

Research



**Cite this article:** Kapheim KM, Jones BM, Søvik E, Stolle E, Waterhouse RM, Bloch G, Ben-Shahar Y. 2020 Brain microRNAs among social and solitary bees. *R. Soc. Open Sci.* **7**: 200517. <http://dx.doi.org/10.1098/rsos.200517>

Received: 27 March 2020

Accepted: 15 June 2020

**Subject Category:**

Genetics and genomics

**Subject Areas:**

evolution/genomics/behaviour

**Keywords:**

gene regulation, small non-coding RNA, microRNA targets, eusociality, lineage-specific

**Author for correspondence:**

Karen M. Kapheim

e-mail: [karen.kapheim@usu.edu](mailto:karen.kapheim@usu.edu)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5046642>.

# Brain microRNAs among social and solitary bees

Karen M. Kapheim<sup>1</sup>, Beryl M. Jones<sup>2</sup>, Eirik Søvik<sup>3</sup>, Eckart Stolle<sup>4</sup>, Robert M. Waterhouse<sup>5</sup>, Guy Bloch<sup>6</sup> and Yehuda Ben-Shahar<sup>7</sup>

<sup>1</sup>Department of Biology, Utah State University, 5305 Old Main Hill, Logan, UT 84322, USA

<sup>2</sup>Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

<sup>3</sup>Department of Science and Mathematics, Volda University College, 6100 Volda, Norway

<sup>4</sup>Centre of Molecular Biodiversity Research, Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

<sup>5</sup>Department of Ecology and Evolution, University of Lausanne and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

<sup>6</sup>Department of Ecology, Evolution and Behavior, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

<sup>7</sup>Department of Biology, Washington University in St Louis, St Louis, MO 63130, USA

KMK, 0000-0002-8140-7712; RMW, 0000-0003-4199-9052; GB, 0000-0003-1624-4926

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 homologues 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.

# 1. 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 forgo 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 probably requires modification of molecular pathways related to development, behaviour, 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 behaviour.

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 (approx. 21–22 nt), non-coding 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].

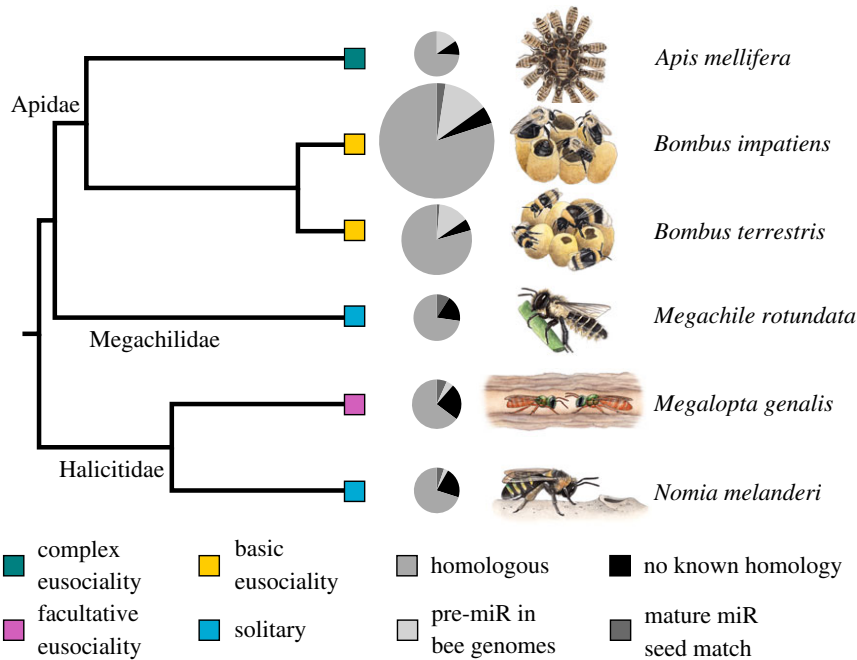
There is accumulating evidence for a role of miRNAs in regulating the social lives of insects. While most miRNAs seem to be conserved in major lineages of insects [14,15], expression levels vary across individuals performing different social functions, such as between workers performing different tasks in honeybees [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 honeybees [19–21] and bumblebees [22]. Additionally, miRNAs play a role in regulating some physiological correlates of social behaviour in honeybees, 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: (i) If lineage-specific miRNAs are assimilated into ancestral gene networks, their predicted target genes should be ancient and conserved. (ii) If lineage-specific miRNAs play a role in social evolution, their predicted targets should be enriched for genes associated with social behaviour (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.

## 2. Material and methods

### 2.1. microRNA diversification

We performed genome scans for small RNAs across 12 bee genomes (electronic supplementary material, table S1) using covariance models implemented with Infernal (v. 1.1) cmsearch using the gathering



**Figure 1.** Diversity of miRNAs expressed in the brains of six bee species. The three types of homology (shades of grey) correspond to those in table 1. Black has not been previously detected in other species. Pie size corresponds to number of miRNAs detected from small RNA sequencing. Boxes indicate social organization (green, complex eusociality; yellow, basic eusociality; pink, facultative eusociality; blue, solitary). Phylogenetic relationships are following previous studies [28–30].

threshold for inclusion ( $-\text{cut\_ga}$ ) [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 eusocial or obligate complex eusocial following Kapheim *et al.* [4]. We used the ape package (v. 3.1) [26] in R (v. 3.5.0) [27] to calculate phylogenetic independent contrasts for both social organization and miRNA copy-number, cor.test to implement Spearman's rank correlations and p.adjust with the Benjamini–Hochberg method to correct for multiple comparisons.

## 2.2. Sample acquisition, RNA isolation and sequencing

We used the brain from a single adult female from six bee species, including both eusocial and solitary species with well-studied behaviour from three bee families (figure 1). Details of sample collection, RNA isolation and sequencing are provided in electronic supplementary material, table S2.

## 2.3. microRNA discovery and quantification

We used miRDeep2 (v. 2.0.0.8) [31] to identify and quantify miRNAs expressed in the brains of each species, with a three-step process of miRNA detection to identify homologous miRNAs between species. First, we gathered mature miRNA sequences previously described in other insect species (electronic supplementary material, table S3). 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 (electronic supplementary material, table S1) with the mapper.pl script. Approximately 61–84% of reads successfully mapped (electronic supplementary material, table S2).

We then identified known and novel miRNAs in each sample with the miRDeep2.pl script, using our set of insect miRNAs (electronic supplementary material, table S3) as known mature sequences. The quantifier.pl script generated sets of known and novel miRNAs in each sample, along with quantified expression information. We filtered novel miRNAs in each species according to the following criteria: no rRNA/tRNA similarities, minimum of five reads each on mature and star strands of the hairpin sequence, and a randfold  $p$ -value  $< 0.05$ . Randfold describes the RNA secondary structure of potential precursor miR (pre-miRs) [31].

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

Some 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 pre-miR sequences in 12 bee genomes (electronic supplementary material, table S1) for each novel miRNA without a matching seed sequence.

## 2.4. microRNA localization

We used `bedtools` (v. 2.27.0) `intersect` [32] to find overlap of miRNAs with predicted gene models (electronic supplementary material, table S4), and repetitive element annotations from previously established repeat libraries that had been generated using `Repeatmasker` [4,28,33–36].

## 2.5. Target prediction

We extracted potential target sites 500 bp downstream from each gene model using `bedtools flank` and `getfasta` [32], following previous studies [19] and an average 3' UTR of 442 nt in *Drosophila melanogaster* [37]. Target prediction was run with `miRanda` (v. 3.3) [38] (minimum energy threshold  $-20$ , minimum score 140, strict alignment to seed region [-en -20 -sc 140 -strict]) and `RNAhybrid` (v. 2.12) [39] (minimum free energy threshold  $-20$ ). We kept only miRNA-target gene pairs that were predicted by both programs with  $p < 0.01$ .

## 2.6. Target age and functional enrichment

Gene ages were determined using orthogroups from OrthoDB (v. 9) [40], which includes *Apis mellifera*, *Bombus impatiens*, *Bombus terrestris*, and *Megachile rotundata*. Gene sets of *Megalopta genalis* and *Nomia 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 orthologues were labelled 'Unique'.

## 2.7. 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 (electronic supplementary material, table S5). For *M. genalis* caste data, RNAseq reads from Jones *et al.* [41] (NCBI PRJNA331103) were trimmed using `Trimmomatic` (v. 0.36) [42] and aligned to a draft genome assembly of *M. genalis* (NCBI PRJNA494872) [35] using `STAR` (v. 2.5.3) [43]. Gene counts were obtained using `featureCounts` in the `Subread` package (v. 1.5.2) [44], and differential expression analysis was conducted using `edgeR` [45] as in Jones *et al.* [41].

We also tested datasets identifying genes under selection in bee species [34,46,47] or across social lineages of bees [4,48] for enrichment of lineage-specific miRNA targets (electronic supplementary material, table S5). When necessary, we used reciprocal `blastp` (evalue  $< 10e^{-5}$ ) to identify orthologous genes, and only genes with putative orthologues 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.)

# 3. Results

## 3.1. Low levels of microRNA 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 (electronic supplementary

material, table S6). The mean copy-number across all miRNAs in all bee genomes was  $1.19 \pm 0.74$ . 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 (electronic supplementary material, table S6).

### 3.2. Expressed microRNA diversity in bee brains

We identified 97–245 known and novel miRNAs expressed in the brains of each of our six species (electronic supplementary material, 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 centre, number of reads, number or proportion of reads mapped to the genome or type of sample from which they were obtained (electronic supplementary material, table S2). This variation in number of expressed miRNAs is similar to that found in other groups of species with shorter divergence times [49].

Most detected miRNAs in each species had known homologues 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 and figure 1), 24–72% of which did not have any known homologues 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. mellifera*, *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 homologues 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 homologues ( $\chi^2$  tests:  $p = 0.05$ – $0.89$ ).

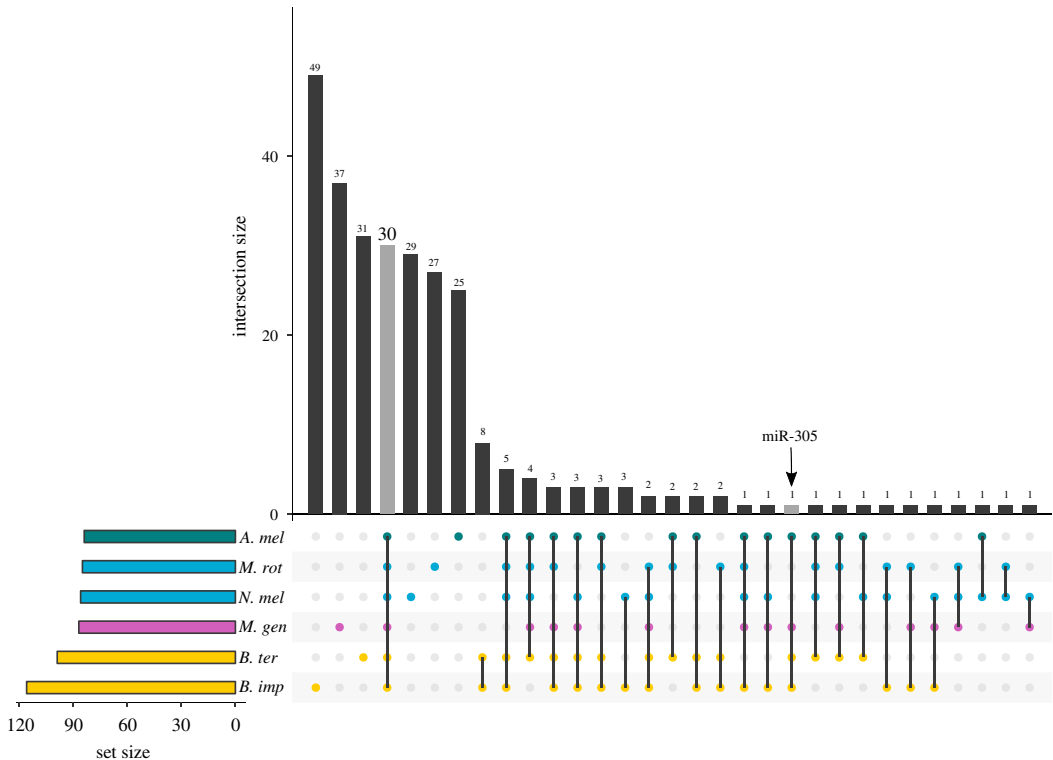
Of miRNAs with homologues, most were expressed in all six species, but one miRNA (miR-305) was expressed in the brains of each of the social, but not the solitary, species (figure 2). Although we did not detect expression of miR-305 in the two solitary species, *M. rotundata* and *N. melanderi*, genome scans of each species against the Rfam database suggested all bee species have one copy of miR-305 (electronic supplementary material, 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*.

### 3.3. Lineage-specific microRNAs preferentially target lineage-specific genes, but not genes with caste-biased expression or genes under positive selection

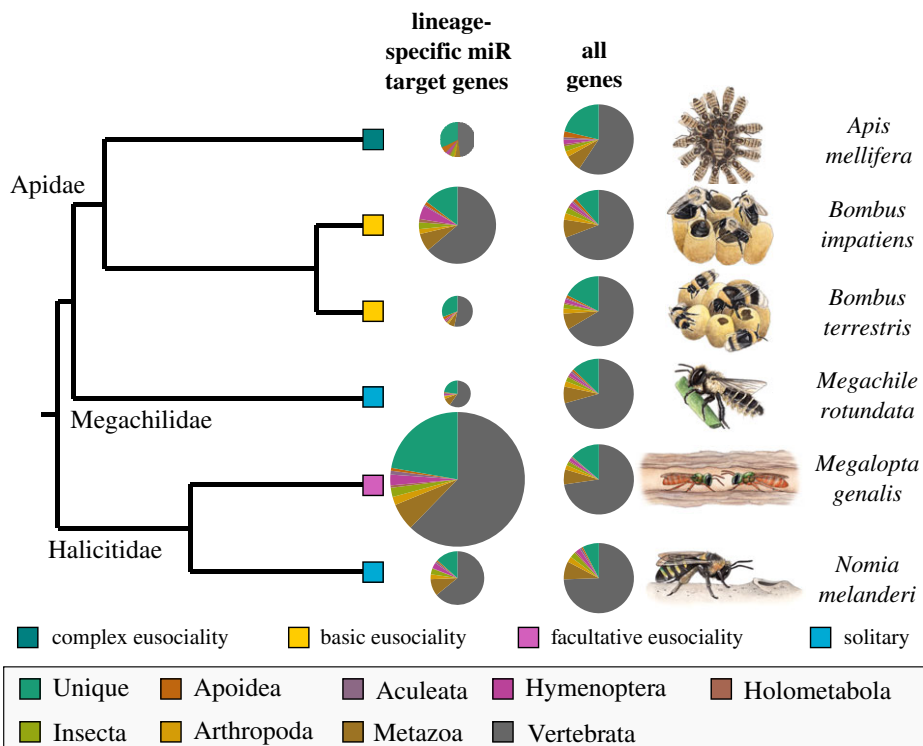
If lineage-specific changes in gene regulatory function associated with social evolution are facilitated by novel miRNAs inserted into existing gene networks, then predicted targets of lineage-specific miRNAs should be highly conserved and enriched for genes with known functions in social evolution. Most predicted mRNA targets of lineage-specific miRNAs were highly conserved and belonged to orthogroups shared by vertebrates (figure 3; electronic supplementary material, 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

**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 homologue was found in other species (seed match to mature or blastn hit to pre-miR).

species	sociality	miRNAs with lineage-specific expression in the brain				location in the genome					uncharacterized repetitive DNA
		expressed miRNAs	total	seed match	pre-miR	unique	intergenic	exon	intron	transposable element	
<i>Apis mellifera</i>	complex eusocial	97	25	0	15	10	45	5	38 (10)	0	0
<i>Bombus impatiens</i>	basic eusocial	245	49	6	31	12	129	4 (1)	89 (27)	7	32
<i>Bombus terrestris</i>		150	31	2	21	8	76	1 (1)	56 (20)	13	36
<i>Megalopta genalis</i>	facultative eusocial	105	37	7	5	25	63	3	30 (10)	2	28
<i>Megachile rotundata</i>	solitary	99	27	9	0	18	48	8 (1)	37 (5)	2	15
<i>Nomia melanderi</i>	solitary	97	29	5	3	21	50	8	34 (8)	2	27



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



**Figure 3.** Age of genes targeted by lineage-specific miRNAs. 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 for lineage-specific miRNAs, but not for all genes. Colour slices indicate orthogroup age for each predicted gene. The green slice (lineage-specific genes) is larger for the set of genes predicted to be targeted by lineage-specific miRNAs than for all genes.

that novel miRNAs are more likely to target novel genes than would be expected by chance (figure 3; electronic supplementary material, table S8).

We did not find support for the prediction that novel miRNAs should target genes that function in social behaviour 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 behavioural 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 (electronic supplementary material, figure S1 and table S5). Also contrary to our prediction, targets of lineage-specific miRNAs were not enriched for genes under positive selection in any species (electronic supplementary material, figure S1 and 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 with solitary species [4,48], but found no significant overlap or depletion with predicted targets of lineage-specific genes (hypergeometric tests:  $p > 0.05$ ; electronic supplementary material, figure S1 and table S5).

## 4. 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 a single bee from three eusocial and three solitary species that span 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 signalling in *D. melanogaster*, both of which are important regulators of social dynamics in insects [50–52]. Interestingly, miR-305 is also upregulated in worker-destined compared with queen-destined honeybee larvae, and may thus play a role in caste differentiation [20]. Further investigation with deeper sampling and additional social and solitary species is necessary to determine whether this miRNA is expressed exclusively in the brains of social species and how it may influence social behaviour.

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 coevolve with novel genes, as has been shown for the evolution of cognitive function in humans [53]. Previous work in honeybees 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) [54], and upregulated in workers [55]. 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 [10,56,57]. 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 [58] and flies [11,59]. Likewise, we find that lineage-specific miRNAs have reduced expression compared with those with homologues. 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), though this could also be an effect of limited sampling. 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 [60]. Previous studies have identified miRNAs that are differentially expressed between queens



and workers in honeybees [19–21] and bumblebees [22]. However, without comparison with 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 [61], most caste-biased genes belong to highly conserved molecular pathways [62].

Lineage-specific miRNAs also showed no evidence for preferential targeting of genes under positive selection—either within or across species. By 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 [63–65]. 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 the functional regulation of these genes are likely to have significant fitness consequences. Also, genes with caste-biased expression and those under positive selection are undergoing rapid evolution [66], and thus may be more likely to 'escape' control by errant miRNAs. Indeed, it is easier for mRNAs to lose miRNA target binding sites, which typically require exact sequence matches, than to gain them [10]. Thus, emergent miRNAs may not be expected to target adaptively or fast evolving genes, regardless of their role in social evolution.

Our analyses reveal important differences in patterns of miRNA evolution between bees and other species. For example, expansion in miRNA repertoire is associated with the evolution of animal complexity in a wide range of species [9,12,13]. The evolution of eusociality from a solitary ancestor is associated with increases in phenotypic complexity, and considered to be one of the major transitions in evolution. We therefore hypothesized that evolutionary increases in social complexity would be associated with expansions in the number of miRNAs found within bee genomes. On the contrary, we find that most bees have a single copy of previously identified miRNAs in their genomes, consistent with results of comparative genome scans across ants [3]. A recent study of miRNA diversity in insects found that morphological innovations such as holometabolous development was accompanied by the acquisition of only three miRNA families [15]. This suggests that insect evolution is not as reliant on major expansions of miRNA families as other taxonomic groups.

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 within ancient genes that have multiple functions and broad expression patterns, which may facilitate persistence of emergent miRNAs by increasing their expression repertoire [56,57]. 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 with 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 coevolution 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.

**Ethics.** Permission to collect samples of *Megalopta genalis* in Panama was granted by the Smithsonian Tropical Research Institute, and samples were exported under permit SEX/A-37-15. Permission to collect samples of *Nomia melanderi* was granted by private landowners.

Data accessibility. Sequences are deposited at NCBI SRA as BioProject PRJNA559906. Other data supporting this article are available in electronic supplementary material, tables S1–S8. Relevant code for this research is stored in GitHub: [www.github.com/kapheimlab/bee\\_microRNA](http://www.github.com/kapheimlab/bee_microRNA) and have been archived within the Zenodo repository: <https://zenodo.org/badge/latestdoi/250347886>.

Authors' contribution. K.M.K. conceived of the study and designed the experiments. K.M.K., E.S., G.B. and Y.B.-S. collected the data. K.M.K., B.M.J., E.S. and R.M.W. analysed the data. K.M.K. wrote the initial draft of the manuscript. All authors edited and approved the final article for publication. All authors agree to be held accountable for the work performed therein.

Competing interests. The authors declare no competing interests.

Funding. Financial support came from the USDA National Institute of Food and Agriculture (2018-67014-27542 to K.M.K.); the Utah Agricultural Experiment Station, Utah State University (Project 1297, journal paper number 9239 to K.M.K.); the U.S.-Israel Binational Science Foundation (BSF 2012807 to G.B. and Y.B.-S.) and the Swiss National Science Foundation (PP00P3\_170664 to R.M.W.).

Acknowledgements. Sequencing was performed at the University of Illinois Roy J. Carver Biotechnology Center. We thank the University of Utah High Performance Computing Center for computational time and assistance. Illustrations were created by J. Johnson (LifeSciences Studios). G.B. thanks the Clark Way Harrison Visiting Professor in Arts and Sciences that supported his stay in Washington University in St Louis. We thank G. Robinson for helpful feedback on an earlier draft of this manuscript.

## References

- Sumner S, Bell E, Taylor D. 2018 A molecular concept of caste in insect societies. *Curr. Opin. Insect Sci.* **25**, 42–50. (doi:10.1016/j.cois.2017.11.010)
- Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE. 2002 Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res.* **12**, 555–566. (doi:10.1101/gr.5302)
- Simola DF *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. (doi:10.1101/gr.155408.113)
- Kapheim KM *et al.* 2015 Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 24–32. (doi:10.1126/science.aaa4788)
- Warner MR, Qiu L, Holmes MJ, Mikheyev AS, Linksvayer TA. 2019 Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. *Nat. Commun.* **10**, 2651. (doi:10.1038/s41467-019-10546-w)
- Bartel DP. 2018 Metazoan microRNAs. *Cell* **173**, 20–51. (doi:10.1016/j.cell.2018.03.006)
- Friedman RC, Farh KK-H, Burge CB, Bartel DP. 2009 Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105. (doi:10.1101/gr.082701.108)
- Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H. 2013 Birth and expression evolution of mammalian microRNA genes. *Genome Res.* **23**, 34–45. (doi:10.1101/gr.140269.112)
- Berezikov E. 2011 Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.* **12**, 846–860. (doi:10.1038/nrg3079)
- Chen K, Rajewsky N. 2007 The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* **8**, 93–103. (doi:10.1038/nrg1990)
- 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. (doi:10.1038/ng.73)
- Christodoulou F *et al.* 2010 Ancient animal microRNAs and the evolution of tissue identity. *Nature* **463**, 1084–1088. (doi:10.1038/nature08744)
- Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ. 2008 MicroRNAs and the advent of vertebrate morphological complexity. *Proc. Natl Acad. Sci. USA* **105**, 2946–2950. (doi:10.1073/pnas.0712259105)
- Søvik E, Bloch G, Ben-Shahar Y. 2015 Function and evolution of microRNAs in eusocial Hymenoptera. *Front. Genet.* **6**, 1–11. (doi:10.3389/fgene.2015.00193)
- Ylla G, Fromm B, Piulachs MD, Belles X. 2016 The microRNA toolkit of insects. *Sci. Rep.* **6**, 1–13. (doi:10.1038/srep37736)
- Greenberg JK *et al.* 2012 Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes, Brain Behav.* **11**, 660–670. (doi:10.1111/j.1601-183X.2012.00782.x)
- Liu F *et al.* 2012 Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol. Biol.* **21**, 297–303. (doi:10.1111/j.1365-2583.2012.01135.x)
- Behura SK, Whitfield CW. 2010 Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol. Biol.* **19**, 431–439. (doi:10.1111/j.1365-2583.2010.01010.x)
- Ashey R, Forêt S, Searle I, Maleszka R. 2016 MicroRNAs in honey bee caste determination. *Sci. Rep.* **6**, 1–15. (doi:10.1038/srep18794)
- 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. (doi:10.1007/s13592-014-0299-9)
- Weaver D *et al.* 2007 Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome Biol.* **8**, R97. (doi:10.1186/gb-2007-8-6-r97)
- Collins DH, Mohorianu I, Beckers M, Moulton V, Dalmay T, Bourke AFG. 2017 MicroRNAs associated with caste determination and differentiation in a primitively eusocial insect. *Sci. Rep.* **7**, 1–9. (doi:10.1038/srep45674)
- MacEdo LMF *et al.* 2016 MicroRNA signatures characterizing caste-independent ovarian activity in queen and worker honeybees (*Apis mellifera* L. *Insect Mol. Biol.* **25**, 216–226. (doi:10.1111/imb.12214)
- Nunes FME, Ihle KE, Mutti NS, Simoes ZLP, Amdam G V. 2013 The gene *vitellogenin* affects microRNA regulation in honey bee (*Apis mellifera*) fat body and brain. *J. Exp. Biol.* **216**, 3724–3732. (doi:10.1242/jeb.089243)
- Cui X, Lu Z, Wang S, Jing-Yan Wang J, Gao X. 2016 CMsearch: simultaneous exploration of protein sequence space and structure space improves not only protein homology detection but also protein structure prediction. *Bioinformatics* **32**, i332–i340. (doi:10.1093/bioinformatics/btw271)
- Paradis E, Claude J, Strimmer K. 2004 APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290. (doi:10.1093/bioinformatics/btg412)
- R Core Team. 2016 R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. See <https://www.R-project.org/>.
- 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)
- Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML, Gates MW, Kula RR, Brady SG. 2017 Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Curr. Biol.* **27**, 1019–1025. (doi:10.1016/j.cub.2017.03.027)

30. Cardinal S, Danforth BBNB. 2011 The antiquity and evolutionary history of social behavior in bees. *PLoS ONE* **6**, e21086. (doi:10.1371/journal.pone.0021086)
31. 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)
32. Quinlan AR, Hall IM. 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842. (doi:10.1093/bioinformatics/btq033)
33. 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)
34. 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)
35. Kapheim KM *et al.* 2020 Developmental plasticity shapes social traits and selection in a facultatively eusocial bee. *Proc. Natl Acad. Sci. USA* **117**, 13 615–13 625. (doi:10.1073/pnas.2000344117)
36. Smith AF, Hubley R, Green P. 2013 RepeatMasker. See <http://www.repeatmasker.org/>.
37. 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)
38. 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)
39. Krüger J, Rehmsmeier M. 2006 RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* **34**, W451–W454. (doi:10.1093/nar/gkl243)
40. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva E V. 2017 OrthoDB v. 9.1: 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)
41. 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* **284**, 20162228. (doi:10.1098/rspb.2016.2228)
42. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
43. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S. 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. (doi:10.1093/bioinformatics/bts635)
44. Liao Y, Smyth GK, Shi W. 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. (doi:10.1093/bioinformatics/btt656)
45. Robinson MD, McCarthy DJ, Smyth GK. 2010 edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. (doi:10.1093/bioinformatics/btp616)
46. 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. USA* **111**, 2614–2619. (doi:10.1073/pnas.1315506111)
47. 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)
48. 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. USA* **108**, 7472–7477. (doi:10.1073/pnas.1103457108)
49. Franchini P, Xiong P, Fruciano C, Schneider RF, Woltering JM, Hulsev 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)
50. 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)
51. 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)
52. 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)
53. 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)
54. 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)
55. Johnson B, Tsutsui N. 2011 Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* **12**, 164. (doi:10.1186/1471-2164-12-164)
56. França GS, Hinske LC, Galante PAFF, Vibration MD. 2017 Unveiling the impact of the genomic architecture on the evolution of vertebrate microRNAs. *Front. Genet.* **8**, 1–8. (doi:10.3389/fgene.2017.00034)
57. França GS, Vibration MD, Galante PAF. 2016 Host gene constraints and genomic context impact the expression and evolution of human microRNAs. *Nat. Commun.* **7**, 11438. (doi:10.1038/ncomms11438)
58. Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RHA. 2006 Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.* **38**, 1375–1377. (doi:10.1038/ng1914)
59. 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. USA* **107**, 12 935–12 940. (doi:10.1073/pnas.1007591107)
60. Patalano S *et al.* 2015 Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc. Natl Acad. Sci. USA* **112**, 13 970–13 975. (doi:10.1073/pnas.1515937112)
61. Sumner S. 2014 The importance of genomic novelty in social evolution. *Mol. Ecol.* **23**, 26–28. (doi:10.1111/mec.12580)
62. Rehan SM, Toth AL. 2015 Climbing the social ladder: the molecular evolution of sociality. *Trends Ecol. Evol.* **30**, 426–433. (doi:10.1016/j.tree.2015.05.004)
63. Chen K, Rajewsky N. 2006 Natural selection on human microRNA binding sites inferred from SNP data. *Nat. Genet.* **38**, 1452–1456. (doi:10.1038/ng1910)
64. Saunders MA, Liang H, Li W-H. 2007 Human polymorphism at microRNAs and microRNA target sites. *Proc. Natl Acad. Sci. USA* **104**, 3300–3305. (doi:10.1073/pnas.0611347104)
65. Franchini P, Xiong P, Fruciano C, Meyer A. 2016 The role of microRNAs in the repeated parallel diversification of lineages of midas cichlid fish from Nicaragua. *Genome Biol. Evol.* **8**, 1543–1555. (doi:10.1093/gbe/eww097)
66. Linksvayer TA, Wade MJ. 2009 Genes with social effects are expected to harbor more sequence variation within and between species. *Evolution* **63**, 1685–1696. (doi:10.1111/j.1558-5646.2009.00670.x)