*, 9(10): e1003872, October 2013. ISSN 1553-7404. doi: 10.1371/journal.pgen.1003872. URL <http://dx.plos.org/10.1371/journal.pgen.1003872>.
- [124] Romain Libbrecht, Peter Robert Oxley, Laurent Keller, and Daniel Jan Christoph Kronauer. Robust DNA Methylation in the clonal raider ant brain. *Current Biology*, 26(3):391–395, February 2016. ISSN 09609822. doi: 10.1016/j.cub.2015.12.040. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982215015717>.
- [125] Yan Wang, Jing Liu, Bo Huang, Yan-Mei Xu, Jing Li, Lin-Feng Huang, Jin Lin, Jing Zhang, Qing-Hua Min, Wei-Ming Yang, and Xiao-Zhong Wang. Mechanism of alternative splicing and its regulation. *Biomedical Reports*, 3(2):

- 152–158, March 2015. ISSN 2049-9434, 2049-9442. doi: 10.3892/br.2014.407. URL <https://www.spandidos-publications.com/10.3892/br.2014.407>.
- [126] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, October 2004. ISSN 0028-0836, 1476-4679. doi: 10.1038/nature03001. URL <http://www.nature.com/doifinder/10.1038/nature03001>.
- [127] Jaime M. Tovar-Corona, Atahualpa Castillo-Morales, Lu Chen, Brett P. Olds, John M. Clark, Stuart E. Reynolds, Barry R. Pittendrigh, Edward J. Feil, and Araxi O. Urrutia. Alternative splice in alternative lice. *Molecular Biology and Evolution*, 32(10):2749–2759, October 2015. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msv151. URL <http://mbe.oxfordjournals.org/lookup/doi/10.1093/molbev/msv151>.
- [128] S. Foret, R. Kucharski, M. Pellegrini, S. Feng, S. E. Jacobsen, G. E. Robinson, and R. Maleszka. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proceedings of the National Academy of Sciences*, 109(13):4968–4973, March 2012. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1202392109. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1202392109>.
- [129] A. Zemach, I. E. McDaniel, P. Silva, and D. Zilberman. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 328(5980):916–919, May 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1186366. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1186366>.
- [130] Daniel Zilberman, Devin Coleman-Derr, Tracy Ballinger, and Steven Henikoff. Histone H2a.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature*, 456(7218):125–129, November 2008. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature07324. URL <http://www.nature.com/doifinder/10.1038/nature07324>.

- [131] Mirko Pegoraro, Emma Picot, Celia N. Hansen, Charalambos P. Kyriacou, Ezio Rosato, and Eran Tauber. Gene expression associated with early and late chronotypes in *Drosophila melanogaster*. *Frontiers in Neurology*, 6, May 2015. ISSN 1664-2295. doi: 10.3389/fneur.2015.00100. URL [http://www.frontiersin.org/Sleep\\_and\\_Chronobiology/10.3389/fneur.2015.00100/abstract](http://www.frontiersin.org/Sleep_and_Chronobiology/10.3389/fneur.2015.00100/abstract).
- [132] C. David Allis, T Jenuwein, D. Reinberg, and ML Caparros. *Epigenetics*. Cold Spring Harbor Laboratory Press, 2007.
- [133] Hye-Young Min, Su-Chan Lee, Jong Kyu Woo, Hyun Jin Jung, Kwan Hee Park, Hae Min Jeong, Seung Yeob Hyun, Jaebeom Cho, Woojin Lee, Ji Eun Park, So Jung Kwon, Hyo-Jong Lee, Xiao Ni, Young Kee Shin, Faye M. Johnson, Madeleine Duvic, and Ho-Young Lee. Essential role of DNA methyltransferase 1-mediated transcription of insulin-like growth factor 2 in resistance to histone deacetylase inhibitors. *Clinical Cancer Research*, 23(5):1299–1311, March 2017. ISSN 1078-0432, 1557-3265. doi: 10.1158/1078-0432.CCR-16-0534. URL <http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-16-0534>.
- [134] Sanjeev Shukla, Ersen Kavak, Melissa Gregory, Masahiko Imashimizu, Bojan Shutinoski, Mikhail Kashlev, Philipp Oberdoerffer, Rickard Sandberg, and Shalini Oberdoerffer. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, 479(7371):74–79, October 2011. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature10442. URL <http://www.nature.com/doifinder/10.1038/nature10442>.
- [135] Roberto Perales and David Bentley. Cotranscriptionality: the transcription elongation complex as a nexus for nuclear transactions. *Molecular Cell*, 36(2):178–191, October 2009. ISSN 10972765. doi: 10.1016/j.molcel.2009.09.018. URL <http://linkinghub.elsevier.com/retrieve/pii/S1097276509006765>.

- [136] J. I. Young, E. P. Hong, J. C. Castle, J. Crespo-Barreto, A. B. Bowman, M. F. Rose, D. Kang, R. Richman, J. M. Johnson, S. Berget, and H. Y. Zoghbi. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proceedings of the National Academy of Sciences*, 102(49):17551–17558, December 2005. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0507856102. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0507856102>.
- [137] Alike K Maunakea, Iouri Chepelev, Kairong Cui, and Keji Zhao. Intra-genic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Research*, 23(11):1256–1269, November 2013. ISSN 1001-0602, 1748-7838. doi: 10.1038/cr.2013.110. URL <http://www.nature.com/doifinder/10.1038/cr.2013.110>.
- [138] Frank Lyko and Ryszard Maleszka. Insects as innovative models for functional studies of DNA methylation. *Trends in Genetics*, 27(4):127–131, April 2011. ISSN 01689525. doi: 10.1016/j.tig.2011.01.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S0168952511000047>.
- [139] J. Park, Z. Peng, J. Zeng, N. Elango, T. Park, D. Wheeler, J. H. Werren, and S. V. Yi. Comparative analyses of DNA methylation and sequence evolution using *Nasonia* genomes. *Molecular Biology and Evolution*, 28(12):3345–3354, December 2011. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msr168. URL <http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/msr168>.
- [140] Kevin Flores, Florian Wolschin, Jason J Corneveaux, April N Allen, Matthew J Huentelman, and Gro V Amdam. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics*, 13(1):480, 2012. ISSN 1471-2164. doi: 10.1186/1471-2164-13-480. URL <http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-480>.
- [141] Karl M. Glastad, Kaustubh Gokhale, Jrgen Liebig, and Michael A. D. Goodman. The caste- and sex-specific DNA methylome of the termite *Zootermopsis*



- nevadensis*. *Scientific Reports*, 6:37110, November 2016. ISSN 2045-2322. doi: 10.1038/srep37110. URL <http://www.nature.com/articles/srep37110>.
- [142] Jianzhen Zhang, Xiaojian Liu, Jianqin Zhang, Daqi Li, Yi Sun, Yaping Guo, Enbo Ma, and Kun Yan Zhu. Silencing of two alternative splicing-derived mRNA variants of chitin synthase 1 gene by RNAi is lethal to the oriental migratory locust, *Locusta migratoria manilensis* (Meyen). *Insect Biochemistry and Molecular Biology*, 40(11):824–833, November 2010. ISSN 09651748. doi: 10.1016/j.ibmb.2010.08.001. URL <http://linkinghub.elsevier.com/retrieve/pii/S0965174810001694>.
- [143] Xianhui Wang, Xiaodong Fang, Pengcheng Yang, Xuanting Jiang, Feng Jiang, Dejian Zhao, Bolei Li, Feng Cui, Jianing Wei, Chuan Ma, Yundan Wang, Jing He, Yuan Luo, Zhifeng Wang, Xiaojiao Guo, Wei Guo, Xuesong Wang, Yi Zhang, Meiling Yang, Shuguang Hao, Bing Chen, Zongyuan Ma, Dan Yu, Zhiqiang Xiong, Yabing Zhu, Dingding Fan, Lijuan Han, Bo Wang, Yuanxin Chen, Junwen Wang, Lan Yang, Wei Zhao, Yue Feng, Guanxing Chen, Jinmin Lian, Qiye Li, Zhiyong Huang, Xiaoming Yao, Na Lv, Guojie Zhang, Yingrui Li, Jian Wang, Jun Wang, Baoli Zhu, and Le Kang. The locust genome provides insight into swarm formation and long-distance flight. *Nature Communications*, 5, January 2014. ISSN 2041-1723. doi: 10.1038/ncomms3957. URL <http://www.nature.com/doifinder/10.1038/ncomms3957>.
- [144] Casper C. Hoogenraad and Frank Bradke. Control of neuronal polarity and plasticity a renaissance for microtubules? *Trends in Cell Biology*, 19(12):669–676, December 2009. ISSN 09628924. doi: 10.1016/j.tcb.2009.08.006. URL <http://linkinghub.elsevier.com/retrieve/pii/S0962892409001834>.
- [145] Ali Shilatifard. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annual Review of Biochemistry*, 75(1):243–269, June 2006. ISSN 0066-4154, 1545-4509. doi: 10.1146/

- annurev.biochem.75.103004.142422. URL <http://www.annualreviews.org/doi/10.1146/annurev.biochem.75.103004.142422>.
- [146] D. F. Simola, R. J. Graham, C. M. Brady, B. L. Enzmann, C. Desplan, A. Ray, L. J. Zwiebel, R. Bonasio, D. Reinberg, J. Liebig, and S. L. Berger. Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science*, 351(6268):aac6633–aac6633, January 2016. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aac6633. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.aac6633>.
- [147] Edward O. Wilson. One giant leap: how insects achieved altruism and colonial life. *BioScience*, 58(1):17, 2008. ISSN 0006-3568. doi: 10.1641/B580106. URL <https://academic.oup.com/bioscience/article-lookup/doi/10.1641/B580106>.
- [148] Kenneth J. Howard and Barbara L. Thorne. Eusocial Evolution in Termites and Hymenoptera. In David Edward Bignell, Yves Roisin, and Nathan Lo, editors, *Biology of Termites: a Modern Synthesis*, pages 97–132. Springer Netherlands, Dordrecht, 2010. ISBN 978-90-481-3976-7 978-90-481-3977-4. URL [http://link.springer.com/10.1007/978-90-481-3977-4\\_5](http://link.springer.com/10.1007/978-90-481-3977-4_5). DOI: 10.1007/978-90-481-3977-4\_5.
- [149] Sydney A. Cameron and Patrick Mardulyn. Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae). *Systematic Biology*, 50(2):194–214, March 2001. ISSN 1063-5157, 1076-836X. doi: 10.1080/10635150120230. URL <https://academic.oup.com/sysbio/article-lookup/doi/10.1080/10635150120230>.
- [150] S. G Brady, S. Sipes, A. Pearson, and B. N Danforth. Recent and simultaneous origins of eusociality in halictid bees. *Proceedings of the Royal Society B: Biological Sciences*, 273(1594):1643–1649, July 2006. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2006.3496. URL <http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2006.3496>.

- [151] H. M. Hines, J. H. Hunt, T. K. O'Connor, J. J. Gillespie, and S. A. Cameron. Multigene phylogeny reveals eusociality evolved twice in vespid wasps. *Proceedings of the National Academy of Sciences*, 104(9):3295–3299, February 2007. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0610140104. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0610140104>.
- [152] S. Cardinal, J. Straka, and B. N. Danforth. Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proceedings of the National Academy of Sciences*, 107(37):16207–16211, September 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1006299107. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1006299107>.
- [153] B. N. Danforth. Evolution of sociality in a primitively eusocial lineage of bees. *Proceedings of the National Academy of Sciences*, 99(1):286–290, January 2002. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.012387999. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.012387999>.
- [154] David C. Queller and Joan E. Strassmann. Kin Selection and Social Insects. *BioScience*, 48(3):165–175, March 1998. ISSN 00063568, 15253244. doi: 10.2307/1313262. URL <http://bioscience.oxfordjournals.org/cgi/doi/10.2307/1313262>.
- [155] J. Jarvis. Eusociality in a mammal: cooperative breeding in naked mole-rat colonies. *Science*, 212(4494):571–573, May 1981. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.7209555. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.7209555>.
- [156] E. O. Wilson. *The Insect Societies*. Cambridge (MA), Harvard University Press, Belknap Press, 1971.
- [157] C. D. Michener. Reproduction and Castes in Social Halictine Bees. In Wolf Engels, editor, *Social Insects: An Evolutionary Approach to Castes and Reproduction*, pages 77–121. Springer, New York, 1990.

- [158] John R. Harbo. Effect of population size on brood production, worker survival and honey gain in colonies of honeybees. *Journal of Apicultural Research*, 25(1):22–29, January 1986. ISSN 0021-8839, 2078-6913. doi: 10.1080/00218839.1986.11100687. URL <http://www.tandfonline.com/doi/full/10.1080/00218839.1986.11100687>.
- [159] H.H.W. Velthuis and M.J. Duchateau. Development and reproductive strategies in *Bombus terrestris* colonies. *Behaviour*, 107(3):186–207, January 1988. ISSN 0005-7959, 1568-539X. doi: 10.1163/156853988X00340. URL <http://booksandjournals.brillonline.com/content/journals/10.1163/156853988x00340>.
- [160] Etya Amsalem and Abraham Hefetz. The effect of group size on the interplay between dominance and reproduction in *Bombus terrestris*. *PLoS ONE*, 6(3): e18238, March 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0018238. URL <http://dx.plos.org/10.1371/journal.pone.0018238>.
- [161] Dave Goulson. *Bumblebees: their Behaviour and Ecology*. Oxford Biology. Oxford Univ. Press, Oxford, repr edition, 2006. ISBN 978-0-19-852607-0. OCLC: 634514469.
- [162] Laurent Keller and Stephanie Jemielity. Social insects as a model to study the molecular basis of ageing. *Experimental Gerontology*, 41(6):553–556, June 2006. ISSN 05315565. doi: 10.1016/j.exger.2006.04.002. URL <http://linkinghub.elsevier.com/retrieve/pii/S0531556506001057>.
- [163] Jurgen Liebig and Hans-Joachim Poethke. Queen lifespan and colony longevity in the ant *Harpegnathos saltator*. *Ecological Entomology*, 29(2):203–207, April 2004. ISSN 0307-6946, 1365-2311. doi: 10.1111/j.1365-2311.2004.00583.x. URL <http://doi.wiley.com/10.1111/j.1365-2311.2004.00583.x>.
- [164] Joo Alpedrinha, Andy Gardner, and Stuart A. West. Haplodiploidy and the evolution of eusociality: worker revolution. *The American Naturalist*, 184(3):

- 303–317, September 2014. ISSN 0003-0147, 1537-5323. doi: 10.1086/677283. URL <http://www.journals.uchicago.edu/doi/10.1086/677283>.
- [165] Xiaoyun Liao, Stephen Rong, and David C. Queller. Relatedness, conflict, and the evolution of eusociality. *PLOS Biology*, 13(3):e1002098, March 2015. ISSN 1545-7885. doi: 10.1371/journal.pbio.1002098. URL <http://dx.plos.org/10.1371/journal.pbio.1002098>.
- [166] Karen M Kapheim. Genomic sources of phenotypic novelty in the evolution of eusociality in insects. *Current Opinion in Insect Science*, 13:24–32, February 2016. ISSN 22145745. doi: 10.1016/j.cois.2015.10.009. URL <http://linkinghub.elsevier.com/retrieve/pii/S2214574515001613>.
- [167] A. J. Berens, J. H. Hunt, and A. L. Toth. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Molecular Biology and Evolution*, 32(3):690–703, March 2015. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msu330. URL <http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/msu330>.
- [168] David Haig. The Kinship Theory of Genomic Imprinting. *Annual Review of Ecology and Systematics*, 31(1):9–32, November 2000. ISSN 0066-4162. doi: 10.1146/annurev.ecolsys.31.1.9. URL <http://www.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.31.1.9>.
- [169] Vishwesha Guttal, Pawel Romanczuk, Stephen J. Simpson, Gregory A. Sword, and Iain D. Couzin. Cannibalism can drive the evolution of behavioural phase polyphenism in locusts. *Ecology Letters*, 15(10):1158–1166, October 2012. ISSN 1461023X. doi: 10.1111/j.1461-0248.2012.01840.x. URL <http://doi.wiley.com/10.1111/j.1461-0248.2012.01840.x>.

- [170] Sidi Ould Ely, Peter G. N. Njagi, Magzoub Omer Bashir, Salah El-Tom El-Amin, and Ahmed Hassanali. Diel behavioral activity patterns in adult solitary desert locust, *Schistocerca gregaria* (Forsk.). *Psyche: A Journal of Entomology*, 2011:1–9, 2011. ISSN 0033-2615, 1687-7438. doi: 10.1155/2011/459315. URL <http://www.hindawi.com/journals/psyche/2011/459315/>.
- [171] B.P. Uvarov. *Grasshoppers and Locusts: a Handbook of General Acridology. Volume II*. Centre for Overseas Pest Research, London, UK., 1977.
- [172] M. L. Anstey, S. M. Rogers, S. R. Ott, M. Burrows, and S. J. Simpson. Serotonin mediates behavioral gregarization underlying swarm formation in desert locusts. *Science*, 323(5914):627–630, January 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1165939. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1165939>.
- [173] M.P. Pener. Locust Phase Polymorphism and its Endocrine Relations. In *Advances in Insect Physiology*, volume 23, pages 1–79. Elsevier, 1991. ISBN 978-0-12-024223-8. URL <http://linkinghub.elsevier.com/retrieve/pii/S0065280608600910>.
- [174] David M Hunter. Advances in the control of locusts (Orthoptera: Acrididae) in eastern Australia: from crop protection to preventive control. *Australian Journal of Entomology*, 43(3):293–303, August 2004. ISSN 1326-6756, 1440-6055. doi: 10.1111/j.1326-6756.2004.00433.x. URL <http://doi.wiley.com/10.1111/j.1326-6756.2004.00433.x>.
- [175] Andy M. Reynolds, Gregory A. Sword, Stephen J. Simpson, and Don R. Reynolds. Predator percolation, insect outbreaks, and phase polyphenism. *Current Biology*, 19(1):20–24, January 2009. ISSN 09609822. doi: 10.1016/j.cub.2008.10.070. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982208015479>.

- [176] Nathaniel J. Balmert, Samuel S.C. Rund, John P. Ghazi, Peng Zhou, and Giles E. Duffield. Time-of-day specific changes in metabolic detoxification and insecticide resistance in the malaria mosquito *Anopheles gambiae*. *Journal of Insect Physiology*, 64:30–39, May 2014. ISSN 00221910. doi: 10.1016/j.jinsphys.2014.02.013. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022191014000316>.
- [177] Julie Tobback, Bart Boerjan, Hans Peter Vandersmissen, and Roger Huybrechts. The circadian clock genes affect reproductive capacity in the desert locust *Schistocerca gregaria*. *Insect Biochemistry and Molecular Biology*, 41(5): 313–321, May 2011. ISSN 09651748. doi: 10.1016/j.ibmb.2011.01.008. URL <http://linkinghub.elsevier.com/retrieve/pii/S0965174811000257>.
- [178] Hadas Keren, Galit Lev-Maor, and Gil Ast. Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics*, 11(5): 345–355, May 2010. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg2776. URL <http://www.nature.com/doifinder/10.1038/nrg2776>.
- [179] Ricardo Palacios, Elodie Gazave, Joaquin Goi, Gabriel Piedrafitra, 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. ISSN 1932-6203. doi: 10.1371/journal.pone.0004150. URL <http://dx.plos.org/10.1371/journal.pone.0004150>.
- [180] K.K. Farh. The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science*, 310(5755):1817–1821, December 2005. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1121158. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1121158>.
- [181] David P. Bartel. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–297, January 2004. ISSN 0092-8674.

- [182] David A. Galbraith, Sarah D. Kocher, Tom Glenn, Istvan Albert, Greg J. Hunt, Joan E. Strassmann, David C. Queller, and Christina M. Grozinger. Testing the kinship theory of intragenomic conflict in honey bees (*Apis mellifera*). *Proceedings of the National Academy of Sciences*, 113(4):1020–1025, January 2016. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1516636113. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1516636113>.
- [183] Alexandra Chittka, Yannick Wurm, and Lars Chittka. Epigenetics: the making of ant castes. *Current Biology*, 22(19):R835–R838, October 2012. ISSN 09609822. doi: 10.1016/j.cub.2012.07.045. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982212008706>.
- [184] W. Reik and J. Walter. Genomic imprinting: Parental influence on the genome. *Nature Reviews Genetics*, 2, 2001. doi: 10.1038/35047554. URL <http://dx.doi.org/10.1038/35047554>.
- [185] Harindra Amarasinghe, Bradley J Toghill, Despina Nathanael, and Eamonn B Mallon. Allele specific expression in worker reproduction genes in the bumblebee *Bombus terrestris*. *PeerJ*, 2015. doi: 10.7287/peerj.preprints.1094v1. URL [10.7287/peerj.preprints.1094v1](http://10.7287/peerj.preprints.1094v1).
- [186] H. E. Amarasinghe, C. I. Clayton, and E. B. Mallon. Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proceedings of the Royal Society B: Biological Sciences*, 281(1780):20132502–20132502, February 2014. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2013.2502. URL <http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2013.2502>.
- [187] S. D. Kocher, J. M. Tsuruda, J. D. Gibson, C. M. Emore, M. E. Arechavaleta-Velasco, D. C. Queller, J. E. Strassmann, C. M. Grozinger, M. R. Gribskov, P. San Miguel, R. Westerman, and G. J. Hunt. A search for parent-of-origin effects on honey bee gene expression. *G3:Genes|Genomes|Genetics*, 5(8):1657–1662, August 2015. ISSN 2160-1836. doi: 10.1534/g3.115.017814. URL <http://g3journal.org/cgi/doi/10.1534/g3.115.017814>.



- [188] Ben M Sadd, Seth M Barribeau, Guy Bloch, Dirk C de Graaf, Peter Dear-  
den, Christine G Elsik, Jrgen Gadau, Cornelis JP Grimmelihuizen, Mar-  
tin 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, Di-  
dac 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 Car-  
olan, Olivier Christiaens, Kate L Ciborowski, David F Clarke, Thomas J Col-  
gan, David H Collins, Andrew G Cridge, Tamas Dalmay, Stephanie Dreier,  
Louis du Plessis, Elizabeth Duncan, Silvio Erler, Jay Evans, Tiago Falcon,  
Kevin Flores, Flvia CP Freitas, Taro Fuchikawa, Tanja Gempe, Klaus Hart-  
felder, Frank Hauser, Sophie Helbing, Fernanda C Humann, Frano Irvine,  
Lars S Jermiin, Claire E Johnson, Reed M Johnson, Andrew K Jones, Tat-  
suhiko Kadowaki, Jonathan H Kidner, Vasco Koch, Arian Khler, F Bernhard  
Kraus, H Michael G Lattorff, Megan Leask, Gabrielle A Lockett, Eamonn B  
Mallon, David S Marco Antonio, Monika Marxer, Ivan Meeus, Robin FA  
Moritz, Ajay Nair, Kathrin Npflin, Inga Nissen, Jinzhi Niu, Francis MF Nunes,  
John G Oakeshott, Amy Osborne, Marianne Otte, Daniel G Pinheiro, Nina  
Rossi, Olav Rueppell, Carolina G Santos, Regula Schmid-Hempel, Bjrn D  
Schmitt, Christina Schulte, Zil LP Simes, Michelle PM Soares, Luc Swevers,  
Eva C Winnebeck, Florian Wolschin, Na Yu, Evgeny M Zdobnov, Peshte-  
wani K Aqrawi, Kerstin P Blankenburg, Marcus Coyle, Liezl Francisco, Al-  
varo G Hernandez, Michael Holder, Matthew E Hudson, LaRonda Jackson,  
Joy Jayaseelan, Vandita Joshi, Christie Kovar, Sandra L Lee, Robert Mata,  
Tittu Mathew, Irene F Newsham, Robin Ngo, Geoffrey Okwuonu, Christo-  
pher Pham, Ling-Ling Pu, Nehad Saada, Jireh Santibanez, DeNard Sim-  
mons, Rebecca Thornton, Aarti Venkat, Kimberly KO Walden, Yuan-Qing  
Wu, Griet Debyser, Bart Devreese, Claire Asher, Julie Blommaert, Ariel D

- Chipman, Lars Chittka, Bertrand Fouks, Jisheng Liu, Meaghan P O'Neill, Seirian Sumner, Daniela Puiu, Jiaxin Qu, Steven L Salzberg, Steven E Scherer, Donna M Muzny, Stephen Richards, Gene E Robinson, Richard A Gibbs, Paul Schmid-Hempel, and Kim C Worley. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biology*, 16(1), December 2015. ISSN 1465-6906. doi: 10.1186/s13059-015-0623-3. URL <http://genomebiology.com/2015/16/1/76>.
- [189] 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, Matia 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 1087-0156, 1546-1696. doi: 10.1038/nbt.1682. URL <http://www.nature.com/doifinder/10.1038/nbt.1682>.
- [190] 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, 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-1031. URL <http://www.biomedcentral.com/1471-2164/15/1031>.
- [191] 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, 2013. ISSN 1465-6906. doi: 10.1186/gb-2013-14-4-r36. URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-4-r36>.
- [192] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, August 2009. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btp352. URL <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btp352>.
- [193] H. Li and R. Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760, July 2009. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btp324. URL <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btp324>.
- [194] Mark A DePristo, Eric Banks, Ryan Poplin, Kiran V Garimella, Jared R Maguire, Christopher Hartl, Anthony A Philippakis, Guillermo del Angel, Manuel A Rivas, Matt Hanna, Aaron McKenna, Tim J Fennell, Andrew M Kernytsky, Andrey Y Sivachenko, Kristian Cibulskis, Stacey B Gabriel, David Altshuler, and Mark J Daly. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5):491–498, April 2011. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.806. URL <http://www.nature.com/doifinder/10.1038/ng.806>.
- [195] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. 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. doi: 10.1101/gr.107524.110. URL

- <http://genome.cshlp.org/cgi/doi/10.1101/gr.107524.110>.
- [196] P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, August 2011. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btr330. URL <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btr330>.
- [197] J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, and J. D. Storey. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, 28(6):882–883, March 2012. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/bts034. URL <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/bts034>.
- [198] W. E. Johnson, C. Li, and A. Rabinovic. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1):118–127, January 2007. ISSN 1465-4644, 1468-4357. doi: 10.1093/biostatistics/kxj037. URL <http://biostatistics.oxfordjournals.org/cgi/doi/10.1093/biostatistics/kxj037>.
- [199] A. R. Quinlan and I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, March 2010. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btq033. URL <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btq033>.
- [200] C. Harvey, G. A. Moyebrailean, O. Davis, X. Wen, F. Luca, and R. Pique-Regi. QuASAR: Quantitative Allele Specific Analysis of Reads. Technical Report 007492, July 2014. URL <http://biorxiv.org/lookup/doi/10.1101/007492>.

- [201] S. Altschul. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402, September 1997. ISSN 13624962. doi: 10.1093/nar/25.17.3389. URL <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/25.17.3389>.
- [202] S. Gotz, J. M. Garcia-Gomez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talon, J. Dopazo, and A. Conesa. High-throughput functional annotation and data mining with the Blast2go suite. *Nucleic Acids Research*, 36(10):3420–3435, April 2008. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkn176. URL <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkn176>.
- [203] Egle Gineikiene, Mindaugas Stoskus, and Laimonas Griskevicius. Single nucleotide polymorphism-based system improves the applicability of quantitative PCR for chimerism monitoring. *The Journal of Molecular Diagnostics*, 11(1): 66–74, January 2009. ISSN 15251578. doi: 10.2353/jmoldx.2009.080039. URL <http://linkinghub.elsevier.com/retrieve/pii/S1525157810602115>.
- [204] M. W. Pfaffl. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29(9):45e–45, May 2001. ISSN 13624962. doi: 10.1093/nar/29.9.e45. URL <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/29.9.e45>.
- [205] 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 00032697. doi: 10.1006/abio.2001.5530. URL <http://linkinghub.elsevier.com/retrieve/pii/S0003269701955307>.
- [206] K. D. Kelley, K. R. Miller, A. Todd, A. R. Kelley, R. Tuttle, and S. J. Berberich. YPEL3, a p53-regulated gene that induces cellular senescence. *Cancer Research*, 70(9):3566–3575, May 2010. ISSN 0008-5472, 1538-7445. doi:

- 10.1158/0008-5472.CAN-09-3219. URL <http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-09-3219>.
- [207] 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.
- [208] Dan Zheng, Bin-Bin Liu, Yin-Kun Liu, Xiao-Nan Kang, Lu Sun, Kun Guo, Rui-Xia Sun, Jie Chen, and Yan Zhao. 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(3):198–202, March 2009. ISSN 1007-3418.
- [209] 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, 2006. URL <http://www.nature.com/onc/journal/v25/n24/abs/1209383a.html>.
- [210] 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(4):1–6, July 2010. ISSN 2041-1723. doi: 10.1038/ncomms1044. URL <http://www.nature.com/doifinder/10.1038/ncomms1044>.
- [211] 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. URL <http://www.bioone.org/doi/abs/10.2108/zsj.24.596>.

- [212] Dries Cardoen, Tom Wenseleers, Ulrich R. Ernst, Ellen L. Danneels, Dries Laget, Dirk C. De Graaf, Liliane Schoofs, and Peter Verleyen. Genome-wide analysis of alternative reproductive phenotypes in honeybee workers: transcriptomic analysis of honeybee worker sterility. *Molecular Ecology*, 20(19): 4070–4084, October 2011. ISSN 09621083. doi: 10.1111/j.1365-294X.2011.05254.x. URL <http://doi.wiley.com/10.1111/j.1365-294X.2011.05254.x>.
- [213] 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, 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-75. URL <http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-75>.
- [214] 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. ISSN 1932-6203. doi: 10.1371/journal.pone.0013455. URL <http://dx.plos.org/10.1371/journal.pone.0013455>.
- [215] D. C. Markesich, K. M. Gajewski, M. E. Nazimiec, and K. Beckingham. bi-caudal 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.
- [216] Michael E. Scharf, Dancia Wu-Scharf, Barry R. Pittendrigh, and Gary Bennett. Caste- and development-associated gene expression in a lower termite. *Genome Biology*, 4(R62), 2003. doi: 10.1186/gb-2003-4-10-r62.
- [217] R. A. Drewell, E. C. Bush, E. J. Remnant, G. T. Wong, S. M. Beeler, J. L. Stringham, J. Lim, and B. P. Oldroyd. The dynamic DNA methylation cycle from egg to sperm in the honey bee *Apis mellifera*. *Development*, 141(13):

- 2702–2711, July 2014. ISSN 0950-1991, 1477-9129. doi: 10.1242/dev.110163. URL <http://dev.biologists.org/cgi/doi/10.1242/dev.110163>.
- [218] Haoxuan Liu, Yanxiao Jia, Xiaoguang Sun, Dacheng Tian, Laurence D. Hurst, and Sihai Yang. Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. *Molecular Biology and Evolution*, 34(1): 119–130, January 2017. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msw226. URL <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw226>.
- [219] Karl M Glastad, Brendan G Hunt, and Michael AD Goodisman. Evolutionary insights into DNA methylation in insects. *Current Opinion in Insect Science*, 1:25–30, July 2014. ISSN 22145745. doi: 10.1016/j.cois.2014.04.001. URL <http://linkinghub.elsevier.com/retrieve/pii/S2214574514000029>.
- [220] Sebastian Alvarado, Rajendhran Rajakumar, Ehab Abouheif, and Moshe Szyf. Epigenetic variation in the Egfr gene generates quantitative variation in a complex trait in ants. *Nature Communications*, 6:6513, March 2015. ISSN 2041-1723. doi: 10.1038/ncomms7513. URL <http://www.nature.com/doifinder/10.1038/ncomms7513>.
- [221] W. Dean, L. Bowden, A. Aitchison, J. Klose, T. Moore, J. J. Meneses, W. Reik, and R. Feil. Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes. *Development (Cambridge, England)*, 125(12):2273–2282, June 1998. ISSN 0950-1991.
- [222] F Pardo-Manuel de Villena. Natural selection and the function of genome imprinting: beyond the silenced minority. *Trends in Genetics*, 16(12):573–579, December 2000. ISSN 01689525. doi: 10.1016/S0168-9525(00)02134-X. URL <http://linkinghub.elsevier.com/retrieve/pii/S016895250002134X>.



- [223] P. Onyango, S. Jiang, H. Uejima, M. J. Shambloott, J. D. Gearhart, H. Cui, and A. P. Feinberg. Monoallelic expression and methylation of imprinted genes in human and mouse embryonic germ cell lineages. *Proceedings of the National Academy of Sciences*, 99(16):10599–10604, August 2002. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.152327599. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.152327599>.
- [224] C. Sapienza. Imprinted gene expression, transplantation medicine, and the “other” human embryonic stem cell. *Proceedings of the National Academy of Sciences*, 99(16):10243–10245, August 2002. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.172384299. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.172384299>.
- [225] Y. Zhang, T. Shields, T. Crenshaw, Y. Hao, T. Moulton, and B. Tycko. 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(1):113–124, July 1993. ISSN 0002-9297.
- [226] Daniela Pignatta, Robert M Erdmann, Elias Scheer, Colette L Picard, George W Bell, and Mary Gehring. Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *eLife*, 3, July 2014. ISSN 2050-084X. doi: 10.7554/eLife.03198. URL <http://elifesciences.org/lookup/doi/10.7554/eLife.03198>.
- [227] Emily J. Remnant, Alyson Ashe, Paul E. Young, Gabriele Buchmann, Madeleine Beekman, Michael H. Allsopp, Catherine M. Suter, Robert A. Drewell, and Benjamin P. Oldroyd. 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(1), December 2016. ISSN 1471-2164. doi: 10.1186/s12864-016-2506-8. URL <http://www.biomedcentral.com/1471-2164/17/226>.

- [228] Daniel Edsgard, Maria Jesus Iglesias, Sarah-Jayne Reilly, Anders Hamsten, Per Tornvall, Jacob Odeberg, and Olof Emanuelsson. GeneiASE: Detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information. *Scientific Reports*, 6(1), August 2016. ISSN 2045-2322. doi: 10.1038/srep21134. URL <http://www.nature.com/articles/srep21134>.
- [229] Emma L. Meaburn, Leonard C. Schalkwyk, and Jonathan Mill. Allele-specific methylation in the human genome: Implications for genetic studies of complex disease. *Epigenetics*, 5(7):578–582, October 2010. ISSN 1559-2294, 1559-2308. doi: 10.4161/epi.5.7.12960. URL <http://www.tandfonline.com/doi/abs/10.4161/epi.5.7.12960>.
- [230] Xu Wang, John H. Werren, and Andrew G. Clark. Allele-specific transcriptome and methylome analysis reveals stable inheritance and cis-regulation of DNA methylation in *Nasonia*. *PLOS Biology*, 14(7):e1002500, July 2016. ISSN 1545-7885. doi: 10.1371/journal.pbio.1002500. URL <http://dx.plos.org/10.1371/journal.pbio.1002500>.
- [231] Laura Wedd, Robert Kucharski, and Ryszard Maleszka. Differentially methylated obligatory epialleles modulate context-dependent *LAM* gene expression in the honeybee *Apis mellifera*. *Epigenetics*, 11(1):1–10, January 2016. ISSN 1559-2294, 1559-2308. doi: 10.1080/15592294.2015.1107695. URL <http://www.tandfonline.com/doi/full/10.1080/15592294.2015.1107695>.
- [232] Douglas L. Black. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, 72(1):291–336, June 2003. ISSN 0066-4154, 1545-4509. doi: 10.1146/annurev.biochem.72.121801.161720. URL <http://www.annualreviews.org/doi/10.1146/annurev.biochem.72.121801.161720>.

- [233] Robin A Weiss. Apes, lice and prehistory. *Journal of Biology*, 8(2):20, 2009. ISSN 1475-4924. doi: 10.1186/jbiol114. URL <http://jbiol.biomedcentral.com/articles/10.1186/jbiol114>.
- [234] Elena Rydkina. Bartonella quintana in Body Lice Collected from Homeless Persons in Russia. *Emerging Infectious Diseases*, 5(1):176–178, February 1999. ISSN 10806040. doi: 10.3201/eid0501.990126. URL <http://www.cdc.gov/ncidod/eid/vol5no1/letters.htm#Rydkina>,.
- [235] A. Jarosch, E. Stolle, R. M. Crewe, and R. F. A. Moritz. Alternative splicing of a single transcription factor drives selfish reproductive behavior in honeybee workers (*Apis mellifera*). *Proceedings of the National Academy of Sciences*, 108(37):15282–15287, September 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1109343108. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1109343108>.
- [236] Ellen M. Formesyn, Dries Cardoen, Ulrich R. Ernst, Ellen L. Danneels, Matthias Van Vaerenbergh, Dieter De Koker, Peter Verleyen, Tom Wenseleers, Liliane Schoofs, and Dirk C. de Graaf. Reproduction of honeybee workers is regulated by epidermal growth factor receptor signaling. *General and Comparative Endocrinology*, 197:1–4, February 2014. ISSN 00166480. doi: 10.1016/j.ygcen.2013.12.001. URL <http://linkinghub.elsevier.com/retrieve/pii/S0016648013004681>.
- [237] H. Li-Byarlay, Y. Li, H. Stroud, S. Feng, T. C. Newman, M. Kaneda, K. K. Hou, K. C. Worley, C. G. Elsik, S. A. Wickline, S. E. Jacobsen, J. Ma, and G. E. Robinson. RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proceedings of the National Academy of Sciences*, 110(31):12750–12755, July 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1310735110. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1310735110>.

- [238] Adam J. Bewick, Kevin J. Vogel, Allen J. Moore, and Robert J. Schmitz. Evolution of DNA Methylation across Insects. *Molecular Biology and Evolution*, page msw264, December 2016. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msw264. URL <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw264>.
- [239] Roberto Bonasio. Emerging topics in epigenetics: ants, brains, and non-coding RNAs: Ants, brains, and noncoding RNAs. *Annals of the New York Academy of Sciences*, 1260(1):14–23, July 2012. ISSN 00778923. doi: 10.1111/j.1749-6632.2011.06363.x. URL <http://doi.wiley.com/10.1111/j.1749-6632.2011.06363.x>.
- [240] R. Bonasio, S. Tu, and D. Reinberg. Molecular signals of epigenetic states. *Science*, 330(6004):612–616, October 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1191078. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1191078>.
- [241] M. J. Duchateau and H. H. W. Velthuis. Ovarian development and egg laying in workers of *Bombus terrestris*. *Entomologia Experimentalis et Applicata*, 51(3):199–213, July 1989. ISSN 00138703. doi: 10.1111/j.1570-7458.1989.tb01231.x. URL <http://doi.wiley.com/10.1111/j.1570-7458.1989.tb01231.x>.
- [242] Sharon Geva, Klaus Hartfelder, and Guy Bloch. Reproductive division of labor, dominance, and ecdysteroid levels in hemolymph and ovary of the bumble bee *Bombus terrestris*. *Journal of Insect Physiology*, 51(7):811–823, July 2005. ISSN 00221910. doi: 10.1016/j.jinsphys.2005.03.009. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022191005000764>.
- [243] Robin L. Foster, Ameilia Brunskill, David Verdirame, and Sean O’Donnell. Reproductive physiology, dominance interactions, and division of labour among bumble bee workers. *Physiological Entomology*, 29(4):327–334, September 2004. ISSN 0307-6962, 1365-3032. doi: 10.1111/j.0307-6962.2004.00388.x. URL <http://doi.wiley.com/10.1111/j.0307-6962.2004.00388.x>.

- [244] C. Trapnell, L. Pachter, and S. L. Salzberg. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9):1105–1111, May 2009. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btp120. URL <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btp120>.
- [245] Cole Trapnell, David G Hendrickson, Martin Sauvageau, Loyal Goff, John L Rinn, and Lior Pachter. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, 31(1):46–53, December 2012. ISSN 1087-0156, 1546-1696. doi: 10.1038/nbt.2450. URL <http://www.nature.com/doifinder/10.1038/nbt.2450>.
- [246] L. Goff, C. Trapnell, and D. Kelley. cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version 2.14.0., 2013.
- [247] F. Krueger and S. R. Andrews. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11):1571–1572, June 2011. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btr167. URL <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr167>.
- [248] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, March 2012. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.1923. URL <http://www.nature.com/doifinder/10.1038/nmeth.1923>.
- [249] Simon Andrews. SeqMonk, 2017. URL <http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>.
- [250] Kasper Daniel Hansen. Bsseq v1.8.2, 2017. URL <https://www.rdocumentation.org/packages/bsseq/versions/1.8.2>.

- [251] Florian Halbritter, Harsh J Vaidya, and Simon R Tomlinson. GeneProf: analysis of high-throughput sequencing experiments. *Nature Methods*, 9(1):7–8, December 2011. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.1809. URL <http://www.nature.com/doifinder/10.1038/nmeth.1809>.
- [252] Jacob Porter, Ming-an Sun, Hehuang Xie, and Liqing Zhang. Investigating bisulfite short-read mapping failure with hairpin bisulfite sequencing data. *BMC Genomics*, 16(Suppl 11):S2, 2015. ISSN 1471-2164. doi: 10.1186/1471-2164-16-S11-S2. URL <http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-16-S11-S2>.
- [253] Kristel R van Eijk, Simone de Jong, Marco PM Boks, Terry Langeveld, Fabrice Colas, Jan H Veldink, Carolien GF de Kovel, Esther Janson, Eric Strengman, Peter Langfelder, Ren S Kahn, Leonard H van den Berg, Steve Horvath, and Roel A Ophoff. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics*, 13(1): 636, 2012. ISSN 1471-2164. doi: 10.1186/1471-2164-13-636. URL <http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-636>.
- [254] L. Verdone. Histone acetylation in gene regulation. *Briefings in Functional Genomics and Proteomics*, 5(3):209–221, May 2006. ISSN 1473-9550, 1477-4062. doi: 10.1093/bfgp/ell028. URL <https://academic.oup.com/bfgp/article-lookup/doi/10.1093/bfgp/ell028>.
- [255] Courtney A. Miller, Susan L. Campbell, and J. David Sweatt. DNA methylation and histone acetylation work in concert to regulate memory formation and synaptic plasticity. *Neurobiology of Learning and Memory*, 89(4): 599–603, May 2008. ISSN 10747427. doi: 10.1016/j.nlm.2007.07.016. URL <http://linkinghub.elsevier.com/retrieve/pii/S1074742707001104>.
- [256] Yuling Qiu, Kwang Sik Lee, Young Moo Choo, Dexin Kong, Hyung Joo Yoon, and Byung Rae Jin. Molecular cloning and antifibrinolytic activity of a serine protease inhibitor from bumblebee (*Bombus terrestris*) venom. *Toxicon*, 63:

- 1–6, March 2013. ISSN 00410101. doi: 10.1016/j.toxicon.2012.11.004. URL <http://linkinghub.elsevier.com/retrieve/pii/S0041010112007994>.
- [257] Young Moo Choo, Kwang Sik Lee, Hyung Joo Yoon, Yuling Qiu, Hu Wan, Mi Ri Sohn, Hung Dae Sohn, and Byung Rae Jin. Antifibrinolytic role of a bee venom serine protease inhibitor that acts as a plasmin inhibitor. *PLoS ONE*, 7(2):e32269, February 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0032269. URL <http://dx.plos.org/10.1371/journal.pone.0032269>.
- [258] T Grunwald, B Bockisch, E Spillner, J Ring, R Bredehorst, and M Ollert. Molecular cloning and expression in insect cells of honeybee venom allergen acid phosphatase (Api m 3). *Journal of Allergy and Clinical Immunology*, 117(4):848–854, April 2006. ISSN 00916749. doi: 10.1016/j.jaci.2005.12.1331. URL <http://linkinghub.elsevier.com/retrieve/pii/S0091674905040856>.
- [259] Bo Yeon Kim and Byung Rae Jin. Molecular characterization of a venom acid phosphatase Acph-1-like protein from the Asiatic honeybee *Apis cerana*. *Journal of Asia-Pacific Entomology*, 17(4):695–700, December 2014. ISSN 12268615. doi: 10.1016/j.aspen.2014.07.002. URL <http://linkinghub.elsevier.com/retrieve/pii/S1226861514000880>.
- [260] Jia-ying Zhu, Gong-yin Ye, and Cui Hu. Molecular cloning and characterization of acid phosphatase in venom of the endoparasitoid wasp *Pteromalus puparum* (Hymenoptera: Pteromalidae). *Toxicon*, 51(8):1391–1399, June 2008. ISSN 00410101. doi: 10.1016/j.toxicon.2008.03.008. URL <http://linkinghub.elsevier.com/retrieve/pii/S0041010108001013>.
- [261] Mary E. Heavner, Gwenaëlle Gueguen, Roma Rajwani, Pedro E. Pagan, Chiyedza Small, and Shubha Govind. Partial venom gland transcriptome of a *Drosophila* parasitoid wasp, *Leptopilina heterotoma*, reveals novel and shared bioactive profiles with stinging Hymenoptera. *Gene*, 526(2):195–204,

- September 2013. ISSN 03781119. doi: 10.1016/j.gene.2013.04.080. URL <http://linkinghub.elsevier.com/retrieve/pii/S0378111913006136>.
- [262] M.P. Dani, J.P. Edwards, and E.H. Richards. Hydrolase activity in the venom of the pupal endoparasitic wasp, *Pimpla hypochondriaca*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 141(3): 373–381, July 2005. ISSN 10964959. doi: 10.1016/j.cbpc.2005.04.010. URL <http://linkinghub.elsevier.com/retrieve/pii/S1096495905000886>.
- [263] D. C. de Graaf, M. Aerts, M. Brunain, C. A. Desjardins, F. J. Jacobs, J. H. Werren, and B. Devreese. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. *Insect Molecular Biology*, 19:11–26, February 2010. ISSN 09621075, 13652583. doi: 10.1111/j.1365-2583.2009.00914.x. URL <http://doi.wiley.com/10.1111/j.1365-2583.2009.00914.x>.
- [264] Martin K. Safo, Faik N. Musayev, and Verne Schirch. Structure of *Escherichia coli* pyridoxine 5-phosphate oxidase in a tetragonal crystal form: insights into the mechanistic pathway of the enzyme. *Acta Crystallographica Section D Biological Crystallography*, 61(5):599–604, May 2005. ISSN 0907-4449. doi: 10.1107/S0907444905005512. URL <http://scripts.iucr.org/cgi-bin/paper?S0907444905005512>.
- [265] Alexandre S. Cristino, Angel R. Barchuk, Flavia C. P. Freitas, Ramesh K. Narayanan, Stephanie D. Biergans, Zhengyang Zhao, Zila L. P. Simoes, Judith Reinhard, and Charles Claudianos. Neuroligin-associated microRNA-932 targets actin and regulates memory in the honeybee. *Nature Communications*, 5:5529, November 2014. ISSN 2041-1723. doi: 10.1038/ncomms6529. URL <http://www.nature.com/doifinder/10.1038/ncomms6529>.
- [266] K. M. Glastad, B. G. Hunt, S. V. Yi, and M. A. D. Goodisman. DNA methylation in insects: on the brink of the epigenomic era. *Insect Molecular Biology*, 20(5):553–565, October 2011. ISSN 09621075. doi: 10.1111/j.1365-2583.2011.



- 01092.x. URL <http://doi.wiley.com/10.1111/j.1365-2583.2011.01092.x>.
- [267] Susan A. Weiner and Amy L. Toth. Epigenetics in social insects: a new direction for understanding the evolution of castes. *Genetics Research International*, 2012:1–11, 2012. ISSN 2090-3154, 2090-3162. doi: 10.1155/2012/609810. URL <http://www.hindawi.com/journals/gri/2012/609810/>.
- [268] 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 22145745. doi: 10.1016/j.cois.2014.04.005. URL <http://linkinghub.elsevier.com/retrieve/pii/S2214574514000066>.
- [269] Gang Li, Jae Hoon Bahn, Jae-Hyung Lee, Guangdun Peng, Zugen Chen, Stanley F. Nelson, and Xinshu Xiao. Identification of allele-specific alternative mRNA processing via transcriptome sequencing. *Nucleic Acids Research*, 40(13):e104–e104, July 2012. ISSN 1362-4962, 0305-1048. doi: 10.1093/nar/gks280. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks280>.
- [270] Martin Kircher, Patricia Heyn, and Janet Kelso. Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*, 12(1), December 2011. ISSN 1471-2164. doi: 10.1186/1471-2164-12-382. URL <http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-12-382>.
- [271] S. Andrews. FastQC: a quality control tool for high throughput sequence data., 2010. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- [272] Qiang Song, Benjamin Decato, Elizabeth E. Hong, Meng Zhou, Fang Fang, Jianghan Qu, Tyler Garvin, Michael Kessler, Jun Zhou, and Andrew D. Smith. A reference methylome database and analysis pipeline to facilitate integrative

- and comparative epigenomics. *PLoS ONE*, 8(12):e81148, December 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0081148. URL <http://dx.plos.org/10.1371/journal.pone.0081148>.
- [273] Martin Vincent, Kamilla Mundbjerg, Jakob Skou Pedersen, Gangning Liang, Peter A. Jones, Torben Falck rntoft, Karina Dalsgaard Srensen, and Carsten Wiuf. epiG: statistical inference and profiling of DNA methylation from whole-genome bisulfite sequencing data. *Genome Biology*, 18(1), December 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1168-4. URL <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1168-4>.
- [274] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, October 1990. ISSN 0022-2836. doi: 10.1016/S0022-2836(05)80360-2.
- [275] R. D. Finn, J. Clements, and S. R. Eddy. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, 39(suppl):W29–W37, July 2011. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkr367. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr367>.
- [276] M. Punta, P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman, and R. D. Finn. The Pfam protein families database. *Nucleic Acids Research*, 40(D1):D290–D301, January 2012. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkr1065. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr1065>.
- [277] Thomas Nordahl Petersen, Sren Brunak, Gunnar von Heijne, and Henrik Nielsen. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10):785–786, September 2011. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.1701. URL <http://www.nature.com/doifinder/10.1038/nmeth.1701>.

- [278] A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580, January 2001. ISSN 0022-2836. doi: 10.1006/jmbi.2000.4315.
- [279] S. Powell, D. Szklarczyk, K. Trachana, A. Roth, M. Kuhn, J. Muller, R. Arnold, T. Rattei, I. Letunic, T. Doerks, L. J. Jensen, C. von Mering, and P. Bork. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Research*, 40(D1):D284–D289, January 2012. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkr1060. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr1060>.
- [280] Minoru Kanehisa, Susumu Goto, Yoko Sato, Miho Furumichi, and Mao Tanabe. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40(Database issue):D109–114, January 2012. ISSN 1362-4962. doi: 10.1093/nar/gkr988.
- [281] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1):25–29, May 2000. ISSN 1061-4036. doi: 10.1038/75556.
- [282] Trees-Juen Chuang, Yu-Hsiang Tseng, Chia-Ying Chen, and Yi-Da Wang. Assessment of imprinting- and genetic variation-dependent monoallelic expression using reciprocal allele descendants between human family trios. *Scientific Reports*, 7(1), December 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-07514-z. URL <http://www.nature.com/articles/s41598-017-07514-z>.
- [283] N. Asomaning and K.J. Archer. High-throughput DNA methylation datasets for evaluating false discovery rate methodologies. *Computational Statistics &*

- Data Analysis*, 56(6):1748–1756, June 2012. ISSN 01679473. doi: 10.1016/j.csda.2011.10.020. URL <http://linkinghub.elsevier.com/retrieve/pii/S0167947311003914>.
- [284] Iksoo Huh, Xin Wu, Taesung Park, and Soojin V. Yi. Detecting differential DNA methylation from sequencing of bisulfite converted DNA of diverse species. *Briefings in Bioinformatics*, July 2017. ISSN 1467-5463, 1477-4054. doi: 10.1093/bib/bbx077. URL <https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbx077>.
- [285] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.
- [286] G. Gill. Regulation of the initiation of eukaryotic transcription. *Essays in Biochemistry*, 37:33–43, 2001. ISSN 0071-1365.
- [287] Robert J. Klose, Eric M. Kallin, and Yi Zhang. JmjC-domain-containing proteins and histone demethylation. *Nature Reviews Genetics*, 7(9):715–727, September 2006. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg1945. URL <http://www.nature.com/doifinder/10.1038/nrg1945>.
- [288] Lei Wang, Jianjun Chang, Diana Varghese, Michael Dellinger, Subodh Kumar, Anne M. Best, Julio Ruiz, Richard Bruick, Samuel Peña-Llopis, Junjie Xu, David J. Babinski, Doug E. Frantz, Rolf A. Brekken, Amy M. Quinn, Anton Simeonov, Johnny Easmon, and Elisabeth D. Martinez. A small molecule modulates Jumonji histone demethylase activity and selectively inhibits cancer growth. *Nature Communications*, 4, June 2013. ISSN 2041-1723. doi: 10.1038/ncomms3035. URL <http://www.nature.com/doifinder/10.1038/ncomms3035>.

- [289] Gianluigi Franci, Alfonso Ciotta, and Lucia Altucci. The Jumonji family: past, present and future of histone demethylases in cancer. *Biomolecular Concepts*, 5(3), January 2014. ISSN 1868-503X, 1868-5021. doi: 10.1515/bmc-2014-0010. URL <https://www.degruyter.com/view/j/bmc.2014.5.issue-3/bmc-2014-0010/bmc-2014-0010.xml>.
- [290] George A. Khoury, Richard C. Baliban, and Christodoulos A. Floudas. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific Reports*, 1(1), December 2011. ISSN 2045-2322. doi: 10.1038/srep00090. URL <http://www.nature.com/articles/srep00090>.
- [291] Ye Zhang, Karen Griffin, Neelima Mondal, and Jeffrey D. Parvin. Phosphorylation of histone H2a inhibits transcription on chromatin templates. *Journal of Biological Chemistry*, 279(21):21866–21872, May 2004. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M400099200. URL <http://www.jbc.org/lookup/doi/10.1074/jbc.M400099200>.
- [292] H. Deng, X. Bao, W. Zhang, J. Girton, J. Johansen, and K. M. Johansen. Reduced levels of Su(var)3-9 but not Su(var)2-5 (HP1) counteract the effects on chromatin structure and viability in loss-of-function mutants of the JIL-1 histone H3s10 kinase. *Genetics*, 177(1):79–87, July 2007. ISSN 0016-6731. doi: 10.1534/genetics.107.075143. URL <http://www.genetics.org/cgi/doi/10.1534/genetics.107.075143>.
- [293] Yoko Ikeda and Taisuke Nishimura. The Role of DNA Methylation in Transposable Element Silencing and Genomic Imprinting. In Olga Pontes and Hailing Jin, editors, *Nuclear Functions in Plant Transcription, Signaling and Development*, pages 13–29. Springer New York, New York, NY, 2015. ISBN 978-1-4939-2385-4 978-1-4939-2386-1. URL [http://link.springer.com/10.1007/978-1-4939-2386-1\\_2](http://link.springer.com/10.1007/978-1-4939-2386-1_2). DOI: 10.1007/978-1-4939-2386-1\_2.

- [294] Drew Schwartz and Elizabeth Dennis. Transposase activity of the Ac controlling element in maize is regulated by its degree of methylation. *Molecular and General Genetics MGG*, 205(3):476–482, December 1986. ISSN 1432-1874. doi: 10.1007/BF00338085. URL <https://doi.org/10.1007/BF00338085>.
- [295] Miho M. Suzuki and Adrian Bird. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, 9(6):465–476, June 2008. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg2341. URL <http://www.nature.com/doifinder/10.1038/nrg2341>.
- [296] S. Feng, S. E. Jacobsen, and W. Reik. Epigenetic reprogramming in plant and animal development. *Science*, 330(6004):622–627, October 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1190614. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1190614>.
- [297] Mirjana B. Colovic, Danijela Z. Krstic, Tamara D. Lazarevic-Pasti, Aleksandra M. Bondzic, and Vesna M. Vasic. Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. *Current Neuropharmacology*, 11(3):315–335, April 2013. ISSN 1570159X. doi: 10.2174/1570159X11311030006. URL <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1570-159X&volume=11&issue=3&spage=315>.
- [298] M.L. Yang, J.Z. Zhang, K.Y. Zhu, T. Xuan, X.J. Liu, Y.P. Guo, and E.B. Ma. Mechanisms of organophosphate resistance in a field population of oriental migratory locust, *Locusta migratoria manilensis* (Meyen). *Archives of Insect Biochemistry and Physiology*, 71(1):3–15, May 2009. ISSN 07394462, 15206327. doi: 10.1002/arch.20254. URL <http://doi.wiley.com/10.1002/arch.20254>.
- [299] A. Steedman. *Locust Handbook*. Overseas Development Natural Resource Institute, London, UK, second edition, 1988.

- [300] R. J. Konopka and S. Benzer. Clock mutants of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 68(9):2112–2116, September 1971. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.68.9.2112. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.68.9.2112>.
- [301] Gustavo B. S. Rivas, Luiz G. S. da R. Bauzer, and Antonio C. A. Meireles-Filho. The environment is everything that isn't me: molecular mechanisms and evolutionary dynamics of insect clocks in variable surroundings. *Frontiers in Physiology*, 6, January 2016. ISSN 1664-042X. doi: 10.3389/fphys.2015.00400. URL <http://journal.frontiersin.org/Article/10.3389/fphys.2015.00400/abstract>.
- [302] Nicholas J McGlincy, Amandine Valomon, Johanna E Chesham, Elizabeth S Maywood, Michael H Hastings, and Jernej Ule. Regulation of alternative splicing by the circadian clock and food related cues. *Genome Biology*, 13(6):R54, 2012. ISSN 1465-6906. doi: 10.1186/gb-2012-13-6-r54. URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2012-13-6-r54>.
- [303] S. Kadener, J. S. Menet, K. Sugino, M. D. Horwich, U. Weissbein, P. Nawathean, V. V. Vagin, P. D. Zamore, S. B. Nelson, and M. Rosbash. A role for microRNAs in the *Drosophila circadian clock*. *Genes & Development*, 23(18):2179–2191, September 2009. ISSN 0890-9369. doi: 10.1101/gad.1819509. URL <http://genesdev.cshlp.org/cgi/doi/10.1101/gad.1819509>.
- [304] Liesbeth Badisco, Jurgen Huybrechts, Gert Simonet, Heleen Verlinden, Elisabeth Marchal, Roger Huybrechts, Liliane Schoofs, Arnold De Loof, and Jozef Vanden Broeck. Transcriptome analysis of the desert locust central nervous system: production and annotation of a *Schistocerca gregaria* EST database. *PLoS ONE*, 6(3):e17274, March 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0017274. URL <http://dx.plos.org/10.1371/journal.pone.0017274>.

- [305] Darron A. Cullen, Gregory A. Sword, Tim Dodgson, and Stephen J. Simpson. Behavioural phase change in the Australian plague locust, *Chortoicetes terminifera*, is triggered by tactile stimulation of the antennae. *Journal of Insect Physiology*, 56(8):937–942, August 2010. ISSN 1879-1611. doi: 10.1016/j.jinsphys.2010.04.023.
- [306] null Hgele and null Simpson. The influence of mechanical, visual and contact chemical stimulation on the behavioural phase state of solitarious desert locusts (*Schistocerca gregaria*). *Journal of Insect Physiology*, 46(9):1295–1301, September 2000. ISSN 1879-1611. doi: 10.1016/S0022-1910(00)00051-2.
- [307] P. Roessingh, A. Bouachi, and S. J. Simpson. Effects of sensory stimuli on the behavioural phase state of the desert locust, *Schistocerca gregaria*. *Journal of Insect Physiology*, 44(10):883–893, October 1998. ISSN 0022-1910.
- [308] D. Obeng-Ofori, P. G. N. Njagi, B. Torto, A. Hassanali, and H. Amiani. Sex differentiation studies relating to releaser aggregation pheromones of the desert locust, *Schistocerca gregaria*. *Entomologia Experimentalis et Applicata*, 73(1): 85–91, October 1994. ISSN 00138703. doi: 10.1111/j.1570-7458.1994.tb01842.x. URL <http://doi.wiley.com/10.1111/j.1570-7458.1994.tb01842.x>.
- [309] Yftach Golov, Jan Rillich, Ally Harari, and Amir Ayali. Precopulatory behavior and sexual conflict in the desert locust. pages –, September 2017. doi: 10.1101/190793. URL <http://biorxiv.org/lookup/doi/10.1101/190793>.
- [310] Arnold De Loof, Jurgen Huybrechts, Marisa Geens, Tim Vandersmissen, Bart Boerjan, and Liliane Schoofs. Sexual differentiation in adult insects: Male-specific cuticular yellowing in *Schistocerca gregaria* as a model for reevaluating some current (neuro)endocrine concepts. *Journal of Insect Physiology*, 56(8):919–925, August 2010. ISSN 00221910. doi: 10.1016/j.jinsphys.2010.02.021. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022191010000703>.



- [311] Hadley Wilson Horch. Molecular approach to the circadian clock mechanism in the cricket. In *The cricket as a model organism: development, regeneration, and behavior*, pages 78–79. Springer Berlin Heidelberg, New York, NY, 2016. ISBN 978-4-431-56476-8.
- [312] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10, May 2011. ISSN 2226-6089. doi: 10.14806/ej.17.1.200. URL <http://journal.embnet.org/index.php/embnetjournal/article/view/200>.
- [313] Gustavo AT Sacomoto, Janice Kielbassa, Rayan Chikhi, Raluca Uricaru, Pavlos Antoniou, Marie-France Sagot, Pierre Peterlongo, and Vincent Lacroix. KIS SPLICE: de-novo calling alternative splicing events from RNA-seq data. *BMC Bioinformatics*, 13(6):S5, 2012. ISSN 1471-2105. doi: 10.1186/1471-2105-13-S6-S5. URL <http://dx.doi.org/10.1186/1471-2105-13-S6-S5>.
- [314] Lilia Brinza, Vincent Lacroix, Frank Picard, Vincent Miele, Camille Marchet, and Clara Benoit-Pilven. kissDE: retrieves condition-specific variants in RNA-seq data. R package version 1.4.0., 2016.
- [315] Michael E. Hughes, John B. Hogenesch, and Karl Kornacker. JTK\_cycle: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of Biological Rhythms*, 25(5):372–380, October 2010. ISSN 0748-7304, 1552-4531. doi: 10.1177/0748730410379711. URL <http://journals.sagepub.com/doi/10.1177/0748730410379711>.
- [316] A. R. Jonckheere. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 41(1/2):133, June 1954. ISSN 00063444. doi: 10.2307/2333011. URL <http://www.jstor.org/stable/2333011?origin=crossref>.

- [317] T.J. Terpstra. The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Proc. Kon. Ned. Akad. v. Wetensch. A*, 55:327–333, 1952.
- [318] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, June 1938. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/30.1-2.81. URL <https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/30.1-2.81>.
- [319] A. Conesa, S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18):3674–3676, September 2005. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/bti610. URL <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bti610>.
- [320] Y. Liu and N. F. Tsinoremas. An unusual gene arrangement for the putative chromosome replication origin and circadian expression of dnaN in *Synechococcus* sp. strain PCC 7942. *Gene*, 172(1):105–109, June 1996. ISSN 0378-1119.
- [321] Xiao Liu, Yunkun Dang, Toru Matsu-ura, Yubo He, Qun He, Christian I. Hong, and Yi Liu. DNA replication is required for circadian clock function by regulating rhythmic nucleosome composition. *Molecular Cell*, 67(2):203–213.e4, July 2017. ISSN 10972765. doi: 10.1016/j.molcel.2017.05.029. URL <http://linkinghub.elsevier.com/retrieve/pii/S1097276517303933>.
- [322] Jerry F. Feldman, GF Gardner, and Richard Denison. Genetic Analysis of the Circadian Clock of *Neurospora*. In M Suda, editor, *Biological Rhythms and their Central Mechanism*, pages 57–66. Elsevier Amsterdam, 1979.

- [323] Jillian M. Emerson, Bradley M. Bartholomai, Carol S. Ringelberg, Scott E. Baker, Jennifer J. Loros, and Jay C. Dunlap. *period* -1 encodes an ATP-dependent RNA helicase that influences nutritional compensation of the *Neurospora* circadian clock. *Proceedings of the National Academy of Sciences*, page 201521918, December 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1521918112. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1521918112>.
- [324] K. Padmanabhan, M. S. Robles, T. Westerling, and C. J. Weitz. Feedback regulation of transcriptional termination by the mammalian circadian clock *period* complex. *Science*, 337(6094):599–602, August 2012. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1221592. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1221592>.
- [325] Oren Froy. Cytochrome P450 and the biological clock in mammals. *Current Drug Metabolism*, 10(2):104–115, February 2009. ISSN 1389-2002.
- [326] Jun Yan, Haifang Wang, Yuting Liu, and Chunxuan Shao. Analysis of gene regulatory networks in the mammalian circadian rhythm. *PLoS Computational Biology*, 4(10):e1000193, October 2008. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000193. URL <http://dx.plos.org/10.1371/journal.pcbi.1000193>.
- [327] Michael B. Cohen and Ren Feyereisen. A cluster of cytochrome P450 genes of the CYP6 family in the house fly. *DNA and Cell Biology*, 14(1):73–82, January 1995. ISSN 1044-5498, 1557-7430. doi: 10.1089/dna.1995.14.73. URL <http://www.liebertonline.com/doi/abs/10.1089/dna.1995.14.73>.
- [328] L. M. Field and A. L. Devonshire. Evidence that the E4 and FE4 esterase genes responsible for insecticide resistance in the aphid *Myzus persicae* (Sulzer) are part of a gene family. *The Biochemical Journal*, 330 ( Pt 1):169–173, February 1998. ISSN 0264-6021.

- [329] Tatiana Varanita, Maria Eugenia Soriano, Vanina Romanello, Tania Zagli, Rubn Quintana-Cabrera, Martina Semenzato, Roberta Menab, Veronica Costa, Gabriele Civiletto, Paola Pesce, Carlo Viscomi, Massimo Zeviani, Fabio Di Lisa, Marco Mongillo, Marco Sandri, and Luca Scorrano. The Opa1-dependent mitochondrial cristae remodeling pathway controls atrophic, apoptotic, and ischemic tissue damage. *Cell Metabolism*, 21(6):834–844, June 2015. ISSN 15504131. doi: 10.1016/j.cmet.2015.05.007. URL <http://linkinghub.elsevier.com/retrieve/pii/S1550413115002193>.
- [330] Fan Wang, Caiyan Li, Yongluo, and Lvan Chen. The circadian gene clock plays an important role in cell apoptosis and the DNA damage response *in vitro*. *Technology in Cancer Research & Treatment*, 15(3):480–486, June 2016. ISSN 1533-0346, 1533-0338. doi: 10.1177/1533034615585433. URL <http://journals.sagepub.com/doi/10.1177/1533034615585433>.
- [331] Valerie L. Kilman, Luoying Zhang, Rose-Anne Meissner, Elyssa Burg, and Ravi Allada. Perturbing dynamin reveals potent effects on the *Drosophila* circadian clock. *PLoS ONE*, 4(4):e5235, April 2009. ISSN 1932-6203. doi: 10.1371/journal.pone.0005235. URL <http://dx.plos.org/10.1371/journal.pone.0005235>.
- [332] M. Mlodzik, Y. Hiromi, U. Weber, C. S. Goodman, and G. M. Rubin. The *Drosophila* seven-up gene, a member of the steroid receptor gene superfamily, controls photoreceptor cell fates. *Cell*, 60(2):211–224, January 1990. ISSN 0092-8674.
- [333] Edouard Jaumouille, Pedro Machado Almeida, Patrick Sthli, Rafael Koch, and Emi Nagoshi. Transcriptional regulation via nuclear receptor crosstalk required for the *Drosophila* circadian clock. *Current Biology*, 25(11):1502–1508, June 2015. ISSN 09609822. doi: 10.1016/j.cub.2015.04.017. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982215004303>.

- [334] A. Sehgal, J. L. Price, B. Man, and M. W. Young. Loss of circadian behavioral rhythms and per RNA oscillations in the *Drosophila* mutant timeless. *Science*, 263(5153):1603–1606, March 1994. ISSN 0036-8075.
- [335] Lu Chen, Stephen J. Bush, Jaime M. Tovar-Corona, Atahualpa Castillo-Morales, and Araxi O. Urrutia. Correcting for Differential Transcript Coverage Reveals a Strong Relationship between Alternative Splicing and Organism Complexity. *Molecular Biology and Evolution*, 31(6):1402–1413, June 2014. ISSN 1537-1719, 0737-4038. doi: 10.1093/molbev/msu083. URL <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu083>.
- [336] Pooja Singh, Christine Brger, Heather More, and Christian Sturmbauer. The Role of Alternative Splicing and Differential Gene Expression in Cichlid Adaptive Radiation. *Genome Biology and Evolution*, 9(10):2764–2781, October 2017. ISSN 1759-6653. doi: 10.1093/gbe/evx204. URL <http://academic.oup.com/gbe/article/9/10/2764/4259059>.
- [337] Grzegorz Buczkowski and Gary Bennett. Colony budding and its effects on food allocation in the highly polygynous ant, *Monomorium pharaonis*. *Ethology*, 115(11):1091–1099, November 2009. ISSN 01791613, 14390310. doi: 10.1111/j.1439-0310.2009.01698.x. URL <http://doi.wiley.com/10.1111/j.1439-0310.2009.01698.x>.
- [338] Deborah M. Gordon. The dynamics of the daily round of the harvester ant colony (*Pogonomyrmex barbatus*). *Animal Behaviour*, 34(5):1402–1419, October 1986. ISSN 00033472. doi: 10.1016/S0003-3472(86)80211-1. URL <http://linkinghub.elsevier.com/retrieve/pii/S0003347286802111>.
- [339] C. Peeters, J. Liebig, and B. Hlldobler. Sexual reproduction by both queens and workers in the ponerine ant *Harpegnathos saltator*. *Insectes Sociaux*, 47(4): 325–332, November 2000. ISSN 0020-1812. doi: 10.1007/PL00001724. URL <http://link.springer.com/10.1007/PL00001724>.

- [340] M. Zavolan. Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Research*, 13(6):1290–1300, June 2003. ISSN 1088-9051. doi: 10.1101/gr.1017303. URL <http://www.genome.org/cgi/doi/10.1101/gr.1017303>.
- [341] B. Modrek. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Research*, 29(13):2850–2859, July 2001. ISSN 13624962. doi: 10.1093/nar/29.13.2850. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/29.13.2850>.
- [342] Manuel Irimia, Jakob Rukov, David Penny, and Scott Roy. Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. *BMC Evolutionary Biology*, 7(1): 188, 2007. ISSN 14712148. doi: 10.1186/1471-2148-7-188. URL <http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-7-188>.
- [343] David H. Oi, Karen M. Vail, David F. Williams, and Donald N. Bieman. Indoor and outdoor foraging locations of pharaoh ants (Hymenoptera: Formicidae) and control strategies using bait stations. *The Florida Entomologist*, 77(1): 85, March 1994. ISSN 00154040. doi: 10.2307/3495874. URL <http://www.jstor.org/stable/3495874?origin=crossref>.
- [344] Xavier Bonnefoy, Helge Kampen, and Kevin Sweeney. *Public health significance of urban pests*. WHO Reginal Office for Europe, Copenhagen, 2008. ISBN 978-92-890-7188-8. OCLC: 254961491.
- [345] Bert Holldobler. Recruitment behavior, home range orientation and territoriality in harvester ants, *Pogonomyrmex*. *Behavioral Ecology and Sociobiology*, 1(1):3–44, 1976. ISSN 0340-5443, 1432-0762. doi: 10.1007/BF00299951. URL <http://link.springer.com/10.1007/BF00299951>.
- [346] C. R. Smith, C. D. Smith, H. M. Robertson, M. Helmkamp, A. Zimin, M. Yandell, C. Holt, H. Hu, E. Abouheif, R. Benton, E. Cash, V. Croset,

- C. R. Currie, E. Elhaik, C. G. Elsik, M.-J. Fave, V. Fernandes, J. D. Gibson, D. Graur, W. Gronenberg, K. J. Grubbs, D. E. Hagen, A. S. I. Viniegra, B. R. Johnson, R. M. Johnson, A. Khila, J. W. Kim, K. A. Mathis, M. C. Munoz-Torres, M. C. Murphy, J. A. Mustard, R. Nakamura, O. Niehuis, S. Nigam, R. P. Overson, J. E. Placek, R. Rajakumar, J. T. Reese, G. Suen, S. Tao, C. W. Torres, N. D. Tsutsui, L. Viljakainen, F. Wolschin, and J. Gadau. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proceedings of the National Academy of Sciences*, 108(14):5667–5672, April 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1007901108. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1007901108>.
- [347] R. H. Crozier and P. Pamilo. In *Evolution of Social Insect Colonies: Sex Allocation and Kin Selection*. Oxford, Oxford University Press, 1996.
- [348] null Bloch. Regulation of queen-worker conflict in bumble-bee (*Bombus terrestris*) colonies. *Proceedings. Biological Sciences*, 266(1437):2465–2469, December 1999. ISSN 1471-2954.
- [349] Charles D. Michener. Comparative social behavior of bees. *Annual Review of Entomology*, 14(1):299–342, January 1969. ISSN 0066-4170, 1545-4487. doi: 10.1146/annurev.en.14.010169.001503. URL <http://www.annualreviews.org/doi/10.1146/annurev.en.14.010169.001503>.
- [350] David R Smith. Systematics, Life History, and Distribution of Sawflies. In *Sawfly Life History Adaptations to Woody Plants*, pages 3–32. Academic Press: New York, NY, USA, 1993.
- [351] Ana Conesa, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Micha Wojciech Szczeniak, Daniel J. Gaffney, Laura L. Elo, Xuegong Zhang, and Ali Mortazavi. A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), December 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-0881-8. URL <http://genomebiology.com/2016/17/1/13>.

- [352] Adam Roberts, Cole Trapnell, Julie Donaghey, John L Rinn, and Lior Pachter. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology*, 12(3):R22, 2011. ISSN 1465-6906. doi: 10.1186/gb-2011-12-3-r22. URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-3-r22>.
- [353] Christine G Elsik, Kim C Worley, Anna K Bennett, Martin Beye, Francisco Camara, Christopher P Childers, Dirk C de Graaf, Griet Debyser, Jixin Deng, Bart Devreese, Eran Elhaik, Jay D Evans, Leonard J Foster, Dan Graur, Roderic Guigo, HGSC production teams, Katharina Hoff, Michael E Holder, Matthew E Hudson, Greg J Hunt, Huaiyang Jiang, Vandita Joshi, Radhika S Khetani, Peter Kosarev, Christie L Kovar, Jian Ma, Ryszard Maleszka, Robin F A Moritz, Monica C Munoz-Torres, Terence D Murphy, Donna M Muzny, Irene F Newsham, Justin T Reese, Hugh M Robertson, Gene E Robinson, Olav Rueppell, Victor Solovyev, Mario Stanke, Eckart Stolle, Jennifer M Tsuruda, Matthias Vaerenbergh, Robert M Waterhouse, Daniel B Weaver, Charles W Whitfield, Yuanqing Wu, Evgeny M Zdobnov, Lan Zhang, Dianhui Zhu, Richard A Gibbs, and on behalf of Honey Bee Genome Sequencing Consortium. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*, 15(1):86, 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-86. URL <http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-86>.
- [354] Roberto Bonasio, Guojie Zhang, Chaoyang Ye, Navdeep S. Mutti, Xiaodong Fang, Nan Qin, Greg Donahue, Pengcheng Yang, Qiye Li, Cai Li, Pei Zhang, Zhiyong Huang, Shelley L. Berger, Danny Reinberg, Jun Wang, and Jrgen Liebig. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science (New York, N. Y.)*, 329(5995):1068–1071, August 2010. ISSN 1095-9203. doi: 10.1126/science.1192428.



- [355] Mauricio Bacci, Scott E. Solomon, Ulrich G. Mueller, Vanderlei G. Martins, Alfredo O.R. Carvalho, Luiz G.E. Vieira, and Ana Carla O. Silva-Pinhati. Phylogeny of leafcutter ants in the genus *Atta Fabricius* (Formicidae: Attini) based on mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, 51(3):427–437, June 2009. ISSN 10557903. doi: 10.1016/j.ympev.2008.11.005. URL <http://linkinghub.elsevier.com/retrieve/pii/S1055790308005265>.
- [356] Alexis Dereeper, Stephane Audic, Jean-Michel Claverie, and Guillaume Blanc. BLAST-explorer helps you building datasets for phylogenetic analysis. *BMC Evolutionary Biology*, 10(1):8, 2010. ISSN 1471-2148. doi: 10.1186/1471-2148-10-8. URL <http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-10-8>.
- [357] A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J.-F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J.-M. Claverie, and O. Gascuel. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, 36(Web Server):W465–W469, May 2008. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkn180. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkn180>.
- [358] Robert C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, 2004. ISSN 1362-4962. doi: 10.1093/nar/gkh340.
- [359] J. Castresana. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4):540–552, April 2000. ISSN 0737-4038.
- [360] Stphane Guindon and Olivier Gascuel. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5):696–704, October 2003. ISSN 1063-5157.

- [361] Maria Anisimova and Olivier Gascuel. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, 55(4):539–552, August 2006. ISSN 1063-5157. doi: 10.1080/10635150600755453.
- [362] Francois Chevenet, Christine Brun, Anne-Laure Bauls, Bernard Jacq, and Richard Christen. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC bioinformatics*, 7:439, October 2006. ISSN 1471-2105. doi: 10.1186/1471-2105-7-439.
- [363] R. Diaz-Uriarte and T. Garland. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from brownian motion. *Systematic Biology*, 45(1):27–47, March 1996. ISSN 1063-5157, 1076-836X. doi: 10.1093/sysbio/45.1.27. URL <https://academic.oup.com/sysbio/article-lookup/doi/10.1093/sysbio/45.1.27>.
- [364] T. Garland, P. H. Harvey, and A. R. Ives. Procedures for the analysis of comparative data Using phylogenetically independent contrasts. *Systematic Biology*, 41(1):18–32, March 1992. ISSN 1063-5157, 1076-836X. doi: 10.1093/sysbio/41.1.18. URL <https://academic.oup.com/sysbio/article-lookup/doi/10.1093/sysbio/41.1.18>.
- [365] Theodore Garland, Jr and Anthony R. Ives. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist*, 155(3):346–364, March 2000. ISSN 0003-0147, 1537-5323. doi: 10.1086/303327. URL <http://www.journals.uchicago.edu/doi/10.1086/303327>.
- [366] F. J. Rohlf. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution; International Journal of Organic Evolution*, 55(11):2143–2160, November 2001. ISSN 0014-3820.
- [367] A. Purvis and A. Rambaut. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data.

- Computer applications in the biosciences: CABIOS*, 11(3):247–251, June 1995. ISSN 0266-7061.
- [368] Kristoffer Vitting-Seerup, Bo Porse, Albin Sandelin, and Johannes Waage. spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics*, 15(1):81, 2014. ISSN 1471-2105. doi: 10.1186/1471-2105-15-81. URL <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-81>.
- [369] L. Sian Gramates, Steven J. Marygold, Gilberto dos Santos, Jose-Maria Urbano, Giulia Antonazzo, Beverley B. Matthews, Alix J. Rey, Christopher J. Tabone, Madeline A. Crosby, David B. Emmert, Kathleen Falls, Joshua L. Goodman, Yanhui Hu, Laura Ponting, Andrew J. Schroeder, Victor B. Strelets, Jim Thurmond, Pinglei Zhou, and the FlyBase Consortium. FlyBase at 25: looking to the future. *Nucleic Acids Research*, 45(D1):D663–D671, January 2017. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkw1016. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1016>.
- [370] L. Keller and N. Perrin. Quantifying the level of eusociality. *Proceedings of the Royal Society B: Biological Sciences*, 260(1359):311–315, June 1995. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.1995.0097. URL <http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.1995.0097>.
- [371] C. W. Sugnet, W. J. Kent, M. Ares, and D. Haussler. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 66–77, 2004. ISSN 2335-6936.
- [372] Eddo Kim, Alon Magen, and Gil Ast. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Research*, 35(1):125–131, January 2007. ISSN

- 1362-4962, 0305-1048. doi: 10.1093/nar/gkl924. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkl924>.
- [373] A. Burt and R. L. Trivers. *Genes in conflict: the biology of selfish genetic elements*. Harvard University Press, Cambridge, MA., 2006.
- [374] R. Duffie, S. Ajjan, M. V. Greenberg, N. Zamudio, M. Escamilla del Arenal, J. Iranzo, I. Okamoto, S. Barbaux, P. Fauque, and D. Bourc'his. The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals. *Genes & Development*, 28(5):463–478, March 2014. ISSN 0890-9369. doi: 10.1101/gad.232058.113. URL <http://genesdev.cshlp.org/cgi/doi/10.1101/gad.232058.113>.
- [375] Maxim V C Greenberg, Juliane Glaser, Mt Borsos, Fatima El Marjou, Marius Walter, Aurlie Teissandier, and Dborah Bourc'his. Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. *Nature Genetics*, November 2016. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3718. URL <http://www.nature.com/doi/10.1038/ng.3718>.
- [376] Ian M. Morison, Joshua P. Ramsay, and Hamish G. Spencer. A census of mammalian imprinting. *Trends in Genetics: TIG*, 21(8):457–465, August 2005. ISSN 0168-9525. doi: 10.1016/j.tig.2005.06.008.
- [377] Michael T. Raissig, Clia Baroux, and Ueli Grossniklaus. Regulation and flexibility of genomic imprinting during seed development. *The Plant Cell*, 23(1):16–26, January 2011. ISSN 1532-298X. doi: 10.1105/tpc.110.081018.
- [378] V. Lloyd. Parental imprinting in *Drosophila*. *Genetica*, 109(1-2):35–44, 2000. ISSN 0016-6707.
- [379] R. A. McGowan and C. C. Martin. DNA methylation and genome imprinting in the zebrafish, *Danio rerio*: some evolutionary ramifications. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire*, 75(5):499–506, 1997. ISSN 0829-8211.

- [380] Christopher J. Bean, Christine E. Schaner, and William G. Kelly. Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. *Nature Genetics*, 36(1):100–105, January 2004. ISSN 1061-4036. doi: 10.1038/ng1283.
- [381] Oluwaseun Ogunwuyi, Ankur Upadhyay, Simeon K Adesina, Reema Puri, Tasha M Foreman, Belinda R Hauser, Juanita Cox, Eric Afoakwah, Andre Porter, Emma Annan, Cheu Manka, Denloye Olatilewa, Bethtrice Thompson, Peter R Kibanyi, Kimberly Miller, and Hemayet Ullah. Genetic imprinting: comparative analysis between plants and mammals. *Plant Tissue Culture and Biotechnology*, 26(2):267, December 2016. ISSN 1818-8745, 1817-3721. doi: 10.3329/ptcb.v26i2.30576. URL <http://www.banglajol.info/index.php/PTCB/article/view/30576>.
- [382] A Mochizuki, Y Takeda, and Y Iwasa. The evolution of genomic imprinting. *Genetics*, 144(3):1283–1295, 1996.
- [383] Sarah P. Otto and Aleeza C. Gerstein. The evolution of haploidy and diploidy. *Current Biology*, 18(24):R1121–R1124, December 2008. ISSN 09609822. doi: 10.1016/j.cub.2008.09.039. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982208012682>.
- [384] M M Patten, L Ross, J P Curley, D C Queller, R Bonduriansky, and J B Wolf. The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity*, 113(2):119–128, August 2014. ISSN 0018-067X, 1365-2540. doi: 10.1038/hdy.2014.29. URL <http://www.nature.com/doifinder/10.1038/hdy.2014.29>.
- [385] T. Day. Intralocus sexual conflict can drive the evolution of genomic imprinting. *Genetics*, 167(4):1537–1546, August 2004. ISSN 0016-6731. doi: 10.1534/genetics.103.026211. URL <http://www.genetics.org/cgi/doi/10.1534/genetics.103.026211>.

- [386] Jason B Wolf and Reinmar Hager. A maternal - offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biology*, 4(12):e380, November 2006. ISSN 1545-7885. doi: 10.1371/journal.pbio.0040380. URL <http://dx.plos.org/10.1371/journal.pbio.0040380>.
- [387] Francisco Ubeda and Andy Gardner. A model for genomic imprinting in the social brain: adults. *Evolution*, 65(2):462–475, February 2011. ISSN 00143820. doi: 10.1111/j.1558-5646.2010.01115.x. URL <http://doi.wiley.com/10.1111/j.1558-5646.2010.01115.x>.
- [388] Yaniv Brandvain, Jeremy Van Cleve, Francisco Ubeda, and Jon F. Wilkins. Demography, kinship, and the evolving theory of genomic imprinting. *Trends in Genetics*, 27(7):251–257, July 2011. ISSN 01689525. doi: 10.1016/j.tig.2011.04.005. URL <http://linkinghub.elsevier.com/retrieve/pii/S0168952511000618>.
- [389] Tomas Babak, Brian DeVeale, Emily K Tsang, Yiqi Zhou, Xin Li, Kevin S Smith, Kim R Kukurba, Rui Zhang, Jin Billy Li, Derek van der Kooy, Stephen B Montgomery, and Hunter B Fraser. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nature Genetics*, 47(5):544–549, April 2015. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3274. URL <http://www.nature.com/doifinder/10.1038/ng.3274>.
- [390] W. D. Hamilton. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, 7(1):1–16, 1964. URL [http://books.google.com/books?hl=en&lr=&id=3ULyAgAAQBAJ&oi=fnd&pg=PA161&dq=%22very+few+exceptions,+the+only+parts+of+the+theory+of+natural%22+%22offspring.+In+certain+circumstances+an+individual+may+leave%22+%22W.+D.%22+%22at+one+quarter%3B+and+so%22+&ots=jZe6AUlwvi&sig=itvHCME8\\_1jGloUHA\\_opJr6vxVc](http://books.google.com/books?hl=en&lr=&id=3ULyAgAAQBAJ&oi=fnd&pg=PA161&dq=%22very+few+exceptions,+the+only+parts+of+the+theory+of+natural%22+%22offspring.+In+certain+circumstances+an+individual+may+leave%22+%22W.+D.%22+%22at+one+quarter%3B+and+so%22+&ots=jZe6AUlwvi&sig=itvHCME8_1jGloUHA_opJr6vxVc).

- [391] David A. Galbraith, Soojin V. Yi, and Christina M. Grozinger. Evaluation of possible proximate mechanisms underlying the kinship theory of intragenomic conflict in social insects. *Integrative and Comparative Biology*, 56(6): 1206–1214, December 2016. ISSN 1540-7063, 1557-7023. doi: 10.1093/icb/icw111. URL <https://academic.oup.com/icb/article-lookup/doi/10.1093/icb/icw111>.
- [392] Kirsten R. McEwen and Anne C. Ferguson-Smith. Genomic Imprinting: A Model for Roles of Histone Modifications in Epigenetic Control. In Anne C. Ferguson-Smith, John M. Greally, and Robert A. Martienssen, editors, *Epigenomics*, pages 235–258. Springer Netherlands, Dordrecht, 2009. ISBN 978-1-4020-9186-5 978-1-4020-9187-2. URL [http://link.springer.com/10.1007/978-1-4020-9187-2\\_14](http://link.springer.com/10.1007/978-1-4020-9187-2_14). DOI: 10.1007/978-1-4020-9187-2\_14.
- [393] Azusa Inoue, Lan Jiang, Falong Lu, Tsukasa Suzuki, and Yi Zhang. Maternal H3k27me3 controls DNA methylation-independent imprinting. *Nature*, 547(7664):419–424, July 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature23262. URL <http://www.nature.com/doifinder/10.1038/nature23262>.
- [394] Heidi Ledford. CRISPR: gene editing is just the beginning. *Nature*, 531(7593): 156–159, March 2016. ISSN 0028-0836, 1476-4687. doi: 10.1038/531156a. URL <http://www.nature.com/doifinder/10.1038/531156a>.
- [395] Hiroki Kohno, Shota Suenami, Hideaki Takeuchi, Tetsuhiko Sasaki, and Takeo Kubo. Production of Knockout Mutants by CRISPR/Cas9 in the European Honeybee, *Apis mellifera* L. *Zoological Science*, 33(5):505–512, October 2016. ISSN 0289-0003. doi: 10.2108/zs160043.
- [396] Ming Li, Lauren Yun Cook Au, Deema Douglass, Abigail Chong, Bradley J. White, Patrick M. Ferree, and Omar S. Akbari. Generation of heritable

- germline mutations in the jewel wasp *Nasonia vitripennis* using CRISPR/-Cas9. *Scientific Reports*, 7(1):901, April 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-00990-3.
- [397] Waring Tribble, Leonora Olivos-Cisneros, Sean K. McKenzie, Jonathan Saragosti, Ni-Chen Chang, Benjamin J. Matthews, Peter R. Oxley, and Daniel J.C. Kronauer. orco mutagenesis causes loss of antennal lobe glomeruli and impaired social behavior in ants. *Cell*, 170(4):727–735.e10, August 2017. ISSN 00928674. doi: 10.1016/j.cell.2017.07.001. URL <http://linkinghub.elsevier.com/retrieve/pii/S0092867417307729>.
- [398] Peter Cheung and Priscilla Lau. Epigenetic regulation by histone methylation and histone variants. *Molecular Endocrinology (Baltimore, Md.)*, 19(3):563–573, March 2005. ISSN 0888-8809. doi: 10.1210/me.2004-0496.
- [399] Miguel R. Branco, Gabriella Ficz, and Wolf Reik. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nature Reviews. Genetics*, 13(1):7–13, November 2011. ISSN 1471-0064. doi: 10.1038/nrg3080.