4-4-2016

The Genome and Methylome of a Subsocial Small Carpenter Bee, Ceratina calcarata

Sandra M. Rehan
University of New Hampshire, Durham, sandra.rehan@unh.edu

Karl M. Glastad
Georgia Institute of Technology

Sarah P. Lawson
University of New Hampshire, Durham

Brendan G. Hunt
University of Georgia

Follow this and additional works at: https://scholars.unh.edu/nhaes

Recommended Citation
The Genome and Methylome of a Subsocial Small Carpenter Bee, Ceratina calcarata

Sandra M. Rehan1,*,†, Karl M. Glastad2,†, Sarah P. Lawson1, and Brendan G. Hunt3

1Department of Biological Sciences, University of New Hampshire, Durham
2School of Biology, Georgia Institute of Technology
3Department of Entomology, University of Georgia
†These authors contributed equally to this work.
*Corresponding author: E-mail: sandra.rehan@unh.edu.
Accepted: March 29, 2016

Abstract

Understanding the evolution of animal societies, considered to be a major transition in evolution, is a key topic in evolutionary biology. Recently, new gateways for understanding social evolution have opened up due to advances in genomics, allowing for unprecedented opportunities in studying social behavior on a molecular level. In particular, highly eusocial insect species (caste-containing societies with nonreproductives that care for siblings) have taken center stage in studies of the molecular evolution of sociality. Despite advances in genomic studies of both solitary and eusocial insects, we still lack genomic resources for early insect societies. To study the genetic basis of social traits requires comparison of genomes from a diversity of organisms ranging from solitary to complex social forms. Here we present the genome of a subsocial bee, Ceratina calcarata. This study begins to address the types of genomic changes associated with the earliest origins of simple sociality using the small carpenter bee. Genes associated with lipid transport and DNA recombination have undergone positive selection in Ceratina calcarata relative to other bee lineages. Furthermore, we provide the first methylome of a noneusocial bee. Ceratina calcarata contains the complete enzymatic toolkit for DNA methylation. As in the honey bee and many other holometabolous insects, DNA methylation is targeted to exons. The addition of this genome allows for new lines of research into the genetic and epigenetic precursors to complex social behaviors.

Key words: small carpenter bee, epigenetics, maternal care, social evolution, insect genome.

Introduction

The goal of sociogenomics is to explore the molecular genetic basis of social life (Robinson et al. 2005). This field was spearheaded by sequencing the first genome of an eusocial species, the honey bee, Apis mellifera (Honeybee Genome Sequencing Consortium 2006). Following the publication of the honey bee genome, researchers sequenced the genomes of many other eusocial insect species and their solitary relatives, including seven ants (Bonasio et al. 2010; Nygaard et al. 2011; Smith, Smith, et al. 2011; Smith, Zimin, et al. 2011; Suen et al. 2011; Wurm et al. 2011) and nine additional bees (Kocher et al. 2013; Kapheim et al. 2015; Sadd et al. 2015). These genomes revealed much about the importance of gene family expansions, chemical communication, and epigenetics to caste determination (Gadau et al. 2012; Simola et al. 2013; Yan et al. 2014). Although we are beginning to elucidate the genetic basis of highly eusocial behaviors, hallmarked by overlapping generations, reproductive division of labor, and cooperative brood care, we still understand very little about the genetic architecture of incipient sociality. Indeed, identification of the genetic precursors to advanced eusociality necessitates a comparison of highly social species with those species possibly transitioning to highly social behaviors.

A prominent theory in the transition to sociality is that incipient groups consisted of a mother and her offspring (Alexander et al. 1991). Maternal care is thus a key step in the evolution of social behavior (Michener 1974; Anderson 1984; Gadagkar 1990). Subsocial species often display behaviors thought to be necessary precursors to highly social behavior, such as maternal care, nest loyalty, cohabitation, and adult longevity (Sakagami and Maeta 1977; Tallamy and Wood 1986; Crespi 1994; Costa 2006; Wilson 2008). To date, there are published genomes of ten bee species (Kocher et al. 2013; Kapheim et al. 2015; Sadd et al. 2015), but there are no sequenced genomes of a subsocial bee species.
Because the evolutionary path from solitary and eusocial behavior likely involves a subsocial intermediate, comparisons between solitary and subsocial species closely related to eusocial taxa may provide evidence of the genetic scaffolding for more complex social behaviors (West-Eberhard 1967; Rehan and Toth 2015). Ceratina small carpenter bees (Hymenoptera: Apidae) are a diverse and cosmopolitan genus (Michener 2007). Most species of this genus are considered solitary, but sociality occurs naturally in some species (Sakagami and Laroca 1971; Sakagami and Maeta 1984; Rehan et al. 2009; Rehan et al. 2014) and can be induced artificially in others (Chandler 1975; Sakagami and Maeta 1987, 1989, 1995). This natural variation provides a unique opportunity to explore the transition from solitary to social life (Michener 1985). Ceratina are members of the subfamily Xylocopinae in the bee family Apidae and provide a phylogenetic contrast and independent origin of eusociality to honey bees and bumble bees, members of the subfamily Apinae (Cardinal and Danforth 2011; Rehan et al. 2012).

Ceratina calcarata (Hymenoptera: Apidae) is a small carpenter bee native to eastern North America (Rehan and Sheffield 2011; Shell and Rehan 2016). Unlike solitary species, such as the leaf cutter bee Megachile rotundata, C. calcarata displays prolonged maternal care, mother–adult offspring interactions, and mutual tolerance (Rehan et al. 2014). Ceratina females make their nests in the pith of dead, broken twigs (Daly 1973). The mother lays eggs within the stem and provides each offspring with its own nutrition-rich pollen ball (McFrederick and Rehan 2016). As the offspring grow, the mother will protect and groom her offspring (Michener 1985; Rehan and Richards 2010a). Ceratina calcarata females are long-lived and nest loyal, and provide care to their offspring throughout development and even into adulthood (Rehan and Richards 2010b). In spring, overwintered females emerge from their hibernacula, dispersing to construct new nests in twigs and stems. In summer, mothers remain in the nest to guard against predators and parasites. In autumn, offspring emerge and are primarily fed by the mother or female siblings prior to overwintering in the natal nest (Rehan et al. 2014). Male and female offspring overwinter together as adults and all individuals disperse and mate in the spring. Ceratina calcarata is of special interest to the study of social evolution because in addition to prolonged mother–offspring interactions and overlapping generations this species also exhibits reproductive division of labor (Rehan et al. 2014). Mothers have complete control over the size and sex of their offspring and produce a special class of dwarf eldest daughters as well as regular daughters (Rehan and Richards 2010b). Regular daughters emerge in the natal nest, do not forage, and wait until spring to disperse and establish their own nests solitarily. Conversely, dwarf eldest daughters are first to emerge in autumn, forage and feed siblings prior to overwintering, and do not survive the winter (Rehan and Richards 2010a). The sibling care without reproduction exhibited in C. calcarata is comparable in some ways with worker behavior exhibited in eusocial species such as the honey bee. However, C. calcarata exhibits alloprenatal care but not cooperative brood care in the strict sense (Michener 1974). The distinction here is that dwarf eldest daughters feed adult siblings but do not contribute to the next generation by provisioning immature offspring. Thus, C. calcarata exhibits two of the three hallmarks of eusociality: Reproductive division of labor and overlapping generations, yet this species does not exhibit cooperative brood care and therefore remains a bee on the brink of eusociality (Rehan et al. 2014). This system provides an excellent model for studying the genetic and epigenetic factors associated with maternal care, sibling care, and transitional stages of social behavior. Here we present the genome of the small carpenter bee, C. calcarata. We also compare these data with the previously published bee genomes to identify genomic regions under selection at the brink of sociality. Finally, we provide the DNA methylome of C. calcarata. The hypothesis that DNA methylation in bee genomes may be correlated with social complexity has garnered recent support from an indirect measure of DNA methylation (Kapheim et al. 2015). However, empirical assessment of DNA methylation has remained limited among bees to A. mellifera. Our study provides the first direct comparison of DNA methylation in A. mellifera with a bee with markedly less complex social interactions.

Methods and Materials

Sample Preparation and Sequencing

DNA was isolated from the whole body of two haploid male C. calcarata collected in Durham, NH, in the winter of 2013. DNA extractions were performed using a Qiagen Genomic-tip 20G kit (Valencia, CA) and standard protocol. Genomic DNA was used to generate nonamplified 150-bp DNA libraries for Illumina sequencing. In order to improve genome assembly, genomic DNA was extracted from the whole body of ten female C. calcarata as above, pooled, and used to construct 3–4 kb and 5–8 kb mate pair libraries. Four paired-end libraries with insert sizes ranging from 150 bp to 8 kb were constructed and sequenced on Illumina HiSeq 2500. Before filtering this produced 45 Gb of raw sequence data (150.5 million read pairs, 2 × 150 bp each). Low quality reads, reads with a high proportion of Ns, or poly-A structures were filtered and paired-end reads whose mates overlapped were merged prior to assembly. PRINSEQ (Schneider and Edwards 2011) was used to filter out reads containing a high number of Ns or poly-A 5’ or 3’ runs (>15 poly-A nts, > 50% Ns). FLASH (Magoc and Salzberg 2011) was used to merge overlapping reads of the same pair (due to fragment size < read length × 2 on many reads). All genomic data have been submitted to public repositories and can be found using the NCBI BioProject number PRJNA299559.
RepeatModeler (Jurka et al. 2005). Functional annotation of custom species-specific repeat library generated by Repbase transposable element library (“species = all”) and a parameters to identify repetitive elements based on the previous publication (Holt and Yandell 2011).

Gene Family Expansion and Contraction
To identify orthologous groups for analysis of gene family expansion and contraction, we used OrthoMCL (Li et al. 2003) to assign proteins from our study species, C. calcarata, and ten other bee species: Apis florea, A. melifera, Bombus impatiens, Bombus terrestris, Dufourea novaangliae, Eufriesia mexicana, Habropoda laboriosa, Lasiosglossum albipes, Melipona quadrid fasciata, and M. rotundata into gene families. Briefly, the longest transcript isoform only for each gene was used and we only considered proteins larger than 50 amino acids. An all-versus-all sequence search was performed using BLASTP with default parameters and cut-off E-values of $10^{-5}$ and 50% match.

The rate optimization was computed by COUNT using the default parameters and posterior probabilities for each gene family were given. These gene families were then described using Pfam (Finn et al. 2014) domains and keywords and subsequently sorted into GO terms (Ashburner et al. 2000). By comparing the GO/IRP/KEGG enrichment results for C. calcarata with the previously published bee genomes, we identified genes specific to C. calcarata. To identify genes lost or absent from the C. calcarata genome, we identified genes found in the other ten bee species, but not in C. calcarata (see Methods in supplementary material S1, Supplementary Material online).

Molecular Evolution
For alignments used in PAML analyses, we identified sequences with homology to putative 1-to-1 ortholog groups previously established among ten bee genomes by OrthoDB (Kriventseva et al. 2015). We first performed reciprocal BLASTP searches between the C. calcarata OGS and published
gene sets of each of the ten bee species. *Ceratina calcarata* genes were then assigned to a given ortholog group if the *C. calcarata* gene was a reciprocal best hit to the respective ortholog in at least 60% of the ten bees. For multiple sequence alignment, PRANK was used to align amino acid sequences then back-translate to coding sequences (CDS) (with the option “+F”) from within the Guidance wrapper (v.2.0; Löytynoja and Goldman 2005; Sela et al. 2015). Initial gene trees produced by guidance were then improved using RAxML (Stamatakis 2014) with the GTRGAMMA model. Gblocks was then used to extract higher confidence alignment blocks, while maintaining reading frame, for downstream analysis (Talavera and Castresana 2007). PAML (v.4.7) was then used to estimate synonymous and nonsynonymous substitution rates (dS and dN) and ω (dN/dS) with free ratios for each branch (Yang 2007). Values corresponding to the *C. calcarata* branch were used in analyses. To test for further evidence of positive selection, we used the branch-site test of positive selection (branch-site model A, test 2; Zhang et al. 2005), which uses likelihood ratio tests (LRTs) to detect positive selection on a small number of sites along a specific lineage. *Ceratina calcarata* was set as the foreground branch, with two dN/dS ratios for branches (model = 2). For the alternative model, ω2 was estimated from the data (fix_omega = 0, omega = 1, NSsites = 2), while the null model fixes ω2 at 1 (fix_omega = 1, omega = 1, NSsites = 2) as described in PAML documentation and sample files (Yang 2007). The log-likelihoods for the null and alternative models were used to calculate an LRT statistic, which was then compared against the $\chi^2_{df} = 1$ distribution. P values were then corrected for multiple testing using the method of Benjamini and Hochberg (1995), with a corrected P value threshold of 0.05 being used to assign significance.

**Species Phylogeny**

Four-fold degenerate sites were extracted from codon sequence alignments after treating with Gblocks. The msa_view utility (from the PhastCons program suite; Siepel and Haussler 2004) was used to extract and concatenate 4-fold degenerate sites from the alignments. PhyloFit (REV model) was then used to fit tree models to subwindows (l = 20,000; step = 10,000) of the concatenated 4-fold degenerate sites, which were then averaged using phyloBoot (supplementary material S1, Supplementary Material online).

**Gene Expression**

RNA-seq reads (Rehan et al. 2014; Durant et al. 2016) were aligned to the reference genome using Tophat2 with default options (bar limiting maximum intron length to ≤ 25000). Cuffnorm was used to produce normalized (geometric) FPKM values for genes (Trapnell et al. 2012). Raw data are available on the NCBI Sequence Read Archive with accession number SRX541305. Six classes of females were assayed and are represented in these studies: Spring mothers, summer mothers, autumn mothers, autumn workers, autumn gynes, and overwintering gynes. Each of the six RNA-seq libraries comprised whole brain tissue samples from 20-pooled females. Reads were filtered for quality (threshold greater than or equal to 20) with a length threshold of 50 bases. After adapter removal, approximately 20% of the reads are represented.
were removed from the libraries after quality filtering (98 million raw Illumina reads; 79 million reads after preprocessing).

DNA Methylation

In order to assess DNA methylation at a single-nucleotide resolution, genomic DNA was extracted from ten whole body females using a standard phenol–chloroform protocol and then pooled into one sample for bisulfite conversion and sequencing. Unmethylated enterobacteria phage lambda DNA (GenBank accession: J02459.1) was added to C. calcarata genomic DNA as a control for bisulfite conversion efficiency. Bisulfite conversion and sequencing on the Illumina HiSeq platform was performed by Beijing Genomics Institute (Shenzhen, China). We obtained 37.9× mean coverage of genomic CpG sites. We used the program Bismark (Krueger and Andrews 2011) to align raw sequencing reads to the C. calcarata reference genome. Apis mellifera bisulfite-sequencing data from a worker whole body (Zemach et al. 2010) and from queen and worker brains (Lyko et al. 2010) were previously published, but raw reads were mapped and processed in the same manner as C. calcarata. For comparison with C. calcarata DNA methylation, we first examined data from an A. mellifera worker whole body library (Zemach et al. 2010) to achieve comparability in tissues assayed. Second, we examined data produced by merging worker whole body, worker brain, and queen brain libraries (Lyko et al. 2010; Zemach et al. 2010) to achieve comparability in read depth between species. Results from each comparison were qualitatively similar.

Significantly methylated CpG sites were assessed using a binomial test, which incorporated deamination rate (from our unmethylated control) as the probability of success, and assigned a significance value to each CpG site related to the number of unconverted reads (putatively methylated Cs) as they compare with the expected number from control (Lyko et al. 2010). Resulting \( P \) values were then adjusted for multiple testing using the method of Benjamini and Hochberg (1995). Only sites with false discovery rate (FDR) corrected binomial \( P \) values <0.05 and \( >3 \) reads sequence coverage were considered “methylated.”

Fractional methylation values were calculated for each CpG site as mCG/CG, where mCG is the number of reads with a methylated cytosine at a CpG site and CG is the total number of reads mapped to the site. Fractional methylation was calculated for specific genomic features (e.g., exons and introns) as the mean of all CpG fractional methylation values within that feature.

Results and Discussion

Genome and Annotation

The C. calcarata genome size is estimated at 364 Mb after using k-mer-based error correction in QUAKE (Kelley et al. 2010). The final assembly has an N50 of 74 kb and a total length of 261.9 Mb (table 1). Completeness of the assembly was first assessed using the CEGMA pipeline (Parra et al. 2007). Of the 248 core eukaryotic genes (CEGs), 247 CEGs were completely assembled in the C. calcarata genome. A combination of RNA-sequencing, de novo, and homology-based predictions generated the unfiltered gene set including 20,693 predicted genes. A more conservative “filtered” official gene set was produced including 11,310 predicted genes. The closest relative to C. calcarata with a genome sequence is the honey bee, A. mellifera, and our results of genome size and gene count are comparable with that of the other ten sequenced bee species (supplementary table S2, Supplementary Material online).

The 11,310 predicted genes comprise 7,888 gene families (supplementary table S3, Supplementary Material online). A total of 5,307 gene families are shared among all bee species and there are 82 predicted orthogroups unique to C. calcarata (fig. 2). The orthogroup and Pfam protein family results for C. calcarata-specific genes are listed in supplementary table S4, Supplementary Material online. Among these 82 unique orthogroups many are associated with zinc finger protein binding and have overrepresentation of protein domain-specific binding (OGS5_15873).

Gene Family Expansion

A total of 306 gene families had been significantly expanded in C. calcarata (supplementary table S5, Supplementary Material online). Two notable gene families appear to be expanded in the C. calcarata lineage: Zinc finger and odorant receptors proteins. Zinc finger proteins have been implicated in gene regulation in female reproduction and are reportedly expanded in the eusocial termite (Zootermopsis nevadensis; Terrapon et al. 2014). We found an expansion of zinc finger proteins in C. calcarata (OGS5_126627) with 50 copies in their

### Table 1

<table>
<thead>
<tr>
<th>Genome Assembly Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contigs (n)</td>
</tr>
<tr>
<td>Largest contig (bp)</td>
</tr>
<tr>
<td>Scaffolds &gt; 1 kb (n)</td>
</tr>
<tr>
<td>N50 scaffolds (bp)</td>
</tr>
<tr>
<td>Scaffolds &gt; N50 (n)</td>
</tr>
<tr>
<td>Largest scaffold (bp)</td>
</tr>
<tr>
<td>Predicted genes</td>
</tr>
<tr>
<td>Ultraconserved core eukaryotic genes (complete/partial, %)</td>
</tr>
<tr>
<td>Arthropod BUSCOs in genome assembly</td>
</tr>
<tr>
<td>BUSCOs in predicted genes</td>
</tr>
</tbody>
</table>

The genome compared with 16 in *L. albipes* and only 2 copies in *A. mellifera*. Zinc finger proteins are known transcription factors and recent analyses of ten bee genomes revealed increasing number of zinc finger binding sites in eusocial versus solitary species (Kapheim et al. 2015). Detailed comparative analyses within and between lineages of bees will be required to reveal the role and regulation of transcription factors across the social spectrum from solitary to eusocial (Rehan and Toth 2015).

Odorant receptors are a large family of transmembrane chemosensory proteins that detect odor molecules and initiate a signal transduction cascade to neurons in the brain (Sanchez-Gracia et al. 2009). The *C. calcarata* genome contains 16 members of this gene family, in contrast to 4 in *M. rotundata* and 2 in *L. albipes* (OG5_168510). Recent analysis of ten bee genomes (Kapheim et al. 2015) found odorant receptors to be expanded in complex eusocial lineages (Apinae) compared with solitary (Megachilidae and Halictidae) species. However, we found expansion of odorant receptors in a subsocial apid bee, consistent with gene family expansion common to the Apidae and likely preceding the evolution of eusociality (Rehan and Toth 2015).

**Gene Family Contraction**

A total of 281 gene families have been significantly contracted in *C. calcarata* (supplementary table S6, Supplementary Material online). One underrepresented protein domain of interest is the Perilipin gene family. *Perilipin* is also known as lipid droplet-associated protein or *PLIN*, a protein that, in *Drosophila*, is encoded by the *PLIN* gene. The perilipins are a family of proteins that associate with the surface of lipid droplets. Phosphorylation of *perilipin* is essential for the mobilization of fats in adipose tissue (Beller et al. 2010). Past studies have implicated lipid storage and metabolism as key to overwintering physiology in this and other Hymenopteran species (Strassmann et al. 1984; Toth et al. 2009; Wasielewski et al. 2014; Durant et al. 2016). Moreover, previous studies have linked honey bee immune status to lipid store gene expression profiling (Mao et al. 2013; Salmela et al. 2015). Future studies on lipid metabolism and the role of fat bodies on bee health are an important line of research and will benefit from additional genome-wide comparisons. *Ceratina calcarata* is an emerging model species to understand the lipid metabolism and overwintering diapause and nutritional regulation of caste determination (Rehan et al. 2014; Durant et al. 2016).

**DNA Methylation**

Epigenetic marks can be influenced by environmental variation, and are capable of influencing postembryonic development (Jaenisch and Bird 2003). Such a role for DNA methylation has been highlighted in the honey bee *A. mellifera*, where larval knockdown of *DNA methyltransferase 3* (*DNMT3*), an enzyme essential to de novo DNA methylation (Klose and Bird 2006), resulted in a developmental shift from the phenotype of workers to the phenotype of queens (Kucharski et al. 2008). Reversible changes in DNA methylation are also associated with changes in behavior linked to the division of labor among workers in *A. mellifera* (Herb et al. 2012). These findings, along with the analysis of mutation signatures of DNA methylation in ten bee genomes, have given rise to the hypothesis that increased gene regulatory capacity may contribute to the evolution of social complexity in bees (Kapheim et al. 2015). One potential mechanism implicating DNA methylation in such regulatory complexity is its role in the regulation of alternative mRNA splicing (Shukla et al. 2011; Li-Byarlay et al. 2013).

Here, we present results of whole genome bisulfite sequencing (WGBS) for *C. calcarata* (supplementary table S7, Supplementary Material online). Like *A. mellifera* (Wang et al. 2006), the genome of *C. calcarata* contains a complete enzymatic toolkit for DNA methylation (fig. 3a; Klose and Bird 2006), comprised in *C. calcarata* of two copies of *DNMT1* (Calc.v2.000349 and Calc.v2.004812) and one copy of *DNMT3* (Calc.v2.012237); *DNMT2* is not known to methylate DNA. Moreover, genes targeted by DNA methylation in *C. calcarata* are enriched for GO biological processes associated with core cellular functions, including translation, and have detectable expression in our RNA-seq data (fractional CDS DNA methylation vs. gene expression level [FPKM], Spearman’s rho = 0.58, *P* < 0.0001). These findings, along with the analysis of mutational signatures of DNA methylation in ten bee genomes, have implicated lipid storage and metabolism as key to overwintering physiology in this and other Hymenopteran species (Strassmann et al. 1984; Toth et al. 2009; Wasielewski et al. 2014; Durant et al. 2016). Moreover, previous studies have linked honey bee immune status to lipid store gene expression profiling (Mao et al. 2013; Salmela et al. 2015). Future studies on lipid metabolism and the role of fat bodies on bee health are an important line of research and will benefit from additional genome-wide comparisons. *Ceratina calcarata* is an emerging model species to understand the lipid metabolism and overwintering diapause and nutritional regulation of caste determination (Rehan et al. 2014; Durant et al. 2016).

**Gene Family Overlap**

A total of 281 gene families have been significantly contracted in *C. calcarata* (supplementary table S6, Supplementary Material online). One underrepresented protein domain of interest is the Perilipin gene family. *Perilipin* is also known as lipid droplet-associated protein or *PLIN*, a protein that, in *Drosophila*, is encoded by the *PLIN* gene. The perilipins are a family of proteins that associate with the surface of lipid droplets. Phosphorylation of *perilipin* is essential for the mobilization of fats in adipose tissue (Beller et al. 2010). Past studies have implicated lipid storage and metabolism as key to overwintering physiology in this and other Hymenopteran species (Strassmann et al. 1984; Toth et al. 2009; Wasielewski et al. 2014; Durant et al. 2016). Moreover, previous studies have linked honey bee immune status to lipid store gene expression profiling (Mao et al. 2013; Salmela et al. 2015). Future studies on lipid metabolism and the role of fat bodies on bee health are an important line of research and will benefit from additional genome-wide comparisons. *Ceratina calcarata* is an emerging model species to understand the lipid metabolism and overwintering diapause and nutritional regulation of caste determination (Rehan et al. 2014; Durant et al. 2016).
methylation is preferentially targeted to exons positioned
proximal to the translation start site (TSS) in C. calcarata (fig.
3b; Wilcoxon rank-sum test between exons 1–3 and 4–n, P < 0.0001). We previously put forth the hypothesis that TSS-prox-
imal DNA methylation plays a role in differentiating states of
transcriptional initiation and elongation (Glastad et al. 2015).

To gain insight into the potential relationship between
social complexity and DNA methylation, we compared
C. calcarata DNA methylation with that observed in orthologs of the
advanced eusocial honey bee A. mellifera using WGBS data
generated from female adults of each species (Lyko et al.
2010; Zemach et al. 2010; supplementary table S9,
Supplementary Material online). We observed that a greater
proportion of CpG dinucleotides were methylated in C. calcar-
ata, which is generally considered subsocial, compared with
A. mellifera, which exhibits complex eusociality (fig. 3c).
Moreover, among 3,841 ortholog pairs with significant DNA
methylation detected in CDS of at least one species, 3,147
(81.9%) were targeted by DNA methylation in both species,
456 (11.9%) were targeted by DNA methylation in C. calcar-
ata only, and 238 (6.2%) were targeted by DNA methylation
in A. mellifera only. Thus, our results stand in contrast with the
prediction that DNA methylation targets are more widespread
in species with complex social organization in bees (Kapheim
et al. 2015). However, we did observe that methylated cyto-
sines in A. mellifera exhibited higher DNA methylation levels
(i.e., were markedly more frequently methylated across se-
quenced cells) than in C. calcarata (fig. 3b and d). Overall,
our results highlight the multidimensionality of DNA methyla-
tion—the number of targets and the frequency with which
cells are methylated each stand to influence gene regulation.
Thus, detailed investigation of DNA methylation patterns in
specific tissues from numerous taxa will be required to resolve
whether generalities exist in the relationship between DNA
methylation and the evolution of sociality.

Molecular Evolution
Ceratina calcarata is the first bee of the subfamily Xylocopinae
and the first subsocial bee species with genome characteriza-
tion. Thus, this species represents an important point of com-
parison for the molecular evolution of sociality (Rehan and
Toth 2015). To identify genes showing signatures of positive
selection in the C. calcarata lineage, we conducted branch-site
tests in PAML. We included all currently published bee ge-
nomes to perform these analyses. These taxa include ten spe-
cies (A. mellifera, A. florea, E. mexicana, B. terrestris,
B. impatiens, Mel. quadriﬁcata, H. laboriosa, M. rotundata, L. albipes, D. novaengliae). We used the “branch-site A” model in PAML to search for sites showing signatures of positive selection in the C. calcarata lineage using a LRT to calculate signiﬁcance. Following correction for multiple testing, 877 genes showed signatures of positive selection in the C. calcarata lineage (FDR $P < 0.05$; supplementary table S10, Supplementary Material online). Among the noteworthy genes with signatures of positive selection were two genes associated with regulation of TOR functioning (Ccalc.v2_007251 and Ccalc.v2_008102; supplementary table S11, Supplementary Material online). TOR genes are highly conserved among bees (Kapheim et al. 2015) and have been shown to play a functionally important role in honey bee caste development (Patel et al. 2007; Mutti et al. 2011).

In general, genes associated with lipid transport, protein binding, and ribonuclease activity appear to have undergone positive selection in C. calcarata relative to the other species examined (table 2). Genes associated with DNA recombination are also evolving more quickly in C. calcarata (Ccalc.v2_012690, Ccalc.v2_001846, Ccalc.v2_013055, Ccalc.v2_004327; supplementary table S11, Supplementary Material online). High rates of DNA recombination are known in advanced eusocial insects (Wilfert et al. 2007; Sirviö et al. 2011; Kapheim et al. 2015) and could be linked to accelerated evolution of caste related traits (Kent et al. 2012). It is possible that the accelerated evolution of DNA recombination genes may reﬂect this group’s transitional form of sociality including the presence of an incipient worker caste (Rehan et al. 2014).

### Conclusions

The maternal care and sibling care expressed in an incipiently social bee make C. calcarata ideal for developmental and sociodemographic studies to characterize the genomic effects of social experience on behavioral plasticity. Our results suggest that mechanisms associated with DNA methylation and nutrition are candidate targets of evolutionary change during the origin of more complex social forms. The addition of C. calcarata to the published bee genomes provides the ﬁrst subsocial bee species to ask questions on the evolutionary origins of social behavior. Moreover, the characterization of a fully functional C. calcarata methylome allows for new lines of research into the epigenetic regulation of maternal care, sibling care, and worker behavior.

### Supplementary Material

Supplementary tables S1–S11 and supplementary material S1 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

### Acknowledgments

We would like to thank Wyatt Shell for assistance with ﬁeld collections and DNA extractions as well as the staff at the Hubbard Center for Genome Studies at the University of New Hampshire for DNA library preparation and sequencing. This work was supported by funding from the University of Georgia to B.G.H., the University of New Hampshire and a National Science Foundation grant I0S-1456296 to S.M.R. Partial funding was provided by the New Hampshire Agricultural Experiment Station. This is Scientiﬁc Contribution Number 2667. This work was also supported by the USDA National Institute of Food and Agriculture Hatch Project 1004515.

### Literature Cited


Sadd BM, et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 16:76.


Associate editor: Dennis Lavrov