Brain microRNA expression associated with social evolution in bees

Karen M. Kapheim[1*], Beryl M. Jones[2ϕ], Eirik Søvik[3], Eckart Stolle[4], Robert M. Waterhouse[5], Guy Bloch[6], Yehuda Ben-Shahar[7]

[1] Department of Biology, Utah State University, Logan, UT 84322, USA

[2] Program in Ecology, Evolution, and Conservation Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[3] Department of Science and Mathematics, Volda University College, 6100 Volda, Norway

[4] Martin-Luther-Universität Halle-Wittenberg, Institut für Biologie, Hoher Weg 8, Halle (Saale), Germany

[5] Department of Ecology and Evolution, University of Lausanne and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

[6] Department of Ecology, Evolution and Behavior, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

[7] Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, USA

[ϕ] Current affiliation: Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

* Corresponding author:
Karen M. Kapheim
Biology Department
Utah State University
5305 Old Main Hill
Logan, UT 84322, U.S.A.
karen.kapheim@usu.edu
+1-435-797-0685

ABSTRACT

Evolutionary transitions to a social lifestyle in insects are associated with lineage-specific changes in gene expression, but the key nodes that drive these regulatory changes are largely unknown. We tested the hypothesis that changes in gene regulatory function associated with social evolution are facilitated by lineage-specific microRNA (miRNA) regulatory function. Genome scans across 12 bees showed that miRNA copy number is highly conserved and is not associated with variation in social organization. However, deep sequencing of small RNAs of six bee species revealed a substantial proportion (20-35%) of miRNAs are expressed in the brains of a single species, and many of these do not have identifiable homologs in any other species. Lineage-specific miRNAs disproportionately target lineage-specific genes, and have lower expression levels than more evolutionarily conserved miRNAs. Consistent with our hypothesis, the predicted targets of lineage-specific miRNAs are enriched for genes related to social behavior, such as caste-biased genes, in social species, but they are either not enriched for or significantly depleted of genes under positive selection. Together, these results suggest that novel miRNAs may contribute to lineage-specific patterns of molecular evolution associated with the origins and elaborations of eusociality. Our analyses also lend support to earlier hypotheses concerning miRNA origins from a relatively understudied taxonomic group, and reveal important differences in the evolution and assimilation of novel miRNAs between mammals and insects.

INTRODUCTION

Eusociality, the most complex form of social organization, has evolved several times in insects from the order Hymenoptera. In its most basic form, this lifestyle involves reproductive queens

living with their worker daughters who forego direct reproduction of their own to cooperatively defend the nest, care for their siblings, and forage for the colony (Wilson 1971). Due to the complex nature of this lifestyle, the evolution of eusociality likely requires modification of molecular pathways related to development, behavior, neurobiology, physiology, and morphology (Robinson and Ben-Shahar 2002; Toth and Robinson 2007; Bloch and Grozinger 2011; Sumner et al. 2018). The evolution of eusociality is thus expected to involve both genetic changes as well as changes in the way the genome responds to the environment (Robinson and Ben-Shahar 2002; Johnson and Linksvayer 2010). It is therefore unsurprising that recent studies aimed at identifying the genomic signatures of eusocial evolution in insects have found that social species share an increased capacity for gene regulation (Simola et al. 2013; Kapheim et al. 2015). Evidence for this comes from signatures of rapid evolution of genes involved in transcription and translation, gene family expansions of transcription factors, and increasing potential for DNA methylation and transcription factor binding activity in conserved genes. Interestingly, while these types of regulatory changes are common to independent origins and elaborations of eusociality, the specific genes and regulatory elements involved are unique to each lineage in which eusociality evolved (Kapheim et al. 2015). This suggests that lineage-specific processes are influential in generating new patterns of gene regulation that contribute to social behavior.

Small, non-coding RNAs such as microRNAs (miRNAs) may be an important source of regulatory novelty associated with the evolution of phenotypic complexity, including eusociality. MiRNAs are short (~21-22 nt) noncoding RNAs that regulate protein-coding genes through post-transcriptional binding to the 3' UTR region of messenger RNA (mRNA) transcripts, in most

cases preventing translation or causing degradation (Bartel 2018). Each miRNA can target dozens to hundreds of mRNAs, and therefore miRNAs have enormous potential to regulate gene networks (Filipowicz et al. 2008; Friedman et al. 2009; Bartel 2018). Like mRNAs, miRNAs are spatially- and temporally-specific in their expression patterns. Thus, complex changes in gene regulation can be achieved with relatively minor changes in miRNA expression. This can result in major phenotypic shifts or fine-tuning of phenotypic optimization (Bartel 2018). Novel miRNAs can originate in a variety of genomic features, including exons and introns of protein-coding and non-coding RNA genes, transposable elements, pseudogenes, or intergenic regions, and thus emerge and disappear over relatively rapid timescales (Chen and Rajewsky 2007; Lu et al. 2008; Berezikov 2011; Zhu et al. 2012; Meunier et al. 2013). It is thus not surprising that expansion of the miRNA repertoire is associated with the evolution of morphological complexity across the tree of life (Grimson et al. 2008; Heimberg et al. 2008; Wheeler et al. 2009; Christodoulou et al. 2010; Berezikov 2011).

There is accumulating evidence for a role of miRNAs in regulating the social lives of insects. MiRNAs have been identified in the genomes of most major insect groups (Asgari 2013; Quah et al. 2015; Ylla et al. 2016), including several ant and bee species (Weaver et al. 2007; Bonasio et al. 2010; Kocher et al. 2013; Patalano et al. 2015; Sadd et al. 2015). Bioinformatic scans of insect genomes have identified candidate miRNAs present in the genomes of social, but not solitary, insects (Søvik et al. 2015). While most miRNAs seem to be conserved in major lineages of insects (Søvik et al. 2015; Ylla et al. 2016), expression levels have been found to vary across individuals performing different social functions, such as between workers performing different tasks in honey bees (*Apis mellifera*) (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et

106 al. 2012). MiRNAs may also play a role in caste determination, as queen- and worker-destined
107 larvae express different sets of miRNAs throughout development in honey bees (Weaver et al.
108 2007; Shi et al. 2015; Ashby et al. 2016) and bumble bees (*Bombus terrestris*) (Collins et al.
109 2017). Additionally, miRNAs play a role in regulating some of the physiological correlates of
110 social behavior in honey bees, including activation of ovaries in queens and workers (MacEdo et
111 al. 2016) and response to the reproductive protein *vitellogenin* (Nunes et al. 2013). Together,
112 these studies suggest that miRNAs could play a role in the evolution of eusociality through their
113 effects on gene regulatory networks that are involved in traits important for social behavior. A
114 rigorous test of this hypothesis requires comparisons of the presence, expression, and function of
115 miRNAs across related species that vary in social organization. However, none of the previous
116 studies of insect miRNAs have included solitary species that are closely related to eusocial
117 insects, and thus representative of the ancestors from which sociality evolved.
118
119 Here we present a comprehensive comparative analysis of miRNAs across bee species that vary
120 in social organization. We first looked for miRNA repertoire expansions associated with
121 eusociality by scanning the genomes of 12 bee species for known miRNAs, and statistically
122 evaluating copy number of each miRNA type with regard to differences in sociality in a
123 phylogenetic model. We then described and compared miRNAs expressed in the brains of six
124 bee species from three families that include repeated origins of eusociality. In our analysis, we
125 identified shared and lineage-specific miRNAs, their evolutionary histories, and their predicted
126 gene targets. We then tested the hypothesis that changes in gene regulatory function associated
127 with social evolution are facilitated by lineage-specific miRNA regulatory function. We tested
128 three predictions of this hypothesis: (1) miRNAs that play a role in social behavior should target

different genes in solitary and social species. (2) If lineage-specific miRNAs are assimilated into ancestral gene networks, then their predicted target genes should be ancient and conserved. (3) If lineage-specific miRNAs play a role in social evolution, then their predicted targets should be enriched for genes that play a role in social behavior (e.g., caste-biased expression) or genes that are under selection in social species.

MATERIALS AND METHODS

Sample Acquisition

We used adult females from six bee species for our study. *Megalopta genalis* samples were collected on Barro Colorado Island, Panama in 2015 and exported to the U.S.A. under permit SEX/A-37-15. *Nomia melanderi* samples were collected in Touchet, WA, U.S.A. with permission from private land owners. *Megachile rotundata* samples were collected from Logan, UT, U.S.A. on the Utah State University campus. *Bombus impatiens* samples were collected from a commercial colony purchased from BioBest. *Bombus terrestris* samples were collected from Pollination Services Yad-Mordechai, Kibbutz Yad-Mordechai,Israel. *Apis mellifera* samples were collected from typical field hives in Urbana-Champaign, IL and the Tyson Research Field Station, MO, U.S.A. All samples were flash-frozen in liquid nitrogen upon collection and stored at -80 °C until dissection.

RNA Isolation and Sequencing

Head capsules from the *B. impatiens*, *M. genalis*, and *N. melanderi* samples were dissected after incubation in RNALater ICE (Ambion) to remove the entire brain. We used the mirVana miRNA Isolation kit with phenol (Ambion) to isolate total RNA from individual brains. Total RNA was

sent to the University of Illinois Roy J. Carver Biotechnology Center for library preparation and sequencing. Libraries were prepared with the Illumina TruSeq Small RNA Sample Preparation kit. Libraries were pooled, quantitated by qPCR, and sequenced on one lane for 51 cycles on a HiSeq 2500, using TruSeq SBS sequencing kit version 2. Sequencing yielded a mean of 19,524,877 (± 2,545,208 s.d.) reads per sample.

Whole brains of *A. mellifera*, *B. terrestris*, and *M. rotundata* were dissected from frozen heads. Total RNA from individual brains were isolated by using the TRIzol reagent (Thermo Fisher Scientific). All subsequent small-RNA sequencing steps were performed by the Genome Technologies Access Center at Washington University, using their Illumina TrueSeq pipeline. In short, total RNA samples were size fractionated and multiplexed. Single-end small RNA libraries were prepared by using the SMARTer kit (Clontech). Up to 12 barcoded libraries from a single species were run on a single Illumina HiSeq 2500 lane.

miRNA Discovery and Quantification

We used miRDeep2 (Friedländer et al. 2012) to identify and quantify miRNAs expressed in the brains of each species. We used a three-step process of miRNA detection in order to identify homologous miRNAs between species. For the first step, we generated a set of mature miRNA sequences previously described in other insect species (Table S1). Reads for each sample were quality filtered (minimum length 18, removal of reads with non-standard bases), adapter-trimmed, and aligned to the species genome (Table S2) with the mapper.pl script. Approximately 60-84% of reads successfully mapped.

We then identified known and novel miRNAs in each sample with the miRDeep2.pl script, using our curated set of insect miRNAs (Table S1) as known mature sequences. We followed this with quantification of the miRNAs using the quantifier.pl script. This generated a set of known and novel miRNAs in each sample, along with quantified expression information for each. We then filtered the novel miRNAs in each species according to the following criteria: no rRNA/tRNA similarities, minimum of five reads each on the mature and star strands of the hairpin sequence, and a significant randfold p-value ($p < 0.05$). Randfold describes the RNA secondary structure of potential pre-miRs (Friedländer et al. 2012).

We used these filtered miRNAs in a second run of detection and quantification. We added the mature sequences of the novel miRNAs from each species to our set of known miRNAs, and repeated the pipeline above. This allowed detection of homologous miRNAs (based on matching seed sequences) that are not represented in miRBase across our species. We applied the same set of filtering criteria as for our first run.

Some of the novel miRNAs may exist in the genomes of other bees, even if they are not expressed. We used blastn (-perc_identity 50 -evalue 1e-5) to search for homologous precursor miR (pre-miR) sequences in 12 bee genomes (Table S2) for each of the novel miRNAs without a matching seed sequence.

miRNA Localization

We characterized the location of each known and novel miRNA in respective genome assembly in relation to genes and transposable elements. We used bedtools intersect (Quinlan and Hall

198 2010) to find overlap with predicted gene models (Table S3), and repetitive element
199 repeatmasker (Smit et al. 2013) annotations from previously established repeat libraries
200 (Kapheim et al. unpublished; Elsik et al. 2014; Kapheim et al. 2015a; Sadd et al. 2015; Kapheim
201 et al. 2019).
202
203 **Target Prediction**
204 We used computational methods to predict targets of each miRNA in each species. We used
205 bedtools flank and getfasta (Quinlan and Hall 2010) to extract a 500 bp region downstream from
206 each gene model, following previous studies (Ashby et al. 2016) and an average 3' UTR region
207 of 442 nt in *Drosophila melanogaster* (Grün et al. 2005). We used these as potential target sites
208 in miRanda (Enright et al. 2004) and RNAhybrid (Krüger and Rehmsmeier 2006) target
209 prediction analyses. miRanda v3.3 was run with a minimum energy threshold of -20, a minimum
210 score of 140, and strict alignment in the seed region (parameters -en -20 -sc 140 -strict). We also
211 used RNAhybrid v2.1.2 with a minimum free energy threshold of -20 and the fly 3' UTR set was
212 used to estimate xi and theta values (These are the position and shape parameters of the value
213 distribution from which p-values are calculated). We kept only miRNA-target gene pairs that
214 were predicted by both programs with $p < 0.01$.
215
216 **Target Age and Functional Enrichment**
217 We explored the taxonomic ages and putative functions of predicted target genes. Gene ages
218 were determined using orthogroups from OrthoDB v9 (Zdobnov et al. 2017), which includes *A.*
219 *mellifera*, *B. impatiens*, *B. terrestris*, and *M. rotundata*. OrthoDB delineates orthologs by
220 clustering gene best reciprocal hits (BRHs) between all species pairs, first with triangulation of

BRHs and then addition of in-paralogous groups and genes to build clusters of orthologs. Gene sets of *M. genalis* and *N. melanderi* were mapped to Metazoa-level (330 species) orthogroups following the same procedure as for BRH clustering allowing genes to join existing orthogroups when all BRH criteria are met. The age of each gene from the six bee species was inferred from the taxonomic breadth of all species in each orthogroup, from Vertebrata (at least one of 172 vertebrates), to Metazoa (at least one of 25 non-arthropod and non-vertebrate metazoans), to Arthropoda (at least one of 17 non-insect arthropods), to Insecta (at least one of 16 non-holometabolous insects), to Holometabola (at least one of 68 non-hymenopteran holometabolous insects), to Hymenoptera (at least one of 7 non-Aculeata hymenopterans), to Aculeata (at least one of 13 non-Apoidea Aculeata), to Apoidea (at least one of 11 other Apoidea), and finally, genes without identifiable orthologs were labeled 'Unique'.

Gene Ontology (GO) terms for each species were derived from a previous study (Kapheim et al. 2015), with the exception of *B. impatiens*, for which GO terms were assigned based on reciprocal blastp (evalue < $1e^{-5}$) between two sets of gene models (OGS v1.2 and OGS v1.0). Functional enrichment was performed with the GOstats package (Gentleman and Falcon 2013) in R (R Core Team 2016). We included all terms enriched at a value of unadjusted $p < 0.1$ to allow for more inclusivity in our cross-species comparisons.

Enrichment tests of lineage-specific miRNA targets with previous studies

For each species, brain or head gene expression datasets related to socially relevant phenotypes (e.g., caste) and genes under selection were compared against targets of lineage-specific miRNAs. Differential expression lists for *A. mellifera* included those from Grozinger et al.

(2007) (queen vs. worker, reproductive worker vs. sterile worker, queen vs. reproductive worker), Whitfield et al. (2003) (nurse vs. forager), Alaux et al. (2009) (nurse vs. forager), and Wheeler et al. (2013) (*vitellogenin* RNAi vs. control). For *B. terrestris,* differential expression lists included three pairwise comparisons from Harrison et al. (2015) (queen vs. worker, reproductive worker vs. sterile worker, queen vs. reproductive worker), as well as a comparison between reproductive and sterile workers from Marshall et al. (2019). (The Harrison et al. dataset was from whole body, rather than just head or brain.) For *M. genalis* caste data, RNAseq reads from Jones et al. (2017) (NCBI PRJNA331103) were trimmed using Trimmomatic (v. 0.36) and aligned to an unpublished genome assembly of *M. genalis* (NCBI PRJNA494872) using STAR (v. 2.5.3). Reads were mapped to gene features using the featureCounts function of the Subread package (v. 1.5.2). Remaining differential expression analysis followed the methods of Jones et al. (2017) using edgeR (Robinson et al. 2010) with tagwise dispersion estimates and FDR correction. The complete list of included studies and gene lists are in Table S4.

In addition to gene expression comparisons, multiple datasets identifying genes under selection in bee species or across multiple social lineages of bees were tested for enrichment of lineage-specific miRNA targets. Species-specific selection datasets were used for *A. mellifera* (Harpur et al. 2014), *B. impatiens* (Harpur et al. 2017), *M. genalis* (Kapheim et al. unpublished), and *N. melanderi* (Kapheim et al. 2019). In addition, genes under selection in social relative to solitary lineages identified by Woodard et al. (2011) and Kapheim et al. (2015) were tested against each species' set of lineage-specific miRNA targets. The complete list of included studies and gene lists are in Table S4.

In cases when we needed to identify orthologous genes across species, we used reciprocal blastp (evalue < $10e^{-5}$). Only genes with putative orthologs were included in the final tested sets of genes. Hypergeometric tests (using phyper in R) were used to test for significance of over- or under-enrichment between each pair of lists. The representation factor (RF) given represents the degree of overlap relative to random expectation (RF=1).

miRNA Diversification

We explored the diversification of miRNAs that have been previously implicated in social behavior (miR-13b, miR-276, miR-6001-3p) or which are expressed in social bees, but not solitary bees (miR-305). We performed multiple sequence alignment with the web version of Clustal Omega with default settings (Sievers et al. 2011), and generated a Neighbour-joining phylogenetic tree in Newick format.

We also performed genome scans for small RNAs across 12 bee genomes (Table S2) using covariance models implemented with Infernal cmsearch using the gathering threshold for inclusion (--cut_ga) (Cui et al. 2016) to find all Rfam accessions in each bee genome. We used bedtools intersect to identify overlap between small RNAs identified through cmsearch and gene models. We then used Spearman rank regressions to test for significant associations between miRNA copy number and social biology. We categorized each species as either solitary, facultative basic eusociality, obligate basic eusociality, or obligate complex eusociality following Kapheim et al. (2015). We used the ape package (Paradis et al. 2004) in R (R Core Team 2016) to calculate phylogenetic independent contrasts for both social organization and miRNA copy

number, cor.test to implement the Spearman's rank correlation, and p.adjust with the Benjamini-Hochberg method (method = "BH") to correct for multiple comparisons.

RESULTS

Low levels of miRNA copy number variation among bee genomes

Our genome scans revealed very little variation in copy number of most miRNAs among bee genomes. Of the 50 miRNA Rfam accessions, half had the same number of copies in all 12 bee genomes (1 or 2 copies) (Table S5). The mean copy number across all miRNAs in all bee genomes was 1.19 ± 0.74. Seven of the Rfam miRNAs were detected in a single bee species, but mostly at low copy numbers (1-3). One exception was miR-1122, for which we found 70 copies in *M. genalis*, but no copies in any of the other species. We did not find any significant associations between miRNA copy number and social organization (Table S5).

Expressed miRNA diversity in bee brains

We identified 97-245 known and novel miRNAs expressed in the brains of each of our six species (Table S6). The majority of these were located in intergenic regions or in introns (Table 1). Each species had at least one miRNA that originated from exons of protein-coding genes and repetitive DNA (Table 1). Most of the overlap between miRNA precursors and repetitive DNA corresponded to uncharacterized repeat elements, with very few overlaps with well characterized transposons or retrotransposons (Table 1).

Most of the detected miRNAs in each species had known homologs in at least one other species. However, each species had a substantial proportion (20-35%) of miRNAs with lineage-specific

expression in the brain (Table 1; Fig. 1A), 24-72% of which did not have any known homologs in other species (Table 1). We defined lineage-specific miRNAs as those miRNAs with lineage-specific expression and for which no seed match with a known mature miRNA was identified (columns 6 and 7 in Table 1), because these show the most evidence of being real miRNAs that are unique to a particular species. (Sequence similarity of pre-miRs in the genome of other bee species is not sufficient evidence that a mature miRNA is transcribed.) Lineage-specific miRNAs had significantly lower expression levels in each species (t-tests: *A. mellifera*, p = 3.81e^{-05}, *B. impatiens*, p = 0.003, *B. terrestris*, p = 0.006, *M. genalis*, p = 0.0003, *M. rotundata*, p = 8.00e-^{05}, *N. melanderi*, p = 0.02.

Table 1. Localization of miRNAs in the genomes of six bee species. Numbers not in parentheses represent features on the same strand as the pre-miR. Numbers in parentheses indicate strand mismatch. Some pre-miRs overlapped with one or more genes on both the same and opposite strands, and are thus counted twice (*A. mellifera* and *M. genalis* – 1, *B. impatiens* – 5, *B. terrestris* – 4, *N. melanderi* – 3). Seed match – Mature miR had a seed match with a known miR; Pre-miR – Successful blastn hit to the pre-miR sequence in at least one other bee genome; Unique – No homolog was found in other species (seed match to mature or blastn hit to pre-miR).

			miRs with lineage-specific expression in the brain				Location in the genome				
Species	**Sociality**	**Expressed miRs**	**Total**	**Seed match**	**Pre-miR**	**Unique**	**Intergenic**	**Exon**	**Intron**	**Transposable element**	**Uncharacterized repetitive DNA**
Apis mellifera	Complex eusocial	97	25	0	15	10	45	5	38 (10)	0	0
Bombus impatiens	Basic eusocial	245	49	6	31	12	129	4 (1)	89 (27)	7	32
Bombus terrestris		150	31	2	21	8	76	1 (1)	56 (20)	13	36
Megalopta genalis	Facultative eusocial	105	37	7	5	25	63	3	30 (10)	2	28
Megachile rotundata	Solitary	99	27	9	0	18	48	8 (1)	37 (5)	2	15
Nomia melanderi	Solitary	97	29	5	3	21	50	8	34 (8)	2	27

Lineage-specific miRNAs were localized both within genes and intergenically. The proportion of lineage-specific miRNAs that are intragenic or intergenic was similar to miRNAs with homologs for every species except *N. melanderi*, for which a disproportionate number of lineage-specific miRNAs were intragenic ($\chi^2 = 4.78$, $p = 0.03$). Genes that serve as hosts for intragenic lineage-specific miRNAs were not significantly older than would be expected by chance (i.e., belong to orthogroups shared with vertebrates) in any species (hypergeometric tests: $p = 0.14$-0.76). Across all species, genes that serve as hosts for intragenic lineage-specific miRNAs were not significantly older than genes that host miRNAs with known homologs (χ^2 tests: $p = 0.05$-0.89).

Of the miRNAs with mature miRNA homologs, most were expressed in all six species, and we detected few family-specific expression patterns of miRNAs (Table S6). For example, miR-3049 and miR-3786 were only detected in the bees from the family Apidae (*A. mellifera*, *B. impatiens*, *B. terrestris*). miR-3049 is predicted to target a neurotransmitter-gated ion-channel in all three species (OG EOG091G0R20), and a histone acetyltransferase (OG EOG091G00D2), a zinc-finger protein (OG EOG091G0N0A), a leucine-rich repeat (OG EOG091G01ZI), and a sodium-potassium-calcium exchanger (OG EOG091G0M5C) in both *Bombus* species. The two *Bombus* species shared a cytochrome P450 (OG EOG091G06KN) as a predicted target of miR-3786, but did not share predicted targets with *A. mellifera*. We identified a miRNA (nmel-scaffold2759_cov63_18669) without a known seed match that was only expressed in the two halictid bees (*N. melanderi*, *M. genalis*) and is predicted to target kinase associated proteins in both species.

We identified one miRNA (miR-305) that was expressed in the brains of each of the social, but not the solitary species. Although we did not detect expression of miR-305 in the two solitary species, *M. rotundata* and *N. melanderi*, genome scans of each species against the Rfam database suggested all bee species have one copy of this miRNA (Table S5). Predicted targets of miR-305 differed across species. *Oxysterol* (OG EOG091G0FV2) was a common target among the (social) Apidae bees, but was not among the targets for *M. genalis*. However, *arylformamidase* (OG EOG091G0KT8), which is also involved in lipid metabolism and transport, was a predicted target in *M. genalis*. *Synaptobrevin* (OG EOG091G0MPE), which is involved in synaptic plasticity and neurotransmitter release, was a predicted target of miR-305 in *B. impatiens*.

miRNAs associated with honey bee social behavior are conserved across bee species

As existing gene networks become co-opted for social evolution, we should expect novel regulatory connections between existing miRNAs and protein-coding genes. If social evolution involves miRNA-mediated changes in gene regulatory networks, then miRNAs with functions in social behavior should target different genes in social and solitary species. We tested this prediction by comparing the diversification and targets of three miRNAs that have been repeatedly associated with social behavior in bees. miR-6001-3p is upregulated in queen-destined larvae, compared to worker-destined larvae, in both *A. mellifera* (Shi et al. 2015) and *B. terrestris* (Collins et al. 2017). However, we only detected brain expression of this miRNA in *M. rotundata*, despite the fact that the honey bee miR-6001-3p was included in our known set of miRNAs. Because this miRNA does not have an accession in Rfam, we were unable to verify presence or copy number of this miRNA across the bee species. Additional targeted sequencing is needed to determine the role of this miRNA in adult bee brains.

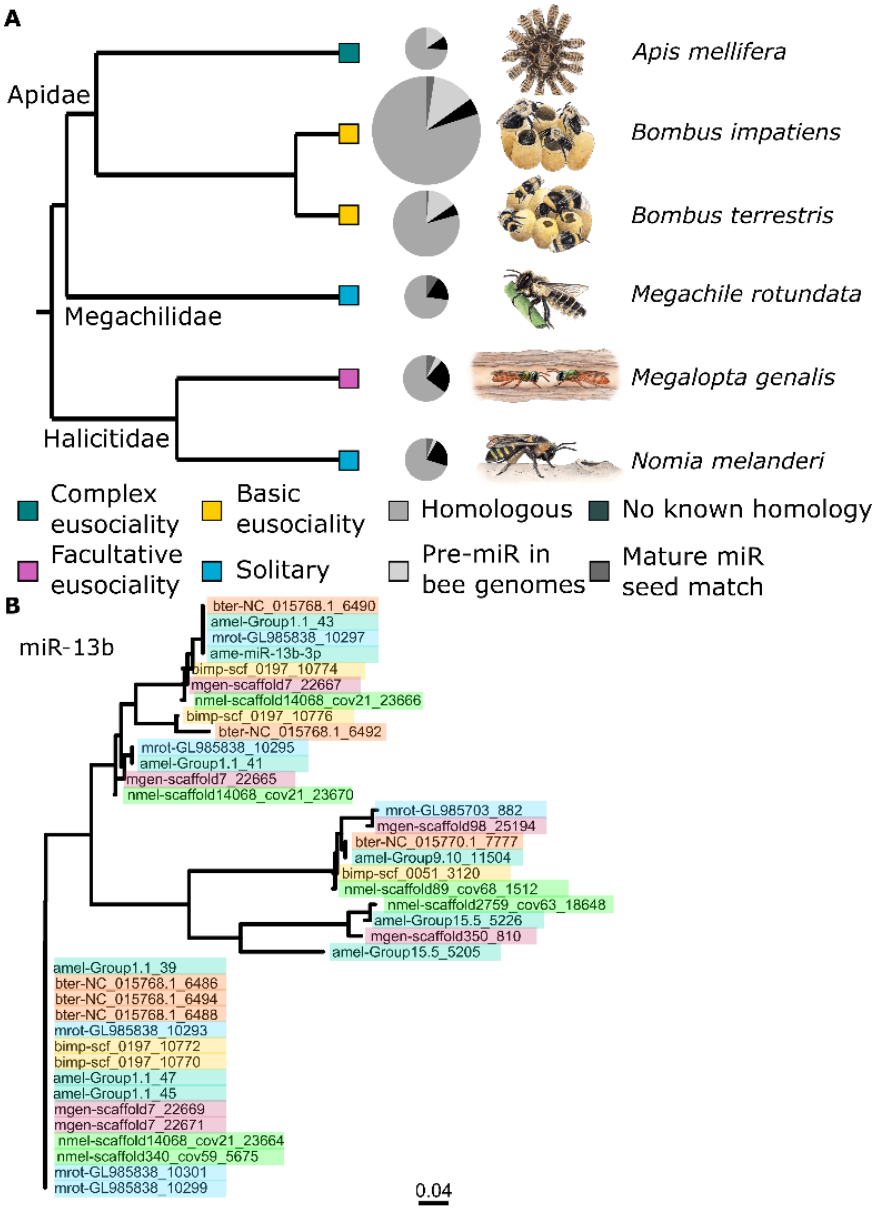
375

376 MiR-276 is upregulated in honey bee worker ovaries (MacEdo et al. 2016) and in queen-destined
377 larvae (Shi et al. 2015; Ashby et al. 2016). Our genome scans revealed that each bee species has
378 a single copy of this miRNA, and we detected expression in the brains of each of our six focal
379 species. Multiple alignment of the expressed mature miRNA revealed almost no sequence
380 variation across the species. There were few shared gene targets across the six species, but the
381 functional enrichment of predicted miR-276 targets was similar among species. Target genes in
382 each species were enriched for cell signaling, though the specific pathways differed. *A. mellifera*,
383 *B. impatiens*, and *M. rotundata* targets were enriched for insulin receptor binding, and the two
384 halictids (*Megalopta genalis* and *N. melanderi*) were enriched for G-protein coupled receptor
385 signaling. All species except *N. melanderi* shared enrichment for functions related to protein
386 translation, including translation, protein-containing complex binding, structural constituent of
387 ribosome, peptide biosynthetic process, peptide metabolic process, and amide biosynthetic
388 process.

389

390 MiR-13b is differentially expressed across honey bee castes in different life stages and tissues. It
391 is upregulated in worker ovaries compared to queen ovaries (MacEdo et al. 2016),
392 downregulated in queen-destined larvae compared to worker-destined larvae (Shi et al. 2015;
393 Ashby et al. 2016), and downregulated in workers performing nursing tasks compared to
394 foragers (Liu et al. 2012). This miRNA has expanded and diversified among the bees, with three
395 major clades containing at least one copy from each species (Fig. 1B). As with miR-276, there
396 was very little overlap in predicted targets among species. Just one gene, the *beta subunit of*
397 *nuclear transcription factor Y*, is predicted to be a target of miR-13b in each of the social

398 species. However, there were more similarities across species at the functional level (Table S7).
399 All species had enrichment for transcription factor activity, and all species except the two
400 halictid bees had significant enrichment for genes involved in ecdysone-mediated signaling.
401 Within the family Halictidae, the predicted targets of miR-13b for *M. genalis* were enriched for
402 genes related to neurotransmitter-gated ion channel activity, while *N. melanderi* targets were
403 enriched for phospholipid metabolic processes.
404



405

406 **Fig. 1. Diversity of miRNAs expressed in the brains of six bee species.** (A) Homologous and
407 lineage-specific miRNAs expressed in the brain of each species. The three types of homology
408 (shades of grey) correspond to those in Table 1. Black – has not been previously detected in
409 other species. Pie size corresponds to number of miRNAs detected from small RNA sequencing.
410 Boxes indicate social organization (green – complex eusociality, yellow – basic eusociality, pink
411 – facultative eusociality, blue – solitary). Phylogenetic relationships are following previous
412 studies (Cardinal and Danforth 2011; Sadd et al. 2015; Branstetter et al. 2017). (B)
413 Diversification of miR-13b across six bee species. Each miRNA has a matching seed sequence to
414 ame-miR-13b-3p (miRBase v21). Tip labels are the name of each miRNA expressed in the brain,
415 beginning with a four letter code for species. Different species are also indicated by colors. Scale
416 indicates substitution rate.
417

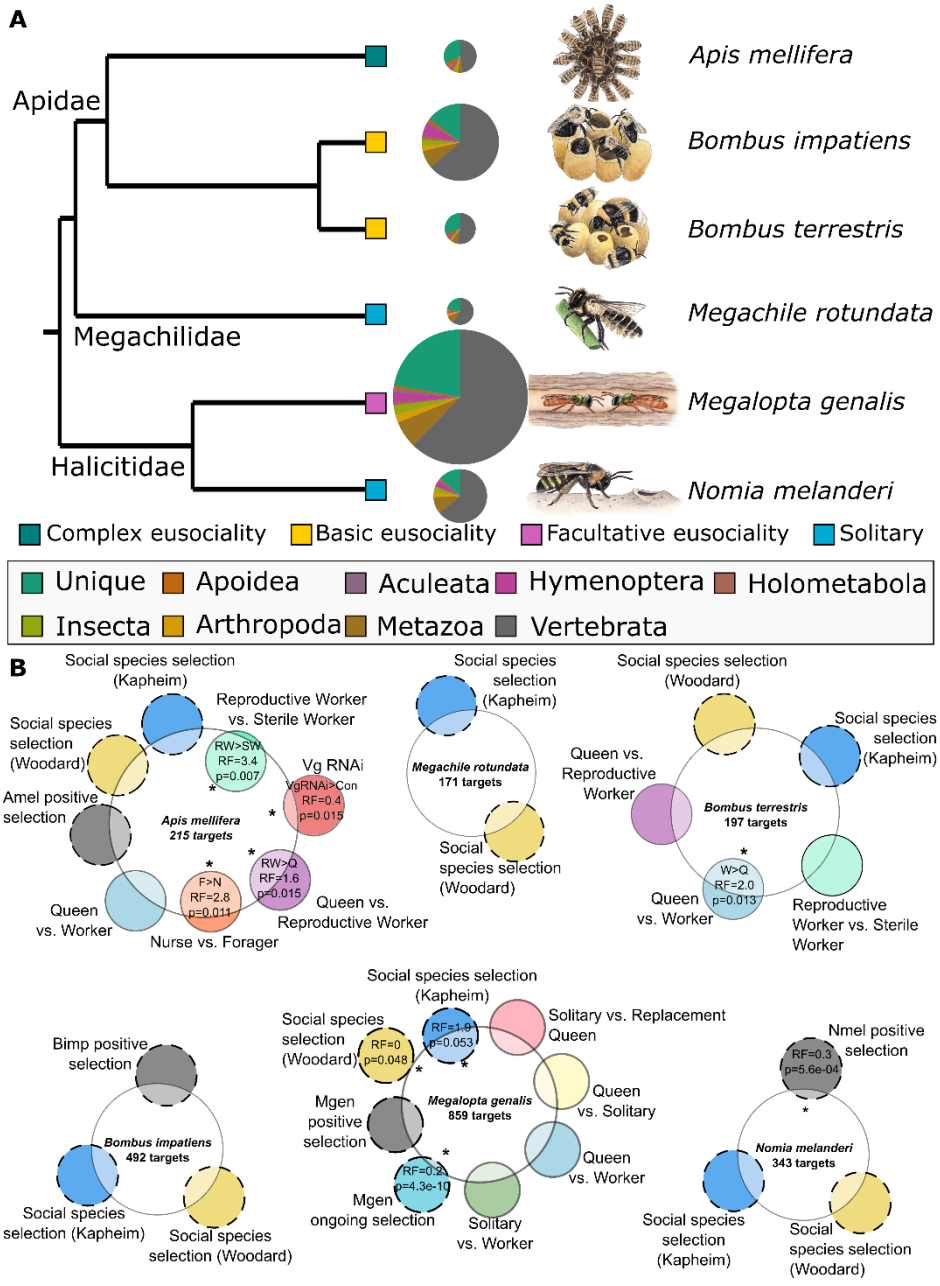
418 **Lineage-specific miRNAs preferentially target lineage-specific genes and genes with caste-**
419 **biased expression, but not genes under selection**

420 If lineage-specific changes in gene regulatory function associated with social evolution are
421 facilitated by novel miRNAs inserted into existing gene networks, then the predicted targets of
422 lineage-specific miRNAs should be highly conserved and enriched for genes with a known
423 function in social evolution. Most of the predicted mRNA targets of lineage-specific miRNAs
424 were highly conserved and belonged to orthogroups shared by vertebrates (Fig. 2A; Table S8).
425 However, most of the genes in each genome are also highly conserved, and there was not a
426 significant enrichment for conserved genes among predicted targets of lineage-specific miRNAs,
427 beyond what would be expected by chance ($p > 0.99$). We did, however, find a significant
428 enrichment for lineage-specific genes that are unique to each species among the predicted targets
429 of lineage-specific miRNAs ($p = 0.02 - 1.48e^{-12}$), indicating that novel miRNAs are more likely
430 to target novel genes than would be expected by chance (Fig. 2A; Table S8).
431

432 We found mixed support for the prediction that novel miRNAs should target genes that function
433 in social behavior and evolution. The predicted targets of lineage-specific miRNAs were
434 enriched for genes that are differentially expressed between castes in the social Apidae (*A.*

435 *mellifera* and *B. terrestris*), but not Halictidae (*M. genalis*) (Fig. 2B; Table S4). In *A. mellifera*,
436 this included genes that are upregulated in the brains of reproductive workers, compared with
437 sterile workers (RF = 3.4, p = 0.007) and queens (RF = 1.6, p = 0.015) (Grozinger et al. 2007), as
438 well as genes upregulated in the brains of foragers compared with nurses (RF = 2.8, p = 0.011)
439 (Whitfield et al. 2003). However, there was not significant enrichment for genes differentially
440 expressed between nurse and forager honey bee brains in a later study (p = 0.09) (Alaux et al.
441 2009). In *B. terrestris*, we find significant overlap between the predicted targets of lineage-
442 specific miRNAs and genes that are upregulated in worker bodies, compared to queens (RF = 2,
443 p = 0.013). We did not find significant overlap with genes differentially expressed between
444 reproductive and sterile workers, but this was a much more limited gene set (p = 0.39) (Marshall
445 et al. 2019). To our knowledge, there are no studies of gene expression differences between *B.*
446 *impatiens* castes, so we could not evaluate target overlap with caste-biased genes in this species.
447 We do not find significant enrichment for caste-biased genes in the brains of the facultatively
448 eusocial halictid *M. genalis* (p = 0.25).
449
450 Contrary to our prediction, targets of lineage-specific miRNAs were not significantly enriched
451 for genes under selection in any species. We assessed overlaps between genes undergoing
452 positive directional selection in *A. mellifera* (Harpur et al. 2014), *B. impatiens* (Harpur et al.
453 2017), *M. genalis* (Kapheim et al., unpublished), and *N. melanderi* (Kapheim et al. 2019) and the
454 predicted targets of lineage-specific miRNAs in each of these species. There was no significant
455 enrichment for predicted targets of lineage-specific miRNAs with genes under positive
456 directional selection in any species (Table S4). In fact, genes undergoing current selection in the
457 halictid bees were significantly depleted for targets of lineage-specific miRNAs (*M. genalis* – RF

458 = 0.2, p = 4.28e^{-10}; *N. melanderi* – RF = 0.3, p = 5.59e^{-4}). We also assessed overlaps with genes
459 previously found to be under selection in social species, compared to solitary species (Woodard
460 et al. 2011; Kapheim et al. 2015), but found only marginally significant overlap (Kapheim et al.
461 2015) or depletion (Woodard et al. 2011) with predicted targets of lineage-specific genes in one
462 species (*M. genalis* – RF = 1.9, p = 0.053; RF = 0, p = 0.05; Table S4).
463



464

Fig. 2. Predicted targets of lineage-specific miRNAs in relationship to social behavior. (A) Genes predicted to be targeted by lineage-specific miRNAs are more likely to be unique to each species than predicted by chance. Pie chart size is scaled to number of predicted target genes. Color slices indicate orthogroup age for each predicted gene. (B) Genes that are both predicted targets of lineage-specific miRNAs and genes with differential expression in a social context (solid outlines) or genes under selection (dashed outlines) are represented by overlapping circles for each study and species. Numbers of lineage-specific miRNA targets are given for each species. Colors indicate different studies. Overlaps not significantly different from random (representation factor, RF=1) are unlabeled, while significant over- or under-enrichments are marked with asterisks with RF and p-value as indicated.

DISCUSSION

Eusociality is a major evolutionary innovation that requires regulatory changes in a wide range of molecular pathways (Robinson and Ben-Shahar 2002; Bloch and Grozinger 2011; Sumner et al. 2018). We tested the hypothesis that miRNAs play a role in the evolution of eusociality via their regulatory effects on gene networks. Our results provide several lines of support for this hypothesis.

First, miRNAs that have been previously implicated in social behavior (i.e., miR-276, miR-13b) do not show significant differences in copy number or diversification patterns consistent with social evolution, but the predicted targets of these miRNAs are largely disparate across species. This is consistent with what has been shown in vertebrates, flies, and nematodes, in which miRNAs are highly conserved at the sequence level, but miRNA-target relationships are highly divergent across clades (Chen and Rajewsky 2006; Chen and Rajewsky 2007). In our analysis, there were similarities in the broad functional categories of predicted targets across closely related species, and these categories were largely consistent with gene networks expected to play a role in social evolution (e.g., hormone and signal transduction pathways). Future research experimentally evaluating the function of these miRNAs in other social species will be necessary

493 to understand the degree to which miRNA function has diversified as a function of social
494 evolution.
495
496 Second, we identified a single miRNA (miR-305) that was expressed exclusively in the brains of
497 the social bees in our study. The presence of this miRNA in the solitary bee genomes suggests
498 that an evolutionary shift in expression pattern has accompanied at least two independent origins
499 of eusociality in bees. This miRNA coordinates Insulin and Notch signaling in *D. melanogaster*,
500 and both of these pathways are important regulators of social dynamics in insects (Wheeler et al.
501 2006; Ament et al. 2008; de Azevedo and Hartfelder 2008; Nilsen et al. 2011; Wang et al. 2013;
502 Hartfelder et al. 2015; Duncan et al. 2016; Chandra et al. 2018; Hartfelder et al. 2018). Some of
503 the predicted targets of miR-305 in bees (e.g., *oxysterol*, *synaptobrevin*) interact with Insulin and
504 Notch pathways in *D. melanogaster* (Stewart et al. 2001; Obniski et al. 2018). This miRNA was
505 also found to be upregulated in worker-destined honey bee larvae, compared to queen-destined
506 larvae, and may thus play a role in caste differentiation (Shi et al. 2015). Interestingly, miR-305
507 also promotes aging in *D. melanogaster* (Ueda et al. 2018). Extreme differences in lifespan are
508 one of the defining characteristics of social insect castes, whereby queens of some ant and bee
509 species live orders of magnitude longer than workers (Wilson 1971). The mechanisms
510 underlying these differences are a topic of interest, and a role for a miRNA that modulates
511 insulin signaling is consistent with current understanding of this phenomenon (Corona et al.
512 2007; Münch 2008; Remolina and Hughes 2008). Additional investigation is necessary to
513 determine how this miRNA may influence social behavior across species.
514

We focused additional attention on miRNAs for which no mature miRNAs with seed matches were detected in any other species, because these have the potential to influence the lineage-specific patterns of gene regulatory changes previously shown to influence social evolution (Simola et al. 2013; Kapheim et al. 2015). We hypothesized that if novel miRNAs are inserted into existing gene networks that become co-opted for social evolution, they should target genes that are highly conserved across species. Instead, we find that the targets of lineage-specific miRNAs are enriched for lineage-specific genes, while genes belonging to ancient orthogroups were not more likely to be targets than expected by chance. This could suggest that novel miRNAs co-evolve with novel genes, as has been shown for the evolution of cognitive function in humans (Barbash et al. 2014). Previous work in honey bees has shown that taxonomically-restricted (i.e., younger) genes play an important role in social evolution. Expression of taxonomically-restricted genes is significantly biased toward glands with specialized functions for life in a social colony (e.g., the hypopharyngeal gland and the sting gland) (Jasper et al. 2015), and toward genes that are upregulated in workers (Johnson and Tsutsui 2011). Thus, it is reasonable to expect that new miRNAs targeting new genes could have important social functions in honey bees.

Alternatively, it is possible that new miRNAs that target lineage-specific genes are transient and more likely to be purged by natural selection because they are less likely to integrate into existing gene networks (Chen and Rajewsky 2007; França et al. 2016; França et al. 2017). Emergent miRNAs are expected to initially have limited expression to mitigate potential deleterious effects on the protein-coding genes that they target. Thus, lineage-specific miRNAs with low levels of expression may be in the process of being purged and may not have

538 accumulated gene targets with important functions (Chen and Rajewsky 2007; Berezikov 2011).
539 Evidence for this model comes from primates (Berezikov et al. 2006) and flies (Lu et al. 2008;
540 Tang et al. 2010). Likewise, we find that lineage-specific miRNAs are expressed at significantly
541 lower levels than those with at least one homolog in another species. A purging process like this
542 could explain why there are large differences in the numbers of miRNAs detected in even closely
543 related species (e.g. the two *Bombus* species). It is also possible that some of this variation comes
544 from sequencing stochasticity. Functional analysis of lineage-specific genes in additional tissues
545 and life stages will help to resolve their role in social evolution.
546
547 We find support for the prediction that lineage-specific miRNAs should target genes with social
548 function in the Apidae (i.e., honey bees and bumble bees), but not the Halictidae (*M. genalis*).
549 There are several potential explanations for this pattern. One explanation is technical. We define
550 genes with social functions as those that are differentially expressed among castes. The genetic
551 basis of social behavior has been much better studied in honey bees than in any other species,
552 and the set of genes known to function in sociality is thus richer for apids than for halictids.
553 Further, not all genes that function in social behavior are expected to be differentially expressed
554 in the brains of different castes, and our analysis is thus likely to exclude some important genes.
555 Nonetheless, our results reflect differences in the degree of social complexity, and thus caste-
556 biased gene expression patterns, between apid and halictid bees. Unlike for honey bees and
557 bumble bees, which cannot live outside of social colonies, eusociality is facultative in *M. genalis*.
558 As such, caste traits are not fixed during development, and females who once served as non-
559 reproductive workers can transform into reproductive queens if given the opportunity (Smith et
560 al. 2009; Kapheim et al. 2012). This flexibility is reflected in the magnitude of differences in

561 brain gene expression patterns between queen and worker honey bees (thousands of genes;
562 Grozinger et al. 2007) and *M. genalis* (dozens of genes; Jones et al. 2017). Previous research
563 suggests that miRNAs increase their functional influence over evolutionary time (Berezikov et
564 al. 2006; Chen and Rajewsky 2007; Lu et al. 2008; Roux et al. 2012; França et al. 2016; Patel
565 and Capra 2017). Thus, emergent miRNAs are more likely to target genes with social function
566 due to chance alone in species with increased social complexity and a larger set of caste-biased
567 genes. Consistent with this explanation, regulatory relationships between miRNAs and genes
568 with caste-biased expression were not found among two other social insect species with reduced
569 social complexity (Patalano et al. 2015).
570
571 An additional explanation for these differences in the function of lineage-specific miRNAs
572 between bee families concerns the role of miRNAs in gene regulatory networks. One of these
573 roles is to stabilize regulatory relationships in the face of environmental variation, thus
574 canalizing phenotypes during development (Hornstein and Shomron 2006; Peterson et al. 2009;
575 Wu et al. 2009; Berezikov 2011). This is likely to be more important in species with obligate
576 eusociality, such as the honey bees and bumble bees for which caste determination is canalized,
577 than in species like *M. genalis*, where plasticity of phenotypes related to eusociality are
578 maintained in totipotent females.
579
580 Contrary to their effects on genes with socially-differentiated expression patterns, lineage-
581 specific miRNAs showed no evidence for preferential targeting of genes under positive selection
582 – either within individual species or across social species. In contrast, we find these emergent
583 miRNAs are less likely than expected by chance to target genes under positive selection in the

584 two halictid bees. A potential explanation for this pattern is that genes that are adaptively
585 targeted by miRNAs tend to be under purifying selection, because this maintains the regulatory
586 relationship between the miRNA and target, thus preventing gene mis-expression (Chen and
587 Rajewsky 2006; Saunders et al. 2007; Sarasin-Filipowicz et al. 2009; Franchini et al. 2016).
588 However, this selective constraint is likely to be most significant in the 3' UTR region, where
589 miRNA binding sites are located.
590
591 A more likely explanation involves the hypothesized pattern of miRNA origins and assimilation,
592 as proposed by Chen and Rajewsky (2007). This model suggests that newly emerged miRNAs
593 are likely to have many targets throughout the genome due to chance. Most of these initial
594 miRNA-target regulatory relationships are likely to have slightly deleterious effects, and would
595 likely be purged through purifying selection very quickly. These deleterious effects could be
596 particularly strong for target genes undergoing positive selection, because changes in the
597 functional regulation of these genes are likely to have significant fitness consequences. Also,
598 genes under positive selection are undergoing rapid evolution, and thus may be more likely to
599 "escape" control by errant miRNAs. Indeed, it is easier for mRNAs to lose miRNA target
600 binding sites, which typically require exact sequence matches, than to gain them (Chen and
601 Rajewsky 2007). Thus, emergent miRNAs may not be expected to target adaptively or fast
602 evolving genes, regardless of their role in social evolution.
603
604 The evolution of eusociality depends on many different tissues and physiological processes, and
605 brain-specific expression patterns are not likely to be representative of the complete role of
606 individual miRNAs in social behavior. Some or all of the predicted miRNA-gene relationships

607 we identified may have evolved to support traits in other cell types or processes unrelated to
608 sociality. Additional sequencing of miRNA and mRNA across tissue-types and stages of
609 development in social and solitary species is necessary to provide a comprehensive assessment
610 of the role of emergent miRNAs in social traits. Nonetheless, the brain is a major focus of
611 research in social evolution because it is the primary source of behavioral and neuroendocrine
612 output. Our results thus provide a good starting place for evaluating the role of miRNAs in
613 lineage-specific processes in the evolution of social behavior.
614
615 Our analyses reveal important differences in patterns of miRNA evolution between bees and
616 other species. For example, expansion in miRNA repertoire is associated with the evolution of
617 animal complexity in a wide range of species (Heimberg et al. 2008; Christodoulou et al. 2010;
618 Berezikov 2011). The evolution of eusociality from a solitary ancestor is associated with
619 increases in phenotypic complexity, and considered to be one of the major transitions in
620 evolution (Maynard Smith and Szathmáry 1995). We therefore hypothesized that evolutionary
621 increases in social complexity would be associated with expansions in the number of miRNAs
622 found within bee genomes. To the contrary, we find that most bees have a single copy of
623 previously identified miRNAs in their genomes. This is consistent with results of comparative
624 genome scans across several ant species (Simola et al. 2013). A recent study of miRNA diversity
625 in insects found that morphological innovations such as holometabolous development was
626 accompanied by the acquisition of only three miRNA families (Ylla et al. 2016). This suggests
627 that insect evolution is not as reliant on major expansions of miRNA families as other taxonomic
628 groups.
629

Additionally, our characterization of lineage-specific miRNAs expressed in the brain of each species reveals that genome structure is not as influential in regulating bee miRNA evolution as has been shown for human miRNAs. Novel human miRNAs tend to arise in ancient genes that have multiple functions and broad expression patterns (França et al. 2016). It is hypothesized that this increases the expression repertoire of emergent miRNAs, and thus facilitates persistence in the population (França et al. 2016; França et al. 2017). Only in one species (*N. melanderi*) were lineage-specific miRNAs more likely to be localized intragenically than previously identified miRNAs, while lineage-specific miRNAs did not differ from previously identified miRNAs in their genomic locations in the other five species. This suggests that emergence patterns for new miRNAs are unique to each lineage in bees. We also do not find a consistent pattern between young, emerging miRNAs and host gene age. There was no significant difference in the age of genes that serve as hosts for established versus lineage-specific miRNAs across all species. This is despite the fact that a similar proportion of bee miRNAs are located within introns (31-43%; Table 1), as compared to in vertebrates (36-65%) (Meunier et al. 2013). However, the fact that 73-88% of miRNAs that are localized to genes are encoded on the sense strand suggests that they would benefit from host transcription, as is observed in vertebrates (Meunier et al. 2013). Additional research with insects will be necessary to identify general patterns of miRNA evolution in relationship to genome structure.

Our study identifies patterns of miRNA evolution in a set of closely related bees that vary in social organization. Our results highlight important similarities and differences in the emergence patterns and functions of mammalian and insect genomes. We also find evidence that emergent miRNAs function in lineage-specific patterns of social evolution, perhaps through co-evolution

of novel miRNAs and species-specific targets. We do not see an overall increase in the number of miRNAs in the genome or expressed in the brains of species with more complex eusociality. However, we do find evidence that the role of miRNAs in social evolution may strengthen with increasing social complexity, perhaps due to an increased need for canalization of caste determination or due to chance, as a function of an increased number of genes with caste-biased expression. Empirical tests of miRNA function across additional species with variable social organization will further improve our understanding of how gene regulatory evolution gives rise to eusociality.

REFERENCES

Alaux C, Le Conte Y, Adams HA, Rodriguez-Zas S, Grozinger CM, Sinha S, Robinson GE. 2009. Regulation of brain gene expression in honey bees by brood pheromone. Genes, Brain Behav. 8:309–319.

Ament S A, Corona M, Pollock HS, Robinson GE. 2008. Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. Proc. Natl. Acad. Sci. U. S. A. 105:4226–4231.

Asgari S. 2013. MicroRNA functions in insects. Insect Biochem. Mol. Biol. 43:388–397.

Ashby R, Forêt S, Searle I, Maleszka R. 2016. MicroRNAs in Honey Bee Caste Determination. Sci. Rep. 6:1–5.

de Azevedo SV, Hartfelder K. 2008. The insulin signaling pathway in honey bee (Apis mellifera) caste development — differential expression of insulin-like peptides and insulin receptors in queen and worker larvae. J. Insect Physiol. 54:1064–1071.

Barbash S, Shifman S, Soreq H. 2014. Global coevolution of human microRNAs and their target

676 genes. Mol. Biol. Evol. 31:1237–1247.

677 Bartel DP. 2018. Metazoan MicroRNAs. Cell 173:20–51.

678 Behura SK, Whitfield CW. 2010. Correlated expression patterns of microRNA genes with age-
679 dependent behavioural changes in honeybee. Insect Mol. Biol. 19:431–439.

680 Berezikov E. 2011. Evolution of microRNA diversity and regulation in animals. Nat. Rev.
681 Genet. 12:846–860.

682 Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RHA.
683 2006. Diversity of microRNAs in human and chimpanzee brain. Nat. Genet. 38:1375–
684 1377.

685 Bloch G, Grozinger CM. 2011. Social molecular pathways and the evolution of bee societies.
686 Philos. Trans. R. Soc. B Biol. Sci. 366:2155–2170.

687 Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C, et al.
688 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*.
689 Science 329:1068–1071.

690 Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML, Gates MW,
691 Kula RR, Brady SG. 2017. Phylogenomic insights into the evolution of stinging wasps and
692 the origins of ants and bees. Curr. Biol. 27:1019–1025.

693 Cardinal S, Danforth BN. 2011. The antiquity and evolutionary history of social behavior in
694 bees. PLoS One 6:e21086.

695 Chandra V, Fetter-Pruneda I, Oxley PR, Ritger AL, McKenzie SK, Libbrecht R, Kronauer DJC.
696 2018. Social regulation of insulin signaling and the evolution of eusociality in ants. Science
697 361:398–402.

698 Chen K, Rajewsky N. 2006. Deep conservation of microRNA-target relationships and 3′UTR

699 motifs in vertebrates, flies, and nematodes. In: Cold Spring Harbor Symposia on
700 Quantitative Biology. Vol. 71. p. 149–156.
701 Chen K, Rajewsky Nikolaus. 2006. Natural selection on human microRNA binding sites inferred
702 from SNP data. Nat. Genet. 38:1452–1456.
703 Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and
704 microRNAs. Nat. Rev. Genet. 8:93–103.
705 Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, Klaus S, Snyman H, Hannon GJ,
706 Bork P, Arendt D. 2010. Ancient animal microRNAs and the evolution of tissue identity.
707 Nature 463:1084–1088.
708 Collins DH, Mohorianu I, Beckers M, Moulton V, Dalmay T, Bourke AFG. 2017. MicroRNAs
709 associated with caste determination and differentiation in a primitively eusocial insect. Sci.
710 Rep. 7:1–9.
711 Corona M, Velarde RA, Remolina S, Moran-Lauter A, Wang Y, Hughes KA, Robinson GE.
712 2007. Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity.
713 Proc. Natl. Acad. Sci. U. S. A. 104:7128–7133.
714 Cui X, Lu Z, Wang S, Jing-Yan Wang J, Gao X. 2016. CMsearch: simultaneous exploration of
715 protein sequence space and structure space improves not only protein homology detection
716 but also protein structure prediction. Bioinformatics 32:i332–i340.
717 D. Münch GVAFW. 2008. Ageing in a eusocial insect: molecular and physiological
718 characteristics of life span plasticity in the honey bee. Funct. Ecol. 22:407–421.
719 Duncan EJ, Hyink O, Dearden PK. 2016. Notch signalling mediates reproductive constraint in
720 the adult worker honeybee. Nat. Commun. 7:1–10.
721 Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP, de Graaf DC, Debyser G,

722 Deng J, Devreese B, et al. 2014. Finding the missing honey bee genes: Lessons learned
723 from a genome upgrade. BMC Genomics 15:1–29.
724 Enright AJ, John B, Gaul U, Tuschl T, Sander C. 2004. MicroRNA targets in *Drosophila*.
725 Genome Biol. 5:R1.
726 Filipowicz W, Bhattacharyya SN, Sonenberg N. 2008. Mechanisms of post-transcriptional
727 regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9:102-114.
728 França GS, Hinske LC, Galante PAFF, Vibranovski MD. 2017. Unveiling the impact of the
729 genomic architecture on the evolution of vertebrate microRNAs. Front. Genet. 8:1–8.
730 França GS, Vibranovski MD, Galante PAF. 2016. Host gene constraints and genomic context
731 impact the expression and evolution of human microRNAs. Nat. Commun. 7:11438.
732 Franchini P, Xiong P, Fruciano C, Meyer A. 2016. The role of microRNAs in the repeated
733 parallel diversification of lineages of midas cichlid fish from Nicaragua. Genome Biol.
734 Evol. 8:1543–1555.
735 Friedländer MR, MacKowiak SD, Li N, Chen W, Rajewsky N. 2012. MiRDeep2 accurately
736 identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic
737 Acids Res. 40:37–52.
738 Friedman RC, Farh KK-H, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved
739 targets of microRNAs. Genome Res. 19:92–105.
740 Gentleman R, Falcon S. 2013. Package "GOstats." 2.26.0.
741 Greenberg JK, Xia J, Zhou X, Thatcher SR, Gu X, Ament SA, Newman TC, Green PJ, Zhang W,
742 Robinson GE, et al. 2012. Behavioral plasticity in honey bees is associated with differences
743 in brain microRNA transcriptome. Genes, Brain Behav. 11:660–670.
744 Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar

745 DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs
746 in animals. Nature 455:1193–1197.
747 Grozinger CM, Fan Y, Hoover SER, Winston ML. 2007. Genome-wide analysis reveals
748 differences in brain gene expression patterns associated with caste and reproductive status
749 in honey bees (Apis mellifera). Mol. Ecol. 16:4837–4848.
750 Grün D, Wang Y-L, Langenberger D, Gunsalus KC, Rajewsky N. 2005. microRNA target
751 predictions across seven *Drosophila* species and comparison to mammalian targets. PLoS
752 Comput. Biol. 1:e13.
753 Harpur BA, Dey A, Albert JR, Patel S, Hines HM, Hasselmann M, Packer L, Zayed A. 2017.
754 Queens and workers contribute differently to adaptive evolution in bumble bees and honey
755 bees. Genome Biol. Evol. 9:2395–2402.
756 Harpur BA, Kent CF, Molodtsova D, Lebon JM, Alqarni AS, Owayss AA, Zayed A. 2014.
757 Population genomics of the honey bee reveals strong signatures of positive selection on
758 worker traits. Proc Natl Acad Sci U S A 111:2614–2619.
759 Harrison MC, Hammond RL, Mallon EB. 2015. Reproductive workers show queenlike gene
760 expression in an intermediately eusocial insect, the buff-tailed bumble bee *Bombus*
761 *terrestris*. Mol. Ecol. 24:3043–3063.
762 Hartfelder K, Guidugli-Lazzarini KR, Cervoni MS, Santos DE, Humann FC. 2015. Old threads
763 make new tapestry-rewiring of signalling pathways underlies caste phenotypic plasticity in
764 the honey bee, *Apis mellifera* L. 1st ed. Elsevier Ltd.
765 Hartfelder K, Tiberio GJ, Lago DC, Dallacqua RP, Bitondi MMG. 2018. The ovary and its
766 genes—developmental processes underlying the establishment and function of a highly
767 divergent reproductive system in the female castes of the honey bee, *Apis mellifera*.

768 Apidologie 49:49–70.

769 Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ. 2008. MicroRNAs and the
770 advent of vertebrate morphological complexity. Proc. Natl. Acad. Sci. U. S. A. 105:2946–
771 2950.

772 Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. Nat. Genet.
773 38:S20–S24.

774 Jasper WC, Linksvayer TA, Atallah J, Friedman D, Chiu JC, Johnson BR. 2015. Large-scale
775 coding sequence change underlies the evolution of postdevelopmental novelty in honey
776 bees. Mol. Biol. Evol. 32:334–346.

777 Johnson B, Tsutsui N. 2011. Taxonomically restricted genes are associated with the evolution of
778 sociality in the honey bee. BMC Genomics 12:164.

779 Johnson BR, Linksvayer TA. 2010. Deconstructing the superorganism: social physiology,
780 groundplans, and sociogenomics. Q. Rev. Biol. 85:57–79.

781 Jones BM, Kingwell CJ, Wcislo WT, Robinson GE. 2017. Caste-biased gene expression in a
782 facultatively eusocial bee suggests a role for genetic accommodation in the evolution of
783 eusociality. Proc. R. Soc. B Biol. Sci. 284.

784 Kapheim KM, Jones BM, Pan H, Harpur BA, Kent CF, Zayed A, Ioannidis P, Waterhouse RM,
785 Kingwell C, Stolle E, et al. Developmental plasticity shapes the genomic signatures of
786 eusociality. Unpublished

787 Kapheim KM, Pan H, Li C, Blatti C, Harpur BA, Ioannidis P, Jones BM, Kent CF, Ruzzante L,
788 Sloofman L, et al. 2019. Draft genome assembly and population genetics of an agricultural
789 pollinator, the solitary alkali bee (Halictidae: *Nomia melanderi*). G3. 9:625–634.

790 Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, Robertson HM, Hudson ME, Venkat

A, Fischman BJ, et al. 2015. Genomic signatures of evolutionary transitions from solitary to group living. Science 348:24–32.

Kapheim KM, Smith AR, Ihle KE, Amdam G V, Nonacs P, Wcislo WT. 2012. Physiological variation as a mechanism for developmental caste-biasing in a facultatively eusocial sweat bee. Proc. R. Soc. B Biol. Sci. 279:1437–1446.

Kocher SD, Li C, Yang W, Tan H, Yi S V, Yang X, Hoekstra HE, Zhang G, Pierce NE, Yu DW. 2013. The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. Genome Biol. 14:R142.

Krüger J, Rehmsmeier M. 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 34:W451-4.

Liu F, Peng W, Li Z, Li W, Li L, Pan J, Zhang S, Miao Y, Chen S, Su S. 2012. Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: Comparison between nurses and foragers. Insect Mol. Biol. 21:297–303.

Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu C-I. 2008. The birth and death of microRNA genes in *Drosophila*. Nat. Genet. 40:351–355.

MacEdo LMF, Nunes FMF, Freitas FCP, Pires C V., Tanaka ED, Martins JR, Piulachs MD, Cristino AS, Pinheiro DG, Simões ZLP. 2016. MicroRNA signatures characterizing caste-independent ovarian activity in queen and worker honeybees (*Apis mellifera* L.). Insect Mol. Biol. 25:216–226.

Marshall H, Lonsdale ZN, Mallon EB. 2019. Methylation and gene expression differences between reproductive castes of bumblebee workers. bioRxiv :517698. Available from: https://www.biorxiv.org/content/early/2019/01/11/517698

Maynard Smith J, Szathmáry E. 1995. The major transitions in evolution. Oxford, England:

814 Oxford University Press

815 Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P,
816 Kaessmann H. 2013. Birth and expression evolution of mammalian microRNA genes.
817 Genome Res. 23:34–45.

818 Nilsen K-A, Ihle KE, Frederick K, Fondrk MK, Smedal B, Hartfelder K, Amdam G V. 2011.
819 Insulin-like peptide genes in honey bee fat body respond differently to manipulation of
820 social behavioral physiology. J. Exp. Biol. 214:1488–1497.

821 Nunes FM, Ihle KE, Mutti NS, Simoes ZL, Amdam G V. 2013. The gene vitellogenin affects
822 microRNA regulation in honey bee (*Apis mellifera*) fat body and brain. J Exp Biol
823 216:3724–3732.

824 Obniski R, Sieber M, Spradling AC. 2018. Dietary lipids modulate Notch signaling and
825 influence adult intestinal development and metabolism in *Drosophila*. Dev. Cell 47:98-
826 111.e5.

827 Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R
828 language. Bioinformatics 20:289–290.

829 Patalano S, Vlasova A, Wyatt C, Ewels P, Camara F, Ferreira PG, Asher CL, Jurkowski TP,
830 Segonds-Pichon A, Bachman M, et al. 2015. Molecular signatures of plastic phenotypes in
831 two eusocial insect species with simple societies. Proc. Natl. Acad. Sci. 112:13970–13975.

832 Patel VD, Capra JA. 2017. Ancient human miRNAs are more likely to have broad functions and
833 disease associations than young miRNAs. BMC Genomics 18:672.

834 Peterson KJ, Dietrich MR, McPeek MA. 2009. MicroRNAs and metazoan macroevolution:
835 insights into canalization, complexity, and the Cambrian explosion. BioEssays 31:736–747.

836 Quah S, Hui JHL, Holland PWH. 2015. A burst of miRNA innovation in the early evolution of

837 butterflies and moths. Mol. Biol. Evol. 32:1161–1174.

838 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic

839 features. Bioinformatics 26:841–842.

840 R Core Team 2016. R: A language and environment for statistical computing. Available from:

841 https://www.r-project.org/

842 Remolina SC, Hughes KA. 2008. Evolution and mechanisms of long life and high fertility in

843 queen honey bees. Age. 30:177–185.

844 Robinson GE, Ben-Shahar Y. 2002. Social behavior and comparative genomics: new genes or

845 new gene regulation? Genes, Brain Behav. 1:197–203.

846 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential

847 expression analysis of digital gene expression data. Bioinformatics 26.

848 Roux J, Gonzàlez-Porta M, Robinson-Rechavi M. 2012. Comparative analysis of human and

849 mouse expression data illuminates tissue-specific evolutionary patterns of miRNAs. Nucleic

850 Acids Res. 40:5890–5900.

851 Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Gadau J,

852 Grimmelikhuijzen CJ, Hasselmann M, Lozier JD, et al. 2015. The genomes of two key

853 bumblebee species with primitive eusocial organization. Genome Biol. 16:76. A

854 Sarasin-Filipowicz M, Krol J, Markiewicz I, Heim MH, Filipowicz W. 2009. Decreased levels of

855 microRNA miR-122 in individuals with hepatitis C responding poorly to interferon therapy.

856 Nat. Med. 15:31–33.

857 Saunders MA, Liang H, Li W-H. 2007. Human polymorphism at microRNAs and microRNA

858 target sites. Proc. Natl. Acad. Sci. U. S. A. 104:3300–3305.

859 Shi YY, Zheng HJ, Pan QZ, Wang ZL, Zeng ZJ. 2015. Differentially expressed microRNAs

between queen and worker larvae of the honey bee (*Apis mellifera*). Apidologie 46:35–45.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539.

Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkampf M, Roux J, Nygaard S, Glastad KM, Hagen DE, Viljakainen L, et al. 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. Genome Res. 23:1235–1247.

Smit AFA, Hubley R, Green P: RepeatMasker. Available from: http://www.repeatmasker.org/

Smith AR, Kapheim KM, O'Donnell S, Wcislo WT. 2009. Social competition but not subfertility leads to a division of labour in the facultatively social sweat bee *Megalopta genalis* (Hymenoptera: Halictidae). Anim. Behav. 78:1043–1050.

Søvik E, Bloch G, Ben-Shahar Y, Sovik E, Bloch G, Ben-Shahar Y, Søvik E, Bloch G, Ben-Shahar Y. 2015. Function and evolution of microRNAs in eusocial Hymenoptera. Front. Genet. 6:1–11.

Stewart BA, Mohtashami M, Zhou L, Trimble WS, Boulianne GL. 2001. SNARE-dependent signaling at the *Drosophila* wing margin. Dev. Biol. 234:13–23.

Sumner S, Bell E, Taylor D. 2018. A molecular concept of caste in insect societies. Curr. Opin. Insect Sci. 25:42–50. Available from: http://dx.doi.org/10.1016/j.cois.2017.11.010

Tang T, Kumar S, Shen Y, Lu J, Wu M-L, Shi S, Li W-H, Wu C-I. 2010. Adverse interactions between micro-RNAs and target genes from different species. Proc. Natl. Acad. Sci. U. S. A. 107:12935–12940.

Toth AL, Robinson GE. 2007. Evo-devo and the evolution of social behavior. Trends Genet.

883 23:334–341.

884 Ueda M, Sato T, Ohkawa Y, Inoue YH. 2018. Identification of *miR-305* , a microRNA that

885 promotes aging, and its target mRNAs in *Drosophila*. Genes to Cells 23:80–93.

886 Wang Y, Azevedo S V, Hartfelder K, Amdam G V. 2013. Insulin-like peptides (AmILP1 and

887 AmILP2) differentially affect female caste development in the honey bee (*Apis mellifera*

888 L.). J Exp Biol 216:4347–4357.

889 Weaver D, Anzola J, Evans J, Reid J, Reese J, Childs K, Zdobnov E, Samanta M, Miller J, Elsik

890 C. 2007. Computational and transcriptional evidence for microRNAs in the honey bee

891 genome. Genome Biol. 8:R97.

892 Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ. 2009.

893 The deep evolution of metazoan microRNAs. Evol. Dev. 11:50–68.

894 Wheeler DE, Buck N, Evans JD. 2006. Expression of insulin pathway genes during the period of

895 caste determination in the honey bee, *Apis mellifera*. Insect Mol. Biol. 15:597–602.

896 Wheeler MM, Ament SA, Rodriguez-Zas SL, Robinson GE. 2013. Brain gene expression

897 changes elicited by peripheral *vitellogenin* knockdown in the honey bee. Insect Mol. Biol.

898 22:562–573.

899 Whitfield CW, Cziko A-. M, Robinson GE. 2003. Gene expression profiles in the brain predict

900 behavior in individual honey bees. Science 302.

901 Wilson EO. 1971. The Insect Societies. Cambridge, Massachusetts: Harvard University Press

902 Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG,

903 Robinson GE. 2011. Genes involved in convergent evolution of eusociality in bees. Proc

904 Natl Acad Sci U S A 108:7472–7477.

905 Wu C-I, Shen Y, Tang T. 2009. Evolution under canalization and the dual roles of microRNAs: a

906 hypothesis. Genome Res. 19:734–743.

907 Ylla G, Fromm B, Piulachs MD, Belles X. 2016. The microRNA toolkit of insects. Sci. Rep.

908 6:1–13.

909 Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M,

910 Loetscher A, Kriventseva E V. 2017. OrthoDB v9.1: cataloging evolutionary and functional

911 annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids

912 Res. 45:D744–D749.

913 Zhu Y, Skogerbø G, Ning Q, Wang Z, Li B, Yang S, Sun H, Li Y. 2012. Evolutionary

914 relationships between miRNA genes and their activity. BMC Genomics13:718.

915

916 **AUTHOR CONTRIBUTIONS**

917 K.M.K. conceived of the study and designed the experiments. K.M.K., E.S., G.B., and Y.B-S.

918 collected the data. K.M.K., B.M.J., E.S., and R.M.W. analyzed the data. K.M.K. wrote the initial

919 draft of the manuscript. All authors edited and approved the article for publication.

920

921 **ACKNOWLEDGMENTS**

922 This study was supported by funds from Utah State University and the Utah Agricultural

923 Experiment Station (Project 1297) and by the US-Israel Binational Science Foundation (BSF

924 2012807) (to G.B.). R.M.W. was supported by Swiss National Science Foundation Grant

925 PP00P3_170664. Sequencing was performed at the University of Illinois Roy J. Carver

926 Biotechnology Center. We thank the University of Utah High Performance Computing Center

927 for computational time and assistance. Illustrations were created by J. Johnson (LifeSciences

928 Studios). We thank G. Robinson for helpful feedback on an earlier draft of this manuscript.

929