bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Parent of origin gene expression in the bumblebee, *Bombus terrestris*, supports Haig's kinship theory for the evolution of genomic imprinting.

Hollie Marshall[1], Jelle S. van Zweden[2], Anneleen Van Geystelen[2], Kristof Benaets[2], Felix Wäckers[3,4], Eamonn B. Mallon[1] and Tom Wenseleers[2]

[1]Department of Genetics and Genome Biology, The University of Leicester, Leicester, UK.
[2]Laboratory of Socioecology and Social Evolution, Department of Biology, KU Leuven, Leuven, Belgium.
[3]Biobest Belgium N.V., Westerlo, Belgium.
[4]The Lancaster Environmental Centre, University of Lancaster, Lancaster, UK.

Submitted: February 28th, 2020

DRAFT

1 **Corresponding authors:** Eamonn Mallon
2 **Tel:** +44(0)1162523488
3
4 **Emails:** hollie_marshall@hotmail.co.uk, js.vanzweden@cbs.nl, anneleen.vangeystelen@gmail.com,
5 kristof_benaets@hotmail.com, felix.wackers@biobest.be, ebm3@le.ac.uk and tom.wenseleers@kuleuven.be.
6
7 **Abstract work count:** 150.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

8 ## Abstract

9 *Genomic imprinting is the differential expression of alleles in diploid individuals, with the*
10 *expression being dependent upon the sex of the parent from which it was inherited. Haig's kinship*
11 *theory hypothesizes that genomic imprinting is due to an evolutionary conflict of interest between*
12 *alleles from the mother and father. In social insects, it has been suggested that genomic imprinting*
13 *should be widespread. One recent study identified parent-of-origin expression in honeybees and*
14 *found evidence supporting the kinship theory. However, little is known about genomic imprinting*
15 *in insects and multiple theoretical predictions must be tested to avoid single-study confirmation*
16 *bias. We, therefore, tested for parent-of-origin expression in a primitively eusocial bee. We*
17 *found equal numbers of maternally and paternally biased expressed alleles. The most highly*
18 *biased alleles were maternally expressed, offering support for the kinship theory. We also found*
19 *low conservation of potentially imprinted genes with the honeybee, suggesting rapid evolution of*
20 *genomic imprinting in Hymenoptera.*

21 **Impact summary**

22 Genomic imprinting is the differential expression of alleles in diploid individuals, with the
23 expression being dependent upon the sex of the parent from which it was inherited. Genomic
24 imprinting is an evolutionary paradox. Natural selection is expected to favour expression of both
25 alleles in order to protect against recessive mutations that render a gene ineffective. What then is
26 the benefit of silencing one copy of a gene, making the organism functionally haploid at that locus?
27 Several explanations for the evolution of genomic imprinting have been proposed. Haig's kinship
28 theory is the most developed and best supported.

29 Haig's theory is based on the fact that maternally (matrigene) and paternally (patrigene) inherited
30 genes in the same organism can have different interests. For example, in a species with multiple
31 paternity, a patrigene has a lower probability of being present in siblings that are progeny of the
32 same mother than does a matrigene. As a result, a patrigene will be selected to value the survival of
33 the organism it is in more highly, compared to the survival of siblings. This is not the case for a
34 matrigene.

35 Kinship theory is central to our evolutionary understanding of imprinting effects in human health
36 and plant breeding. Despite this, it still lacks a robust, independent test. Colonies of social bees
37 consist of diploid females (queens and workers) and haploid males created from unfertilised eggs.
38 This along with their social structures allows for novel predictions of Haig's theory.

39 In this paper, we find parent of origin allele specific expression in the important pollinator, the
40 buff-tailed bumblebee. We also find, as predicted by Haig's theory, a balanced number of genes
41 showing matrigenic or patrigenic bias with the most extreme bias been found in matrigenically biased
42 genes.

43 ***Keywords***— reciprocal cross, bumblebee, genomic imprinting, Hymenoptera, kinship theory,
44 RNA-seq

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

45 # Introduction

46 Genomic imprinting is the differential expression of alleles in diploid individuals, with the expression
47 being dependent upon the sex of the parent from which it was inherited [1]. Multiple evolutionary
48 theories attempt to explain its existence [reviewed in 2]. The most widely accepted explanation is
49 the kinship theory developed by Haig [3]. This theory predicts genomic imprinting arose due to
50 natural selection acting differently on the matrigenes (maternal alleles) and the patrigenes (paternal
51 alleles) of an individual for given processes. For example, in a polyandrous mating system with
52 maternal care (e.g. mammals), patrigenes are predicted to be subject to selection pressures which
53 increase resource allocation from the mother at the expense of siblings. Whereas matrigenes in this
54 scenario are predicted to be selected for equal resource distribution amongst offspring.

55 The majority of support for this theory comes from studies based on mammals and flowering
56 plant systems [2]. However, it has been suggested haplodiploid social insects can provide an ideal
57 system to independently test Haig's kinship theory [4]. Colonies of social bees consist of diploid
58 females (queens and workers) and haploid males created from unfertilised eggs. This along with
59 their social structures allows for novel predictions of Haig's theory.

60 Research exploring parent-of-origin effects in social insects has focused on the behavioural
61 and physiological outputs of genetic crosses. In the Argentine ant (*Linepithema humile*) paternal
62 effects were observed in care-giving associated behaviours and in sex allocation of offspring [5, 6].
63 Paternal effects on dominance and stinging behaviour have also been observed in crosses of European
64 and Africanized honeybees [7]. Additionally, Oldroyd *et al.* [8] found a parent-of-origin effect of
65 increased ovary size in honeybees but could not definitively determine which parent this effect was
66 driven by.

67 More recently, reciprocal crosses and next generation sequencing technologies have been
68 used to identify genes with parent-of-origin allele specific expression patterns in honeybees [9, 10].
69 Both groups used RNA-Seq to study parent-of-origin gene expression in hybrid crosses of honeybee

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

70 sub-species. The logic of their test was that as honeybee queens are multiply mated, matrigenes can
71 occur in half sisters and therefore should be selected to moderate worker reproduction in queenless
72 colonies. Patrigenes, on the other hand will not be in half sisters and will be selected to reproduce at
73 any cost [10]. Therefore, the prediction is that parent of origin allele specific expression should exist
74 in honeybees and patrigenic expression will dominate [10]. Surprisingly, Kocher *et al*. [9] found
75 a matrigenic bias in gene expression however, it was later shown that sub-species incompatibility
76 effects influenced the results obtained [11]. Showing support for the kinship theory, Galbraith *et al*.
77 [10] found greater patrigenic expression in reproductive workers compared to sterile workers, with
78 increased patrigenic expression in the reproductive tissues.

79 The lack of agreement between these studies weakens their support for Haig's theory. Another
80 weakness of the evidence is that it is limited to only one species and tests only one prediction from the
81 many predictions made [4] for Haig's theory and genomic imprinting's role in social insect biology.

82 To test the robustness of Haig's kinship theory we present gene expression data (RNA-seq) of
83 reproductive and sterile workers from reciprocal crosses of sub-species of the primitively eusocial
84 bumblebee, *Bombus terrestris*. This species is naturally singly-mated. As such, the predictions for
85 the presence of matrigenic/patrigenic expression bias are different of those predicted for the naturally
86 multiply-mated honeybee [4]. As in Galbraith *et al*. [10] we test the effect of queenless conditions.
87 The same patrigene is present in all sisters (Fig.1), therefore patrigenes should be selected to moderate
88 worker reproduction to reduce the cost to nephews. This is the opposite of the honeybee prediction.
89 Matrigenes, similar to honeybees, have a non-zero probability of being in any given sister (Fig.1),
90 therefore matrigenes should also be selected to moderate worker reproduction, although possibly at
91 a lower level that patrigenes. Therefore we make three predictions; 1) that parent-of-origin allele
92 specific expression exists in bumblebees, 2) that parent-of-origin allele specific expression should
93 be balanced between genes biased patrigenically and matrigenically, with perhaps a slight surplus
94 of matrigenically expressed genes due to the decreased probability of a matrigene occurring in a
95 nephew as opposed to a son and 3) that genes showing parent-of-origin expression will be enriched

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

96 for reproductive related processes.

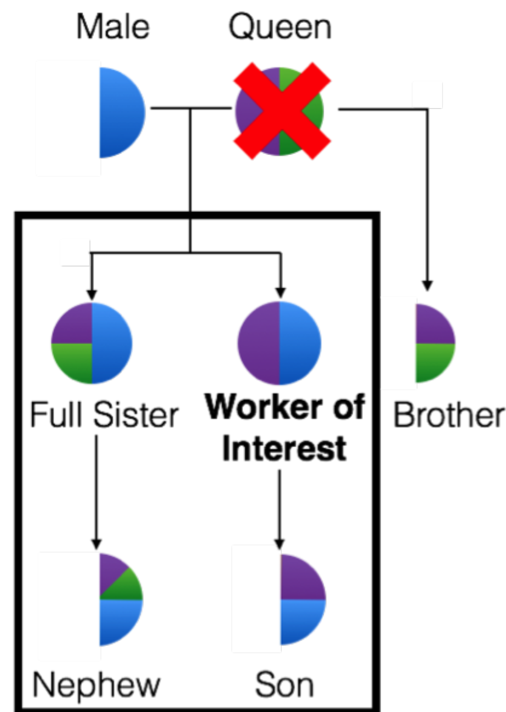

Figure 1: **Parental genome inheritance probabilities under queenless conditions in bumblebees**. Schematic adapted from Drewell *et al*. [12]. As the bumblebee is singly mated, there is only one patrigene (blue) shared between all workers. This means there is an equal probability of the patrigene ending up in the worker's son or nephew. The worker of interest has inherited the purple matrigene. Her sister has a 50% change of inheriting the purple or the green matrigene. For the purple matrigene, there is a 50% chance of ending up in the son but a 25% chance of ending up in the nephew.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

97 # Results

98 ## Differential expression between reproductive castes

99 Ten bases were trimmed from the 3’ end of each worker RNA-seq read due to base bias generated by
100 the Illumina protocol [13]. The mean number of uniquely mapped reads to the maternal genomes
101 was 91.31% ± 0.57% (mean ± standard deviation) for the head samples and 91.53% ± 1.86% for the
102 abdomen samples. This equated to a mean of 13,602,926 ± 2,774,040 and 14,181,003 ± 1,303,486
103 uniquely mapped reads respectively (supplementary 1.0.1). The mean number of uniquely mapped
104 reads to the paternal genomes was 91.50% ± 0.68% (mean ± standard deviation) for the head samples
105 and 91.70% ± 1.83% for the abdomen samples. This equated to a mean of 13,630,611 ± 2,776,779
106 and 14,208,273 ± 1,307,054 uniquely mapped reads respectively (supplementary 1.0.1).

107 Tissue type explains the majority of variation within all of the RNA-seq samples, followed by
108 reproductive status; either reproductive or sterile (supplementary 2.0: Fig.S1). Following differential
109 expression analysis a total of 3,505 genes were up-regulated in the abdomen of reproductive workers
110 compared to sterile workers and 4,069 genes were down-regulated ($q<0.01$) (supplementary 1.0.2).

111 The enriched GO terms for the differentially expressed genes between reproductive and sterile
112 castes in the abdomen included mostly regulatory processes but also "*reproduction*" (GO:0000003)
113 (supplementary 1.0.3). Enriched GO terms associated specifically with up-regulated genes in
114 reproductive workers in the abdomen also included: "*reproduction*" (GO:0000003) and "*DNA*
115 *methylation*" (GO:0006306) (supplementary 1.0.4), these terms were not found in the enriched GO
116 terms for genes up-regulated in sterile workers (supplementary 1.0.5).

117 Considerably fewer genes were differentially expressed in the head samples; 86 up-regulated
118 genes in reproductive compared to sterile workers and 41 down-regulated genes ($q<0.01$) (supple-
119 mentary 1.0.6). The majority of the GO terms associated with these differentially expressed genes
120 involved biosynthetic processes (supplementary 1.0.7). Up-regulated genes in the head tissue of

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

121 reproductive workers also included "*reproduction*" (GO:0000003) whereas the up-regulated genes in
122 the head tissue of sterile workers consisted of mostly metabolic processes (supplementary 1.0.8 and
123 1.0.9).

124 Parent-of-origin gene expression

125 A total of 10,211 genes had a minimum of two SNPs with at least a coverage of five reads each. Of
126 those, 7,508 genes occurred in every cross, worker-type and tissue type. 700 genes had significant
127 maternal/paternal expression bias in both cross-directions for reproductive workers (q <0.05), and
128 747 were significant for sterile workers. The expression bias was averaged across; tissue type, worker
129 type, family and direction of cross to obtain an extremely conservative expression proportion. The
130 significant genes were then filtered to also have an average maternal expression proportion of >0.6 or
131 <0.4 to give a final confident list of genes showing parent-of-origin expression.

132 Reproductive workers have 163 genes showing significant parent-of-origin expression (Fig.2)
133 (supplementary 1.1.0). Sterile workers have 170 genes showing significant parent-of-origin expression
134 (Fig.2) (supplementary 1.1.1). There is no significant difference between the number of genes
135 showing maternal expression bias compared to paternal expression bias based on reproductive status
136 (chi-squared test of independence, χ-squared = 0, df = 1, p-value = 1). There is also no difference
137 in the number of genes showing paternal expression bias compared to maternal expression bias in
138 reproductive and sterile workers, assessed independently (chi-squared goodness of fit, reproductive:
139 χ-squared = 1.3804, df = 1, p-value = 0.24, sterile: χ-squared = 1.5059, df = 1, p-value = 0.2198).
140 The most extreme expression bias is seen in the maternally expressed alleles in both castes, with 17
141 genes showing a maternal expression proportion of >0.9 in both reproductive and sterile workers
142 (Fig.2). There were no genes showing >0.9 paternal expression bias. Additionally we did not find
143 any genes with significant sub-species expression bias.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

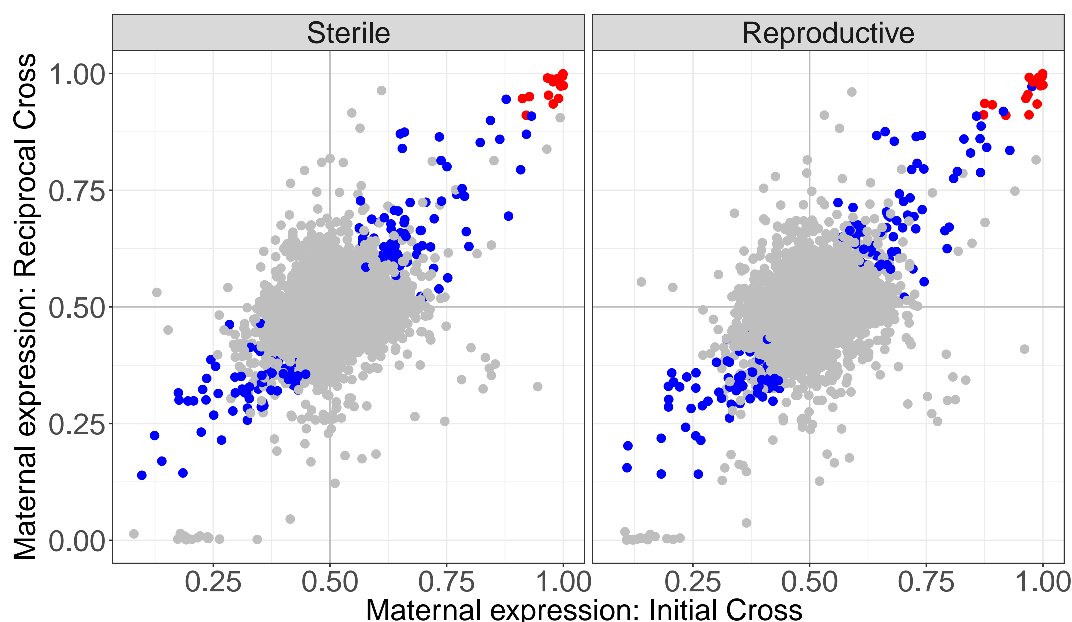

Figure 2: **Maternal expression proportion of all genes by worker caste.** Each point represents a gene. Blue points are genes with significant parental expression bias (q<0.05 and expression proportion >0.6). Red points are genes with significant parental expression bias (q<0.05) with the proportion of expression >0.9. The top left quadrant of each plot represents genes with a *B. terrestris audax* expression bis, the bottom right quadrant represents gene with a *B. terrestris dalmatinus* expression bias. The top right represents genes with a maternal expression bias and the bottom left represents genes with a paternal expression bias.

144 Reproductive and sterile workers share a significant number of genes showing parent-of-origin
145 expression with the same parental bias (Fig.3, maternal expression bias: hypergeometric test, p =
146 9.20×10^{-108}, paternal expression bias: hypergeometric test, p = 7.66×10^{-90}). There were no
147 genes with maternal/paternal bias in one caste which also had the opposite bias in the other caste.
148 Additionally the majority of genes identified as showing parental expression bias show the same
149 bias in both abdomen and head tissue as well as across behaviourally defined castes (dominant and
150 subordinate reproductives and sterile foragers and nurses), see supplementary 2.0: Fig.S2-S7.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

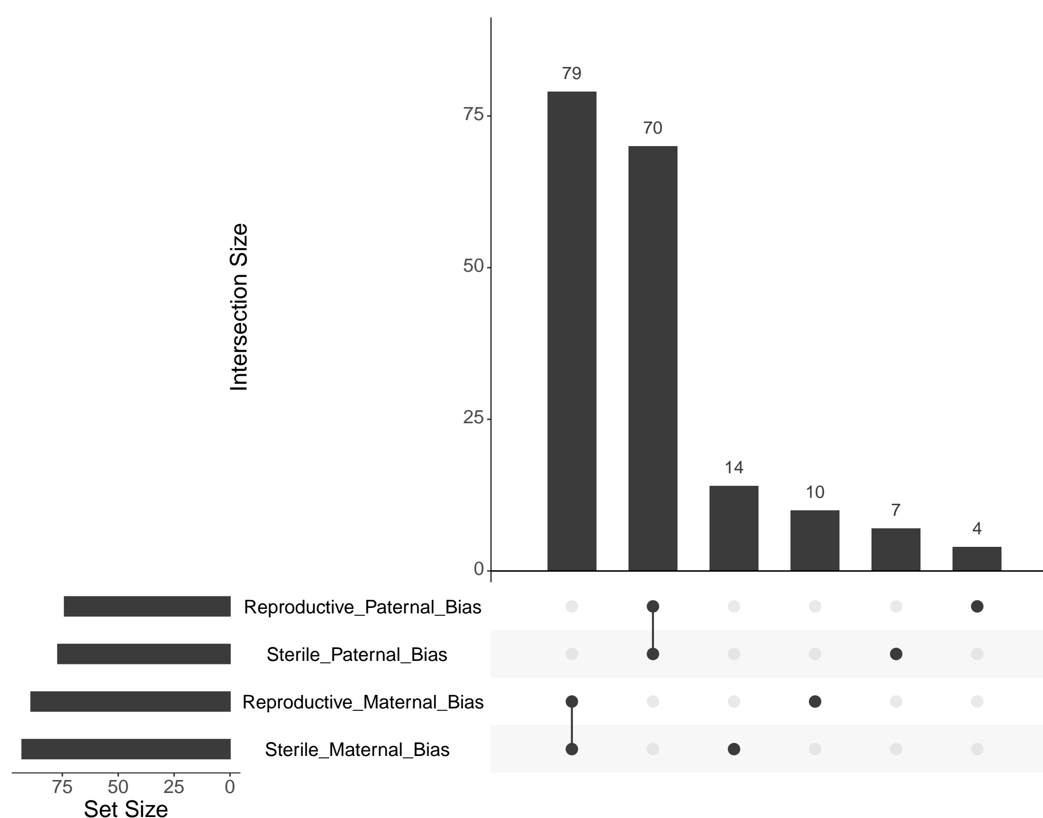

Figure 3: **Overlapping genes showing parent-of-origin expression in reproductive and sterile workers.** The set size indicates the number of genes in each list. The intersection size shows how many genes the corresponding lists have in common. A single dot refers to the number of genes unique to each list.

151 Overall genes showing parent-of-origin expression have enriched GO terms for multiple
152 biological processes (supplementary 1.1.2), specifically the GO terms "*negative regulation of*
153 *reproductive processes*" (GO:2000242) and "*female germ-line sex determination*" (GO:0019099) are
154 enriched. Genes with maternal and paternal bias in both reproductive and sterile workers also have
155 enriched GO terms for multiple biological processes (supplementary 1.1.3 and 1.1.4). Specifically,
156 paternally expressed genes in both reproductive and sterile workers are enriched for the GO term;
157 "*behaviour*" (GO:0007610).

158 GO terms for genes showing parent-of-origin expression in only reproductive or sterile
159 workers were also enriched for various biological processes (supplementary 1.1.5). Including

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

160 "*histone ubiquitination*" (GO:0016574) and "*histone H2A monoubiquitination*" (GO:0035518) in
161 reproductive paternally expressed genes.

162 Genes showing maternal parent-of-origin expression are enriched for genes which are also
163 differentially expressed in head tissue (Supplementary 2.0: Fig.S8, hypergeometric test, p = 0.004).
164 Specifically *Serine Protease Inhibitor 3/4* (LOC100652301) shows maternal expression bias in both
165 reproductive and sterile workers and is up-regulated in the head tissue of reproductive workers.

166 Genes with paternal parent-of-origin expression do not significantly overlap with differentially
167 expressed genes (supplementary 2.0: Fig.S8, hypergeometric test, p = 0.148). There is also no
168 significant overlap between genes showing parent-of-origin expression bias and differential expression
169 in the abdomen (supplementary 2.0: Fig.S9, hypergeometric test, p = 0.996).

170 Honeybee homology

171 A custom database of putative orthologs was made between *A. mellifera* and *B. terrestris*, as in [14],
172 containing 6,539 genes. 68% of differentially expressed abdominal genes were identified within the
173 database, 43% of differentially expressed head genes and 46% of genes showing parent-of-origin
174 expression bias (supplementary 1.1.6). Gene lists were obtained from Galbraith *et al*. [10], the
175 ortholog database contained 38% of the differentially expressed genes identified between honeybee
176 reproductive worker castes and 53% of the genes found to show parent-of-origin expression bias
177 (supplementary 1.1.6).

178 There was no significant overlap between the genes identified as showing parent-of-origin
179 expression between both studies (hypergeometric p = 0.64), with only two genes overlapping
180 (supplementary 2.0 Fig.S10). One of these is uncharacterised in both species (honeybee id:
181 LOC552195, bumblebee id: LOC100648162) and the second is a serine protease inhibitor (honeybee
182 id: LOC411889, bumblebee id: LOC100644680). The serine protease inhibitor shows paternal
183 expression bias in honeybees and maternal bias in both bumblebee reproductive and sterile castes.
184 It is not differentially expressed in the honeybee but it shows up-regulation in the abdomen tissue

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

185 of reproductive bumblebee workers compared to sterile workers. Additionally, [10] identified
186 numerous genes of interest which are involved in reproductive behaviour (*vitellogenin, yolkless,*
187 *ecdysone-receptor and ecdysone-induced protein*) which show paternal expression bias in honeybees,
188 none of which show significant parent-of-origin expression in the bumblebee (supplementary 2.0
189 Fig.S11).

190 There was also no significant overlap between differentially expressed genes identified in
191 head tissue and abdomen tissue between *B. terrestris* reproductive castes with those identified as
192 differentially expressed between reproductive castes of *A. mellifera* from [15] (supplementary 2.0
193 Fig.S12, head: hypergeometric $p = 1$, abdomen hypergeometric $p = 1$).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

Discussion

Using parental genome sequencing and offspring RNA-seq we have identified genes showing parent-of-origin allele specific expression in a primitively eusocial bumblebee species. There was no difference in the number of genes showing maternal or paternal expression bias in either reproductive or sterile workers. The genes showing the highest proportion of expression bias were all maternally expressed in both worker castes. Additionally, reproductive related GO terms were enriched in both maternally and paternally biased genes.

Reproductive and sterile workers were chosen in order to provide a robust test for the kinship theory. Imprinted genes should maintain their expression bias regardless of the current reproductive state of the individual, i.e. higher matrigenic expression compared to patrigenic expression should be present in queenless workers regardless of whether they have become reproductive or remained sterile. Additionally the use of sterile and reproductive workers allowed us to assess differential expression between castes. This meant we could investigate if there was a relationship between differentially expressed genes between reproductive castes and genes showing parent-of-origin expression. A high overlap of differentially expressed genes between reproductive castes with genes showing parent-of-origin expression would suggest a role for imprinted genes in reproductive caste determination.

As in Galbraith *et al*. [10] we also found no significant overlap of paternally expressed genes with differentially expressed genes between reproductive and sterile workers. However, we did find a significant overlap with maternally expressed genes and genes differentially expressed in head tissue between reproductive and sterile castes. This significant overlap should be interpreted cautiously however as only seven of the 103 unique maternally biased genes were differentially expressed in head tissue between reproductive castes. The lack of overlap of differentially expressed genes with paternally expressed genes and the small overlap with maternally expressed genes suggests parent-of-origin expression may not directly influence reproductive status in bumblebee workers.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

219 One of the overlapping genes found to be differentially expressed in head tissue between
220 reproductive castes which also shows maternal expression bias is a serine protease inhibitor. One of
221 the two homologous genes identified as showing parent-of-origin expression in this study and in
222 honeybees was also a serine protease inhibitor. Serine proteases (also know as serpins) have been
223 shown to be involved in insect immunity in various species including; the silkworm *Bombyx mori*
224 [16], another species of silk producing moth *Antheraea pernyi* [17] and the mosquito *Anopheles*
225 *gambiae* [18]. Most recently a kazal-type serine protease inhibitor has been directly linked to
226 oocyte development in the desert locust *Schistocerca gregaria* [19]. Future work identifying the
227 function of serine protease inhibitors in social insects is needed to better understand the function of
228 parent-of-origin expression of these genes in both *B. terrestris* and *A. mellifera*.

229 We found genes involved in histone modifications to be paternally expressed in reproductive
230 workers. Histone modifications have been identified as an imprinting mark in plants [20] and thought
231 to be involved in imprinting maintenance in mammals [21]. Histone modifications can alter gene
232 expression by affecting gene accessibility via chromatin [22]. Chromatin modifications have been
233 associated with parent-of-origin expression in the fruit fly *Drosophilia melanogaster* [23]. Galbraith
234 *et al*. [10] also found genes involved in histone modifications showing parent-of-origin expression in
235 the honeybee.

236 Only two genes were found in common between those found in Galbraith *et al*. [10] as
237 showing parent-of-origin gene expression in the honeybee *A. mellifera* and those identified here in
238 *B. terrestris*. Galbraith *et al*. [10] used ovaries and fat bodies as their tissues samples whereas we
239 selected to test whole head and whole abdomen samples for strong signals of expression bias. Some
240 imprinted genes in mammals are known to be tissue specific, such as *GBR10* which has been found
241 to be maternally expressed in brain and muscle tissue but not in growth plate cartilage [24]. Tissue
242 specificity could account for the lack of concordance in parentally expressed genes found between *B.*
243 *terrestris* and *A. mellifera*. Additionally 51% of bumblebee genes and 52% of honeybee genes were
244 not present in the homology database created.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

245 Imprinted genes in mammals show much more consistency across species, with mice and
246 humans reportedly sharing around 50 imprinted genes out of around 150 and 100 characterised in
247 each respectively [25]. Additionally, domestic cattle and pigs have been shown to share 14 imprinted
248 genes out of 26 and 18 respectively [26]. However, imprinted genes in plants generally show less
249 conservation, with one study reporting 14% of maternally expressed genes and 29% of paternally
250 expressed genes in *Capsella rubella* show the same imprinting status in *Arabidopsis thaliana*, even
251 though both species belong to the Brassicaceae family [27]. Hatorangan *et al*. [27] suggest the lack
252 of consistency between species could be the result of a historical shift in mating-systems. Given the
253 differences in mating-systems between honeybees and bumblebees, and that variable predictions
254 from the kinship theory apply to each species, rapid evolution of imprinted genes in Hymenoptera is
255 a feasible explanation for the lack of consistency in potentially imprinted genes identified here and in
256 honeybees.

257 The GO terms associated with both maternally and paternally expressed genes are diverse.
258 It has been suggested imprinted genes can function as a mechanism for plasticity, allowing gene
259 regulation to change depending on environmental conditions by activating the silenced allele and
260 increasing dosage of that gene [28]. Social insects display, sometimes extreme, phenotypic plasticity,
261 where multiple discrete phenotypes (castes) can arise from a single genome within a colony. In some
262 species this is genetically determined [29], however there is growing evidence epigenetic factors
263 may play a role in caste determination in some species [14, 30, 31]. Matsuura *et al*. [32] modelled a
264 genomic imprinting mediated caste determination system in the termite *Reticulitermes speratus* and
265 found this better explained the influence of parental phenotype on offspring than a purely genetic
266 model. Given the diversity of genes found here showing both maternal and paternal expression
267 bias we believe, along with Matsuura [33], that further experimental investigation into the role of
268 genomic imprinting in caste determination in social insects is needed.

269 The identification of genes showing parent-of-origin expression in this study lays the ground
270 work for future research to identify potential epigenetic mechanisms of allele specific expression in

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

271 social insects. Genes showing allele specific expression and DNA methylation have been previously
272 identified in *B. terrestris* [34], and genes involved in the reproductive process have been shown to be
273 differentially methylated between reproductive castes [14, 35]. DNA methylation is the mechanism by
274 which some genes are imprinted in mammals and plants [36] and so investigation of parent-of-origin
275 methylation in *B. terrestris* may be fruitful.

276 These results provide support for Haig's kinship theory. We found as predicted a balanced
277 number of genes showing matrigenic or patrigenic bias. Also as predicted, the most extreme bias
278 was found in matrigenically biased genes. This is novel, independent support for this important
279 evolutionary theory. The results of this study create a base for many future avenues of research
280 including; gene function analysis of serine protease inhibitors in Hymenoptera, epigenetic mechanisms
281 of imprinting in insects and imprinted genes as a mechanism for plasticity, caste determination and
282 social evolution.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

283 # Methods

284 ## Sample collection

285 Reciprocal crosses of *B. terrestris dalmatinus* (native to southern Europe) and *B. terrestris audax*
286 (native to the UK) were carried out by Biobest, Leuven. Reciprocal crosses allow subspecies-of-origin
287 effects (i.e. the effect of genotype) to be disentangled from parent-of-origin effects [9, 10]. In
288 order to obtain enough successful colonies multiple males and females for each cross (Fig. 4) were
289 released into cages to mate. Once mating had occurred the male and female were removed. The
290 male was immediately frozen at -80°C and the female was placed in cold conditions for eight weeks
291 to induce diapause. Ten matings were carried out for each cross.

292 Four successful colonies (one of each cross direction) from two 'families' (Fig.4) were housed
293 at the University of Leuven and kept in 21°C with red light conditions, they were fed *ad libitum* with
294 pollen and a sugar syrup. Callow workers were labelled with numbered disks in order to determine
295 age and allow behaviour to be recorded. Once each colony contained approximately 30 workers the
296 queen was removed. The colonies were then filmed under queenless conditions for 30mins per day
297 for 14 days in order to score individual behaviour. The following behaviours were used to classify
298 workers: incubating, feeding larvae, inspecting brood cells, building egg cups, ventilation, biting,
299 pushing, egg-laying, egg-eating, foraging, feeding and grooming. Workers were classified based on
300 the frequency of each of the above behaviours as either: sterile foragers, sterile nurses, dominant
301 reproductives or subordinate reproductives (supplementary 1.0.0).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

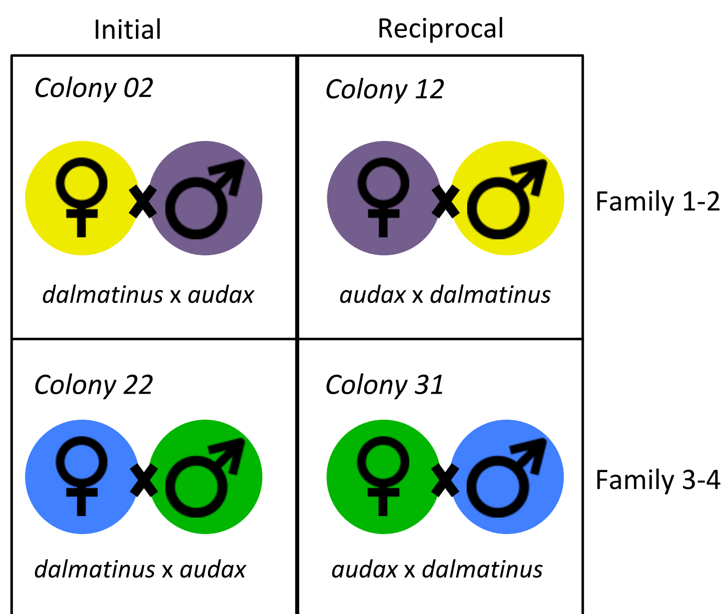

Figure 4: **Graphic display of the family-wise reciprocal crosses carried out.** Each colour refers to related individuals, i.e. the queen from colony 02 is the sister of the male used in colony 12. This design reduces genetic variability between the initial and reciprocal crosses.

302 Worker reproductive status was confirmed by ovary dissection, ovaries were scored on a
303 0-4 scale as in Duchateau and Velthuis [37], entire bodies were then stored at -80°C along with
304 the original queen mothers and male fathers. Workers were selected for sequencing based on their
305 behavioural classification and ovary status. Two of each behavioural type per colony were selected
306 (i.e. two dominant reproductives, two subordinate reproductives, two nurses and two foragers),
307 with the exception of colony 22 (Fig.4) which contained three subordinate reproductives and one
308 dominant reproductive. All sterile and reproductive samples were age matched. This gave a total of
309 32 samples, 8 per colony, four of each reproductive status, reproductive or sterile. See supplementary
310 1.0.0 for behavioural and ovary scoring per sample.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

311 ## DNA and RNA extraction and sequencing

312 DNA was extracted from the mother and father of each colony using the Qiagen DNeasy® Blood &
313 Tissue Kit. Thirty-two workers were selected for RNA sequencing. The head and abdomen were
314 dissected and RNA extracted separately for each, using the Qiagen RNeasy® Lipid Tissue Kit, giving
315 64 total RNA samples. The quality of the DNA and RNA extraction were measured by Nanodrop
316 and Qubit® fluorometer. Whole genome parental DNA was sequenced using 91bp paired-end reads
317 and worker RNA was sequenced using 90bp paired-end reads on an Illumina HiSeq 2000 by BGI,
318 China. Lane effects were minimised for the RNA samples by spreading colony and tissue samples
319 across five lanes.

320 ## Generation of alternative reference genomes

321 Whole genome sequencing data was checked using Fastqc (v.0.11.05) (www.bioinformatics.ba-
322 braham.ac.uk/projects/fastqc/) and adapters and low quality reads were trimmed using Cutadapt
323 v.1.11 [38]. Reads were aligned to the bumblebee reference genome (Bter_1.0, Refseq accession no.
324 GCF_000214255.1 [39]) using BWA-mem v.0.7.15 [40] with standard parameters. SNPs were then
325 called using freebayes v.1.1.0 [41] which can account for the difference in ploidy between males and
326 females, a minimum of five reads per SNP were required. Queen SNPs were then filtered so only the
327 homozygous alternative SNPs remained. The subtractbed command from BEDtools v.2.25.0 [42]
328 was then used to create files containing SNPs unique to either the mother/father of each colony. The
329 individual parental SNP files were then used to create alternate reference genomes for each parent
330 using the 'fasta alternate reference maker' command in GATK v.3.6 [43].

331 ## Identification of parent-of-origin expression

332 RNA-Seq data was quality checked and trimmed as above. STAR v.2.5.2b [44] was used to align
333 worker RNA-seq reads to each of that colony's specific parental genomes with zero mismatches

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

334 allowed. This ensures any reads containing a SNP will only be matched to the parent that allele was
335 inherited from. Alignment files were then filtered using BEDtools v.2.25.0 [42] so only alignments
336 which contain an informative SNP (a unique SNP from either the mother or father) were kept. Reads
337 were counted for the maternal/paternal alignments, also using BEDtools v.2.25.0 [42] and SNP
338 positions were annotated with a gene ID taken from the Bter_1.0 annotation file (Refseq accession
339 no. GCF_000214255.1) using a custom R script. SNPs which had zero maternal reads in at least
340 one sample were removed completely from the analysis to avoid possible inflation of paternal counts.
341 This would occur if the queen position was mis-called as homozygous with the missing allele
342 matching that of the male [10].

343 Genes showing parent-of-origin expression were determined using a logistic regression model
344 in R v.3.4.0 (https://cran.r-project.org). Only genes occurring in both cross directions and in both
345 family combinations, with a minimum of two SNPs per gene were analysed, this left a total of 7508
346 genes. If any gene showed zero reads for paternal counts this was changed to 1 to avoid complete
347 separation. A quasibionimal distribution was also used to account for overdispersion within the data.
348 Fixed factors included the direction of the cross, family and reproductive status (reproductive or
349 sterile). Correction for multiple testing was carried out using the Benjamini–Hochberg method [45].
350 Genes were determined as showing parent-of-origin expression if the allelic ratio (maternal/paternal)
351 corrected p-value was <0.05 and the parental expression proportion was >0.6.

352 Differential expression

353 All RNA-seq samples were aligned to the reference genome (Bter_1.0, Refseq accession no.
354 GCF_000214255.1 [39]) using STAR v.2.5.2b [44] with standard parameters. HTseq v.0.8.0 [46]
355 was then used to count the number of reads per gene for each sample. Differential gene expression
356 between reproductive and sterile workers for head and abdomen samples was assessed using the
357 DESeq2 package v.1.16.1 [47] in R. DESeq2 allows the incorporation of a general linear model
358 (GLM) to identify differential expression; family, age, weight, direction of the cross, tissue type

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

359 and reproductive status were all factors. P-values were corrected for multiple testing using the the
360 Benjamini–Hochberg method [45].

361 Gene ontology enrichment

362 GO enrichment analysis was carried out using the hypergeometric test with Benjamini-Hochberg
363 [45] multiple-testing correction in a custom R script, (q <0.05). This script utilised a previously
364 made GO database for the *B. terrestris* genome, Bter_1.0 [48]. GO terms for differentially expressed
365 genes were tested for enrichment against GO terms associated with all genes identified in either the
366 the RNA-seq data from the abdomen or head. Up-regulated genes in either reproductive or sterile
367 workers were tested for GO term enrichment against all differentially expressed genes from the
368 respective tissue type as a background set.

369 Genes showing parent-of-origin expression were tested for enrichment against GO terms
370 associated with all genes identified in both abdomen and head RNA-Seq data sets. Genes maternally
371 or paternally biased were checked for GO term enrichment against all genes showing parent-of-origin
372 expression as a background set. REVIGO [49] was used to obtain the GO descriptions from the GO
373 identification numbers.

374 Comparative analyses

375 A hypergeometric test was applied to gene lists from the differential expression analysis and the
376 parent-of-origin expression analysis to identify potential enrichment. *B. terrestis* and *A. mellifera*
377 orthologous genes were determined as in Marshall *et al*. [14] and a custom R script was then
378 used to check for overlap between genes identified as showing parent-of-origin expression here and
379 orthologous *A. mellifera* genes identified in Galbraith *et al*. [10].

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

380 # Acknowledgements

381 We thank Biobest N.V. for providing us with the forty bumblebee colonies of the desired reciprocal
382 crosses. Hollie Marshall had a Ph.D-scholarship from the Central England NERC Training Alliance
383 (NERC, UK). Kristof Benaets had a Ph.D.-scholarship of the Flemish Agency for Innovation by
384 Science and Technology (IWT-Vlaanderen). This study was funded in the context of the KU
385 Leuven BOF Centre of Excellence Financing on 'Eco- and socio-evolutionary dynamics' (Project
386 number PF/2010/07) and a grant from the Research Foundation-Flanders (FWO-Vlaanderen, grant
387 G.0463.12). E.B.M. was funded by NERC grant NE/N010019/1. This research used the ALICE2
388 High Performance Computing Facility at the University of Leicester.

389 # Author Contributions

390 T.W. and E.B.M. conceived the study. The reciprocal crosses were carried out by Biobest (Westerlo,
391 Belgium) under supervision of F.W. J.S.Z. and K.B. carried out the behavioural observations and lab
392 work. H.M. and T.W. carried out the analyses with input from A.V.G. H.M. and E.B.M. wrote the
393 initial manuscript. All authors contributed to and reviewed the manuscript.

394 # Data Accessibility

395 All RNA-seq and whole genome sequencing data is available under NCBI BioProject number
396 PRJNA329487. Custom scripts are available at the following DOI: `https://zenodo.org/`
397 `record/3235636`.

398 # Conflict of Interest

399 All authors declare no conflict of interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

400 # References

401 [1] Pegoraro, M., Marshall, H., Lonsdale, Z.N. and Mallon, E.B. Do social insects support Haig's
402 kin theory for the evolution of genomic imprinting? *Epigenetics*, 12(9):725–742, (2017).

403 [2] Patten, M.M., Ross, L., Curley, J.P., Queller, D.C., Bonduriansky, R. and Wolf, J.B. The
404 evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity*, 113(2):
405 119–128, (2014).

406 [3] Haig, D. The Kinship Theory of Genomic Imprinting. *Annual Review of Ecology, Evolution,*
407 *and Systematics*, (31):9–32, (2000).

408 [4] Queller, D.C. Theory of genomic imprinting conflict in social insects. *BMC Evolutionary*
409 *Biology*, 3(15), (2003).

410 [5] Libbrecht, R. and Keller, L. Genetic compatibility affects division of labor in ants. *Evolution*,
411 67(2):517–524, (2012).

412 [6] Libbrecht, R., Schwander, T. and Keller, L. Genetic Components to Caste Allocation in a
413 Multiple-Queen Ant Species. *Evolution*, 65(10):2907–2915, (2011).

414 [7] Guzman-Novoa, E., Hunt, G.J., Page, R.E., Uribe-Rubio, J.L., Prieto-Merlos, D. and Becerra-
415 Guzman, F. Paternal effects on the defensive behavior of honeybees. *Journal of Heredity*, 96
416 (4):376–380, (2005).

417 [8] Oldroyd, B.P., Allsopp, M.H., Roth, K.M., Remnant, E.J., Drewell, R.A. and Beekman, M. A
418 parent-of-origin effect on honeybee worker ovary size. *Proceedings of the Royal Society B:*
419 *Biological Sciences*, 281(1775):20132388, (2014).

420 [9] Kocher, S.D. et al. A Search for Parent-of-Origin Effects on Honey Bee Gene Expression. *G3*,
421 5(8):1657–1662, (2015).

422 [10] Galbraith, D.A. et al. Testing the kinship theory of intragenomic conflict in honey bees (*Apis*
423 *mellifera*). *Proceedings of the National Academy of Sciences*, 113(4):1020–1025, (2016).

424 [11] Gibson, J.D., Arechavaleta-Velasco, M.E., Tsuruda, J.M. and Hunt, G.J. Biased allele expression
425 and aggression in hybrid honeybees may be influenced by inappropriate nuclear-cytoplasmic
426 signaling. *Frontiers in Genetics*, 6(343), (2015).

427 [12] Drewell, R., Lo, N., Oxley, P.R. and Oldroyd, B.P. Kin conflict in insect societies: A new
428 epigenetic perspective. *Trends in Ecology and Evolution*, 27(7):367–373, (2012).

429 [13] Krueger, F., Andrews, S.R. and Osborne, C.S. Large scale loss of data in low-diversity illumina
430 sequencing libraries can be recovered by deferred cluster calling. *PLoS ONE*, 6(1):4–10, (2011).

431 [14] Marshall, H., Lonsdale, Z.N. and Mallon, E.B. Methylation and gene expression differences
432 between reproductive and sterile bumblebee workers. *Evolution Letters*, 3(5):485–499, (8
433 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

434 [15] Galbraith, D.A., Yi, S.V. and Grozinger, C.M. Evaluation of possible proximate mechanisms
435 underlying the kinship theory of intragenomic conflict in social insects. *Integrative and*
436 *Comparative Biology*, 56(6):1206–1214, (2016).

437 [16] Zou, Z., Picheng, Z., Weng, H., Mita, K. and Jiang, H. A comparative analysis of serpin genes
438 in the silkworm genome. *Genomics*, 93(4):367–375, (2009).

439 [17] Yu, H.M. et al. Characterization and functional analysis of serpin-1 like gene from oak silkworm
440 *Antheraea pernyi*. *Bulletin of Entomological Research*, (2017):1–7, (2017).

441 [18] Gorman, M.J. and Paskewitz, S.M. Serine proteases as mediators of mosquito immune
442 responses. *Insect Biochemistry and Molecular Biology*, 31(3):257–262, (2001).

443 [19] Guo, W., Wu, Z., Yang, L., Cai, Z., Zhao, L. and Zhou, S. Juvenile hormone-dependent
444 Kazal-type serine protease inhibitor Greglin safeguards insect vitellogenesis and egg production.
445 *FASEB Journal*, 33(1):917–927, (2019).

446 [20] Rodrigues, J.A. and Zilberman, D. Evolution and function of genomic imprinting in plants.
447 *Genes and Development*, 29(24):2517–2531, (2015).

448 [21] Delaval, K. and Feil, R. Epigenetic regulation of mammalian genomic imprinting. *Current*
449 *Opinion in Genetics & Development*, 14:188–195, (2004).

450 [22] Bannister, A.J. and Kouzarides, T. Regulation of chromatin by histone modifications. *Cell*
451 *Research*, 21(3):381–395, (2011).

452 [23] Joanis, V. and Lloyd, V. Genomic imprinting in Drosophila is maintained by the products of
453 Suppressor of variegation and trithorax group, but not Polycomb group, genes. *Molecular*
454 *Genetics and Genomics*, 268:103–112, (2002).

455 [24] McCann, J.A., Zheng, H., Islam, A., Goodyer, C.G. and Polychronakos, C. Evidence against
456 GRB10 as the gene responsible for Silver-Russell syndrome. *Biochemical and Biophysical*
457 *Research Communications*, 286(5):943–948, (2001).

458 [25] Babak, T. et al. Genetic conflict reflected in tissue-specific maps of genomic imprinting in
459 human and mouse. *Nature Genetics*, 47(5):544–549, (2015).

460 [26] Tian, X.C. Genomic Imprinting in Farm Animals. *Annual Review of Animal Biosciences*, 2(1):
461 23–40, (2013).

462 [27] Hatorangan, M.R., Laenen, B., Steige, K.A., Slotte, T. and Köhler, C. Rapid Evolution of
463 Genomic Imprinting in Two Species of the Brassicaceae. *The Plant Cell*, 28(8):1815–1827,
464 (2016).

465 [28] Radford, E.J., Ferrón, S.R. and Ferguson-Smith, A.C. Genomic imprinting as an adaptative
466 model of developmental plasticity. *FEBS Letters*, 585(13):2059–2066, (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

467 [29] Mott, B.M., Gadau, J. and Anderson, K.E. Phylogeography of Pogonomyrmex barbatus and
468 *P. rugosus* harvester ants with genetic and environmental caste determination. *Ecology and*
469 *Evolution*, 5(14):2798–2826, (2015).

470 [30] Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C. and Maleszka, R. The honey bee
471 epigenomes: Differential methylation of brain DNA in queens and workers. *PLoS Biology*, 8
472 (11), (2010).

473 [31] Bonasio, R. et al. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus*
474 *floridanus* and *Harpegnathos saltator*. *Current Biology*, 22(19):1755–1764, (2012).

475 [32] Matsuura, K. et al. A Genomic Imprinting Model of Termite Caste Determination: Not Genetic
476 but Epigenetic Inheritance Influences Offspring Caste Fate. *The American Naturalist*, 191(6),
477 (2018).

478 [33] Matsuura, K. Genomic imprinting and evolution of insect societies. *Population Ecology*, pages
479 1–15, (2019).

480 [34] Lonsdale, Z. et al. Allele specific expression and methylation in the bumblebee, *Bombus*
481 *terrestris*. *PeerJ*, 5(e3798), (2017).

482 [35] Amarasinghe, H.E., Clayton, C.I. and Mallon, E.B. Methylation and worker reproduction in
483 the bumble-bee (*Bombus terrestris*). *Proceedings of the Royal Society B: Biological Sciences*,
484 281(20132502), (2014).

485 [36] Scott, R.J. and Spielman, M. Genomic imprinting in plants and mammals: How life history
486 constrains convergence. *Cytogenetic and Genome Research*, 113(1-4):53–67, (2006).

487 [37] Duchateau, M.J. and Velthuis, H.H.W. Development and reproductive strategies in *Bombus*
488 *terrestris* colonies. *Behaviour*, 107(3):186–207, (1988).

489 [38] Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
490 *EMBnet.journal*, 17(1):10, (2011).

491 [39] Sadd, B.M. et al. The genomes of two key bumblebee species with primitive eusocial
492 organization. *Genome Biology*, 16(1):76, (2015).

493 [40] Li, H. and Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
494 *Bioinformatics*, 25(14):1754–1760, (2009).

495 [41] Garrison, E. and Marth, G. Haplotype-based variant detection from short-read sequencing.
496 *arXiv*, 1207.3907, (2012).

497 [42] Quinlan, A.R. and Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic
498 features. *Bioinformatics*, 26(6):841–842, (2010).

499 [43] McKenna, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing
500 next-generation DNA sequencing data. *Genome Research*, 20:1297–1303, (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.17.909168. this version posted February 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

501 [44] Dobin, A., Gingeras, T.R. and Spring, C. Mapping RNA-seq Reads with STAR Alexander.
502 *Current Protocols in Bioinformatics*, (51):1–11, (2016).

503 [45] Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful
504 approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1):289–300, (1995).

505 [46] Anders, S. et al. Count-based differential expression analysis of RNA sequencing data using R
506 and Bioconductor. *Nature Protocols*, 8(9):1765–1786, (2013).

507 [47] Love, M.I., Huber, W. and Anders, S. Moderated estimation of fold change and dispersion for
508 RNA-seq data with DESeq2. *Genome biology*, 15(12):550, (2014).

509 [48] Bebane, P.S.A. et al. The effects of the neonicotinoid imidacloprid on gene expression and DNA
510 methylation in the buff-tailed bumblebee *Bombus terrestris*. *Proc. R. Soc. B*, 286:20190718,
511 (2019).

512 [49] Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. Revigo summarizes and visualizes long lists
513 of gene ontology terms. *PLoS ONE*, 6(7), (2011).