Brain microRNAs among social and solitary bees

Karen M. Kapheim[1*], Beryl M. Jones[2], 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] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

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

[4] Centre of Molecular Biodiversity Research, Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, 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, 91904, Israel

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

* Author for Correspondence:
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 unknown. We examined the relationship between social organization and lineage-specific microRNAs (miRNAs). Genome scans across 12 bee species showed that miRNA copy-number is mostly conserved and not associated with sociality. However, deep sequencing of small RNAs in six bee species revealed a substantial proportion (20-35%) of detected miRNAs had lineage-specific expression in the brain, 24-72% of which did not have homologs in other species. Lineage-specific miRNAs disproportionately target lineage-specific genes, and have lower expression levels than shared miRNAs. The predicted targets of lineage-specific miRNAs are not enriched for genes with caste-biased expression or genes under positive selection in social species. Together, these results suggest that novel miRNAs may coevolve with novel genes, and thus contribute to lineage-specific patterns of evolution in bees, but do not appear to have significant influence on social evolution. Our analyses also support the hypothesis that many new miRNAs are purged by selection due to deleterious effects on mRNA targets, and suggest genome structure is not as influential in regulating bee miRNA evolution as has been shown for mammalian miRNAs.

Keywords: Gene regulation; small non-coding RNA; microRNA targets; eusociality; lineage-specific

INTRODUCTION

Eusociality has evolved several times in the hymenopteran insects. In its most basic form, this lifestyle involves reproductive queens living with their worker daughters who forego direct reproduction to cooperatively defend the nest, care for their siblings, and forage for the colony. 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 [1]. 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 [2]. Recent studies have found that social insect species share evolutionary genomic changes that may reflect an increased capacity for gene regulation [3,4]. 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 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 [3–5]. 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 of messenger RNA (mRNA) transcripts, in most cases preventing translation or causing mRNA degradation [6]. Each miRNA can target dozens to hundreds of mRNAs, and may therefore regulate multiple gene networks [6,7]. 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 [6]. Novel miRNAs 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 [8–11]. It is thus not surprising that expansion of the miRNA repertoire is associated with the evolution of morphological complexity across the tree of life [9,12,13].

There is accumulating evidence for a role of miRNAs in regulating the social lives of insects. While most miRNAs seem to be conserved in major lineages of insects [14,15], expression levels vary across individuals performing different social functions, such as between workers performing different tasks in honey bees [16–18]. MiRNAs may also play a role in caste determination, as queen- and worker-destined larvae express different sets of miRNAs throughout development in honey bees [19–21] and bumble bees [22]. Additionally, miRNAs play a role in regulating some physiological correlates of social behavior in honey bees, including activation of ovaries in queens and workers [23] and response to the reproductive protein *vitellogenin* [24]. Together, these studies suggest that miRNAs could play a role in the evolution of eusociality through their effects on gene regulatory networks involved in socially-relevant traits. A rigorous test of this hypothesis requires comparisons of the presence, expression, and function of miRNAs across related species that vary in social organization.

Here we present a comparative analysis of miRNAs across bee species with variable social organization. We first looked for miRNA repertoire expansions associated with eusociality by scanning 12 bee genomes for known miRNAs, and statistically evaluating copy-number of each miRNA type with regard to differences in sociality in a phylogenetic model. We then described and compared miRNAs expressed in the brains of six bee species from three families that include repeated origins of eusociality. We tested the hypothesis that changes in gene regulatory function associated with social evolution are facilitated by lineage-specific miRNAs with two predictions: (1) If lineage-specific miRNAs are assimilated into ancestral gene networks, their predicted target genes should be ancient and conserved. (2) If lineage-specific miRNAs play a role in social evolution, their predicted targets should be enriched for genes associated with social behavior (e.g., caste-biased expression) or genes that are under selection in social species. We do not find evidence for a role of lineage-specific miRNAs in social evolution. However, we do identify unexpected patterns of coevolution between miRNAs and their putative target genes. We interpreted our results in light of current hypotheses for patterns of miRNA evolution in vertebrates.

MATERIALS AND METHODS

miRNA Diversification

We performed genome scans for small RNAs across 12 bee genomes (Table S1) using covariance models implemented with Infernal (v1.1) cmsearch using the gathering threshold for inclusion (--cut_ga) [25] to find all Rfam (release 29.0) accessions in each genome. We used Spearman rank regressions to test for associations between miRNA copy-number and social biology. We categorized each species as solitary, facultative basic eusocial, obligate basic

125 eusocial, or obligate complex eusocial following Kapheim et al. [4]. We used the ape package
126 (v3.1) [26] in R (v3.5.0) [27] to calculate phylogenetic independent contrasts for both social
127 organization and miRNA copy-number, cor.test to implement Spearman's rank correlations, and
128 p.adjust with the Benjamini-Hochberg method to correct for multiple comparisons.
129
130 **Sample Acquisition, RNA isolation, and Sequencing**
131 We used a single adult female from six bee species, including both eusocial and solitary
132 species with well-studied behavior from three bee families (Fig. 1). Details of sample collection,
133 RNA isolation, and sequencing are provided in Table S2.
134
135 **miRNA Discovery and Quantification**
136 We used miRDeep2 (v2.0.0.8) [28] to identify and quantify miRNAs expressed in the
137 brains of each species, with a three-step process of miRNA detection to identify homologous
138 miRNAs between species. First, we gathered mature miRNA sequences previously described in
139 other insect species (Table S3). Reads for each sample were quality filtered (minimum length 18,
140 removal of reads with non-standard bases), adapter-trimmed, and aligned to the species' genome
141 (Table S1) with the mapper.pl script. Approximately 61-84% of reads successfully mapped
142 (Table S2).
143
144 We then identified known and novel miRNAs in each sample with the miRDeep2.pl
145 script, using our set of insect miRNAs (Table S3) as known mature sequences. The quantifier.pl
146 script generated sets of known and novel miRNAs in each sample, along with quantified
147 expression information. We filtered novel miRNAs in each species according to the following

148 criteria: no rRNA/tRNA similarities, minimum of five reads each on mature and star strands of
149 the hairpin sequence, and a randfold p-value<0.05. Randfold describes the RNA secondary
150 structure of potential precursor miR (pre-miRs) [28].
151
152 We used these filtered miRNAs in a second run of detection and quantification, repeating
153 the pipeline above after adding mature sequences of novel miRNAs from each species to our set
154 of known miRNAs. This allowed detection of homologous miRNAs (based on matching seed
155 sequences) not represented in miRBase across our species. We applied the same set of filtering
156 criteria as above.
157
158 Some novel miRNAs may exist in the genomes of other bees, even if they are not
159 expressed. We used blastn (-perc_identity 50 -evalue 1e-5) to search for homologous pre-miR
160 sequences in 12 bee genomes (Table S1) for each novel miRNA without a matching seed
161 sequence.
162
163 **miRNA Localization**
164 We used bedtools (v2.27.0) intersect [29] to find overlap of miRNAs with predicted gene
165 models (Table S4), and repetitive element annotations from previously established repeat
166 libraries that had been generated using Repeatmasker [4,30–34].
167
168 **Target Prediction**
169 We extracted potential target sites 500 bp downstream from each gene model using
170 bedtools flank and getfasta [29], following previous studies [19] and an average 3' UTR of 442

nt in *Drosophila melanogaster* [35]. Target prediction was run with miRanda (v3.3) [36] (minimum energy threshold -20, minimum score 140, strict alignment to seed region [-en -20 -sc 140 –strict]) and RNAhybrid (v2.12) [37] (minimum free energy threshold -20). We kept only miRNA-target gene pairs that were predicted by both programs with $p<0.01$.

Target Age and Functional Enrichment

Gene ages were determined using orthogroups from OrthoDB (v9) [38], which includes *A. mellifera*, *B. impatiens*, *B. terrestris*, and *M. rotundata*. Gene sets of *M. genalis* and *N. melanderi* were mapped to Metazoa-level (330 species) orthogroups. Gene ages were inferred from the taxonomic breadth of all species in each orthogroup, with at least one representative from each of the following groups which does not belong to the next lower group: Vertebrata, Metazoa, Arthropoda, Insecta, Holometabola, Hymenoptera, Aculeata, Apoidea. Genes without identifiable orthologs were labeled 'Unique'.

Enrichment tests of lineage-specific miRNA targets

For each species, gene expression datasets related to socially relevant phenotypes (e.g., caste) were compared against targets of lineage-specific miRNAs (Table S5). For *M. genalis* caste data, RNAseq reads from Jones et al. [39] (NCBI PRJNA331103) were trimmed using Trimmomatic (v0.36) [40] and aligned to an unpublished genome assembly of *M. genalis* (NCBI PRJNA494872) using STAR (v2.5.3) [41]. Gene counts were obtained using featureCounts in the Subread package (v1.5.2) [42], and differential expression analysis was conducted using edgeR [43] as in Jones et al. [39].

We also tested datasets identifying genes under selection in bee species [32,44,45] or across social lineages of bees [4,46] for enrichment of lineage-specific miRNA targets (Table S5). When necessary, we used reciprocal blastp (evalue<$10e^{-5}$) to identify orthologous genes, and only genes with putative orthologs were included. Hypergeometric tests (using phyper in R) were used to test for 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). RF is calculated as RF=x/E, where x is the number of genes in common between two lists and E is the expected number of shared genes (E = nD/N, where n is the number of genes in list 1, D is the number of genes in list 2, and N is the total number of genes.)

RESULTS

Low levels of miRNA copy-number variation among bee genomes

Our genome scans revealed very little variation in copy-number of most miRNAs. Of the 50 miRNA Rfam accessions, half had the same number of copies (1 or 2) in all 12 bee genomes (Table S6). The mean copy-number across all miRNAs in all bee genomes was 1.19±0.74. One exception was miR-1122, for which we found 70 copies in *M. genalis*, but no copies in the other species. We did not find any significant associations between miRNA copy-number and social organization (Table S6).

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 S7). The majority of these were intergenic or within introns (Table 1). Each species had at least one miRNA originating 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 few overlaps with well-characterized transposons or retrotransposons (Table 1). Variation in number of expressed miRNAs in each species was not related to observable technical variation, such as sequencing center, number of reads, number or proportion of reads mapped to the genome, or type of sample from which they were obtained (Table S2). This variation in number of expressed miRNAs is similar to that found in other groups of species with shorter divergence times [47].

Most detected miRNAs in each species had known homologs in at least one other species. However, each species had a substantial proportion (20-35%) of detected miRNAs with lineage-specific expression in the brain (Table 1; Fig. 1), 24-72% of which did not have any known homologs in other species (Table 1). We defined lineage-specific miRNAs as those with lineage-specific expression and with no seed match to a known mature miRNA (Table 1, columns 6-7), because these show the most evidence of being real miRNAs that are unique to a particular species. (Sequence similarity of pre-miRs in other bee genomes is not evidence that a mature miRNA is transcribed.) Lineage-specific miRNAs had significantly lower expression levels compared with homologous miRNAs in each species (t-tests: *A. mellfera, M. rotundata, M. genalis* $p<0.001$, *B. impatiens, B. terrestris* $p<0.01$, *N. melanderi* $p<0.05$).

Lineage-specific miRNAs were localized both within genes and intergenically. The proportion of lineage-specific miRNAs that were intra- 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 serving as hosts for intragenic lineage-specific miRNAs were not significantly older than genes hosting miRNAs with known homologs (χ^2 tests: p=0.05-0.89).

Of miRNAs with homologs, most were expressed in all six species, but one miRNA (miR-305) was expressed in the brains of each of the social, but not the solitary, species (Fig. 2). 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 miR-305 (Table S6). 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*.

Lineage-specific miRNAs preferentially target lineage-specific genes, but not genes with caste-biased expression or genes under positive selection

If lineage-specific changes in gene regulatory function associated with social evolution are facilitated by novel miRNAs inserted into existing gene networks, then predicted targets of lineage-specific miRNAs should be highly conserved and enriched for genes with known functions in social evolution. Most predicted mRNA targets of lineage-specific miRNAs were

highly conserved and belonged to orthogroups shared by vertebrates (Fig. 3; Table S8), but not significantly more than expected given the large number of conserved genes in each genome (hypergeometric tests: $p>0.99$). We did, however, find significant enrichment for genes unique to each species among the predicted targets of lineage-specific miRNAs (hypergeometric tests: *A. mellifera* – RF=1.51, $p=5.44e^{-5}$; *B. impatiens* – RF=1.28, p=0.02; *B. terrestris* – RF=1.78, $p=1.90e^{-6}$; *M. rotundata* – RF=1.79, p=0.0002; *M. genalis* – RF=1.62, $p=1.48e^{-12}$; *N. melanderi* – RF=1.78, $p=9.02e^{-5}$), indicating that novel miRNAs are more likely to target novel genes than would be expected by chance (Fig. 3; Table S8).

We did not find support for the prediction that novel miRNAs should target genes that function in social behavior and evolution. We first considered the genes that are differentially expressed between castes in social species, because these are likely to be involved in regulating behavioral and physiological aspects of sociality. Predicted targets of lineage-specific miRNAs were not significantly enriched for genes with caste-biased expression in the social species (Fig. S1; Table S5). Also contrary to our prediction, targets of lineage-specific miRNAs were not enriched for genes under positive selection in any species (Fig. S1; Table S5). In fact, genes under positive selection in the halictid bees were significantly depleted for targets of lineage-specific miRNAs (hypergeometric tests: *M. genalis* – RF=0.3, $p=3.72e^{-7}$; *N. melanderi* – RF=0.3, $p=9.79e^{-4}$). We also assessed overlaps with genes previously found to be under positive selection in social species, compared to solitary species [4,46], but found no significant overlap or depletion with predicted targets of lineage-specific genes (hypergeometric tests: $p>0.05$; Fig. S1; Table S5).

DISCUSSION

Eusociality is a major evolutionary innovation that requires regulatory changes in a wide range of molecular pathways [1]. We tested the hypothesis that miRNAs play a role in the evolution of eusociality via their regulatory effects on gene networks by comparing miRNA expression in three eusocial and three solitary bee species from three families. Our results provide very limited support for this hypothesis.

We identified a single miRNA (miR-305) that was expressed exclusively in the brains of social bees in our study. The presence of this miRNA in the solitary bee genomes suggests that an evolutionary shift in expression pattern may have accompanied at least two independent origins of eusociality in bees. This miRNA coordinates Insulin and Notch signaling in *D. melanogaster*, both of which are important regulators of social dynamics in insects [48–50]. Interestingly, miR-305 is also upregulated in worker-destined compared to queen-destined honey bee larvae, and may thus play a role in caste differentiation [20]. Further investigation with additional social and solitary species is necessary to determine if this miRNA is expressed exclusively in the brains of social species and how it may influence social behavior.

We focused attention on miRNAs for which no mature miRNAs with seed matches were detected in any other species, because these may influence the lineage-specific patterns of gene regulatory changes previously shown to influence social evolution [3,4]. 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. Instead, we find that 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 suggests that novel miRNAs co-evolve with novel genes, as has been shown for the evolution of cognitive function in humans [51]. Previous work in honey bees has shown that taxonomically-restricted genes play an important role in social evolution, with expression of these genes biased toward glands with specialized functions for life in a social colony (e.g., the hypopharyngeal and sting glands) [52], and upregulated in workers [53]. Thus, it is reasonable to expect that new miRNAs targeting new genes could have important social functions.

Alternatively, it is possible that new miRNAs targeting lineage-specific genes are transient and will be purged by natural selection because they are less integrated into existing gene networks [10,54,55]. Emergent miRNAs are expected to initially have limited expression to mitigate potential deleterious effects on their target genes. Thus, lineage-specific miRNAs with low levels of expression may be in the process of being purged and may not have accumulated gene targets with important functions [9,10]. Evidence for this model comes from primates [56] and flies [11,57]. Likewise, we find that lineage-specific miRNAs have reduced expression compared to those with homologs. A purging process could explain why there are large differences in the numbers of miRNAs detected in even closely related species (e.g., the two *Bombus* species). Functional analysis of lineage-specific genes in additional tissues and life stages will help to resolve their roles in social evolution.

We do not find support for the prediction that lineage-specific miRNAs should target genes associated with caste in social bees. Consistent with this observation, regulatory relationships between miRNAs and genes with caste-biased expression were not found among

two other social insect species [58]. Previous studies have identified miRNAs that are differentially expressed between queens and workers in honey bees [19–21] and bumble bees [22]. However, without comparison to other bee species, it was unknown if these caste-biased miRNAs were unique to social species. Our results suggest this is not the case. This is perhaps unsurprising in light of our finding that lineage-specific miRNAs target an unexpectedly high proportion of lineage-specific genes, potentially through coevolution. Although lineage-specific genes play an important role in sociality [59], most caste-biased genes belong to highly conserved molecular pathways [60].

Lineage-specific miRNAs also showed no evidence for preferential targeting of genes under positive selection – either within or across species. In contrast, we find these emergent miRNAs are less likely than expected by chance to target genes under positive selection in the two halictid bees. A potential explanation is that genes adaptively targeted by miRNAs tend to be under purifying selection to maintain regulatory relationships with their targets, preventing gene mis-expression [61–63]. This selective constraint is likely to be most significant in the 3' UTR, where miRNA binding sites are located.

A more likely explanation for both of these negative results involves the hypothesized pattern of miRNA origins and assimilation [10]. This model suggests that new miRNAs are likely to have many targets throughout the genome due to chance. Most initial miRNA-target regulatory relationships are likely to have slightly deleterious effects, and would be quickly purged through purifying selection. These deleterious effects could be particularly strong for target genes with caste-biased expression or undergoing positive selection, because changes in

355 the functional regulation of these genes are likely to have significant fitness consequences. Also,
356 genes with caste-biased expression and those under positive selection are undergoing rapid
357 evolution [64], and thus may be more likely to "escape" control by errant miRNAs. Indeed, it is
358 easier for mRNAs to lose miRNA target binding sites, which typically require exact sequence
359 matches, than to gain them [10]. Thus, emergent miRNAs may not be expected to target
360 adaptively or fast evolving genes, regardless of their role in social evolution.
361
362 Our analyses reveal important differences in patterns of miRNA evolution between bees
363 and other species. For example, expansion in miRNA repertoire is associated with the evolution
364 of animal complexity in a wide range of species [9,12,13]. The evolution of eusociality from a
365 solitary ancestor is associated with increases in phenotypic complexity, and considered to be one
366 of the major transitions in evolution. We therefore hypothesized that evolutionary increases in
367 social complexity would be associated with expansions in the number of miRNAs found within
368 bee genomes. To the contrary, we find that most bees have a single copy of previously identified
369 miRNAs in their genomes, consistent with results of comparative genome scans across ants [3].
370 A recent study of miRNA diversity in insects found that morphological innovations such as
371 holometabolous development was accompanied by the acquisition of only three miRNA families
372 [15]. This suggests that insect evolution is not as reliant on major expansions of miRNA families
373 as other taxonomic groups.
374
375 Additionally, our characterization of lineage-specific miRNAs expressed in the brain of
376 each species reveals that genome structure is not as influential in regulating bee miRNA
377 evolution as has been shown for human miRNAs. Novel human miRNAs tend to arise within

ancient genes that have multiple functions and broad expression patterns, which may facilitate persistence of emergent miRNAs by increasing their expression repertoire [54,55]. In our study, lineage-specific miRNAs did not differ from previously identified miRNAs in their genomic locations in all but one species (*N. melanderi*). We also do not find a consistent pattern between new miRNAs and host gene age, even though a similar proportion of bee miRNAs are located within introns (31-43%; Table 1), compared to in vertebrates (36-65%) [8]. However, the fact that 73-88% of bee miRNAs localized to genes are encoded on the sense strand suggests that they would benefit from host transcription, as is observed in vertebrates [8]. 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 bees that vary in social organization, and highlights important similarities and differences in the emergence patterns and functions of mammalian and insect genomes. We find no evidence that emergent miRNAs function in lineage-specific patterns of social evolution, but we do find evidence of potential 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 one miRNA (miR-305) expressed in the brains of social, but not solitary, species. 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.

400 **REFERENCES**

401 1. Sumner S, Bell E, Taylor D. 2018 A molecular concept of caste in insect societies. *Curr.*
402 *Opin. Insect Sci.* **25**, 42–50. (doi:10.1016/j.cois.2017.11.010)
403 2. Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM,
404 Soares MB, Robinson GE. 2002 Annotated expressed sequence tags and cDNA
405 microarrays for studies of brain and behavior in the honey bee. *Genome Res.* **12**, 555–566.
406 3. Simola DF *et al.* 2013 Social insect genomes exhibit dramatic evolution in gene
407 composition and regulation while preserving regulatory features linked to sociality.
408 *Genome Res.* **23**, 1235–1247. (doi:10.1101/gr.155408.113)
409 4. Kapheim KM *et al.* 2015 Genomic signatures of evolutionary transitions from solitary to
410 group living. *Science* **348**, 24–32. (doi:10.1126/science.aaa4788)
411 5. Warner MR, Qiu L, Holmes MJ, Mikheyev AS, Linksvayer TA. 2019 Convergent
412 eusocial evolution is based on a shared reproductive groundplan plus lineage-specific
413 plastic genes. *Nat. Commun.* **10**, 2651. (doi:10.1038/s41467-019-10546-w)
414 6. Bartel DP. 2018 Metazoan microRNAs. *Cell* **173**, 20–51. (doi:10.1016/j.cell.2018.03.006)
415 7. Friedman RC, Farh KK-H, Burge CB, Bartel DP. 2009 Most mammalian mRNAs are
416 conserved targets of microRNAs. *Genome Res.* **19**, 92–105. (doi:10.1101/gr.082701.108)
417 8. Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H,
418 Khaitovich P, Kaessmann H. 2013 Birth and expression evolution of mammalian
419 microRNA genes. *Genome Res.* **23**, 34–45. (doi:10.1101/gr.140269.112)
420 9. Berezikov E. 2011 Evolution of microRNA diversity and regulation in animals. *Nat. Rev.*
421 *Genet.* **12**, 846–860. (doi:10.1038/nrg3079)
422 10. Chen K, Rajewsky N. 2007 The evolution of gene regulation by transcription factors and

423 microRNAs. *Nat. Rev. Genet.* **8**, 93–103. (doi:10.1038/nrg1990)

424 11. Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu C-I. 2008 The
425 birth and death of microRNA genes in *Drosophila*. *Nat. Genet.* **40**, 351–355.
426 (doi:10.1038/ng.73)

427 12. Christodoulou F *et al.* 2010 Ancient animal microRNAs and the evolution of tissue
428 identity. *Nature* **463**, 1084–1088. (doi:10.1038/nature08744)

429 13. Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ. 2008 MicroRNAs
430 and the advent of vertebrate morphological complexity. *Proc. Natl. Acad. Sci. U. S. A.*
431 **105**, 2946–50. (doi:10.1073/pnas.0712259105)

432 14. Søvik E, Bloch G, Ben-Shahar Y. 2015 Function and evolution of microRNAs in eusocial
433 Hymenoptera. *Front. Genet.* **6**, 1–11. (doi:10.3389/fgene.2015.00193)

434 15. Ylla G, Fromm B, Piulachs MD, Belles X. 2016 The microRNA toolkit of insects. *Sci.*
435 *Rep.* **6**, 1–13. (doi:10.1038/srep37736)

436 16. Greenberg JK *et al.* 2012 Behavioral plasticity in honey bees is associated with
437 differences in brain microRNA transcriptome. *Genes, Brain Behav.* **11**, 660–670.
438 (doi:10.1111/j.1601-183X.2012.00782.x)

439 17. Liu F *et al.* 2012 Next-generation small RNA sequencing for microRNAs profiling in *Apis*
440 *mellifera*: Comparison between nurses and foragers. *Insect Mol. Biol.* **21**, 297–303.
441 (doi:10.1111/j.1365-2583.2012.01135.x)

442 18. Behura SK, Whitfield CW. 2010 Correlated expression patterns of microRNA genes with
443 age-dependent behavioural changes in honeybee. *Insect Mol. Biol.* **19**, 431–439.
444 (doi:10.1111/j.1365-2583.2010.01010.x)

445 19. Ashby R, Forêt S, Searle I, Maleszka R. 2016 MicroRNAs in honey bee caste

446 determination. *Sci. Rep.* **6**, 1–15. (doi:10.1038/srep18794)

447 20. Shi YY, Zheng HJ, Pan QZ, Wang ZL, Zeng ZJ. 2015 Differentially expressed
448 microRNAs between queen and worker larvae of the honey bee (*Apis mellifera*).
449 *Apidologie* **46**, 35–45. (doi:10.1007/s13592-014-0299-9)

450 21. Weaver D *et al.* 2007 Computational and transcriptional evidence for microRNAs in the
451 honey bee genome. *Genome Biol.* **8**, R97.

452 22. Collins DH, Mohorianu I, Beckers M, Moulton V, Dalmay T, Bourke AFG. 2017
453 MicroRNAs associated with caste determination and differentiation in a primitively
454 eusocial insect. *Sci. Rep.* **7**, 1–9. (doi:10.1038/srep45674)

455 23. MacEdo LMF *et al.* 2016 MicroRNA signatures characterizing caste-independent ovarian
456 activity in queen and worker honeybees (*Apis mellifera* L.). *Insect Mol. Biol.* **25**, 216–226.
457 (doi:10.1111/imb.12214)

458 24. Nunes FMF, Ihle KE, Mutti NS, Simoes ZLP, Amdam G V. 2013 The gene *vitellogenin*
459 affects microRNA regulation in honey bee (*Apis mellifera*) fat body and brain. *J Exp Biol*
460 **216**, 3724–3732. (doi:10.1242/jeb.089243)

461 25. Cui X, Lu Z, Wang S, Jing-Yan Wang J, Gao X. 2016 CMsearch: simultaneous
462 exploration of protein sequence space and structure space improves not only protein
463 homology detection but also protein structure prediction. *Bioinformatics* **32**, i332–i340.
464 (doi:10.1093/bioinformatics/btw271)

465 26. Paradis E, Claude J, Strimmer K. 2004 APE: analyses of phylogenetics and evolution in R
466 language. *Bioinformatics* **20**, 289–290.

467 27. Team RC. 2016 R: A language and environment for statistical computing.

468 28. Friedländer MR, MacKowiak SD, Li N, Chen W, Rajewsky N. 2012 MiRDeep2

accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* **40**, 37–52. (doi:10.1093/nar/gkr688)

29. Quinlan AR, Hall IM. 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842. (doi:10.1093/bioinformatics/btq033)

30. Elsik CG *et al.* 2014 Finding the missing honey bee genes: Lessons learned from a genome upgrade. *BMC Genomics* **15**, 1–29. (doi:10.1186/1471-2164-15-86)

31. Sadd BM *et al.* 2015 The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol.* **16**, 76. (doi:10.1186/s13059-015-0623-3)

32. Kapheim KM *et al.* 2019 Draft genome assembly and population genetics of an agricultural pollinator, the solitary alkali bee (Halictidae: *Nomia melanderi*). *G3* **9**, 625–634. (doi:10.1534/g3.118.200865)

33. Kapheim KM *et al.* In press. Developmental plasticity shapes the genomic signatures of eusociality. *unpublished*

34. Smith AF., Hubley R, Green P. 2013 RepeatMasker. See http://www.repeatmasker.org/.

35. Grün D, Wang Y-L, Langenberger D, Gunsalus KC, Rajewsky N. 2005 microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput. Biol.* **1**, e13. (doi:10.1371/journal.pcbi.0010013)

36. Enright AJ, John B, Gaul U, Tuschl T, Sander C. 2004 MicroRNA targets in *Drosophila*. *Genome Biol.* **5**, R1. (doi:10.1186/gb-2003-5-1-r1)

37. Krüger J, Rehmsmeier M. 2006 RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* **34**, W451-4. (doi:10.1093/nar/gkl243)

38. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva E V. 2017 OrthoDB v9.1: cataloging evolutionary

and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744–D749. (doi:10.1093/nar/gkw1119)

39. Jones BM, Kingwell CJ, Wcislo WT, Robinson GE. 2017 Caste-biased gene expression in a facultatively eusocial bee suggests a role for genetic accommodation in the evolution of eusociality. *Proc. R. Soc. B Biol. Sci.* **284**. (doi:10.1098/rspb.2016.2228)

40. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**. (doi:10.1093/bioinformatics/btu170)

41. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S. 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinforma Oxf Engl* **29**. (doi:10.1093/bioinformatics/bts635)

42. Liao Y, Smyth GK, Shi W. 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**. (doi:10.1093/bioinformatics/btt656)

43. Robinson MD, McCarthy DJ, Smyth GK. 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**. (doi:10.1093/bioinformatics/btp616)

44. Harpur BA, Kent CF, Molodtsova D, Lebon JM, Alqarni AS, Owayss AA, Zayed A. 2014 Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci U S A* **111**, 2614–2619. (doi:10.1073/pnas.1315506111)

45. Harpur BA, Dey A, Albert JR, Patel S, Hines HM, Hasselmann M, Packer L, Zayed A. 2017 Queens and workers contribute differently to adaptive evolution in bumble bees and honey bees. *Genome Biol. Evol.* **9**, 2395–2402. (doi:10.1093/gbe/evx182)

46. Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE. 2011 Genes involved in convergent evolution of eusociality in bees. *Proc*

Natl Acad Sci U S A **108**, 7472–7477. (doi:10.1073/pnas.1103457108)

47. Franchini P, Xiong P, Fruciano C, Schneider RF, Woltering JM, Hulsey CD, Meyer A. 2019 MicroRNA Gene Regulation in Extremely Young and Parallel Adaptive Radiations of Crater Lake Cichlid Fish. *Mol. Biol. Evol.* **36**, 2498–2511. (doi:10.1093/molbev/msz168)

48. Duncan EJ, Hyink O, Dearden PK. 2016 Notch signalling mediates reproductive constraint in the adult worker honeybee. *Nat. Commun.* **7**, 1–10. (doi:10.1038/ncomms12427)

49. Hartfelder K, Tiberio GJ, Lago DC, Dallacqua RP, Bitondi MMG. 2018 The ovary and its genes—developmental processes underlying the establishment and function of a highly divergent reproductive system in the female castes of the honey bee, *Apis mellifera*. *Apidologie* **49**, 49–70. (doi:10.1007/s13592-017-0548-9)

50. Chandra V, Fetter-Pruneda I, Oxley PR, Ritger AL, McKenzie SK, Libbrecht R, Kronauer DJC. 2018 Social regulation of insulin signaling and the evolution of eusociality in ants. *Science* **361**, 398–402. (doi:10.1126/science.aar5723)

51. Barbash S, Shifman S, Soreq H. 2014 Global coevolution of human microRNAs and their target genes. *Mol. Biol. Evol.* **31**, 1237–1247. (doi:10.1093/molbev/msu090)

52. Jasper WC, Linksvayer TA, Atallah J, Friedman D, Chiu JC, Johnson BR. 2015 Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. *Mol. Biol. Evol.* **32**, 334–346. (doi:10.1093/molbev/msu292)

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

54. França GS, Hinske LC, Galante PAFF, Vibranovski MD. 2017 Unveiling the impact of

538 the genomic architecture on the evolution of vertebrate microRNAs. *Front. Genet.* **8**, 1–8.
539 (doi:10.3389/fgene.2017.00034)
540 55. França GS, Vibranovski MD, Galante PAF. 2016 Host gene constraints and genomic
541 context impact the expression and evolution of human microRNAs. *Nat. Commun.* **7**.
542 (doi:10.1038/ncomms11438)
543 56. Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk
544 RHA. 2006 Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.* **38**,
545 1375–1377. (doi:10.1038/ng1914)
546 57. Tang T, Kumar S, Shen Y, Lu J, Wu M-L, Shi S, Li W-H, Wu C-I. 2010 Adverse
547 interactions between micro-RNAs and target genes from different species. *Proc. Natl.*
548 *Acad. Sci. U. S. A.* **107**, 12935–40. (doi:10.1073/pnas.1007591107)
549 58. Patalano S *et al.* 2015 Molecular signatures of plastic phenotypes in two eusocial insect
550 species with simple societies. *Proc. Natl. Acad. Sci.* **112**, 13970–13975.
551 (doi:10.1073/pnas.1515937112)
552 59. Sumner S. 2014 The importance of genomic novelty in social evolution. *Mol. Ecol.* **23**,
553 26–28. (doi:10.1111/mec.12580)
554 60. Rehan SM, Toth AL. 2015 Climbing the social ladder: The molecular evolution of
555 sociality. *Trends Ecol. Evol.* **30**, 426–433. (doi:10.1016/j.tree.2015.05.004)
556 61. Chen K, Rajewsky N. 2006 Natural selection on human microRNA binding sites inferred
557 from SNP data. *Nat. Genet.* **38**, 1452–1456. (doi:10.1038/ng1910)
558 62. Saunders MA, Liang H, Li W-H. 2007 Human polymorphism at microRNAs and
559 microRNA target sites. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3300–5.
560 (doi:10.1073/pnas.0611347104)

561 63. Franchini P, Xiong P, Fruciano C, Meyer A. 2016 The role of microRNAs in the repeated
562 parallel diversification of lineages of midas cichlid fish from Nicaragua. *Genome Biol.*
563 *Evol.* **8**, 1543–1555. (doi:10.1093/gbe/evw097)
564 64. Linksvayer TA, Wade MJ. 2009 Genes with social effects are expected to harbor more
565 sequence variation within and between species. *Evolution* **63**, 1685–1696.
566 (doi:10.1111/j.1558-5646.2009.00670.x)
567 65. Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML, Gates
568 MW, Kula RR, Brady SG. 2017 Phylogenomic insights into the evolution of stinging
569 wasps and the origins of ants and bees. *Curr. Biol.* **27**, 1019–1025.
570 (doi:10.1016/j.cub.2017.03.027)
571 66. Cardinal S, Danforth BNBN. 2011 The antiquity and evolutionary history of social
572 behavior in bees. *PLoS One* **6**, e21086. (doi:10.1371/journal.pone.0021086)

573

574 **ACKNOWLEDGEMENTS**

575 This work was supported by the USDA National Institute of Food and Agriculture [2018-67014-
576 27542 to K.M.K.]; the Utah Agricultural Experiment Station, Utah State University [Project
577 1297, journal paper number 9239 to K.M.K.]; the U.S.-Israel Binational Science Foundation
578 [BSF 2012807 to G.B. and Y.B.S.]; and the Swiss National Science Foundation
579 [PP00P3_170664 to R. M. W.]. G.B. thanks the Clark Way Harrison Visiting Professor in Arts
580 and Sciences that supported his stay in Washington University in St. Louis. Sequencing was
581 performed at the University of Illinois Roy J. Carver Biotechnology Center. We thank the
582 University of Utah High Performance Computing Center for computational time and assistance.

583 Illustrations were created by J. Johnson (LifeSciences Studios). We thank G. Robinson for
584 helpful feedback on an earlier draft of this manuscript.

585

586 **FIGURE CAPTIONS**

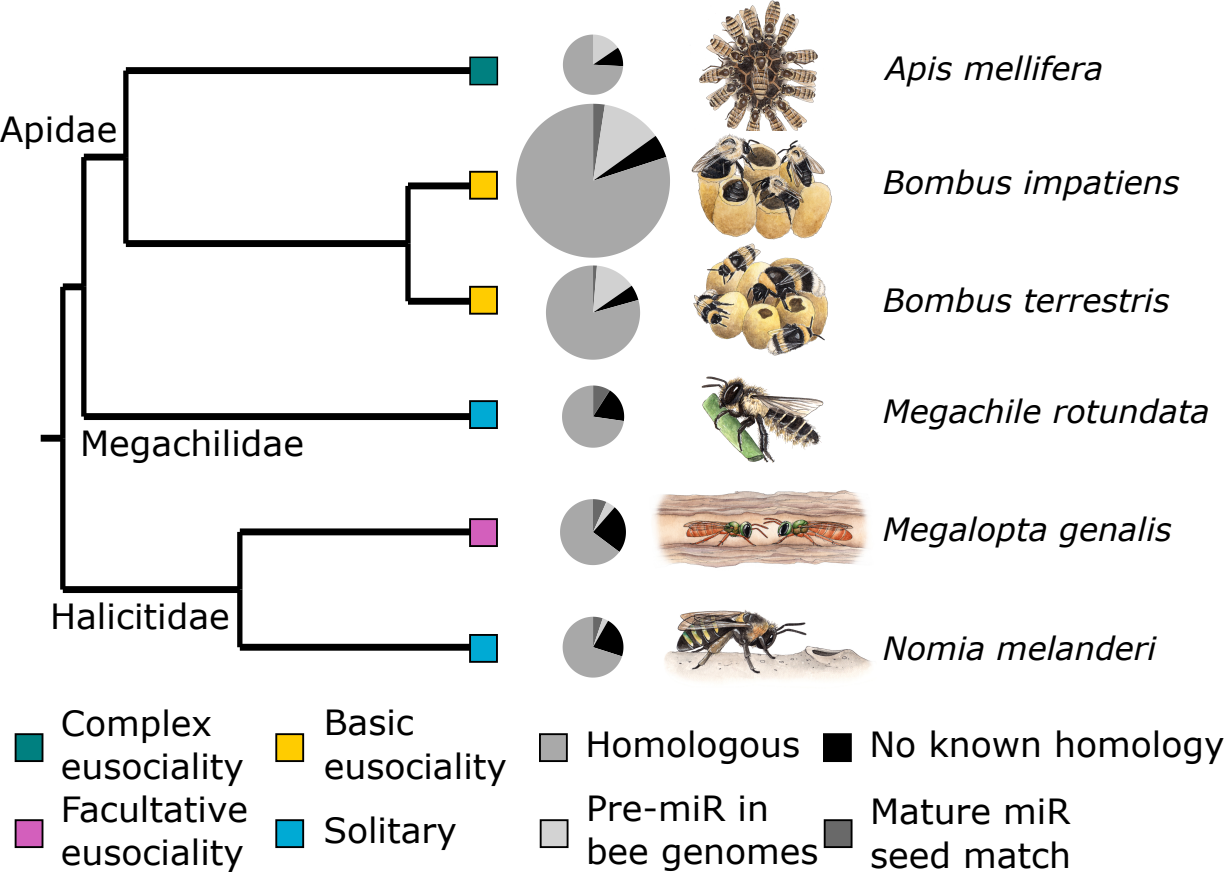
587 **Fig. 1. Diversity of miRNAs expressed in the brains of six bee species.** The three types of
588 homology (shades of grey) correspond to those in Table 1. Black – has not been previously
589 detected in other species. Pie size corresponds to number of miRNAs detected from small RNA
590 sequencing. Boxes indicate social organization (green – complex eusociality, yellow – basic
591 eusociality, pink – facultative eusociality, blue – solitary). Phylogenetic relationships are
592 following previous studies [31,65,66].

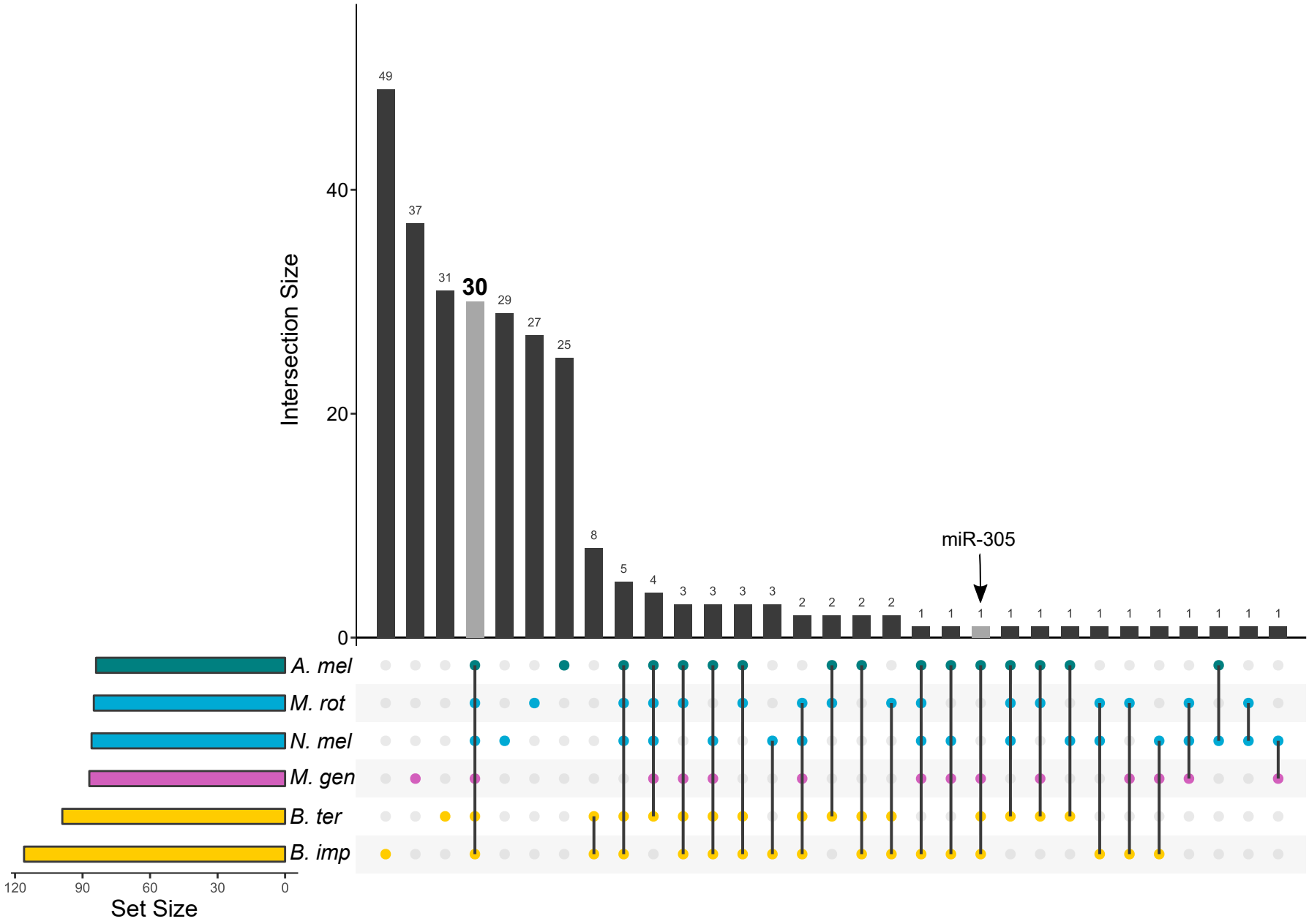
593

594 **Fig. 2.** Overview of miRNAs shared by one or more of six bee species. Dots connected by lines
595 indicate which species each set of miRNAs is shared between. The size of each miRNA set is
596 indicated above the corresponding bar. The grey bar highlights that only 30 miRNAs were
597 expressed in the brains of all six species. The arrow labels miR-305, the only miRNA expressed
598 in the brains of all social species, but none of the solitary species.

599

600 **Fig. 3.** Age of genes targeted by lineage-specific miRNAs. Genes predicted to be targeted by
601 lineage-specific miRNAs are more likely to be unique to each species than predicted by chance.
602 Pie chart size is scaled to number of predicted target genes for lineage-specific miRNAs, but not
603 for all genes. Color slices indicate orthogroup age for each predicted gene. The green slice
604 (lineage-specific genes) is larger for the set of genes predicted to be targeted by lineage-specific
605 miRNAs than for all genes.


Apidae
Megachilidae
Halicitidae
Apis mellifera
Bombus impatiens
Bombus terrestris
Megachile rotundata
Megalopta genalis
Nomia melanderi
Complex eusociality
Basic eusociality
Facultative eusociality
Solitary
Homologous
Pre-miR in bee genomes
No known homology
Mature miR seed match


Intersection Size
40
20
0
49
37
31
30
29
27
25
8
5
4
3
3
3
3
2
2
2
2
1
1
1
1
1
1
1
1
1
1
1
1
1
miR-305
A. mel
M. rot
N. mel
M. gen
B. ter
B. imp
120
90
60
30
0
Set Size

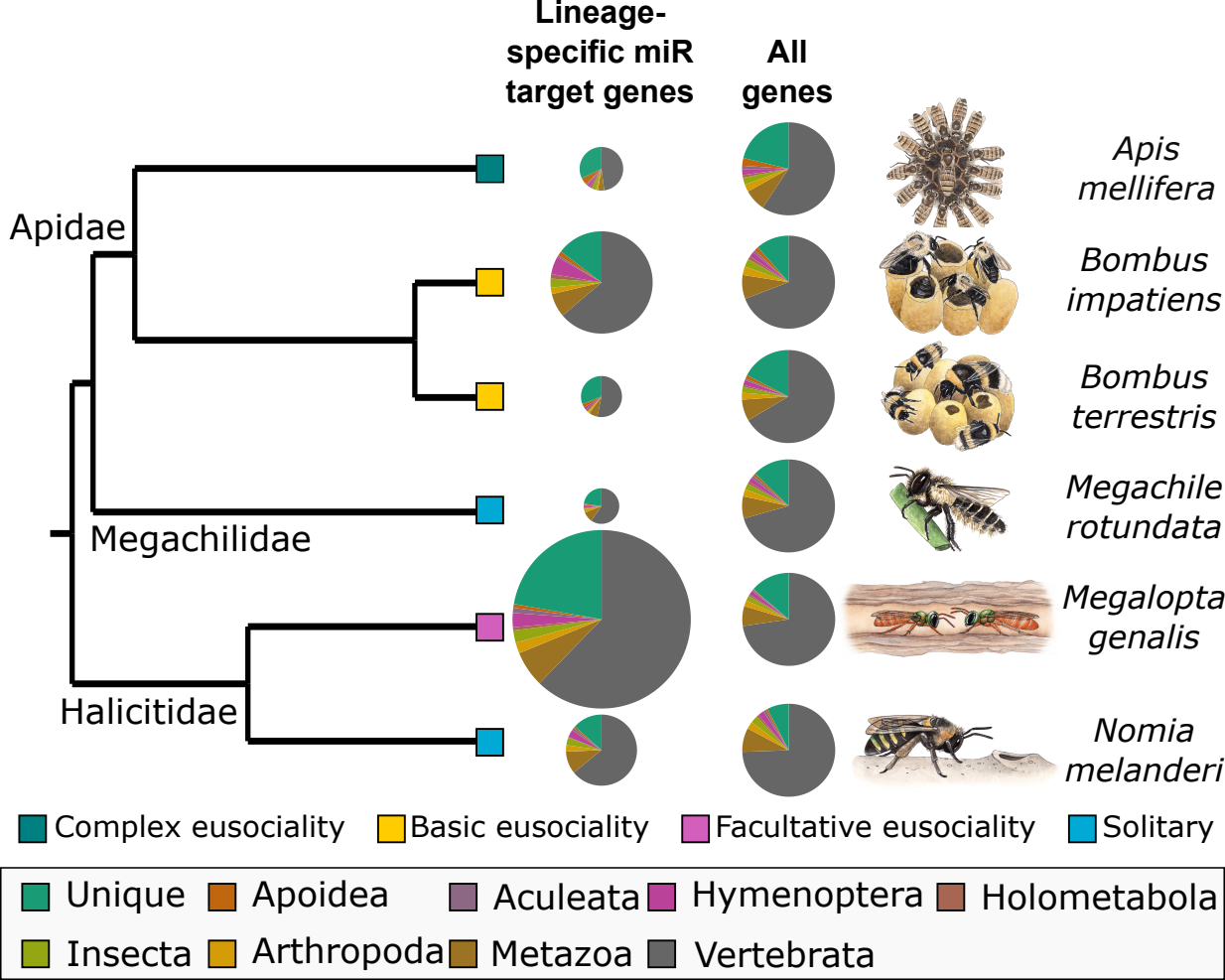

Lineage-specific miR target genes
All genes
Apidae
Megachilidae
Halicitidae
Apis mellifera
Bombus impatiens
Bombus terrestris
Megachile rotundata
Megalopta genalis
Nomia melanderi
Complex eusociality
Basic eusociality
Facultative eusociality
Solitary
Unique
Apoidea
Aculeata
Hymenoptera
Holometabola
Insecta
Arthropoda
Metazoa
Vertebrata

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