Rate variation in conserved noncoding DNA reveals regulatory pathways associated with social evolution

Benjamin E.R. Rubin[1*], Beryl M. Jones[2], Brendan G. Hunt[3], Sarah D. Kocher[1*]

[1]Department of Ecology and Evolutionary Biology; Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA
[2]Program in Ecology, Evolution, and Conservation Biology, University of Illinois, Urbana, IL, USA
[3]Department of Entomology, University of Georgia, Griffin, GA, USA
[*]corresponding authors: skocher@princeton.edu, berubin@princeton.edu

Keywords: conserved non-coding sequence, convergent evolution, eusociality, bees, genomics

Abstract

The evolution of eusociality represents an increase in complexity from individual to caste-based, group reproduction. These behavioral transitions have been hypothesized to go hand-in-hand with an increased ability to regulate when and where genes are expressed. Bees have convergently evolved eusociality up to five times, providing a powerful framework to test this hypothesis. Here, we compare conserved, non-coding sequences in eleven bee species, encompassing three independent origins of reproductive division of labor and two elaborations of eusocial complexity to examine potential links between these putatively regulatory sequences and social evolution. We find that rates of evolution in a number of these loci are correlated with social transitions, suggesting they have played a role in the evolution of these behaviors. Interestingly, loci associated with social origins represent distinct molecular pathways to those associated with subsequent elaborations. We also find many novel non-coding regions that appear to have been recruited alongside the origin of sociality in corbiculate bees; these regions are enriched for cell development and nervous system functions. Thus, our results highlight the potential importance of non-coding change in the evolution of eusociality and are consistent with the idea that regulatory innovations play a key role in insect behavioral complexity.

Introduction

Many aspects of the gene regulatory landscape are highly conserved across animals [1], including some conserved non-coding elements, or CNEs, that exhibit near perfect conservation across deep evolutionary divergences, such as the entire mammalian clade [2]. The breadth of knowledge that has arisen from studies of CNEs has been built primarily upon the large number of vertebrate genomes that have been sequenced over the past decade. The number of sequenced insect genomes has recently expanded but, as of yet, few studies have examined the evolution of CNEs in this group.

Here, we take advantage of eleven publicly-available bee genomes [3–7] to examine how CNEs change as eusociality – an extreme form of social behavior found primarily in insects and

mammals – has convergently evolved and increased in complexity within bees. Eusociality is of particular interest to evolutionary biologists because it represents an increase in complexity from individual to group-level reproduction and includes the evolution of a non-reproductive worker caste [8]. Along with reproductively-specialized castes, many of these societies have also evolved elaborate communication systems used to identify group members and to coordinate and divide labor among individuals [9]. These evolutionary innovations have afforded the social insects major ecological success – they are estimated to make up over 50% of the insect biomass on the planet even though they only account for ~2% of the insect species worldwide [10].

A great deal of effort has been focused on understanding the mechanisms that have enabled the multiple evolutionary origins of sociality [11]. Just like multiple cell types and tissues are derived from the same individual genome, the queen and worker castes are generated from a shared genomic background. This means that just as changes in gene expression drive cell type specifications, they should also drive developmentally-determined caste differentiation in the social insects. There is growing evidence to support this assertion, including well-documented changes in gene expression in developing larvae and in adults [12–22], as well as changes in DNA methylation [23–26], post-translational histone modifications, and chromatin accessibility [27–29].

It is hypothesized that an expansion in the regulatory capacity of eusocial genomes may have been a key mechanism enabling transitions to eusociality in Hymenoptera [4], though coding sequences have also clearly contributed [30]. This hypothesis has largely been based on a comparative study of 10 bee genomes that uncovered expansions in transcription factor binding sites in lineages where social behavior has evolved [4]. Similar observations have also been made in ants, both through examinations of non-coding sequence evolution [31] and by comparisons of gene expression patterns that have begun to uncover signatures of ancestral gene regulatory networks that may underlie caste determination [32]. Suprisingly, however, direct comparisons of conserved, non-coding sequences have not yet been leveraged to assess the contributions of non-coding sequence evolution to the origins and elaborations of social behavior in bees. In fact, surprisingly few studies have taken advantage of available genomic resources to characterize the regulatory landscape in social insects and to determine if, as in the radiation of the mammals [33] and our own human origins [34], there appear to be major gene regulatory innovations associated with the evolution of social behavior.

Here, we use similar approaches to those implemented in mammals [35] and ants [31] to identify conserved non-coding regions in eleven bee species that span three independent origins [36] and two independent elaborations of sociality [4]. Unlike ants, which all share a common eusocial ancestor, eusociality has evolved up to five times within the bees. We have taken advantage of these convergent origins of sociality to identify concordant evolutionary signatures associated with transitions in social behavior in bees. In general, we find that there are many shared characteristics of the regulatory landscapes of mammals and bees. We then examine if and how this regulatory landscape has changed alongside behavioral and reproductive innovations

associated with the evolution of eusociality. We present evidence that there have indeed been convergent changes in different subsets of non-coding regions both across the independent origins of sociality and in association with the larger colonies and increased caste differentiation found within the stingless bees and honey bees. Finally, we discuss how these patterns of regulatory evolution compare to patterns of coding sequence evolution and highlight future areas of research that can help to further illuminate the role of gene regulatory change in the evolution of eusociality.

Methods

Bee genomes and alignments

We used previously published genomes for twelve bee species (Fig. 1; see Supplementary Information for detailed information on genome releases). For each species, we performed whole-genome alignments (see below) to identify conserved non-coding sequences (excluding *E. dilemma*, for which the available genome sequence is highly fragmented; see Supplementary Information). We also used this set of genomes to examine the evolution of protein coding sequence. Classifications of social behavior were drawn from previous studies [4], with respect to reproductive division of labor, as detailed in the Supplementary Information. Species were split into four different behavioral categories: (1) solitary (*Dufourea novaeangliae, Habropoda laboriosa, Megachile rotundata*), (2) facultative simple sociality (*Ceratina calcarata, Eufriesea mexicana, Euglossa dilemma,* and *Lasioglossum albipes*), (3) obligate simple eusociality (*Bombus impatiens* and *Bombus terrestris*), and (4) obligate complex eusociality (*Apis florea, Apis mellifera,* and *Melipona quadrifasciata*; Fig. 1). (Note that because of the variation in behaviors among species considered to have facultative simple social behavior, we refrain from using the more specific term “eusocial” to describe these species.) Both obligate simple and obligate complex eusociality involve the presence of a queen and non-reproductive workers. However, the transition to complex eusociality, as designated here, involves an increase in the number of workers up to at least several thousand, morphological specialization of castes, and vastly more complex systems of communication [37]. Simple sociality occurs in both obligately social and facultative forms wherein individuals vary in their expression of sociality across the species [38].

Identification of non-coding, alignable regions (NCARs)

In order to characterize the landscape of conserved, non-coding regions in our dataset, we used whole-genome alignments to identify NCARs present across at least eight bee species. These were generated using pairwise alignments to *A. mellifera* as a scaffold. These alignments were split into 500 base non-coding alignable regions (NCARs) for analyses. Branch lengths were estimated for each taxon in each NCAR with BASEML [39,40] using the REV model of nucleotide substitution (model = 7) on a previously determined topology [41]. Full details of alignment methodology are provided in the Supplementary Information.

Evolutionary patterns of NCAR evolution

To assess conservation and origins of novel conserved, non-coding elements in bees, we identified sets of NCARs that were present in all bee species, and those unique to particular clades. For an NCAR to be considered unique to a clade, it had to be present in all species within that clade and have no recovered ortholog from any taxa outside of that clade.

Previous studies in vertebrates have revealed that conserved, non-coding sequences tend to occur in genomic clusters [42]. To determine whether the same types of patterns are present in bees, we used permutation tests to identify significant clustering. These tests compared the number of 200kb windows with a minimum of 5, 6, 7, or 10 NCARs to the number of windows with a minimum of the given number of NCARs when locations of all NCARs were randomized within chromosomes (using 1,000 random permutations).

Functional classification and ortholog assignment

In order to assign putative functions to NCARs and to compare changes in the non-coding landscape to coding sequence evolution, we also identified single-copy orthologous genes in our dataset. These orthogroups were identified using ProteinOrtho v.5.15 [43] with a minimum connectivity of 0.5. Gene ontology terms were assigned to orthogroups using Trinotate [44] on *A. mellifera* representative sequences and functions determined by orthology to *Drosophila melanogaster* genes found in OrthoDB [45]. When paralogous sequences from a species were detected in an orthogroup, all sequences from that species were discarded. Genes were aligned using the coding sequence aware implementation of FSA [46]. Branch lengths for translations of all orthologous groups were then estimated with AAML [39,40] using the Empirical+F model of evolution (model=3). We estimated these branch lengths on the topology inferred previously [41]. We also used HOMER [47] to search for putative binding motifs present in NCARs (see Supplementary Information for more details).

Differences in evolutionary rates among bee species

We performed evolutionary rate tests on both NCARs and coding sequences to identify genomic regions that showed consistent changes in evolutionary rates associated with the evolution of social behavior. We implemented these tests with RERconverge [48,49] (downloaded from https://github.com/nclark-lab/RERconverge on March 12, 2018). RERconverge calculates *relative* branch lengths by normalizing branches for a focal locus to the distribution of branch lengths across all loci. This enables analyses that look for convergent changes in evolutionary rates across different taxa while accounting for differences in phylogenetic divergence and in baseline rates of evolution across taxa. RERconverge compares rates of change in focal/foreground branches and the rest of the tree, and identifies loci that have a significant correlation between relative rates and a phenotype of interest.

We made two different comparisons between social and solitary taxa. (1) We tested all taxa with any degree of reproductive division of labor against all other taxa (Fig. S1a). Note the inclusion of ancestral branches in these tests. (2) We identified NCARs and genes associated with the

complex eusociality of *Apis* and *Melipona* by designating these terminal branches and the internal branch representing the ancestral *Apis* lineage as focal branches (Fig. S1b).

Associations between NCARs and caste-biased gene expression

To investigate the relationship between conserved non-coding regions and genes with caste-biased expression in *Apis mellifera*, we examined genes that are expressed at different levels in virgin queens versus sterile workers for both adults [50] and larvae [51] as well as those differentially expressed between nurses and foragers, which represent categories of age-based worker polyethism [52]. Of the 3,610 genes compared between workers and queen adults, 587 are worker-biased and 649 are queen-biased. In the comparison of 4-day old larvae, 276 of the 15,314 genes in the *A. mellifera* OGS v3.2 were found to be worker-biased and 209 were queen-biased. For the nurse and forager study, 9,637 unique genes were compared, 434 of which are expressed at greater levels in foragers and 464 of which are expressed at greater levels in nurses. Hypergeometric tests were used to test for enrichment of particular sets of genes.

Results

The conserved regulatory landscape of bees

Based on the results of our whole-genome alignments, we obtained 3,463 non-overlapping NCARs. We used the genome of *A. mellifera* to examine the distribution of NCARs, finding that the vast majority (3,233) were present on scaffolds grouped into the 16 chromosomes of this species (Table S1) and were found in many regions associated with gene regulatory functions (Fig. 2a). In total, NCARs were within 10kb of 1,275 different genes. They were heavily enriched for proximity to coding sequence (hypergeometric test, $p < 1x10^{-20}$), falling into one of the gene-associated categories more often than expected by chance based on the proportion of the genome represented. 1,144 NCARs were in introns, 552 in downstream regions, 368 in promoters, 348 in 3'-UTRs, 249 in upstream regions, and 164 in 5'-UTRs (Fig. 2a). The remaining 638 NCARs were intergenic. There were 532 NCARs with striking levels of conservation: these were present in all 11 bee species and were associated with 295 genes. Like the ultra-conserved elements (UCEs) in mammals [42], the associated genes were enriched for developmental processes and transcription (Table S2). Similarly, NCARs showed a clustering pattern (Figs. 2b, S2, Table S3) that is also typical of conserved non-coding elements in mammals [42] and plants [53]. All NCARs contained 56.6% AT on average, compared to the mean genomic background of 61.8% AT (Table S4).

The most rapidly evolving NCARs are functionally distinct from the most conserved

To examine general patterns of regulatory evolution across all bee species without considering differences in social behavior, we calculated the total standardized divergence (Supplementary Information) for each NCAR and identified the top 100 fastest and slowest evolving regions (Fig. 2c). The 100 fastest evolving regions were associated with genes enriched for metabolic functions, while the 100 slowest evolving regions were associated with genes enriched for

regulation of gene expression (hypergeometric test, $p < 0.01$; Table S5), including several transcription factors.

There were also differences in the types of genomic features associated with faster or slower evolving NCARs. The fastest-evolving NCARs were enriched for presence in 5'UTR sequence (hypergeometric test, $p = 4.3x10^{-9}$), while the slowest-evolving NCARs were enriched in regions downstream of genes (hypergeometric test, $p = 0.024$). The fastest-evolving regions also contained 30% more binding motifs (n=3763 occurrences of motifs proximal to 80 genes) than the slowest-evolving NCARs, which encompassed 2973 motifs proximal to 69 genes. There were no major differences in which binding motifs were present in these sequences.

Novel NCARs emerge alongside eusociality

Previous work in vertebrates has suggested that the origin of novel phenotypes is correlated with the appearance of novel clusters of CNEs associated with distinct types of genes. For example, before mammals split from reptiles and birds, CNEs were recruited near transcription factors and their developmental targets, but CNEs that arose in placental mammals are enriched near genes that play roles in post-translational modification and intracellular signaling [33].

To determine if similar recruitment processes have played a role in the evolution of eusociality, we identified NCARs that are unique among the social corbiculates (*Apis, Bombus, Melipona,* and *Eufriesea*). These species all share an evolutionary origin of sociality, and the recruitment of novel NCARs in this group is likely to be associated with this major evolutionary transition. We found 1,476 NCARs associated with 605 genes that are shared among all of these species and unique to this clade (Fig. 3a). This group contained the largest number of clade-specific NCARs, suggesting that there was an expansion in regulatory regions at the origin of this clade. These regions are proximal to genes enriched in cell development and nervous system functions (hypergeometric test, $p < 0.01$; Table S6), and are located primarily in introns (35%) and intergenic regions (21%), similar to the distribution of all NCARs.

A subset of NCARs show concordant rates of change associated with sociality

Convergence across bee species that exhibit any form of reproductive division of labor. Our dataset encompasses three independent origins of reproductive division of labor (sociality) in bees (Fig. 1). Using the relative rates test, we found 90 NCARs with signatures of accelerated evolution in all social species relative to non-social bees and 81 with deceleration (relative rates test, $p < 0.05$). The distributions of mean relative rates for all social and all solitary taxa across all NCARs were similar, showing that our tests were not biased to find significance in a particular direction (Fig. 3b). The genes proximal to the NCARs evolving at different rates in social species were not enriched for particular functional gene classes, although NCARs evolving faster in social taxa were found more frequently in promoters than expected by chance (hypergeometric test, $p=4.5x10^{-5}$; Fig. 3b inset). No association with gene features was found for the NCARs evolving at a slower rate in social taxa. In general, promoters are thought to experience greater levels of evolutionary constraint relative to other regulatory features, and this higher degree of conservation

may help to explain why we can identify larger numbers of loci with concordant signatures of selection across the largest evolutionary divergences in these regions.

NCARs showing concordant rate changes across all forms of social behavior contained a similar number of known motif occurrences regardless of whether these regions were faster or slower evolving in social relative to solitary lineages (n=3235 in faster NCARs and 2842 in slower NCARs). However, there were 6 motifs that were significantly more abundant in the NCAR sequences of social bee taxa relative to other branches (Table S7). One of these was a binding motif for the Fragile X protein gene, *Fmr1*, a gene known to play a key role in brain development across a wide range of animals [54] and previously associated with social evolution in bees [4].

Bee species representing independent origins of complex eusociality. The honey bees (*Apis*) and the stingless bees (*Melipona*) share a eusocial ancestor, but most likely represent two, independent transitions from simple to complex eusociality (i.e. with morphologically specialized castes, swarm-founding, and large colony sizes) within the social corbiculates [4,55]. Again, distributions of mean relative rates were not skewed by behavioral type, so our results should not be biased to identify changes in evolutionary rates in one particular direction (Fig. 3c). We found 240 NCARs with faster rates of evolution in these complex eusocial species relative to all other bees (relative rates test, $p < 0.05$; Table S8). These were associated with genes enriched for neuron fate and differentiation (hypergeometric test, $p < 0.01$; Table S9, S10) and were found more often than expected by chance in 5' and 3' UTRs (hypergeometric test, $p < 0.01$; Fig. 3c). Similarly, there were 237 NCARs evolving at significantly slower rates in complex eusocial taxa compared to all other bees (relative rates test, $p < 0.05$; Table S8). These were proximal to genes enriched for various functions related to cellular organization (hypergeometric test, $p < 0.01$; Tables S9, S10) and were found more often than chance in intergenic and upstream regions (hypergeometric test, $p < 0.01$). Two of these NCARs were located within introns of genes known to play key roles in neuronal remodeling and development of the mushroom bodies [56,57] (Fig. 4). This is a brain region crucial for sensory integration and learning and memory in insects, and is thought to play an important role in caste differentiation in honey bees [58,59].

As we found with all social lineages, NCARs associated with the evolution of complex eusociality contained a similar number of sequence motif occurrences regardless of whether they were fast or slow evolving (n=9677 for faster regions and 9311 for slower regions). Thus, there is not likely to be a simple increase in the number of motifs present in accelerated regions relative to those that show increased constraint. However, there were 20 unique motifs present at significantly higher frequencies in regions evolving more rapidly in complex eusocial taxa, as well as 19 different motifs that occurred at higher frequencies in regions evolving more slowly in complex eusocial taxa relative to all other species (Wilcoxon test, $p < 0.05$; Table S11). Higher abundance of these unique motifs suggests that there may be some convergence in the transcription factors co-opted as social behavior has become more elaborate in each lineage.

We did not apply the relative rates test comparing obligately eusocial lineages (complex eusocial taxa + *Bombus*) to non-obligate eusocial lineages because the shared ancestry and single origin of eusociality is likely to generate a shared signal that would not be independent, reducing our confidence in the association between eusociality and the genes identified.

A subset of NCARs are associated with gene expression differences among castes

There were 15 NCARs that showed convergent acceleration associated with the elaborations of eusociality in honey bees (*Apis*) and stingless bees (*Melipona*) that are proximal to 11 genes shown to be differentially expressed between castes in honey bees (Table S12). Many of these genes are predicted to be associated with neuronal fate and brain development, including *Sema-1a*, *babo*, *ftz-f1,* and *shep*. Similarly, there were 21 NCARs with slower rates of evolution on the branches associated with the elaboration of eusociality that have also been shown to be differentially expressed in socially-relevant phenotypes in honey bees. Many of the associated genes are also predicted to be involved in axon and synapse growth (Table S12), including several heat shock proteins and *Fmr1*. Interestingly, binding motifs for *Fmr1* are overrepresented in taxa with any degree of sociality across all NCARs (Wilcoxon test, $p < 0.05$; Table S7).

It is important to note that these overlaps are not more than expected by chance (hypergeometric test, $p > 0.05$; Supplementary Information). This could be due to a few reasons. First, only a small subset of the tested differentially expressed (DE) genes in honey bees were also associated with NCARs and included in our dataset, making it difficult to generate a robust statistical inference. While this could represent a true lack of overlap, it could also be an artifact of the EST-based microarrays that several of these DE sets used, and coupled with the conservation-based approaches we implemented to identify NCARs, we may be missing substantial proportions of genes that would show these concordant signatures. Alternatively, because the gene expression datasets available are primarily limited to honey bees while the comparisons we are making are across multiple species, it could be that many of the genes we identify may not have as large-scale expression differences as those that are species-specific. Finally, within the honey bees, most large studies compare differences between adult bees [50,52], while the NCARs we have identified could affect gene expression at any point throughout development, and it is impossible to predict when, where, and in what context gene expression changes may occur. Although we did examine overlap with genes differentially expressed between worker and queen larvae (Supplementary Information), these results were based on a relatively small dataset and may have only captured those genes with the most extreme expression differences [51].

Both NCARs and coding-sequences show signatures of convergent evolution, but on different functions

The same relative rates tests used to identify changes in NCARs can also be used to identify changes in coding sequence, and we uncovered a number of genes that showed concordant rate changes on all social branches and on all complex eusocial branches (Table S13). There is a significant overlap in both genes and GO terms between our study and a previous study [4] that used different methods to identify signatures of selection across this group of bees (calculated

based on the number of overlapping genes showing concordant changes on complex eusocial branches; hypergeometric test, $p = 0.00039$, 1.98-fold enrichment).

Overall, we find that coding sequence and NCAR sequence evolution appear to be quite distinct. We find no correlation between total standardized branch lengths between NCARs and proximal genes ($\log_2$-transformed $R^2 = 0.08$, $p = 0.18$; Fig. S3). Moreover, the genes and functional terms associated with changes in NCAR rates are quite distinct from the genes and functional terms associated with evolutionary changes in coding sequence. For example, protein-coding genes evolving more rapidly in complex eusocial lineages are associated with cell projections and small molecule transport (Table S13), while NCARs evolving more rapidly in complex eusocial lineages are associated with cell fate commitment and neuron differentiation (Table S9). This suggests that the changing selective pressures that occur during the evolution of eusociality act on the regulatory elements and protein sequences of different sets of genes.

However, of the 317 genes included in both the NCAR and coding sequence relative rates tests of rate differences in complex eusocial lineages, there were 9 genes that showed consistent faster or slower rates of change in both; this is more than expected by chance (hypergeometric test, $p = 0.012$; Table S14), and the mean relative rates in CDS and NCARs in these genes were significantly correlated in both the complex eusocial taxa ($R^2 = 0.91$, $p = 6.8\times10^{-5}$) and all other taxa ($R^2 = 0.61$, $p = 0.013$).

No apparent bias of selection on regulatory versus coding sequence

It is possible that the origins of sociality are associated primarily with changes in gene regulation rather than with changes in coding sequence evolution [60]. However, we did not find any evidence that the proportion of NCARs with evolutionary rate changes associated with sociality was greater than that found in coding sequences (Table S15). As expected from relative rates inferences, we did not find any apparent differences in the distributions of evolutionary rates in the focal or background lineages of either sequence type (Figs. 3, S4). However, the total standardized divergence (total branch lengths in a locus standardized by number of loci and nucleotides) was greater in NCARs than in CDS, as expected when comparing non-coding to coding sequence evolution (Wilcoxon rank sum test, $p < 1\times10^{-10}$). That said, non-coding and coding sequences do overlap in their distributions (Fig. S5), demonstrating that in bees, as in other taxa [61], some non-coding sequences can experience the same level of constraint as protein-coding sequences.

Discussion

The conserved regulatory landscape of bees is similar to mammals and plants

We have characterized a conserved landscape of non-coding sequences in bees. Consistent with the theory that these non-coding landscapes may have ancient, metazoan origins [1], we have found that the features of this landscape are very similar to the conserved, non-coding landscapes described in vertebrates [42] and plants [53]. We find that NCARs are distributed through the

genome in clusters, and we have also identified a subset of these regions that are present in all bee species examined and enriched for developmental functions.

Regulatory innovations are associated with the evolution of eusociality

Many of the major evolutionary innovations in vertebrates have been linked to the appearance of novel clusters of conserved non-coding elements [33], and each innovation appears to be associated with different types of gene functions. We have characterized a set of NCARs that are recruited alongside the origins of eusociality in corbiculate bees. These NCARs are unique to the corbiculates (which share a simple eusocial ancestor) and they are associated with genes enriched in cell development and nervous system functions (Table S6), suggesting that the origins of eusociality may have gone hand-in-hand with changes in when and where these pathways are deployed.

There are concordant changes in conserved, non-coding sequences associated with sociality

By examining the NCARs that are commonly found across all bees, we have also identified NCARs that show concordant rate changes across the independent origins and elaborations of sociality. The regions that show concordant rate shifts on all social lineages may represent changes that are important in the establishment of sociality. NCARs evolving more quickly on all social branches were enriched for promoters, and interestingly, a number of binding motifs were more abundant in the NCARs of social species. These results suggest that certain transcription factors or binding proteins (such as *fmr1*) may have been independently co-opted by different bee lineages across these different origins of sociality to build or modify connections in gene regulatory networks that enable social living [62]. It should be noted that our analyses do not distinguish between types of selection that could be responsible for increased rates of evolution. While slower rates likely indicate increased strength of purifying selection, elevated rates could be indicative of either directional or relaxed selection. Thus, elevated rates in regulatory NCARs could represent selection pressures acting to change expression or simply that strict regulation is no longer necessary.

The NCARs associated with the elaborations of eusociality in honey bees (*Apis*) and stingless bees (*Melipona*) are different from those associated with all forms of social behavior in bees. Faster evolving sequences on these branches were enriched for 5' and 3'UTR sequences and were associated with genes that play important roles in neuron fate commitment. We were able to identify binding motifs present at significantly higher frequencies in regions evolving more rapidly in complex eusocial taxa, as well as motifs that occurred at higher frequencies in regions evolving more slowly in complex eusocial taxa relative to all other species. As with the above results examining the origins of sociality in bees, these results also provide evidence that similar transcription factors or binding proteins may have been co-opted by both honey bees and stingless bees as eusociality increased in complexity in each of these groups.

Evolutionary dynamics of conserved non-coding and protein-coding sequences

We have used the same statistical analyses to examine and compare both coding sequence and NCAR sequence evolution. In general, we find no evidence to support the idea that a greater proportion of NCARs than coding sequences have experienced novel selective pressures associated with the evolution of sociality. Remarkably, some NCARs are evolving at the same overall rate as the most conserved coding sequences, suggesting that, at least for some of the conserved non-coding regions that we can align across species, negative selection may be just as strong as it is for some proteins. These results echo the results of mammalian studies, where non-coding, ultra-conserved elements (UCEs) show similar or stronger levels of negative selection than many coding sequences [61].

Regulatory novelty may play a key role in the evolution of sociality

It should be noted that our analyses focus on concordant evolutionary signatures in conserved regions that are alignable across species. As a result, our dataset and analyses cannot examine the role that novel regulatory regions (i.e. regions that are unique to individual taxa) may play in the evolution of sociality. This kind of regulatory innovation could indeed be a key feature associated with the origins of sociality, but it is beyond the ability of our current datasets and analyses to detect. We did observe an increased number of conserved, non-coding sequences associated with the origin of eusociality in the corbiculates, providing a glimpse into the potential role that regulatory novelty may play in this process. However, future work is needed to better characterize the role of genetic novelty in this process, the nature of which is likely to be taxon-specific.

Similarly, this study has focused on a small subset of bee species for which genomic resources have already been developed. These species are heavily biased towards social lineages, and thus most of the comparative power comes from the corbiculate bees, which share a single origin of sociality. Moreover, these taxa span large periods of evolutionary divergence, and the analyses we have implemented here have been based primarily on sequence conservation among these different taxa. There are over 20,000 bee species on this planet, and there have been up to 5 independent origins of sociality within this clade [63]. Future work focused on more closely-related lineages that encompass more of these evolutionary transitions can help provide greater insight into the role of gene regulation in the origins of sociality.

Conclusions

Overall, we find that changes in non-coding sequences are likely to play an important role in the evolution of sociality. The conserved regulatory landscape in bees is similar to mammals and plants. We find that non-coding regions have been recruited alongside the origin of sociality in corbiculate bees and are associated with distinct molecular processes, similar to patterns of non-coding sequence evolution in vertebrates. Moreover, we observe concordant changes in conserved non-coding sequences associated with three origins of sociality in bees as well as with two transitions from simple to complex eusociality, and we present evidence that the molecular pathways and types of regulatory regions associated with the origins of sociality and the elaborations of eusociality appear to be distinct. Thus, the analyses of non-coding regions in this

study have helped to uncover convergent signatures of social evolution that would have otherwise been overlooked by examination of coding sequence alone. Our results highlight the utility and importance of examining both coding and non-coding change to understand the molecular mechanisms underlying phenotypic evolution.

Acknowledgements
We thank Micah Fletcher, David Galbraith, Luke Henry, Luisa Pallares, Lance Parsons, Serge Picard, and Eli Wyman for their feedback and support. Nathan Clark, Wynn Meyer, and Raghavendran Partha provided valuable guidance for using the relative rates test.

Data accessibility
The main analytical pipeline is available from GitHub: https://github.com/berrubin/BeeGenomeAligner.

Authors' contributions
BERR, BGH, and SDK conceived the project. BERR performed computational analyses. BMJ compiled gene expression datasets. BERR and SDK drafted the manuscript, and all authors revised and approved the final version.

Competing interests
We have no competing interests.

Funding
BERR was funded in part by postdoctoral fellowship grant no. 2018-67012-28085 from the USDA National Institute of Food and Agriculture, additional funds came from NSF DEB-1754476 awarded to SDK and BGH.

References
1. Polychronopoulos D, King JWD, Nash AJ, Tan G, Lenhard B. 2017 Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Res.* **45**, 12611–12624. (doi:10.1093/nar/gkx1074)

2. Bejerano G. 2004 Ultraconserved elements in the human genome. *Science* **304**, 1321–1325. (doi:10.1126/science.1098119)

3. Brand P, Saleh N, Pan H, Li C, Kapheim KM, Ramírez SR. 2017 The nuclear and mitochondrial genomes of the facultatively eusocial orchid bee *Euglossa dilemma*. *G3-Genes Genomes Genet.* **7**, 2891–2898. (doi:10.1534/g3.117.043687)

4. Kapheim KM *et al.* 2015 Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1143. (doi:10.1126/science.aaa4788)

5. Rehan SM, Glastad KM, Lawson SP, Hunt BG. 2016 The genome and methylome of a subsocial small carpenter bee, *Ceratina calcarata*. *Genome Biol. Evol.* **8**, 1401–1410. (doi:10.1093/gbe/evw079)

6. 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)

7. Weinstock GM *et al.* 2006 Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**, 931–949. (doi:10.1038/nature05260)

8. Maynard Smith J, Szathmáry E. 1995 *The Major Transitions in Evolution*. New York: W.H. Freeman.

9. Wilson EO. 1971 *The Insect Societies*. Cambridge, MA: Harvard University Press.

10. Wilson EO. 1990 *Success and Dominance in Ecosystems: The Case of the Social Insects*. Oldendorf/Luhe, Federal Republic of Germany: Ecology Institute.

11. Toth AL, Robinson GE. 2007 Evo-devo and the evolution of social behavior. *Trends Genet.* **23**, 334–341. (doi:10.1016/j.tig.2007.05.001)

12. Whitfield CW. 2003 Gene expression profiles in the brain predict behavior in individual honey bees. *Science* **302**, 296–299. (doi:10.1126/science.1086807)

13. Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE. 2010 Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc. R. Soc. B Biol. Sci.* **277**, 2139–2148. (doi:10.1098/rspb.2010.0090)

14. Patalano S *et al.* 2015 Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc. Natl. Acad. Sci.* **112**, 13970–13975. (doi:10.1073/pnas.1515937112)

15. 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**, 20162228. (doi:10.1098/rspb.2016.2228)

16. Rittschof CC, Robinson GE. 2016 Behavioral genetic toolkits. In *Current Topics in Developmental Biology*, pp. 157–204. Elsevier. (doi:10.1016/bs.ctdb.2016.04.001)

17. Gospocic J *et al.* 2017 The neuropeptide corazonin controls social behavior and caste identity in ants. *Cell* **170**, 748-759.e12. (doi:10.1016/j.cell.2017.07.014)

18. Woodard SH, Bloch GM, Band MR, Robinson GE. 2014 Molecular heterochrony and the evolution of sociality in bumblebees (*Bombus terrestris*). *Proc. R. Soc. B Biol. Sci.* **281**, 20132419. (doi:10.1098/rspb.2013.2419)

19. Schrader L, Simola DF, Heinze J, Oettler J. 2015 Sphingolipids, transcription factors, and conserved toolkit genes: developmental plasticity in the ant *Cardiocondyla obscurior*. *Mol. Biol. Evol.* **32**, 1474–1486. (doi:10.1093/molbev/msv039)

20. Mikheyev AS, Linksvayer TA. 2015 Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. *eLife* **4**, e04775. (doi:10.7554/eLife.04775)

21. Smith CR, Toth AL, Suarez AV, Robinson GE. 2008 Genetic and genomic analyses of the division of labour in insect societies. *Nat. Rev. Genet.* **9**, 735–748. (doi:10.1038/nrg2429)

22. 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)

23. Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010 The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* **8**, e1000506. (doi:10.1371/journal.pbio.1000506)

24. Glastad KM, Gokhale K, Liebig J, Goodisman MAD. 2016 The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. *Sci. Rep.* **6**, 37110. (doi:10.1038/srep37110)

25. Bonasio R *et al.* 2012 Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* **22**, 1755–1764. (doi:10.1016/j.cub.2012.07.042)

26. Foret S, Kucharski R, Pellegrini M, Feng S, Jacobsen SE, Robinson GE, Maleszka R. 2012 DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc. Natl. Acad. Sci.* **109**, 4968–4973. (doi:10.1073/pnas.1202392109)

27. Simola DF *et al.* 2016 Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science* **351**, aac6633. (doi:10.1126/science.aac6633)

28. Simola DF, Ye C, Mutti NS, Dolezal K, Bonasio R, Liebig J, Reinberg D, Berger SL. 2012 A chromatin link to caste identity in the carpenter ant *Camponotus floridanus*. *Genome Res.* **23**, 486–496. (doi:10.1101/gr.148361.112)

29. Wojciechowski M, Lowe R, Maleszka J, Conn D, Maleszka R, Hurd PJ. 2018 Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development. *Genome Res.* **28**, 1532–1542. (doi:10.1101/gr.236497.118)

30. 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.* **108**, 7472–7477. (doi:10.1073/pnas.1103457108)

31. 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)

32. Qiu B, Larsen RS, Chang N-C, Wang J, Boomsma JJ, Zhang G. 2018 Towards reconstructing the ancestral brain gene-network regulating caste differentiation in ants. *Nat. Ecol. Evol.* **2**, 1782. (doi:10.1038/s41559-018-0689-x)

33. Lowe CB, Kellis M, Siepel A, Raney BJ, Clamp M, Salama SR, Kingsley DM, Lindblad-Toh K, Haussler D. 2011 Three periods of regulatory innovation during vertebrate evolution. *Science* **333**, 1019–1024. (doi:10.1126/science.1202702)

34. Doan RN *et al.* 2016 Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341-354.e12. (doi:10.1016/j.cell.2016.08.071)

35. Villar D *et al.* 2015 Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566. (doi:10.1016/j.cell.2015.01.006)

36. Rehan SM, Leys R, Schwarz MP. 2012 A mid-cretaceous origin of sociality in Xylocopine bees with only two origins of true worker castes indicates severe barriers to eusociality. *PLOS ONE* **7**, e34690. (doi:10.1371/journal.pone.0034690)

37. Hölldobler H, Wilson EO. 2009 *The superorganism: the beauty, elegance and strangeness of insect societies.* New York, NY: Norton Press.

38. Michener C. 1974 *The Social Behavior of the Bees: A Comparative Study*. Cambridge: Harvard University Press.

39. Yang Z. 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci. CABIOS* **13**, 555–556.

40. Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)

41. 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)

42. Woolfe A *et al.* 2004 Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7. (doi:10.1371/journal.pbio.0030007)

43. Lechner M, Findeis S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011 Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124. (doi:10.1186/1471-2105-12-124)

44. Haas BJ. 2003 Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666. (doi:10.1093/nar/gkg770)

45. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV. 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)

46. Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, Holmes I, Pachter L. 2009 Fast statistical alignment. *PLoS Comput. Biol.* **5**, e1000392. (doi:10.1371/journal.pcbi.1000392)

47. Heinz S *et al.* 2010 Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589. (doi:10.1016/j.molcel.2010.05.004)

48. Chikina M, Robinson JD, Clark NL. 2016 Hundreds of genes experienced convergent shifts in selective pressure in marine mammals. *Mol. Biol. Evol.* **33**, 2182–2192. (doi:10.1093/molbev/msw112)

49. Partha R, Chauhan BK, Ferreira Z, Robinson JD, Lathrop K, Nischal KK, Chikina M, Clark NL. 2017 Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *eLife* **6**, e25884. (doi:10.7554/eLife.25884)

50. Grozinger CM, Fan Y, Hoover SER, Winston ML. 2007 Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol. Ecol.* **16**, 4837–4848. (doi:10.1111/j.1365-294X.2007.03545.x)

51. He X-J, Jiang W-J, Zhou M, Barron AB, Zeng Z-J. 2017 A comparison of honeybee (*Apis mellifera*) queen, worker and drone larvae by RNA-Seq. *Insect Sci.* **0**. (doi:10.1111/1744-7917.12557)

52. Alaux C, Le Conte Y, Adams HA, Rodriguez-Zas S, Grozinger CM, Sinha S, Robinson GE. 2009 Regulation of brain gene expression in honey bees by brood pheromone. *Genes Brain Behav.* **8**, 309–319. (doi:10.1111/j.1601-183X.2009.00480.x)

53. Burgess D, Freeling M. 2014 The most deeply conserved noncoding sequences in plants serve similar functions to those in vertebrates despite large differences in evolutionary rates. *Plant Cell* **26**, 946–961. (doi:10.1105/tpc.113.121905)

54. Lessing D, Bonini NM. 2009 Maintaining the brain: insight into human neurodegeneration from *Drosophila melanogaster* mutants. *Nat. Rev. Genet.* **10**, 359–370. (doi:10.1038/nrg2563)

55. Bossert S, Murray EA, Almeida EAB, Brady SG, Blaimer BB, Danforth BN. 2019 Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. *Mol. Phylogenet. Evol.* **130**, 121–131. (doi:10.1016/j.ympev.2018.10.012)

56. Boulanger A, Clouet-Redt C, Farge M, Flandre A, Guignard T, Fernando C, Juge F, Dura J-M. 2011 *ftz-f1* and *Hr39* opposing roles on *EcR* expression during *Drosophila* mushroom body neuron remodeling. *Nat. Neurosci.* **14**, 37–44. (doi:10.1038/nn.2700)

57. Michel CI, Kraft R, Restifo LL. 2004 Defective neuronal development in the mushroom bodies of *Drosophila Fragile X mental retardation 1* mutants. *J. Neurosci.* **24**, 5798–5809. (doi:10.1523/JNEUROSCI.1102-04.2004)

58. O'Donnell S, Bulova S, Barrett M, von Beeren C. 2018 Brain investment under colony-level selection: soldier specialization in *Eciton* army ants (Formicidae: Dorylinae). *BMC Zool.* **3**, 3. (doi:10.1186/s40850-018-0028-3)

59. O'Donnell S, Bulova SJ, DeLeon S, Barrett M, Fiocca K. 2017 Caste differences in the mushroom bodies of swarm-founding paper wasps: implications for brain plasticity and brain evolution (Vespidae, Epiponini). *Behav. Ecol. Sociobiol.* **71**, 116. (doi:10.1007/s00265-017-2344-y)

60. 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)

61. Margulies EH *et al.* 2007 Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Res.* **17**, 760–774. (doi:10.1101/gr.6034307)

62. Carroll SB. 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36. (doi:10.1016/j.cell.2008.06.030)

63. Kocher SD, Paxton RJ. 2014 Comparative methods offer powerful insights into social evolution in bees. *Apidologie* **45**, 289–305. (doi:10.1007/s13592-014-0268-3)

64. Ramírez SR, Roubik DW, Skov C, Pierce NE. 2010 Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). *Biol. J. Linn. Soc.* **100**, 552–572. (doi:10.1111/j.1095-8312.2010.01440.x)

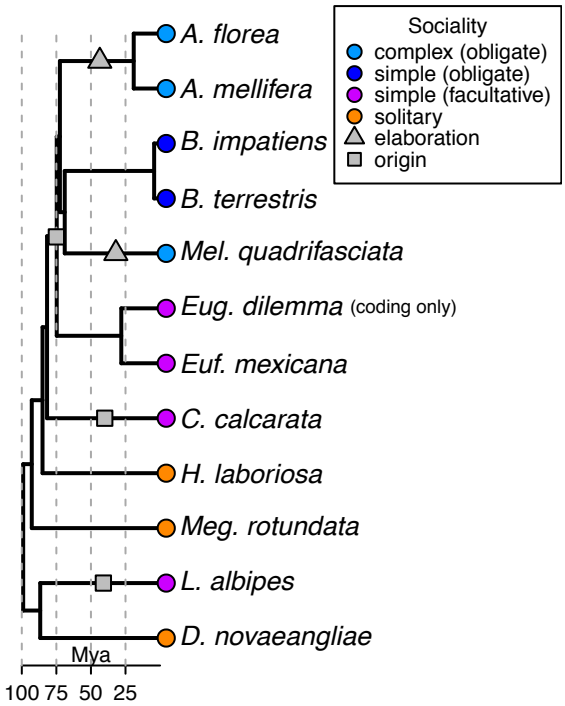

Figure 1. Phylogeny of bee species targeted in this study. These taxa span a range of behavioral forms, from solitary species that live and reproduce independently (orange), to eusocial species with a reproductive division of labor characterized by a queen and worker caste. Simple eusocial societies can either be facultative (purple) or obligate (dark blue). Complex eusocial (light blue) species contain nests made up of hundreds to thousands of individuals with morphological specializations between queen and worker castes. The species examined here encompass two independent origins of simple facultative eusociality and two independent origins of complex eusociality. Topology and dates are drawn from previous studies [41,64].

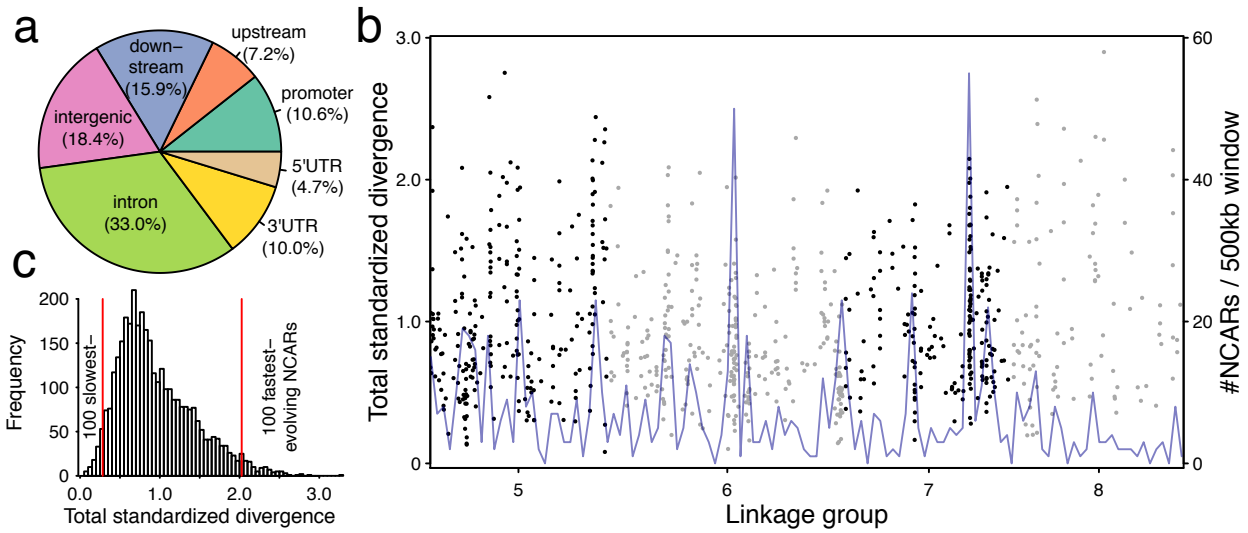

Figure 2. The landscape of non-coding alignable regions (NCARs) in bees. 3,463 NCARs were identified across bee genomes. Locations were mapped using the *A. mellifera* genome as a reference. (a) They were located in several genomic regions associated with gene regulation (see methods for classification scheme). (b) NCARs were distributed across all chromosomes, but are present in clusters on each chromosome. *A. mellifera* linkage groups 5-8 are represented on the x-axis; black and gray dots are used to denote each of these groups; each dot represents a single NCAR, and the y-axis signifies a standardized measure of divergence for each region (detailed in methods). The blue line denotes the # NCARs present in each 500kb window. NCARs occur in clusters across each chromosome, consistent with patterns observed in vertebrates and in plants. (c) NCARs exhibit substantial variation in evolutionary rates of change. The 100 slowest-evolving regions are associated primarily with regulation of gene expression, and the 100 fastest-evolving regions are associated with metabolism.

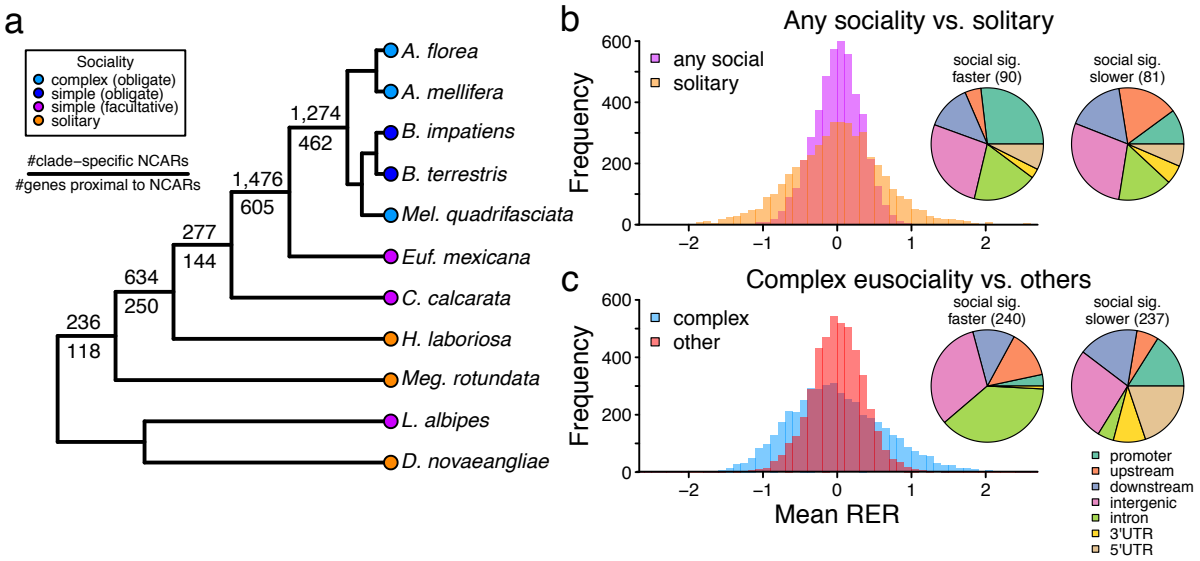

Figure 3. NCAR evolution correlates with evolution of eusociality. (a) Novel NCAR recruitment is associated with the emergence of obligate eusociality in the corbiculate bees. 1,476 NCARs are shared uniquely among these species and are enriched for gene functions associated with cell and nervous system development. (b) Distribution of mean relative rates among taxa with complex sociality vs. others in all NCARs. There are 477 NCARs that show convergent rate changes on complex eusocial branches. Rapidly evolving regions are associated with neuronal fate, and are located in 5' and 3'UTRs more often than predicted by chance (hypergeometric test, $p<1.0\times10^{-5}$). (c) Similarly, distribution of mean relative rates among taxa with any degree of sociality vs. strictly solitary taxa in all NCARs. 171 NCARs show signatures of convergent evolution across all social bee species relative to solitary taxa. 90 of these are evolving more rapidly while 81 are changing more slowly. Fast-evolving regions are enriched for promoter sequences (inset; hypergeometric test, $p=4.5\times10^{-5}$), and contain a surplus of *Fmr1* binding motifs.

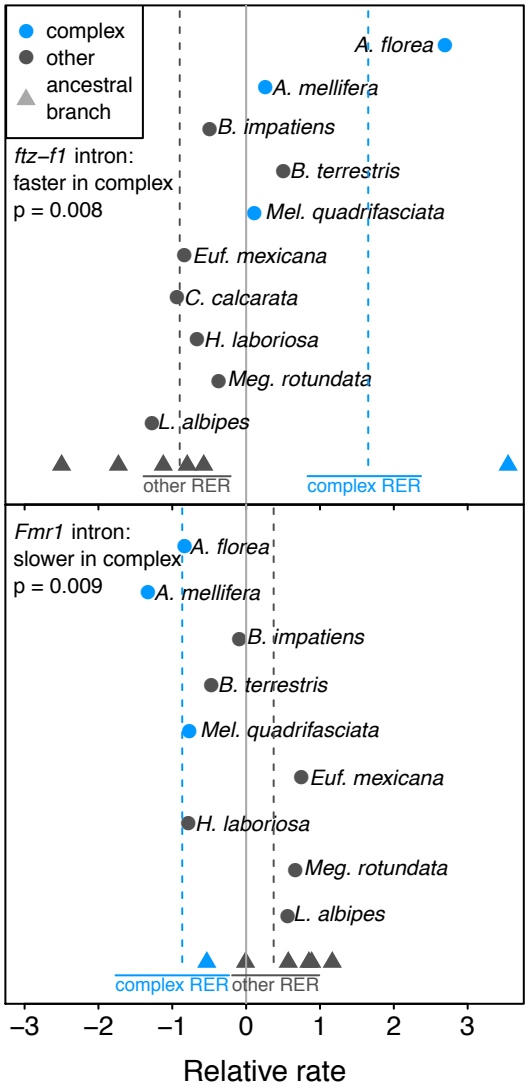

Figure 4. Two NCARs involved in brain development with convergent rate changes in complex eusocial taxa. Two intronic NCARs associated with complex social behavior are key regulators of mushroom body neuronal remodeling (*ftz-f1*; [56]) and development (*Fmr1*; [57]). *ftz-f1* shows accelerated rates of change on complex social branches relative to the remaining branches in the tree (Relative rates tests; τ=0.58, p=0.008). *Fmr1* shows significantly slower evolution on complex social branches (Relative rates tests; τ =0.59, p=0.009).

Supplementary figures

Figure S1. Phylogenies used to conduct relative rates tests with focal lineages colored in red.

Figure S2. Distribution of standardized total evolutionary change in all NCARs.

Figure S3. Log-transformed total branch length of coding sequences and proximal NCARs standardized to the branch lengths inferred from all a concatenation of all protein sequences. When multiple NCARs were associated with individual genes, mean standardized branch lengths were used.

Figure S4. (a) Distribution of mean relative rates among taxa with complex sociality vs. others in all coding sequences. (b) Distribution of mean relative rates among taxa with any degree of sociality vs. strictly solitary taxa in all coding sequences.

Figure S5. Distribution of total evolutionary change in all CDS's and NCARs analyzed. To make these measures comparable across loci and sequence classes, the standardized total evolutionary change was additionally divided by the number of bases in each locus.

Supplementary tables

Table S1. Distribution of NCARs across the 16 *Apis mellifera* chromosomes.
Table S2. GO terms enriched in genes proximal to NCARs present in all 11 bee taxa.
Table S3. Permutation tests of NCAR clustering in 200kb windows.
Table S4. Mean AT-content of NCARs.
Table S5. GO terms enriched in genes proximal to the 100 fastest- and slowest-evolving NCARs.
Table S6. GO enrichment in clade-specific NCARs.
Table S7. Sequence motifs that differ in abundance in species with any degree of sociality.
Table S8. Sequence features of NCARs identified as associated with the evolution of sociality.
Table S9. GO enrichment in genes proximal to NCARs associated with sociality using RER tests.
Table S10. Genes involved in neuron differentiation proximal NCARs evolving faster in taxa with complex sociality.
Table S11. Motifs that differ in abundance in NCARs associated with complex social taxa by RER test.
Table S12. Genes with both caste-biased expression and proximal NCARs with exceptional rates of evolution.
Table S13. GO enrichment in genes associated with sociality using RER tests.
Table S14. Genes with significantly different rates of evolution in both coding and proximal noncoding sequence.
Table S15. Numbers of coding and noncoding sequences evolving at significantly different rates.
Table S16. *De novo* sequence motifs that differ in abundance in species with complex sociality.