

ARTICLE

DOI: 10.1038/s41467-018-05903-0

OPEN

# The effect of maternal care on gene expression and DNA methylation in a subsocial bee

Samuel V. Arsenault <sup>1</sup>, Brendan G. Hunt <sup>1</sup> & Sandra M. Rehan<sup>2</sup>

Developmental plasticity describes the influence of environmental factors on phenotypic variation. An important mediator of developmental plasticity in many animals is parental care. Here, we examine the consequences of maternal care on offspring after the initial mass provisioning of brood in the small carpenter bee, *Ceratina calcarata*. Removal of the mother during larval development leads to increased aggression and avoidance in adulthood. This corresponds with changes in expression of over one thousand genes, alternative splicing of hundreds of genes, and significant changes to DNA methylation. We identify genes related to metabolic and neuronal functions that may influence developmental plasticity and aggression. We observe no genome-wide association between differential DNA methylation and differential gene expression or splicing, though indirect relationships may exist between these factors. Our results provide insight into the gene regulatory context of DNA methylation in insects and the molecular avenues through which variation in maternal care influences developmental plasticity.

---

<sup>1</sup>Department of Entomology, University of Georgia, Athens, GA 30602, USA. <sup>2</sup>Department of Biological Sciences, University of New Hampshire, Durham, NH 03824, USA. Correspondence and requests for materials should be addressed to B.G.H. (email: [huntbg@uga.edu](mailto:huntbg@uga.edu)) or to S.M.R. (email: [sandra.rehan@unh.edu](mailto:sandra.rehan@unh.edu))

Developmental plasticity is the capacity of an organism to produce multiple stable phenotypes from a single genotype given differing environmental conditions<sup>1</sup>. Examples include changes in stress responses<sup>2</sup> and caste differentiation<sup>3</sup>. Social experiences occurring early in development, such as maternal care, are particularly influential on the developing brain<sup>4</sup>. Variation in maternal care has been linked to variation in stress responses of offspring in rats;<sup>5</sup> this effect is reversible by cross-fostering and is mediated by changes in epigenetic information, such as DNA methylation<sup>5,6</sup>. In bees, variation in the social environment during larval development also alters the expression of genes that regulate behavioral and endocrine responses to stress<sup>7</sup>. Moreover, a recent comparison of vertebrate and honey bee responses to social challenge suggests that fundamental commonalities exist in the molecular mediation of variation in social responses across diverse animal taxa<sup>8</sup>. We were curious as to whether variation in parental care in insects results in long-term effects on offspring behavior, as in mammals, and whether common molecular mediators may exist for such behavioral convergence.

The subsocial small carpenter bee *Ceratina calcarata* is an ideal system for addressing questions regarding the effect of developmental environment and parental care on behavior and gene regulation. Unlike its solitary relatives, *C. calcarata* exhibits maternal care behaviors that consist of mass provisioning prior to oviposition, offspring grooming, and nest defense<sup>9</sup>. Mothers of this species are long lived and nest loyal, producing only a single nest in their lifetime and surviving to interact with adult offspring<sup>10</sup>. Adult offspring are typically non-aggressive in the presence of mothers and exhibit mutual tolerance prior to overwintering<sup>11,12</sup>. All of these behaviors take place in the isolated confines of a narrow, compartmentalized twig. Here, we explore the effects of maternal care on offspring development after the initial mass provisioning of brood.

We compare phenotypic and gene regulatory differences between maternally cared for and orphaned individuals, assay aggression and avoidance behaviors, perform RNA-sequencing to measure changes in transcription, and perform bisulfite sequencing to quantify changes in DNA methylation. Similar studies, performed in mammals, have shown that varying maternal care conditions can have effects on DNA methylation and gene expression that persist throughout the lives of offspring<sup>2,5</sup>. However, the gene regulatory consequences of DNA methylation are more poorly understood in insects than in mammals, and it is presently unknown whether DNA methylation mediates potential long-term effects of variation in parental care in *C. calcarata* or other subsocial insects.

Previous work in *C. calcarata* has shown that its genome contains the requisite machinery to establish and maintain DNA methylation<sup>13</sup>. Additionally, the patterning of DNA methylation is similar to what has been observed in many hymenopteran insects, with almost exclusively intragenic DNA methylation, biased to constitutively expressed genes<sup>14</sup>. Two primary mechanistic explanations have been proposed to link gene body DNA methylation to the regulation of environmental effects on development in insects<sup>15–17</sup>. The first is that DNA methylation is involved in mediating changes in gene expression levels. The effects of DNA methylation on gene expression have been extensively studied in mammals since the 1970s<sup>18</sup>. It has long been thought that methylation in promoter regions leads to suppression of gene expression<sup>18</sup>, though this idea has recently been challenged<sup>19</sup>. Despite its supposed suppressive role in the promoter region of genes, DNA methylation within the gene body is, overall, positively correlated with transcription in mammals<sup>20</sup> as well as insects<sup>14</sup>. This positive correlation could be explained by preferential targeting of DNA methylation to

actively transcribed genes rather than a direct effect of DNA methylation on transcription. Indeed, the presence of gene body DNA methylation may actually reduce gene expression levels by impeding transcriptional elongation efficiency<sup>15,21</sup>. Thus, the relationship between gene body DNA methylation and the regulation of gene expression levels is complex and remains poorly understood. A second proposed mechanism by which gene body DNA methylation may affect gene function is through the regulation of alternative splicing<sup>22</sup>. It is known that DNA methylation affects transcription factor binding<sup>23</sup>, which in gene bodies affects RNA Polymerase II processivity and the recognition of variably utilized splice sites<sup>22</sup>. While this phenomenon has not been well studied in invertebrates beyond genome-wide correlations, a few studies have supported a link between DNA methylation and alternative splicing in insects<sup>24–27</sup>. Either or both of these proposed mechanisms could feasibly link DNA methylation to the regulation of developmental plasticity.

The purpose of our study is twofold. First, we investigate the effects of variation in maternal care on offspring behavior and gene regulation, using the subsocial bee *C. calcarata* as a model. Second, we investigate whether stable differences in gene regulation observed between adults from distinct developmental environments are associated with, and potentially driven by, differences in DNA methylation.

We hypothesized that differences in maternal care may lead to changes in DNA methylation and consequently transcriptional differences in the form of differential gene expression or alternative splicing. We found significant changes in gene expression and splicing between individuals raised with a mother and those raised without, and comparatively small changes in DNA methylation. Additionally, we found very little overlap between the changes in DNA methylation and the changes in gene expression or splicing. We hypothesize that the role of DNA methylation is more subtle and nuanced than previously appreciated and may work in concert with other factors to yield context-specific, non-global effects on gene regulation.

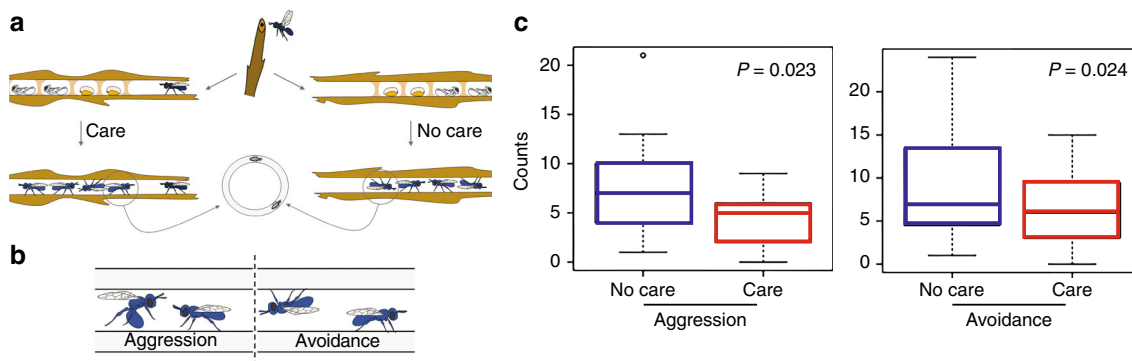
## Results

### Orphan offspring exhibit increased aggression and avoidance.

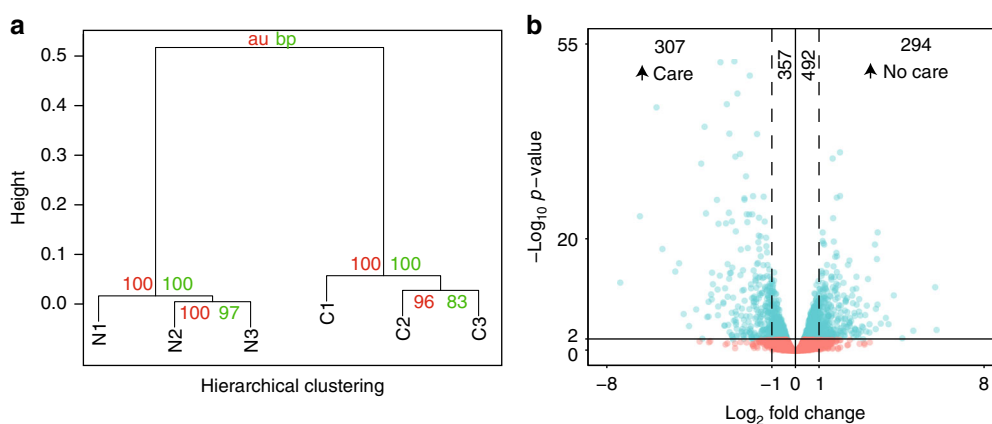
To test the effect of maternal care on behavior, we performed circle tube assays measuring the behaviors of newly eclosed adults. Recently eclosed adult bees raised from nests with and without mothers were placed in circle tubes and aggressive and avoidant behaviors were scored (Fig. 1b). Bees raised in the absence of their mother showed a significant increase in aggression and avoidance when compared to bees raised in the presence of a mother (Mann–Whitney *U*-tests  $p < 0.03$ ; Fig. 1c).

### Loss of maternal care leads to many transcriptional changes.

Maternal care leaves a distinct mark on *C. calcarata* offspring, lasting into adulthood. In association with the behavioral changes described previously, 1450 genes of the 11,068 tested (13.1%) showed significant differential expression between pooled offspring raised with and without a mother (Fig. 2b; FDR-corrected  $p < 0.01$ ). Functional enrichment of Gene Ontology (GO) biological process annotations were observed among differentially expressed genes for terms related to oxidation–reduction reactions, chitin, starch, and sucrose metabolic processes, and neurological functions (ionotropic glutamate receptor signaling pathway) (Table 1, Supplementary Data 1). These pathways represent possible genetic modules involved in offspring response to maternal care. Given the importance of maternal care in *C. calcarata*, we hypothesized that genes developmentally regulated by maternal care may also be subject to positive selection within the *C. calcarata* lineage. We compared differentially expressed



**Fig. 1** Behavioral response of adult offspring to early termination of maternal care. **a** Mothers establish nests in dead broken stems and produce all offspring in a linear hollow. Mothers typically remain on the nest to guard and groom immature offspring (care) or were removed from nests (no care). **b** Drawings of adult offspring engaging in aggression and avoidance behaviors in circle tube arenas. Illustrations in **a** and **b** courtesy of Wyatt Shell (used with permission). **c** Behaviors exhibited by offspring in circle tube assay in the presence and absence of maternal care ( $p$ -values are generated from two-tailed Mann–Whitney  $U$ -tests,  $n = 52$ ; whiskers show observations within  $1.5 \times$  IQR of the lower and upper quartiles)



**Fig. 2** Loss of maternal care leads to extensive changes in gene expression. **a** Hierarchical clustering of RPKM values ( $n = 11,068$  genes) using the “complete” method with correlation-based distances indicates that global gene expression levels cluster into “No Care” (N) and “Care” (C) groups, with high support values using both the approximately unbiased (au) and bootstrap probability (bp) statistics. **b** Volcano plot showing the genes considered differentially expressed (blue) and those not significantly differentially expressed (red). Genes that had more than a twofold difference in gene expression are shown outside of the dashed lines. The numbers of differentially expressed genes are indicated above

genes against a previously published list of genes showing signs of positive selection in the *C. calcarata* lineage, relative to other sequenced bee species, according to the branch-site test A in PAML<sup>13,28</sup>. We found no significant overlap between genes under positive selection and those that are differentially expressed when maternal care is removed (Chi-square test,  $p = 0.72$ ; Supplementary Table 1).

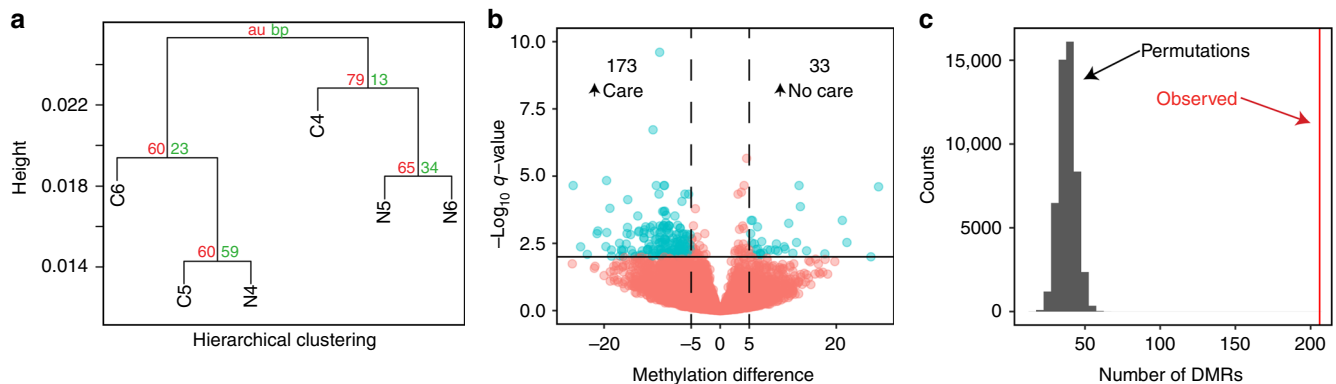
In addition to identifying differentially expressed genes, we also searched for genes that showed signs of differential exon usage and other forms of alternative splicing, between the individuals raised with and without a mother. Production of different isoforms of various genes via alternative splicing is a well-established mechanism for generating developmentally plastic phenotypes<sup>29</sup>. We observed differential exon usage for 1090 exons spread across 373 genes (Supplementary Data 2). Two hundred and thirty-four genes were identified as alternatively spliced using rMATS, with these genes being largely mutually exclusive with the DEXseq-derived genes (Supplementary Figure 4). These results are indicative of changes in splicing regulation between the care and no-care individuals. Genes containing alternative splicing events were enriched for GO terms related to proteolysis, oxidation–reduction processes, and fatty acid biosynthetic processes to name a few (Supplementary Data 1; Fisher’s Exact Test,  $p$ -value  $< 0.01$ ). Similar to results from the differentially

expressed genes, we found no significant overlap between alternatively spliced genes and genes that are putatively subject to positive selection in *C. calcarata* (Chi-square test,  $p = 0.66$ ; Supplementary Table 2).

**Loss of maternal care subtly changes DNA methylation.** Using bisulfite sequencing, we measured DNA methylation at the single-nucleotide scale. For each of our samples of pooled individuals, we observed similar patterns to those found in the previously generated *C. calcarata* methylome<sup>13</sup> and methylomes from many other hymenopteran taxa<sup>14,30–32</sup> (Supplementary Figure 1, Supplementary Table 8). In particular, all of the newly generated *C. calcarata* methylomes show a bias towards exons at the 5’ end of genes (Supplementary Figure 1). Methylated genes ( $>5\%$  mCG/CG) showed GO biological process functional enrichment for ribosome biogenesis, translation, and protein folding (Supplementary Data 1) and were less likely to be under positive selection when compared against unmethylated genes (Chi-square test,  $p = 2.43 \times 10^{-7}$ ; Supplementary Table 3). This corroborates previous results that DNA methylation is biased towards constitutively expressed genes, which tend to be subject to strong purifying selection<sup>14,32,33</sup>.

**Table 1 Gene Ontology (GO) biological process term enrichment among genes differentially expressed between “care” and “no care” offspring (Fisher’s exact test  $p$ -value <0.001)**

GO accession	Term	Number of differentially expressed genes	Expected number of genes	$p$ -Value
GO:0055114	Oxidation–reduction process	115	69.81	9.10E-10
GO:0006030	Chitin metabolic process	29	7.84	3.90E-09
GO:0005982	Starch metabolic process	9	1.96	1.60E-05
GO:0005985	Sucrose metabolic process	9	1.96	1.60E-05
GO:0034220	Ion transmembrane transport	43	31.81	9.30E-05
GO:0005975	Carbohydrate metabolic process	55	29.55	0.00014
GO:0035235	Ionotropic glutamate receptor signaling pathway	11	3.62	0.00032
GO:0016042	Lipid catabolic process	11	7.84	0.00033
GO:0006032	Chitin catabolic process	6	1.36	0.00064
GO:0016998	Cell wall macromolecule catabolic process	6	1.36	0.00064



**Fig. 3** Loss of maternal care leads to localized, significant changes in DNA methylation. **a** Hierarchical clustering of gene level measures of DNA methylation ( $n = 19,361$  genes) using the “complete” method with correlation-based distances reveals no global clustering by sample type (N, no care; C, care), with low support values using both the approximately unbiased (au) and bootstrap probability (bp) statistics. Thus, this tree can be considered one polytomy. Similar results are obtained when clustering only methylated genes. **b** Volcano plot showing the 200 bp regions considered differentially methylated (blue) and those not considered significant (red), as well as the thresholds used to define these categories (black lines). Methylation difference is given as a percentage. **c** Quadratic Assignment Protocol computed null distribution for the number of DMRs possibly occurring randomly using 50,000 randomized permutations and the associated number of experimental DMRs. The red line indicates the observed number of DMRs

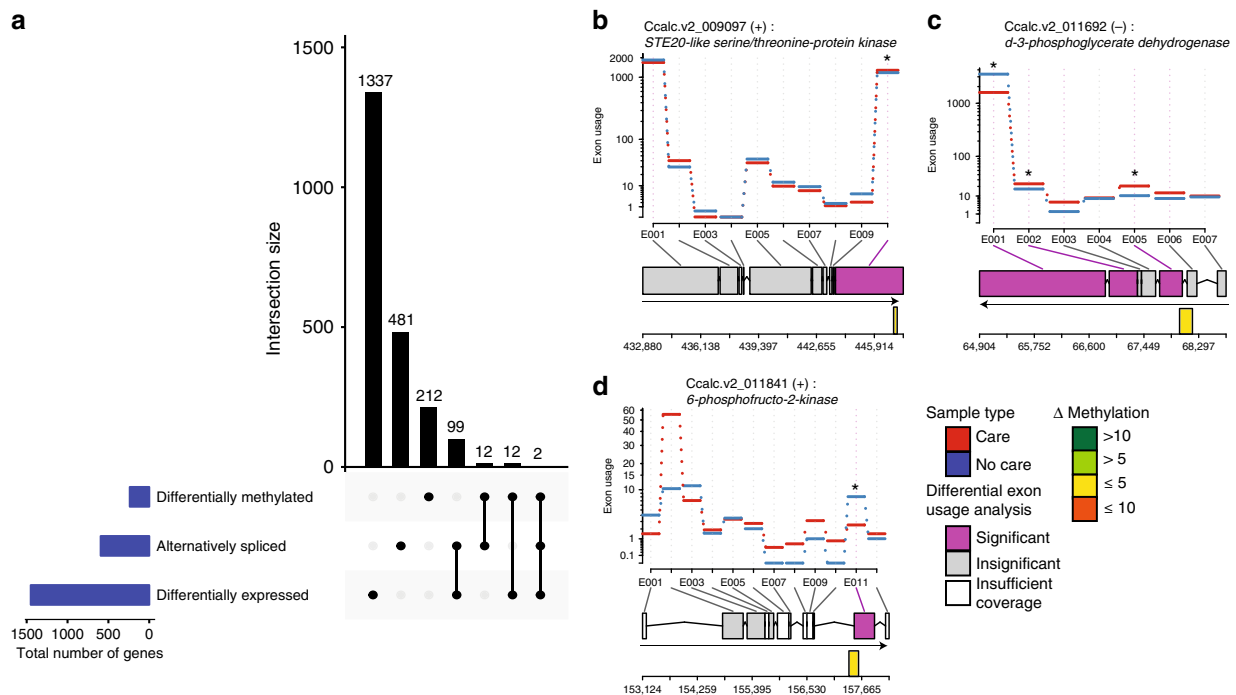
Clustering analysis demonstrated that “no care” and “care” samples did not cluster by sample type based on global measures of DNA methylation (Fig. 3a). However, we also examined DNA methylation differences at a fine spatial scale using a differentially methylated region (DMR) approach. We searched for 200 bp windows throughout the genome that showed significant differential methylation between sample types (MethylKit, FDR-corrected  $p$ -value <0.01, >5% change in methylation). A total of 206 DMRs were identified between the individuals raised with maternal care and those raised without, spanning 238 gene models, with a substantial bias towards elevated DNA methylation in the individuals reared with a mother (84% of DMRs; Fig. 3b; Supplementary Data 2). The genes found to contain DMRs showed functional enrichment for translational elongation and metabolic processes (fructose, mannose, and superoxide) (Supplementary Data 1).

To further validate the presence of biologically meaningful DMRs, we utilized an adapted quadratic assignment method (QAM) to determine the number of DMRs we would expect to find if DNA methylation were distributed randomly with respect to sample type (e.g., if maternal care had no effect on DNA methylation). We found that the number of DMRs produced by proper sample assignment in our analysis (206) was substantially higher than the highest number of DMRs generated by QAM permutations of our data (65; mean = 38.2; Fig. 3c). This further

supported our finding that loss of maternal care itself led to significant, albeit localized, changes in DNA methylation. Additionally, it allows us to estimate how likely any of our individual DMRs occurred by chance (18.5%). We did not find significant overlap between differentially methylated genes and genes subject to positive selection (Chi-square test,  $p = 0.54$ ; Supplementary Table 4).

#### Independent variation in transcription and DNA methylation.

After identifying genes that have undergone changes in expression, splicing, or DNA methylation, we compared the relationship between these processes by characterizing overlap among gene sets (Fig. 4a). As observed previously in *C. calcarata* and other hymenopterans<sup>13,32</sup>, we found a significant correlation between genic methylation levels and gene expression levels when averaged across all samples (Spearman  $\rho = 0.24$ ,  $p < 0.0001$ ). However, we found no evidence of greater overlap than expected by chance between genes containing DMRs and differentially expressed genes (Chi-square test,  $p = 0.31$ ; Supplementary Table 5). Similarly, there was no significant overlap between genes containing DMRs and genes containing alternative splicing events (Chi-square test,  $p = 0.4108$ ; Supplementary Table 6). Taken together, these findings indicate that there is not a consistent, genome-wide association between differential DNA



**Fig. 4** Overlap among differentially expressed, alternatively spliced, and differentially methylated genes. **a** UpSet plot illustrating the intersection between genes containing differentially methylated regions called by MethylKit, alternatively spliced genes called using DEXseq and rMATS, and differentially expressed genes computed using edgeR. The nature of a given intersection is indicated by the dots below the bar plot. For instance, the 99 genes in the fourth column are alternatively spliced and differentially expressed but not differentially methylated. **b–d** DEXseq plots showcasing the instances where a DMR directly overlapped a differentially utilized exon. The top panel shows average exon usage for each treatment group. The gene model is represented below each exon usage plot with purple exons indicating statistical support for differential usage (FDR-corrected  $p$ -value  $< 0.01$ ). Direction of transcription is indicated by an arrow and DMRs are overlaid beneath the gene model with colors indicating the change in methylation level

methylation and alternative splicing or differential expression in our data.

To further examine the potential relationship between DNA methylation and alternative splicing, we next investigated the possibility of spatial concordance between differential DNA methylation and alternative splicing by looking for direct overlap of 200 bp DMRs and various types of splicing events (Supplementary Figures 2 and 3). Among the eight genes that were both differentially methylated and alternatively spliced in our data, three had a DMR directly overlapping a differentially utilized exon (Fig. 4b–d, Supplementary Figure 2) while none of the alternative splicing events discovered using rMATS overlapped directly with a DMR (Supplementary Figure 3). These genes exhibit homology to genes encoding a STE20-like serine/threonine-protein kinase (SLK; Ccalc.v2\_009097, XP\_017892454), a D-3-phosphoglycerate dehydrogenase (PHGDH; Ccalc.v2\_011692, XP\_017888762), and a 6-phosphofructo-2-kinase (PFKFB1; Ccalc.v2\_011841, XP\_017879076). Each of these represents a candidate for localized involvement of DNA methylation in gene regulation associated with maternal care in *C. calcarata*.

## Discussion

In this study, we aimed to identify behavioral and molecular changes influenced by maternal care in developing *C. calcarata*. First, we quantified behavioral changes in newly eclosed adult offspring using circle tube assays. Then we measured changes in gene expression and alternative splicing using RNA-sequencing. Finally, we characterized DMRs in the genome using bisulfite sequencing. Taken together, these data point to a complex interplay of mechanisms underlying the effects of maternal care on developing offspring.

We observed differences in aggression and avoidance behaviors in *C. calcarata* offspring reared with and without maternal care. This pattern is reminiscent of patterns observed when maternal care varies in mammalian systems<sup>34</sup> and is consistent with effects of colony disturbance in honey bees<sup>35</sup>. Adaptive and non-adaptive explanations exist for these results. For example, divergent behaviors observed between bees reared with and without maternal care may simply reflect non-adaptive effects of a sub-optimal developmental environment. However, plasticity in aggression may also be influenced by natural selection if maternal death occurs regularly in the wild or variation in maternal care is prevalent. It is possible that bees reared in the absence of adult kin may incur fitness benefits from increased aggression and avoidance behaviors relative to bees that benefit directly from cooperative defense and care by kin as adults<sup>12</sup>.

To further explore the relationship between maternal care and offspring development, we performed RNA-sequencing on the heads of newly eclosed bees that had been reared with or without a mother. Among the differentially expressed genes, we found enrichment of functional annotations related to metabolic processes, including oxidation–reduction reactions, and neurological processes, among others. These categories point to potential mechanisms for the increased aggression and avoidance behavior that we observed in offspring reared without maternal care.

Many of the differentially expressed genes we identified had functions relating to various forms of carbohydrate metabolism. While this is a very general functional category, variation in behavior has previously been linked to variation in carbohydrate metabolic activity in the brains of insects<sup>36</sup>. In particular, variation in brain glucose metabolism is apparently linked to variation in aggressive behavior in both *A. mellifera* and *Drosophila melanogaster*<sup>37</sup>. Sucrose metabolism has also been found to play a role in

the regulation of aggressive behavior in the Argentine ant, *Linepithema humile*<sup>38</sup>. It has been proposed that these fluctuations in aggression in response to metabolic changes could be mediated through variation in mitochondrial energetics in the brain<sup>36</sup>. The role of carbohydrate metabolism in mediating aggression has even been observed in vertebrate systems such as sticklebacks and mouse<sup>8</sup> indicating the possibility of conserved or convergent gene networks underlying aggression. Our differential expression data are consistent with and further support the interplay between metabolism and aggression in insects.

The most significantly enriched GO term among differentially expressed genes was “oxidation–reduction process”. Enrichment of this term has been observed among genes with expression differences linked to nourishment level and caste in *Polistes* wasps<sup>39,40</sup>. Genes relating to oxidation–reduction processes have also consistently been implicated in differential aggression in the honey bee *A. mellifera*<sup>35,41</sup>. Our size-matched *C. calcarata* offspring raised with and without a mother both received equivalent mass provisioning at the beginning of development, but nourishment may still be a confounding factor (prior experiments have demonstrated changes in aggression in response to diminished nutrient rations in *C. calcarata*<sup>42</sup>).

Eleven of the differentially expressed genes we identified are affiliated with the ionotropic glutamate receptor (IGR) signaling pathway, roughly three times the number expected by chance. The IGR pathway is involved in synaptic communication generally as well as odor perception in specific IGR subfamilies<sup>43</sup>. This class of genes offers a possible neurological explanation for the change in behavior observed in offspring raised without a mother. Fluctuations in IGR expression could be a sign of neural plasticity<sup>44</sup>, manipulating how bees respond to environmental stimuli. IGRs have even been linked to aggression phenotypes specifically in *A. mellifera*<sup>41</sup>.

We further leveraged our RNA-sequencing data to search for alternatively spliced transcripts between care and no-care individuals. We found hundreds of genes containing significantly differentially utilized exons, a simple marker of alternative splicing. Among these genes, we observed functional enrichment of oxidation–reduction processes as well as spermine biosynthesis and protein phosphorylation (Supplementary Data 1). These alternatively spliced genes may play important roles in the behavioral plasticity we observed in *C. calcarata*, as a role for alternative splicing in developmental plasticity has been documented in many species<sup>29</sup>.

In conjunction with RNA-sequencing analysis, we utilized bisulfite sequencing to quantify changes in DNA methylation in response to the loss of maternal care. We detected localized but significant changes in DNA methylation between the treatment groups, demonstrating that such differences can be detected among behaviorally distinct conspecific hymenopterans. We also confirmed previous findings that genes containing DNA methylation are more likely to be alternatively spliced than genes lacking DNA methylation (Chi-square test,  $p = 4.95e-06$ ; Supplementary Table 7). However, we did not detect a significant association between differential methylation and alternative splicing (Supplementary Table 6). This suggests that a dynamic relationship between DNA methylation and splicing may not exist in *C. calcarata*. At minimum, it is clear from our data that changes in splicing corresponding to changes in DNA methylation are the exception rather than the rule. Changes in DNA methylation do not appear to consistently lead to changes in splicing, nor vice versa, at a level observable in our data set.

What explains the lack of global concordance between variation in DNA methylation and variation in transcription in our data? One factor may be that DNA methylation acts through intermediaries to affect transcription. In particular, methylation-

sensitive transcription factors have been documented in mammalian systems<sup>23,45,46</sup> and have been shown to influence both gene expression and splicing<sup>22,45–47</sup>. In such instances, transcription factor expression profiles or splicing patterns could have a profound effect on when and in which cell types the effect of DNA methylation at a binding site would be observed. Our study investigates these associations in a restricted subset of tissues at a single developmental time point. Furthermore, many sites of DNA methylation may not bind transcription factors in any context, which would preclude gene regulation by transcription factor-mediated mechanisms at such loci.

When we examined the spatial relationship between DMRs and alternative splicing events, we found that three of eight genes with both DMRs and differentially utilized exons exhibited spatial overlap between the DMRs and alternative exon usage (Fig. 4b–d, Supplementary Figure 2). We found no genes in which a DMR directly overlapped any of the other splicing events we assessed (Supplementary Figure 3). Thus, our integrative approach to transcriptome and methylome analysis revealed three candidates for the potential interaction of DNA methylation and alternative splicing. These particular genes encode proteins with putative roles in fructose metabolism (PFKFB1<sup>48</sup>), apoptosis and stress (SLK<sup>49</sup>), and brain development (PHGDH<sup>50</sup>), each of which stands to influence organismal behavior. One of these genes, 6-phosphofructo-2-kinase (PFKFB1), was the sole gene to exhibit significant differential expression, alternative splicing, and differential methylation in response to removal of maternal care. A second gene, *sh2b adaptor protein 1-like* (SH2B1<sup>51</sup>), also exhibits significant differential expression, alternative splicing, and differential methylation in response to removal of maternal care but does not exhibit direct overlap between differential methylation and alternative splicing event locations. SH2B1 has a putative role in growth factor signaling. These loci represent prime candidates for follow-up investigations of tissue-specific transcription and methylation over the course of *C. calcarata* development.

In conclusion, removal of the mother during early *C. calcarata* development culminates in offspring that exhibit increased aggression and avoidance behaviors, have over one thousand differentially expressed genes, have hundreds of alternatively spliced genes, and exhibit subtle changes in DNA methylation. We identified specific functional groups of genes that may play a role in developmental plasticity and aggression: carbohydrate metabolism genes, oxidation–reduction process genes, and IGR genes. We observed no genome-wide association between differential DNA methylation and differential gene expression or splicing, though we cannot rule out more complex, indirect relationships between these factors. Our study offers unique insights into the effects of maternal care on offspring development and explores understudied questions about the role of DNA methylation in transcriptional regulation in invertebrate taxa. Further exploration of the candidate genes we identified may help to explain the genetic underpinnings of social behaviors in *C. calcarata* and provide insight into how these genes are regulated in response to environmental stimuli.

## Methods

**Sample collection.** Nests of our study species, *C. calcarata*, were collected from naturally occurring dead broken sumac (*Rhus typhina*) stems in Durham, NH in June 2015 during the active brood rearing early phase of the colony cycle<sup>9,10</sup>. Nests were collected from the field so that mothers could forage and feed offspring as they normally would in the wild. Stems were collected prior to 8 am to ensure the mother would be home. Nest entrances were covered with masking tape and transported to the lab where they were split lengthwise to record all nest contents. During this early active brood phase mothers provision eggs with pollen balls and offspring were no older than small larvae. Pollen masses with eggs were reared in lab with entire nests at room temperature until eclosion of adult offspring in

August 2015. To test the effect of maternal care on developing offspring, we generated “care” and “no care” treatment groups among the lab-reared nests for comparison. In the “care” treatment, mothers were left with the nest and offspring were raised from egg to adulthood in the presence of maternal care. In the “no care” treatment, mothers were removed from the nest after the initial mass provisioning and offspring were raised from egg to adulthood in the absence of maternal care. Pollen and eggs were retained so this was not a manipulation of maternal investment, but rather in half of the lab-reared nests mothers were removed and in the other half mothers remained, such that they could engage in offspring grooming, a stereotypical behavior they complete each evening (reviewed in Rehan et al.<sup>10</sup>). In the following experiments, we assayed one age- and size-matched adult female offspring per nest to determine the behavioral effects of early termination of maternal care.

**Behavioral assays.** Circle tubes are a well-established method for quantifying behavior in this and other bee species<sup>11,52</sup>. Interactions among individuals in circle tube arenas (e.g. aggression, avoidance, and tolerance behaviors) parallel those expressed in observation colonies<sup>53</sup>.

Two dark-winged (five-day-old) offspring bees from different nests were paint marked for unique identification and introduced simultaneously into opposite ends of a clean plastic circle tube<sup>54</sup>. A new piece of tubing was used for each dyad. Observations of each pair lasted for 20 min. An encounter between bees was recorded when individuals came within one body length of each other<sup>55</sup>, and aggression and avoidance behaviors were scored<sup>11</sup>. We conducted 52 circle tube trials. Two females from different nests were paired and no female was assayed twice. Therefore, 104 female offspring in total were assayed in this experiment.

**RNA-sequencing and quantification.** For each sample, RNA was extracted from the heads of three adult female bees and pooled together. Heads were utilized to ensure comparability with bisulfite sequencing data, which have high DNA input demands (see below), while a pooling strategy was leveraged to minimize individual variation in gene expression and maximize our ability to detect the effect of maternal care specifically. All bees utilized for this analysis were from different nests. Each pool was then subject to an Illumina Truseq library preparation protocol to provide a genome-wide profile of mRNA transcript abundance. Three biological replicates were prepared from individuals reared in the presence of the mother and three biological replicates were prepared from individuals reared in the absence of a mother. We note that our statistical power and ability to detect weakly differentially expressed genes is dampened relative to an experimental design featuring more biological replicates, and the sequencing of RNA and DNA from multiple tissues (heads) dampens our data’s sensitivity to tissue-specific differences. Library sequencing was performed on an Illumina HiSeq platform by Genome Quebec (Montreal). Paired-end, 100 bp reads were trimmed using Trimmomatic (v0.32)<sup>56</sup>. Reads were then aligned to the *C. calcarata* reference assembly (v1.1, GCA\_001652005.1)<sup>13</sup> using Tophat<sup>57</sup> with parameters “-no-mixed-no-discordant” to filter out read pairs that did not have both reads mapping or had a read pair mapping discordantly. Each sample produced between 92 and 109 million aligned reads. A target of one hundred million mapped reads per sample is considered much greater depth than necessary for differential expression analysis, but was chosen for this project to enhance detection of alternative splicing<sup>58</sup>. At this point, multiple pipelines were utilized, each using the Tophat output as input.

**Computing differential expression with edgeR.** Aligned reads from each library were mapped to the existing gene models<sup>13</sup> using Rsubread featureCounts<sup>59</sup>. Resulting counts files were then merged and input into edgeR<sup>60</sup>. Genes with more than three libraries being represented by 0.18 counts per million reads, translating to at least 10 reads aligned to a gene, were kept for further analysis. RPKM values were calculated for the remaining genes and the edgeR Exact Test was used to determine which genes were differentially expressed. Differences were considered significant if they had an FDR-corrected  $p$ -value  $< 0.01$ .

**Computing differential exon usage with DEXseq.** Aligned reads from each library were mapped to the existing gene models using the script provided by DEXseq<sup>61</sup>. DEXseq was then used with the standard implementation to compute differential exon usage between the maternal care and no care groups. Exons were considered differentially utilized if they had an FDR-corrected  $p$ -value  $< 0.01$ . The native plotting function in DEXseq was used to plot the exon usage of genes that contained a differentially utilized exon as well as a DMR.

**Additional alternative splicing metrics from rMATS.** Aligned reads spanning junctions were quantified using rMATS<sup>62</sup>. Genes containing an alternative splicing event with an FDR-corrected  $p$ -value  $< 0.01$  were considered significantly alternatively spliced. Alternative splicing events in genes that also contained DMRs were visualized using the DEXseq plotting function. rMATS is ideal for identifying a variety of alternative splicing event types, but its dependence on junction-spanning reads makes it liable to miss alternative splicing events that occur in low proportion. DEXseq, while only able to detect specific types of splicing events, leverages more total reads to yield a more thorough analysis of differential exon usage.

**Bisulfite sequencing and quantification.** Genomic DNA was extracted and pooled from the heads of nine adult females from different nests for each sample used for bisulfite conversion and Illumina sequencing, in order to provide a nucleotide-resolution genomic profile of DNA methylation. Pooling was used to minimize the effects of individual variation while heads were used to accommodate the high DNA input requirement for bisulfite conversion. Three biological replicates each were sequenced (90 bp, paired-end reads after barcode removal) from individuals raised in the presence of maternal care and from individuals raised in the absence of maternal care. Unmethylated Enterobacteria phage lambda DNA (GenBank accession: J02459.1) was added to each sample as a control for bisulfite conversion. Bisulfite conversion and sequencing were performed by BGI (Shenzhen, China) on the Illumina HiSeq platform. We trimmed adapters and removed low-quality reads using Trimmomatic (v0.32)<sup>56</sup>. Reads were then aligned to the reference genome using Bismark (v0.14.5)<sup>63</sup>. Bismark was also used to remove duplicate reads and to extract and quantify the per-site methylation level. We determined that any strand-asymmetric methylation (hemimethylation) was likely due to low coverage (Supplementary Figure 1A, B), and with this information in mind, we merged the strands so that each CpG dinucleotide was represented by one methylation value. We generated an average coverage depth between 21X and 37X reads in each sample at each strand merged CpG. The nonconversion rate was then computed based upon the methylation level in the lambda DNA and this value was used to compute the methylation status for each site based upon an FDR-corrected Binomial test performed in R.

A global methylation level for each gene was computed as the total number of methylated (unconverted) reads summed over all CpG sites divided by the total number of reads summed over all CpG sites. These levels were then plotted to determine a reasonable cutoff for a gene to be considered “methylated” (Supplementary Figure 1C). We chose global CpG methylation (all methylated CpG reads/all CpG reads)  $> 5\%$  for a given gene as a reasonable cutoff. All genes passing this cutoff in at least one of our six samples were compared against a background of the full *C. calcarata* gene set using TopGO. Global methylation and CpG density (total number of CpGs/total number of nucleotides) were computed for each gene feature and a metaplot was created (Supplementary Figure 1D).

**Differential methylation analysis.** After extraction of the methylation levels for each CpG, the values were filtered down to only those CpGs that had significant methylation in at least one sample (FDR-corrected binomial test  $p < 0.05$  with null expectation set by the nonconversion rate of the bisulfite reaction). This was done to diminish background noise in the differential methylation analysis. Following this filtration, MethylKit<sup>64</sup> was used to compute the differential methylation between the “no maternal care” and “maternal care” samples. CpGs with no coverage in any sample were removed, as were CpGs with coverage  $> 170$  in any sample to remove potential PCR artifacts. CpGs were then binned into 200 bp windows and windows that contained fewer than 2 CpGs were removed. The median exon length in *C. calcarata* is 176 bp, while the mean is 355 bp; we used 200 bp windows to approximate exon size so that we could readily map differential methylation to genomic features. Finally, differential methylation was computed for each region using Benjamini–Hochberg FDR correction<sup>65</sup> with a threshold set as minimum methylation difference of 5% and maximum FDR-corrected  $p$ -value of 0.01. Significantly DMRs were then checked for genic localization and a list of genes containing DMRs was created.

In order to ensure that the differential DNA methylation we detected was driven by biological factors, differential methylation analysis was also performed on the nine possible permutations of sample ID assignment, shuffling “maternal care” and “no maternal care” identity using a modified QAM<sup>66</sup>. For this QAM analysis, each 200 bp genomic region was given a probability of being a DMR based on how many of the nine permutations yielded a significant result at that locus. Then 50,000 random, region-independent permutations of the analyses were generated, and the number of DMRs in each iteration was tallied. This method is similar to the permutation method used in Libbrecht et al.<sup>67</sup>, except we utilized 200 bp windows rather than individual CpGs and computed differential methylation for each permutation using MethylKit<sup>64</sup> rather than a series of binomial tests on each sample. These modifications allow us to robustly identify DMRs even when coverage is low in some samples at the given region.

**Exploring and comparing the significant gene sets.** The package UpSet<sup>68</sup> was used to compute the overlap between the lists of significant genes from analyses of differential expression, alternative splicing, and differential methylation. The R package “pvclust”<sup>69</sup> was used to determine the degree of clustering between the two treatment groups based on gene expression and genic methylation. GO term annotations from a previous study<sup>13</sup> were used to assess GO term enrichment for each list of significant genes using the default topGO<sup>70</sup> “weight01” method and the “fisher” statistic. GO terms were considered significantly enriched if they had a  $p$ -value  $< 0.05$ . We also mapped *D. melanogaster* orthologs of *C. calcarata* genes using a reciprocal best BLAST hit approach and imported Online Mendelian Inheritance in Man phenotype information associated with these fly orthologs for perusal (Supplementary Data 2).

**Data availability.** Raw reads and analyzed data for both the RNA-sequencing and bisulfite sequencing are available at the Gene Expression Omnibus accession [GSE111611](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111611). All other data supporting the findings of this study are available within the paper and its supplementary files.

Received: 20 March 2018 Accepted: 31 July 2018

Published online: 27 August 2018

## References

- West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford University Press, 2003).
- Fish, E. W. et al. Epigenetic programming of stress responses through variations in maternal care. *Ann. NY Acad. Sci.* **1036**, 167–180 (2004).
- Miura, T. Developmental regulation of caste-specific characters in social-insect polyphenism. *Evol. Dev.* **7**, 122–129 (2005).
- Curley, J. P., Jensen, C. L., Mashoodh, R. & Champagne, F. A. Social influences on neurobiology and behavior: epigenetic effects during development. *Psychoneuroendocrinology* **36**, 352–371 (2011).
- Weaver, I. C. G. et al. Epigenetic programming by maternal behavior. *Nat. Neurosci.* **7**, 847–854 (2004).
- Weaver, I. C. G., Meaney, M. J. & Szyf, M. Maternal care effects on the hippocampal transcriptome and anxiety-mediated behaviors in the offspring that are reversible in adulthood. *Proc. Natl. Acad. Sci. USA* **103**, 3480–3485 (2006).
- Rittschof, C. C., Coombs, C. B., Frazier, M., Grozinger, C. M. & Robinson, G. E. Early-life experience affects honey bee aggression and resilience to immune challenge. *Sci. Rep.* **5**, 1–8 (2015).
- Rittschof, C. C. et al. Neuromolecular responses to social challenge: common mechanisms across mouse, stickleback fish, and honey bee. *Proc. Natl. Acad. Sci. USA* **111**, 17929–17934 (2014).
- Rehan, S. M. & Richards, M. H. Nesting biology and subsociality in *Ceratina calcarata* (Hymenoptera: Apidae). *Can. Entomol.* **142**, 65–74 (2010).
- Rehan, S. M., Berens, A. J. & Toth, A. L. At the brink of eusociality: transcriptomic correlates of worker behaviour in a small carpenter bee. *BMC Evol. Biol.* **14**, 260 (2014).
- Rehan, S. M. & Richards, M. H. Reproductive aggression and nestmate recognition in a subsocial bee. *Anim. Behav.* **85**, 733–741 (2013).
- Mikát, M., Franchino, C. & Rehan, S. M. Sociodemographic variation in foraging behavior and the adaptive significance of worker production in the facultatively social small carpenter bee, *Ceratina calcarata*. *Behav. Ecol. Sociobiol.* **71**, 135 (2017).
- Rehan, S. M., Glastad, K. M., Lawson, S. P. & Hunt, B. G. The genome and methylome of a subsocial small carpenter bee, *Ceratina calcarata*. *Genome Biol. Evol.* **8**, 1401–1410 (2016).
- Glastad, K. M. et al. Variation in DNA methylation is not consistently reflected by sociality in Hymenoptera. *Genome Biol. Evol.* **9**, 1687–1698 (2017).
- Herb, B. R. et al. Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nat. Neurosci.* **15**, 1371–1373 (2012).
- Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**, 1827–1830 (2008).
- Yan, H. et al. DNA methylation in social insects: how epigenetics can control behavior and longevity. *Annu. Rev. Entomol.* **60**, 435–452 (2015).
- Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).
- Bestor, T. H., Edwards, J. R. & Boulard, M. Notes on the role of dynamic DNA methylation in mammalian development. *Proc. Natl. Acad. Sci. USA* **112**, 6796–6799 (2014).
- Laurent, L. et al. Dynamic changes in the human methylome during differentiation. *Genome Res.* 320–331 (2010). <https://doi.org/10.1101/gr.101907.109>
- Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **39**, 61–69 (2007).
- Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
- Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, pii: eaaj2239 (2017).
- Foret, S. et al. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc. Natl. Acad. Sci. USA* **109**, 4968–4973 (2012).
- Li-Byarlay, H. et al. RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proc. Natl. Acad. Sci. USA* **110**, 12750–12755 (2013).
- Flores, K. et al. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genom.* **13**, 1–9 (2012).
- Glastad, K. M., Gokhale, K., Liebig, J. & Goodisman, M. A. D. The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. *Sci. Rep.* **6**, 1–14 (2016).
- Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
- Kalsotra, A. & Cooper, T. A. Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* **12**, 715–729 (2011).
- Wang, X. et al. Function and evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genet.* **9**, e1003872 (2013).
- Bonasio, R. et al. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* **22**, 1755–1764 (2012).
- Hunt, B. G., Glastad, K. M., Yi, S. V. & Goodisman, M. A. D. Patterning and regulatory associations of DNA methylation are mirrored by histone modifications in insects. *Genome Biol. Evol.* **5**, 591–598 (2013).
- Foret, S., Kucharski, R., Pittelkow, Y., Lockett, G. A. & Maleszka, R. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. *BMC Genom.* **10**, 472 (2009).
- Meaney, M. J. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annu. Rev. Neurosci.* **24**, 1161–1192 (2001).
- Rittschof, C. C. & Robinson, G. E. Manipulation of colony environment modulates honey bee aggression and brain gene expression. *Genes, Brain Behav.* **12**, 802–811 (2013).
- Rittschof, C. C., Vekaria, H. J., Palmer, J. H. & Sullivan, P. G. Brain mitochondrial bioenergetics change with rapid and prolonged shifts in aggression in the honey bee, *Apis mellifera*. *J. Exp. Biol.* (2018). <https://doi.org/10.1242/jeb.167304>
- Li-Byarlay, H., Rittschof, C. C., Massey, J. H., Pittendrigh, B. R. & Robinson, G. E. Socially responsive effects of brain oxidative metabolism on aggression. *Proc. Natl. Acad. Sci. USA* **111**, 12533–12537 (2014).
- Grover, C. D., Kay, A. D., Monson, J. A., Marsh, T. C. & Holway, D. A. Linking nutrition and behavioural dominance: carbohydrate scarcity limits aggression and activity in Argentine ants. *Proc. R. Soc. B Biol. Sci.* **274**, 2951–2957 (2007).
- Berens, A. J., Hunt, J. H. & Toth, A. L. Nourishment level affects caste-related gene expression in *Polistes* wasps. *BMC Genom.* **16**, 1–12 (2015).
- Toth, A. L. et al. Shared genes related to aggression, rather than chemical communication, are associated with reproductive dominance in paper wasps (*Polistes metricus*). *BMC Genom.* **15**, 1–14 (2014).
- Alaux, C. et al. Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proc. Natl. Acad. Sci. USA* **106**, 15400–15405 (2009).
- Lawson, S. P., Helmreich, S. L. & Rehan, S. M. Effects of nutritional deprivation on development and behavior in the subsocial bee *Ceratina calcarata* (Hymenoptera: Apidae). *J. Exp. Biol.* (2017). <https://doi.org/10.1242/jeb.160531>
- Croset, V. et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS. Genet.* **6**, e1001064 (2010).
- Asztely, F. & Gustafsson, B. Ionotropic glutamate receptors. Their possible role in the expression of hippocampal synaptic plasticity. *Mol. Neurobiol.* **12**, 1–11 (1996).
- Maunakea, A. K., Chepelev, I., Cui, K. & Zhao, K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23**, 1256–1269 (2013).
- Kim, J., Kollhoff, A., Bergmann, A. & Stubbs, L. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, *Peg3*. *Hum. Mol. Genet.* **12**, 233–245 (2003).
- Wan, J. et al. Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs. *Nucleic Acids Res.* **41**, 8503–8514 (2013).
- Pilkis, S. J., Claus, T. H., Kurland, I. & Lange, A. J. 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase: a metabolic signaling enzyme. *Annu. Rev. Biochem.* **64**, 799–835 (1995).
- Dan, I., Watanabe, N. M. & Kusumi, A. The Ste20 group kinases as regulators of MAP kinase cascades. *Trends Cell Biol.* **11**, 220–230 (2001).
- Jaeken, J. et al. 3-Phosphoglycerate dehydrogenase deficiency: inborn error of serine biosynthesis. *Arch. Dis. Child.* **74**, 542–545 (1996).
- Nishi, M. et al. Kinase activation through dimerization by human SH2-B. *Mol. Cell Biol.* **25**, 2607–2621 (2005).
- Richards, M. H. & Packer, L. Social behaviours in solitary bees: Interactions among individuals in *Xeralictus bicuspidariae* Snelling (Hymenoptera: Halictidae: Rophitinae). *J. Hymenopt. Res.* **19**, 66–76 (2010).
- Brothers, D. J. & Michener, C. D. Interactions in colonies of primitively social bees—III. Ethometry of division of labor in *Lasiglossum zephyrum* (Hymenoptera: Halictidae). *J. Comp. Physiol.* **90**, 129–168 (1974).



54. Wcislo, W. T. Social interactions and behavioral context in a largely solitary bee, *Lasioglossum (Dialictus) figueresi* (Hymenoptera, Halictidae). *Insectes Soc.* **44**, 199–208 (1997).
55. Packer, L. The influence of marking upon bee behaviour in circle tube experiments with a methodological comparison among studies. *Insectes Soc.* **52**, 139–146 (2005).
56. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
57. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
58. Conesa, A. et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 1–19 (2016).
59. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, 1–17 (2013).
60. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
61. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017 (2012).
62. Shen, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. USA* **111**, E5593–E5601 (2014).
63. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
64. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
65. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
66. Cheverud, J. M., Wagner, G. P. & Dow, M. M. Methods for the comparative analysis of variation patterns. *Syst. Zool.* **38**, 201–213 (1989).
67. Libbrecht, R., Oxley, P. R., Keller, L. & Kronauer, D. J. C. Robust DNA methylation in the clonal raider ant brain. *Curr. Biol.* **26**, 391–395 (2016).
68. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2–5** (2017). <https://doi.org/10.1101/120600>
69. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
70. Alexa, A. & Rahnenfuhrer, J. *topGO: Enrichment Analysis for Gene Ontology*. R package (2016).

## Acknowledgements

This work was supported by the University of New Hampshire and the University of Georgia College of Agricultural and Environmental Sciences. We thank Wyatt Shell for the illustrations used in Fig. 1 and Karl Glastad for bioinformatic consultation.

## Author contributions

S.M.R. conceived of the study. S.M.R. and B.G.H. designed the study. S.M.R. collected specimens, performed rearing experiments, performed behavioral assays, and prepared samples for sequencing. S.V.A., B.G.H., and S.M.R. conceived of bioinformatic analyses; S.V.A. performed them. S.V.A. wrote the manuscript with feedback and contributions from B.G.H. and S.M.R.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-05903-0>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018