

# Alternative splicing associated with phenotypic plasticity in the bumble bee *Bombus terrestris*

---

J. PRICE <sup>1</sup>, M.C. HARRISON <sup>2</sup>, R.L. HAMMOND <sup>3</sup>, S. ADAMS <sup>1</sup>, J.F. GUTIERREZ-MARCOS <sup>1</sup> & E.B. MALLON <sup>3\*</sup>

1) School of Life Sciences  
Gibbet Hill Campus  
University of Warwick  
Coventry, CV4 7AL  
United Kingdom

2) Institute for Evolution and Biodiversity  
University of Münster  
Hüfferstraße 1  
48149 Münster  
Germany

3) Department of Genetics and Genome Biology  
University of Leicester  
University Road  
Leicester, LE1 7RH  
Phone: ++ (0)116 252 3488  
Fax: +44 (0)116 252 3330  
E-mail: ebm3@le.ac.uk

**Keywords:** worker, queen, isoform, eusociality, transcriptomics, reproductive division of labour

Running title: "Alternative splicing in the bumble bee."

## Abstract

Phenotypic plasticity is when one genome can produce more than one phenotype. The caste system found in many social insects is an important example of plasticity. Several studies have examined gene expression in social insect developmental and caste differences. Changes in gene expression, however, are not the only source of phenotypic plasticity. Here we investigate the role of alternative splicing in the buff tailed bumble bee *Bombus terrestris*. We found that 5458 genes in *Bombus terrestris* (40%) express more than one isoform. Larvae have the lowest level of splicing events, followed by adults and then pupae. We found that when an isoform is expressed in a given caste in the larval stage, it tends to be expressed in all castes at the larval stage. The same is true at the pupal stage. However, we see more complicated interactions between the adult castes with reproductive females showing different isoform expression compared to non-reproductive females and male adults showing the most distinct patterns. We found 455 isoform switching genes, that is genes, where one developmental stage, sex or caste uses a specific isoform and another type uses a different isoform. Among genes displaying isoform switching are some involved in the ecdysteroid pathway, an important system in insect behaviour.

# 1 Introduction

2 A fundamental aim of evolutionary biology is to understand the link between genotype, phenotype and  
3 fitness (de Visser and Krug, 2014). A powerful way to understand this is to investigate how a single  
4 genome can give rise to multiple phenotypes. This question can be approached at the cellular level, for  
5 example, how a single genome can make both a skin cell and a blood cell, but also at the societal level,  
6 for example, how a single genome can make both the queen and worker castes in a eusocial insect colony.  
7 Indeed, recent models suggest that these different levels could even be functionally similar (Patalano  
8 *et al.*, 2012).

9 In eusocial insect colonies, this differentiation between castes is fundamental, with queens carrying  
10 out most of the reproduction while workers carry out other tasks required of the colony e.g. brood  
11 care, foraging and nest maintenance. Queens are also generally bigger, live longer and have a host of  
12 behavioural and physiological specialisations compared to workers. These distinct morphological and  
13 behavioural castes are usually alternative expressions of the same genome. We previously found that the  
14 main difference in gene expression, in the buff-tailed bumble bee *Bombus terrestris*, was between castes at  
15 the adult stage rather than at larval or pupal stages (Harrison *et al.*, 2015). A striking difference between  
16 *B. terrestris* and higher eusocial hymenopterans is that reproductive worker bees have gene expression  
17 profiles much more similar to queens than to their non-reproductive sister workers (Harrison *et al.*, 2015).

18 Changes in gene expression are not the only way a single genome can give rise to multiple phenotypes.  
19 For example, the difference between head lice (harmless) and body lice, (disease vector) is caused by the  
20 alternative splicing of an identical genome not by differential gene expression (Tovar-Corona *et al.*, 2015).  
21 Alternative splicing results in a single gene coding for multiple processed messenger RNA (mRNA),  
22 and potentially multiple proteins with distinct functions. Genome-wide analyses suggest alternative  
23 splicing as a source of phenotypic variation in eusocial insects, where a single genome must encode for  
24 numerous castes (honeybees (Foret *et al.*, 2012; Li-Byarlay *et al.*, 2013), possibly ants (Bonasio *et al.*,  
25 2012) and termites (Terrapon *et al.*, 2014)). Bumble bees, as well as being one of the most ecologically and  
26 economically important pollinators (Woodard *et al.*, 2015), are an interesting taxon to study the role of  
27 alternative splicing in eusociality as they possess both highly eusocial and more primitive characteristics  
28 (Harrison *et al.*, 2015).

29 The diversity of forms and behaviours within a eusocial insect species is much greater than that found  
30 within solitary insect species. Yet eusocial insects have gene numbers within the range for solitary insects  
31 (Simola *et al.*, 2013). Tovar-Corona *et al.* 2015 recently found the rate of alternative splicing in eight  
32 arthropod species ranging from 23% to 44%. Do bumble bees use alternative splicing more than other  
33 studied, non-social insects, in order to increase phenotypic diversity?

34 Here, using RNA-seq, we ask how alternative splicing is associated with caste, sex and development  
35 within the buff-tailed bumble bee, *Bombus terrestris*. We compare splicing patterns of larval (male  
36 versus workers), pupal (male versus workers) and adult (male versus non-reproductive worker versus  
37 reproductive worker versus queen) stages. Using the results of this analysis we answer a number of  
38 questions. Do bumble bees have increased rates of alternative splicing compared to studied non-social  
39 insects? Do splicing patterns follow those of gene expression with relatively fewer differences between  
40 juvenile stages compared to adults? Are reproductive workers' splicing patterns more similar to queens'  
41 or non-reproductive workers'? Do different castes, sexes or developmental stages use different isoforms  
42 of the same genes?

## 43 **Methods**

### 44 *Colony sampling, RNA libraries and sequencing*

45 Data for this study were obtained from Harrison *et al.* (2015). Reads were downloaded from the ENA  
46 under accession no. PRJEB9366. This contained a total of 469.3 million reads from 27 libraries. These  
47 libraries are made up of larval, pupal and adults stages of males, queens, workers (reproductive and  
48 non-reproductive). Whole bodies were used for sampling for two reasons. First, we had no prior as-  
49 sumptions regarding the tissues, within which genes would be alternatively spliced between castes, sexes  
50 and developmental stages. Second, this would allow us to detect as many alternatively spliced genes as  
51 possible across all comparisons. Most sample types are replicated over three colonies except for adult  
52 nonreproductive workers (two colonies), reproductive workers (four colonies) and larval,pupal and virgin  
53 queens (one colony). Sample production is described fully by Harrison *et al.* (2015). Due to the lack  
54 of replicates, queen larvae, pupae and virgin queen were not analysed in our current alternative splicing  
55 study.

### 56 *Read processing and analysis*

57 Raw reads were trimmed for adaptor sequence and low quality at the 5' and 3' end, and across a slid-  
58 ing window throughout the reads (PHRED >24) using Trimmomatic (v0.33) (Bolger *et al.*, 2014), then  
59 visualised using FastQC (v0.11.5). Reads were aligned to the Bter\_1.0 genome (Refseq accession no.  
60 GCF\_000214255.1 (Sadd *et al.*, 2015)) using Tophat (v2.1.0) (Trapnell *et al.*, 2009), with an average  
61 of 93.3% efficiency (82.8%-95.8%). After mapping, the protocol for the tuxedo package was followed  
62 as outlined in Trapnell *et al.* (2012). In brief, cufflinks (v2.2.1) was performed on the alignments from  
63 tophat, then cuffmerge (v1.0.0) was used to create a consensus transcriptome. Finally cuffdiff (v2.2.1)  
64 was used to quantify isoform expression and identify statistically significant changes in isoform expres-  
65 sion between the samples. For this study we have analysed isoforms at the CD level because they  
66 represent a change in the protein translated. Further downstream analysis of these isoforms was then  
67 performed in R 3.2.3 (core Team, 2016) using the cummeRbund (v2.16.0) package and custom R scripts  
68 (<https://dx.doi.org/10.6084/m9.figshare.3201355>). rMATs (v3.2.5) was run with the alignments from  
69 tophat to identify the types of alternative splicing present (Shen *et al.*, 2014)

### 70 *Isoform switching and Clustering*

71 Relative isoform contributions for each gene were tested for differences pairwise between the samples util-  
72 ising the jensen-shannon divergence test implemented by the cuffdiff program (FDR < 0.01). Allometric

73 differences in the amount of tissue in different castes, sexes or reproductive status could produce signif-  
74 icant differences in isoform expression if judged solely on statistical testing. We therefore only consider  
75 isoforms with at least a log2-fold change of one in order to reduce the false-positive effects of tissue scal-  
76 ing (Montgomery and Mank, 2016). To identify those genes that display biologically significant isoform  
77 switching, the isoforms from significant genes identified by the jensen-shannon divergence test were sub-  
78 ject to K-means clustering (Hartigan and Wong, 1979) (k=30-150 step of 10, Iterations = 500) based on  
79 their scaled relative abundance. The most distinct abundance patterns along with the fewest duplicated  
80 clusters were found at K of 110. Each cluster was then plotted separately to visualize isoform expression.  
81 As described by Aghamirzaie *et al.* 2015, clusters were then manually combined into super-clusters where  
82 they had similar relative abundance patterns.

### 83 *Gene ontology (GO) analysis*

84 We extracted the nucleotide sequence for all isoforms found in each one of the chosen super-clusters and  
85 searched for any matching sequence on NCBI using BLASTx (Altschul *et al.*, 1990) with an E-value  
86 cutoff of 0.001. Using Blast2Go (Gotz *et al.*, 2008), we identified gene ontology (GO) terms associated  
87 with these loci. We carried out an enrichment analysis (Fisher exact test) using a custom R script  
88 (<https://dx.doi.org/10.6084/m9.figshare.3201355.v1>) on each of these lists of super-cluster associated GO  
89 terms. This identified GO terms that are overrepresented (FDR < 0.1) relative to the entire transcriptome  
90 (<https://dx.doi.org/10.6084/m9.figshare.3458828.v1>).

## 91 Results

### 92 *Isoform, CD and gene level quantification*

93 We measured expression level in terms of fragments per kilobase of exon per million reads mapped  
94 (FPKM) of all isoforms present in published RNAseq libraries of several castes, sexes and developmental  
95 stages of *Bombus terrestris* (Figure 1). After removing genes with FPKM  $< 1$  in all libraries there are  
96 13548 distinct loci with expression in at least one sample. Alternative splicing produces different mRNA  
97 (isoforms) from the same gene. A subset of these isoforms will differ in their conserved protein domains  
98 (CD isoforms). At the isoform level, 8090 of the loci produce only one isoform, the remaining 5458 (40%)  
99 genes express 22786 isoforms (FPKM  $> 0.5$  in at least one library), averaging 4.2 isoforms per gene. At  
100 the CD level 2833 (20.9%) show expression of more than one CD isoform in at least one sample, totalling  
101 7726 CD isoforms (2.7 CD isoforms/gene). To avoid problems with assembly and annotation previously  
102 well described in alternative splicing studies (Hiller *et al.*, 2009; Jiang and Wong, 2009), we concentrated  
103 on the 2833 genes expressing more than one CD isoform (see Table 1).

### 104 *CD isoform expression varies mainly between developmental stages, but also sex and caste*

105 Expression (FPKM) for each of the 7726 CD isoforms is highly correlated between replicates of each  
106 biological type (mated queen, reproductive adult worker, non-reproductive adult worker, worker pupae,  
107 worker larvae, male adult, male pupae and male larvae) (Figure 1a) indicating that the patterns we  
108 observe here are of biological origin.

109 Principle component analysis of the 7726 CD isoform's expression (Figure 1b) shows that there is  
110 variance in the expression of these isoforms which is associated primarily with developmental stage but  
111 also sex and caste. PC1 principally describes variance in the isoform expression of different developmental  
112 stages by separating adult and pupal replicates on the left and larval replicates on the right, this difference  
113 explains 36% of the variance. PC2 is explained also by differences in developmental stages but this time  
114 separating the pupae samples from the rest (22% variance). PC3 displays the variance that exists between  
115 the sexes in the adult stages by dividing male adults and the female adult samples but also separating  
116 the female adult samples (13% variance).

117 Alternative splicing events can come in different forms; exon skipping, mutually exclusive exons,  
118 alternative 5' splice junction, alternative 3' splice junction and intron retention (Black, 2003). The  
119 proportions of the splicing event forms stay relatively constant between samples with only a small increase  
120 in skipped exon splice events for non-reproductive workers (see Figure S1). We observed differences in  
121 the numbers of spliced genes and the number of isoforms expressed in our samples (Table 1). Male and



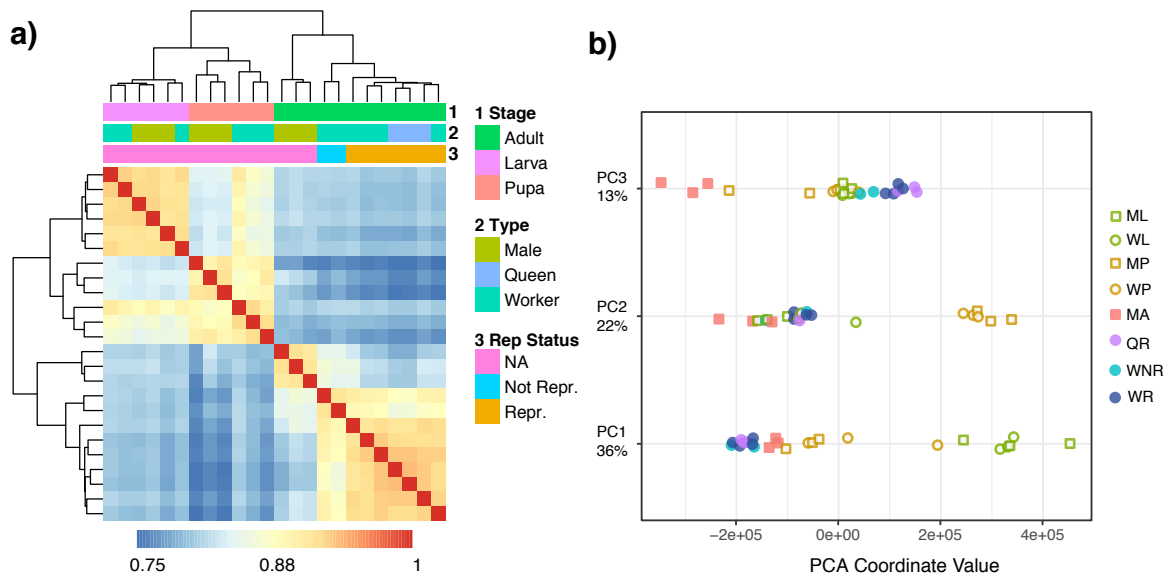


Figure 1: Global splicing analysis **a**. Correlogram of Pearson correlations of expression (FPKM) between replicates for the 7726 coding isoforms from multiple isoform loci. Clustering is produced based on Euclidean distances. This clustering separates replicates into stage, sex and reproductive status (coloured bars on top). **b**. A principal component analysis (PCA) of the 7726 isoforms from multiple isoform loci. QR = Queen Reproductive, WR = adult Worker Reproductive, WNR = adult Worker Non-Reproductive, WP = Worker Pupae, WL = Worker Larvae, MA = Male Adult, MP = Male Pupae, ML = Male Larvae.

Table 1: Total numbers of expressed features in the eight biological types examined

Feature	QR	MA	ML	MP	WR	WNR	WL	WP
Genes	9954	10043	9685	10756	10249	10418	9762	9969
Genes With Multiple Isoforms	5233	5295	5226	5338	5282	5311	5250	5265
Isoforms from Genes with multiple isoforms	16210	16840	16048	16844	16909	17090	16071	16274
Genes with multiple CD isoforms	2792	2801	2760	2800	2812	2818	2772	2782
CD isoforms from genes with multiple CD isoforms	6265	6326	6092	6349	6446	6472	6107	6208

122 worker larvae have the lowest number of transcripts expressed in each category (genes, multiple isoform  
123 genes, multiple isoforms, multiple CD genes, multiple CD isoforms). The pupae shows a discrepancy  
124 between the sexes with the male pupae expressing more transcripts of each type than the worker pupae.  
125 Male pupae express the most genes and the most multiple isoform genes but adult workers (WR and  
126 WNR) express the most isoforms, multiple CD isoform genes and the most CD isoforms.

### 127 *455 genes with varied functions display significant isoform switching*

128 The 2833 multiple isoform loci equate to the total amount of alternative splicing at the CD level observed  
129 in our samples but not all of these loci express different proportions of isoforms in different samples, that  
130 is, isoform switching. To find these loci, we tested the 2833 multiple isoform loci pairwise for differences

131 in their relative isoform abundance between samples ( $FDR < 0.01$ ). 455 genes, 16% of the multiple  
132 isoform loci (2833) display significant isoform switching between samples. These 1303 CD isoforms from  
133 455 genes were analysed further to identify biologically significant isoform switching events.

134 We used K-means clustering to identify clusters of isoforms displaying similar expression patterns  
135 across biological types. 297 (from 1303) were not clustered because they have small changes in their  
136 relative abundance between samples (range  $< 0.2$ ) and would be misrepresented when their isoform  
137 abundance was scaled before clustering. K-means clustering ( $K = 110$ ) of the 1006 remaining isoforms  
138 achieved separation of most isoforms by their relative abundance (Figure S2). By manually combining  
139 clusters with the same pattern of relative abundance (Figure 2) we identified 26 super-clusters that show  
140 a high degree of sample-specific expression (658 isoforms). The remaining 348 isoforms showed cryptic  
141 abundances and could not be assigned to a super-cluster, these isoforms are included in the variable cluster  
142 (Figure 2). 22 of the 26 super-clusters form antagonistic pairs, where both super-clusters represent the  
143 high and low abundance isoforms for that particular sample or group of samples. Most super-clusters  
144 (491 isoforms) show preferential high or low expression in samples(s) from one developmental stage and  
145 four clusters (167 isoforms) have preferential expression in samples from more than one developmental  
146 stage. The majority of isoforms show adult sample(s) specific expression (374) with 105 for pupae and  
147 76 for larvae (supplementary data: isoform\_cluster\_database). Of the 455 isoform switching genes; 328  
148 have multiple isoforms in multiple super-clusters, 39 have isoforms solely in the variable cluster, 25 genes  
149 have one or zero isoforms in a super-cluster and 63 are uncharacterised loci. From the 328 genes that  
150 have multiple isoforms in multiple super-clusters 80 genes have their isoforms clustered in antagonistic  
151 pairs.

152 For each super-cluster and groups of super-clusters relating to each developmental stage we tested  
153 for enriched gene ontology (GO) terms against their background value in the transcriptome. No discrete  
154 cluster had significant enriched GO terms at the  $FDR < 0.1$ . However, twenty three GO terms  
155 were associated with isoform switching genes found in adults (supplementary data: cluster\_GO\_terms).  
156 Two of these, steroid hormone mediated signalling pathway (GO:0043401) and steroid hormone receptor  
157 activity (GO0003707) relate to important functions in social insect biology. A number of isoform switch-  
158 ing genes are associated with these terms, including; ecdysone inducible protein 74EF (LOC100651988),  
159 E75 (LOC100644185), E78 (LOC100646780), ecdysone receptor (LOC100646757) and farnesyl pyrophos-  
160 phatase synthase-like (LOC00165005). The majority of these genes differ in their isoform composition  
161 between one or more adult sample. As an example of one of these genes, we show the expression,  
162 exon map and protein domains for ecdysone inducible protein 75 (Figure 3). In this gene, transcript  
163 variant X1 (XM\_003398944.2) has high levels of expression especially for non-reproductive females com-  
164 pared to transcript variant X2 (XM\_003398942.2) (see Figure 3b). Expression plots and exon maps

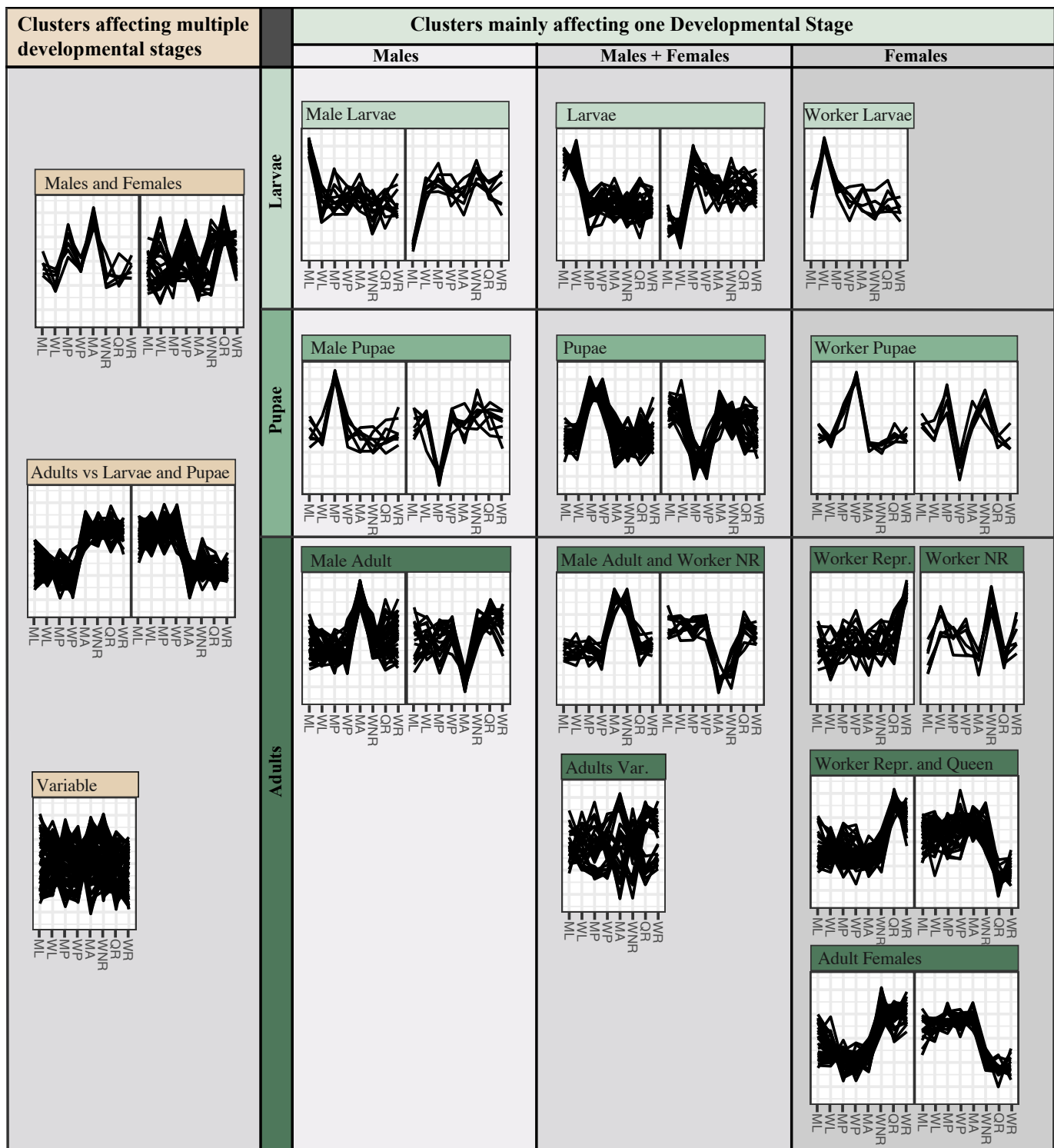


Figure 2: Super-clusters showing abundance patterns around particular biological sample types. K-means clustering for the 1006 isoforms from differentially spliced genes. Y-axis displays scaled relative abundance (-2 to 2). QR = Queen Reproductive, WR = adult Worker Reproductive, WNR = adult Worker Non-Reproductive, WP = Worker Pupae, WL = Worker Larvae, MA = Male Adult, MP = Male Pupae, ML = Male Larvae. The labels in the banners above figures identify the biological types in which the cluster is differentially expressed in. For example, Male Pupae (second column, second row) shows two clusters, one where there is a spike in expression in male pupae and one where there is a drop.

165 (10.6084/m9.figshare.5208940) are available in the supplementary data for all 455 isoform switching genes  
166 along with a database containing their annotation (supplementary data: isoform\_cluster\_database),  
167 their isoform's super-cluster membership (supplementary data: Clusters in super clusters) and GO terms  
168 (supplementary data: cluster\_GO\_terms).

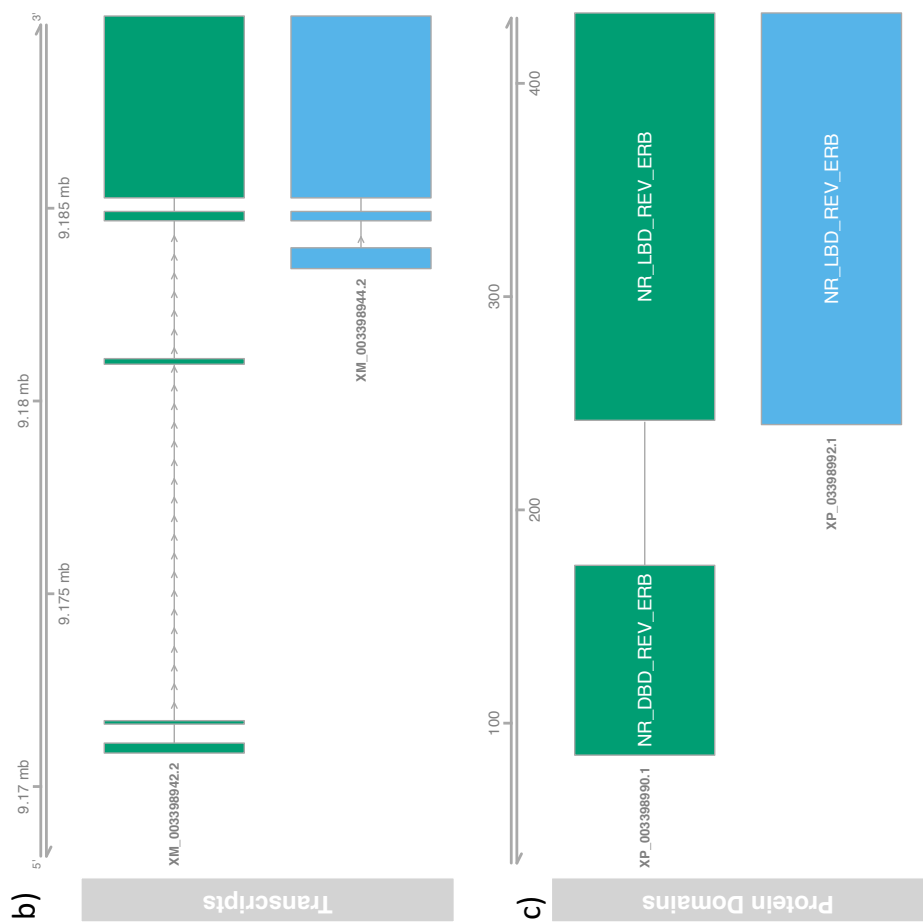


Figure 3: **a**. Expression (FPKM) of the two transcripts of *Ecdyso-inducible protein E75* over the different castes and developmental stages. Circles represent the means, the error bars are standard errors of the mean. QR = Queen Reproductive, WR = adult Worker Reproductive, WNR = adult Worker Non-Reproductive, WP = Worker Pupae, MA = Male Adult, MP = Male Pupae, ML = Male Larvae **b** Exon map of the two transcripts. **c** Protein domains encoded by the two transcripts. NR\_LBD\_REV\_ERB = ligand binding domain of REV-ERB receptors, NR\_DBD\_REV\_ERB = DNA-binding domain of REV-ERB receptor-like.

## 169 Discussion

170 Alternative splicing is a way to increase the variety of proteins produced by a given number of genes. One  
171 aim of this paper was to ask do bumble bees have increased rates of alternative splicing to account for the  
172 increased phenotypic diversity within a eusocial insect species despite them having gene numbers within  
173 the range for solitary insects (Simola *et al.*, 2013). We observed 2833 genes that have expression of more  
174 than one CD isoform in the samples examined. This corresponds to 20% (2833/13548) of the bumble  
175 bee genes producing different proteins from the same genotype. To compare *Bombus terrestris* with  
176 current literature we also looked at the number of genes that express more than one isoform regardless  
177 of the protein produced (Tovar-Corona *et al.*, 2015). Compared to the rate of alternative splicing in  
178 eight arthropod species recently studied (23% to 44%) (Tovar-Corona *et al.*, 2015), bumble bees' rate  
179 of 40% is at the higher end. However this rate is equivalent to those of two other insects (*Drosophila*  
180 *melanogaster* 42% and the pea aphid, *Acyrtosiphon pisum* 44%). Further work in other insects will be  
181 needed to confirm this, but our data indicates that an increased rate of alternative splicing is not required  
182 by bumble bees.

183 We then examined how the bumble bee uses these isoforms. We found four hundred and fifty five  
184 genes express different proportions of isoforms in different samples, that is, isoform switching. We have  
185 annotated and ranked these genes according to their super cluster membership, conserved domain changes  
186 in their isoforms, GO terms and gene expression status, providing a rich resource for future studies in  
187 this and related species. Patterns of alternative splicing follow those previously found for gene expression  
188 (Harrison *et al.*, 2015). This is not due to the effect of overall expression itself, as we only looked at  
189 genes that have expression of at least two isoforms, removing almost 80% of the observed genes (13548  
190 - 2,833). Even still, within the 455 isoform switching genes, 45% also show gene expression changes.  
191 This exemplifies how alternative splicing studies can both add to the knowledge of genes known to be  
192 differentially expressed and also find novel source of plasticity that could not be found by only examining  
193 gene level expression.

194 We found relatively little differentiation between the sexes at the larval stage when compared to  
195 adult stages (see Figures 1 and 2). These findings suggest a comparatively low number of isoforms are  
196 associated with the creation of distinct morphological sexes compared to the high number involved in  
197 distinct behaviours between adult males and females. Reproductive workers show more similarity to  
198 queens than to their fellow, but non-reproductive, workers (see horizontal dendrograms Figure 1a). We  
199 have hypothesised that this is due to the plastic nature of bumble bee workers which allow them to  
200 compete with the queen at the end of the annual colony cycle (Harrison *et al.*, 2015). Male adults are  
201 the most distinct group in terms of isoform expression (Figure 1), a finding that reflects their similarly

202 distinct behaviour, physiology and life history.

203 No GO terms were found to be significantly associated with any of the super clusters. However, twenty  
204 three GO terms were associated with isoform switching genes found in adults. Two of these, steroid hor-  
205 mone mediated signalling pathway (GO:0043401) and steroid hormone receptor activity (GO0003707)  
206 relate to the ecdysteroid pathway, an important system in insect biology (Niwa and Niwa, 2016). Ecdys-  
207 teroids are key compounds involved in ovary activation, regulating agonistic behaviour and establishing  
208 the dominance hierarchy in bumble bee workers and queens (Geva *et al.*, 2005).

209 In the results, as an example of an isoform switching gene, we expanded our analysis for one ecdysone  
210 related gene (Ecdysone-inducible protein E75 (LOC100644185)). Ecdysone-inducible protein E75 is a  
211 transcription factor in the ecdysone cascade (Li *et al.*, 2015). Transcript variant X1 (XM\_003398944.2)  
212 has high levels of expression especially for non-reproductive females compared to transcript variant X2  
213 (XM\_003398942.2) (see Figure 3b). In *Drosophila melanogaster*, different isoforms of E75 have oppo-  
214 site roles in controlling female reproduction with, for example, DmE75A inducing apoptosis in the egg  
215 chamber, while DmE75B inhibits DmE75A function allowing egg development (Terashima and Bownes,  
216 2006). The difference between the transcripts is a number of exons encoding for binding domains of the  
217 nuclear receptor REV-ERB (Figure 3c). REV-ERB receptors are transcriptional regulators belonging  
218 to the nuclear receptor superfamily. They regulate a number of physiological functions including the  
219 circadian rhythm, lipid metabolism, and cellular differentiation. The ligand binding domain (LBD) of  
220 REV-ERB receptor-like, found in both isoforms, is responsible for coactivator interaction with steroid  
221 hormones (Raghuram *et al.*, 2007). The DNA-binding domain (DBD), found only in transcript variant  
222 X2, interacts with specific DNA sites upstream of the target gene and modulates the rate of transcrip-  
223 tional initiation. This truncation of a binding domain is a very simple way for a gene to go from  
224 being expressed specifically to a more global pattern of expression. We found this truncation in two  
225 other genes we examined in depth, the CXXC zinc finger domain in Methylcytosine dioxygenase TET2  
226 (LOC100642293 Figure S3) and the RNA recognition motif in Insulin-like growth factor 2 mRNA-binding  
227 protein 1 (LOC100647990 Figure S4). Further study of our data set and others could test if this is a  
228 common phenomenon.

229 We have shown that alternative splicing of genes is important for bumble bee biology. How is this  
230 alternative splicing is controlled? Alternative splicing in some social insect species has been shown to be  
231 affected by DNA methylation (Bonasio *et al.*, 2012; Foret *et al.*, 2012). Methylation has also been shown  
232 to play a role in social insect biology (Yan *et al.*, 2015) including bumble bee reproductive division of  
233 labour (Amarasinghe *et al.*, 2014). Together these observations suggest that a fruitful area of research  
234 for social insect biologists would be in understanding the role of methylation in controlling alternative  
235 splicing.

## 236 Acknowledgements

237 JP was supported by a a PhD research grant from the BBSRC. MCH was supported by a PhD research  
238 grant from the Natural Environment Research Council (NERC). EBM was supported by research grant  
239 NE/N010019/1 from the Natural Environment Research Council, UK. The raw data used in this study was  
240 funded by a NERC Biomolecular Analysis Facility research grant (NBAF 829). Illumina library prepara-  
241 tion and sequencing were carried out by Edinburgh Genomics, The University of Edinburgh. Edinburgh  
242 Genomics is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1)  
243 and BBSRC (BB/J004243/1).



## 244 References

- 245 Aghamirzaie, D., Batra, D., Heath, L. S., Schneider, A., Grene, R., and Collakova, E. 2015.  
246 Transcriptome-wide functional characterization reveals novel relationships among differentially ex-  
247 pressed transcripts in developing soybean embryos. *BMC Genomics*, 16.
- 248 Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. 1990. Basic local alignment search tool.  
249 *Journal of Molecular Biology*, 215(3): 403–410.
- 250 Amarasinghe, H. E., Clayton, C. I., and Mallon, E. B. 2014. Methylation and worker reproduction in the  
251 bumble-bee (*Bombus terrestris*). *Proceedings of the Royal Society B: Biological Sciences*, 281(1780):  
252 20132502.
- 253 Black, D. L. 2003. Mechanisms of Alternative Pre-Messenger RNA Splicing.
- 254 Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence  
255 data. *Bioinformatics*, 30(15): 2114–2120.
- 256 Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., Ding, Y.,  
257 Jin, Z., Shen, S. S., Wang, Z., Wang, W., Wang, J., Berger, S. L., Liebig, J., Zhang, G., and Reinberg,  
258 D. 2012. Genome-wide and Caste-Specific DNA Methylomes of the Ants *Camponotus floridanus* and  
259 *Harpegnathos saltator*. *Current Biology*, 22(19): 1755–1764.
- 260 core Team, R. 2016. R: A language and environment for statistical computing.
- 261 de Visser, J. A. G. M. and Krug, J. 2014. Empirical fitness landscapes and the predictability of evolution.  
262 *Nature Reviews Genetics*, 15(7): 480–490.
- 263 Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E., and Maleszka, R.  
264 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey  
265 bees. *Proceedings of the National Academy of Sciences*, 109(13): 4968–4973.
- 266 Geva, S., Hartfelder, K., and Bloch, G. 2005. Reproductive division of labor, dominance, and ecdysteroid  
267 levels in hemolymph and ovary of the bumble bee *Bombus terrestris*. *Journal of Insect Physiology*,  
268 51(7): 811–823.
- 269 Gotz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M.,  
270 Talón, M., Dopazo, J., and Conesa, A. 2008. High-throughput functional annotation and data mining  
271 with the Blast2go suite. *Nucleic Acids Research*, 36(10): 3420–3435.
- 272 Harrison, M. C., Hammond, R. L., and Mallon, E. B. 2015. Reproductive workers show queen-like gene  
273 expression in an intermediately eusocial insect, the buff-tailed bumble bee *Bombus terrestris*. *Molecular  
274 Ecology*, 24: 121–129.
- 275 Hartigan, J. A. and Wong, M. A. 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal  
276 of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1): 100–108.
- 277 Hiller, D., Jiang, H., Xu, W., and Wong, W. H. 2009. Identifiability of isoform deconvolution from  
278 junction arrays and RNA-Seq. *Bioinformatics (Oxford, England)*, 25(23): 3056–3059.
- 279 Jiang, H. and Wong, W. H. 2009. Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics  
280 (Oxford, England)*, 25(8): 1026–1032.
- 281 Li, K., Guo, E., Hossain, M. S., Li, Q., Cao, Y., Tian, L., Deng, X., and Li, S. 2015. Bombyx E75 isoforms  
282 display stage- and tissue-specific responses to 20-hydroxyecdysone. *Scientific Reports*, 5: 12114.
- 283 Li-Byarlay, H., Li, Y., Stroud, H., Feng, S., Newman, T. C., Kaneda, M., Hou, K. K., Worley, K. C.,  
284 Elsik, C. G., Wickline, S. A., Jacobsen, S. E., Ma, J., and Robinson, G. E. 2013. RNA interference  
285 knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proceedings  
286 of the National Academy of Sciences*, 110(31): 12750–12755.
- 287 Montgomery, S. H. and Mank, J. E. 2016. Inferring regulatory change from gene expression: the con-  
288 founding effects of tissue scaling. *Molecular Ecology*, 25(20): 5114–5128.

- 289 Niwa, Y. S. and Niwa, R. 2016. Transcriptional regulation of insect steroid hormone biosynthesis and  
 290 its role in controlling timing of molting and metamorphosis. *Development, Growth & Differentiation*,  
 291 58(1): 94–105.
- 292 Patalano, S., Hore, T., Reik, W., and Sumner, S. 2012. Shifting behaviour: Epigenetic reprogramming  
 293 in eusocial insects. *Current Opinion in Cell Biology*, 24(3): 367–373.
- 294 Raghuram, S., Stayrook, K. R., Huang, P., Rogers, P. M., Nosie, A. K., McClure, D. B., Burris, L. L.,  
 295 Khorasanizadeh, S., Burris, T. P., and Rastinejad, F. 2007. Identification of heme as the ligand for the  
 296 orphan nuclear receptors REV-ERBalpha and REV-ERBbeta. *Nature Structural & Molecular Biology*,  
 297 14(12): 1207–1213.
- 298 Sadd, B. M., Barribeau, S. M., Bloch, G., Graaf, D. C. d., Dearden, P., Elsik, C. G., Gadau, J., Grim-  
 299 melikhuijzen, C. J., Hasselmann, M., Lozier, J. D., Robertson, H. M., Smagghe, G., Stolle, E., Vaeren-  
 300 bergh, M. V., Waterhouse, R. M., Bornberg-Bauer, E., Klasberg, S., Bennett, A. K., Câmara, F.,  
 301 Guigó, R., Hoff, K., Mariotti, M., Munoz-Torres, M., Murphy, T., Santesmasses, D., Amdam, G. V.,  
 302 Beckers, M., Beye, M., Biewer, M., Bitondi, M. M., Blaxter, M. L., Bourke, A. F., Brown, M. J.,  
 303 Buechel, S. D., Cameron, R., Cappelle, K., Carolan, J. C., Christiaens, O., Ciborowski, K. L., Clarke,  
 304 D. F., Colgan, T. J., Collins, D. H., Cridge, A. G., Dalmay, T., Dreier, S., Plessis, L. d., Duncan, E.,  
 305 Erler, S., Evans, J., Falcon, T., Flores, K., Freitas, F. C., Fuchikawa, T., Gempe, T., Hartfelder, K.,  
 306 Hauser, F., Helbing, S., Humann, F. C., Irvine, F., Jermini, L. S., Johnson, C. E., Johnson, R. M.,  
 307 Jones, A. K., Kadowaki, T., Kidner, J. H., Koch, V., Köhler, A., Kraus, F. B., Lattorff, H. M., Leask,  
 308 M., Lockett, G. A., Mallon, E. B., Antonio, D. S. M., Marxer, M., Meeus, I., Moritz, R. F., Nair, A.,  
 309 Näpflin, K., Nissen, I., Niu, J., Nunes, F. M., Oakeshott, J. G., Osborne, A., Otte, M., Pinheiro, D. G.,  
 310 Rossié, N., Rueppell, O., Santos, C. G., Schmid-Hempel, R., Schmitt, B. D., Schulte, C., Simões, Z. L.,  
 311 Soares, M. P., Swevers, L., Winnebeck, E. C., Wolschin, F., Yu, N., Zdobnov, E. M., Aqrawi, P. K.,  
 312 Blankenburg, K. P., Coyle, M., Francisco, L., Hernandez, A. G., Holder, M., Hudson, M. E., Jackson,  
 313 L., Jayaseelan, J., Joshi, V., Kovar, C., Lee, S. L., Mata, R., Mathew, T., Newsham, I. F., Ngo, R.,  
 314 Okwuonu, G., Pham, C., Pu, L.-L., Saada, N., Santibanez, J., Simmons, D., Thornton, R., Venkat,  
 315 A., Walden, K. K., Wu, Y.-Q., Debyser, G., Devreese, B., Asher, C., Blommaert, J., Chipman, A. D.,  
 316 Chittka, L., Fouks, B., Liu, J., O’Neill, M. P., Sumner, S., Puiu, D., Qu, J., Salzberg, S. L., Scherer,  
 317 S. E., Muzny, D. M., Richards, S., Robinson, G. E., Gibbs, R. A., Schmid-Hempel, P., and Worley,  
 318 K. C. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome*  
 319 *Biology*, 16(1): 76.
- 320 Shen, S., Park, J. W., Lu, Z.-x., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., and Xing, Y. 2014.  
 321 rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data.  
 322 *Proceedings of the National Academy of Sciences*, 111(51): E5593–E5601.
- 323 Simola, D. F., Wissler, L., Donahue, G., Waterhouse, R. M., Helmkampf, M., Roux, J., Nygaard, S.,  
 324 Glastad, K., Hagen, D. E., Viljakainen, L., Reese, J. T., Hunt, B. G., Graur, D., Elhaik, E., Kriventseva,  
 325 E., Wen, J., Parker, B. J., Cash, E., Privman, E., Childers, C. P., Munos-Torres, M. C., Boomsma, J. J.,  
 326 Bornberg-Bauer, E., Currie, C., Elsik, C. G., Suen, G., Goodisman, M. A., Keller, L., Liebig, J., Rawls,  
 327 A., Reinberg, D., Smith, C. D., Smith, C. R., Tsutsui, N., Wurm, Y., Zdobnov, E. M., Berger, S. L., and  
 328 Gadau, J. 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation  
 329 while preserving regulatory features linked to sociality. *Genome Research*, page gr.155408.113.
- 330 Terashima, J. and Bownes, M. 2006. E75a and E75b have opposite effects on the apoptosis/development  
 331 choice of the *Drosophila* egg chamber. *Cell Death and Differentiation*, 13(3): 454–464.
- 332 Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C. P., Glastad,  
 333 K. M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R. A., Hu, H., Hunt, B. G., Huylmans,  
 334 A. K., Khalil, S. M. S., Mitchell, R. D., Munoz-Torres, M. C., Mustard, J. A., Pan, H., Reese, J. T.,  
 335 Scharf, M. E., Sun, F., Vogel, H., Xiao, J., Yang, W., Yang, Z., Yang, Z., Zhou, J., Zhu, J., Brent,  
 336 C. S., Elsik, C. G., Goodisman, M. A. D., Liberles, D. A., Roe, R. M., Vargo, E. L., Vilcinskas, A.,  
 337 Wang, J., Bornberg-Bauer, E., Korb, J., Zhang, G., and Liebig, J. 2014. Molecular traces of alternative  
 338 social organization in a termite genome. *Nature Communications*, 5: 3636.

- 339 Tovar-Corona, J. M., Castillo-Morales, A., Chen, L., Olds, B. P., Clark, J. M., Reynolds, S. E., Pitten-  
340 drigh, B. R., Feil, E. J., and Urrutia, A. O. 2015. Alternative Splice in Alternative Lice. *Molecular*  
341 *Biology and Evolution*, 32(10): 2749–2759.
- 342 Trapnell, C., Pachter, L., and Salzberg, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq.  
343 *Bioinformatics*, 25(9): 1105–1111.
- 344 Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L.,  
345 Rinn, J. L., and Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq  
346 experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3): 562–578.
- 347 Woodard, S. H., Lozier, J. D., Goulson, D., Williams, P. H., Strange, J. P., and Jha, S. 2015. Molecular  
348 tools and bumble bees: revealing hidden details of ecology and evolution in a model system. *Molecular*  
349 *Ecology*, 24(12): 2916–36.
- 350 Yan, H., Bonasio, R., Simola, D. F., Liebig, J., Berger, S. L., and Reinberg, D. 2015. DNA Methylation in  
351 Social Insects: How Epigenetics Can Control Behavior and Longevity. *Annual Review of Entomology*,  
352 60(1): 435–452.

353 **Data Accessibility,**

354 All sequence data for this study are archived at European Nucleotide Archive (ENA); Accession no. PR-  
355 JEB9366 (<http://www.ebi.ac.uk/ena/data/view/PRJEB9366>). GO-analysis results and lists of isoforms  
356 are available as Supporting Information.

357 **Author contributions**

358 All authors developed the project idea and were involved in the interpretation of data and finalization of  
359 the manuscript. JP and EBM analyzed the data. EBM and JP drafted the manuscript.