



# SCIENTIFIC REPORTS



OPEN

## Insects with similar social complexity show convergent patterns of adaptive molecular evolution

Kathleen A. Dogantzis<sup>1</sup>, Brock A. Harpur <sup>1,2</sup>, André Rodrigues<sup>3</sup>, Laura Beani<sup>4</sup>, Amy L. Toth<sup>5</sup> & Amro Zayed <sup>1</sup>

Eusociality has independently evolved multiple times in the Hymenoptera, but the patterns of adaptive molecular evolution underlying the evolution and elaboration of eusociality remain uncertain. Here, we performed a population genomics study of primitively eusocial *Polistes* (paper wasps), and compared their patterns of molecular evolution to two social bees; *Bombus* (bumblebees), and *Apis* (honey bees). This species triad allowed us to study molecular evolution across a gradient of social complexity (*Polistes* < *Bombus* < *Apis*) and compare species pairs that have similar (i.e. *Polistes* and *Bombus*) or different (i.e. *Polistes* and *Apis*) life histories, while controlling for phylogenetic distance. We found that regulatory genes have high levels of positive selection in *Polistes*; consistent with the prediction that adaptive changes in gene regulation are important during early stages of social evolution. *Polistes* and *Bombus* exhibit greater similarity in patterns of adaptive evolution including greater overlap of genes experiencing positive selection, and greater positive selection on queen-biased genes. Our findings suggest that either adaptive evolution of a few key genes underlie the evolution of simpler forms of eusociality, or that the initial stages of social evolution lead to selection on a few key traits orchestrated by orthologous genes and networks.

Understanding the origin and elaboration of eusociality is a major goal of evolutionary biology. Eusociality has independently evolved several times in the Hymenoptera<sup>1</sup> and is characterized by overlapping generations, cooperative brood care, and reproductive division of labour<sup>2–4</sup>. The expression of eusociality is highly variable and ranges from subsocial species characterized by small nests and no caste system but well developed parental care, to highly advanced eusocial species that are characterized by large colonies and highly specialised castes<sup>3,5</sup>. Eusocial species can thus be categorized along a gradient of social complexity typified by traits that include colony size, extent of caste divergence, behavioural specialization, and mode of colony reproduction. However, this gradient does not necessarily indicate a tendency for social lineages to become increasingly complex over evolutionary time.

There has been growing interest in identifying the mechanisms and patterns of molecular evolution associated with the rise and elaboration of eusociality. Some researchers have postulated that changes in gene regulation, particularly regulating phenotypic plasticity, are essential for eusocial evolution since it allows individuals to maximize their fitness by adapting to fluctuating environmental and social conditions<sup>6–12</sup>. Other researchers have raised the importance of protein coding evolution, especially of novel genes that underlie novel phenotypic traits found in some social species<sup>13–16</sup>. More recently, it was postulated that different patterns of molecular evolution can be involved during the different stages of social evolution; termed the “social ladder hypothesis”<sup>17,18</sup>. Earlier stages of sociality are predicted to involve prominent changes in gene regulation, while intermediate stages are predicted to involve fixed genomic changes in regulatory sequences and protein coding genes. Finally, later stages of sociality are expected to show greater genome and adaptive changes including the appearance and diversification of novel genes that control specialized social traits<sup>17,18</sup>.

<sup>1</sup>Department of Biology, York University, 4700, Keele St., Toronto, Ontario, Canada. <sup>2</sup>Donnelly Centre, University of Toronto, Toronto, Canada. <sup>3</sup>Departamento de Biologia, Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil. <sup>4</sup>Dipartimento di Biologia, Università di Firenze, Florence, Italy. <sup>5</sup>Department of Ecology, Evolution & Organismal Biology and Entomology, Iowa State University, Ames, USA. Correspondence and requests for materials should be addressed to A.Z. (email: [zayed@yorku.ca](mailto:zayed@yorku.ca))

Trait	<i>Apis</i>	<i>Bombus</i>	<i>Polistes</i>
Reproductive Division of Labour	Present	Present	Present
Cooperative Brood Care	Present	Present	Present
Overlapping Generations	Present	Present	Present
Founding	Swarm	Foundress	Foundress
Colony Size	Large	Small	Small
Gynes	Absent	Present	Present
Dominance Hierarchy	Absent	Present	Present
Overwintering	Colony	Solitary	Aggregation
Caste Divergence	High	Intermediate	Low
Totipotent Workers	Absent	Absent	Present

**Table 1.** Traits associated with social complexity. Similarities and differences between colony and life history traits of *Apis*, *Bombus*, and *Polistes*.

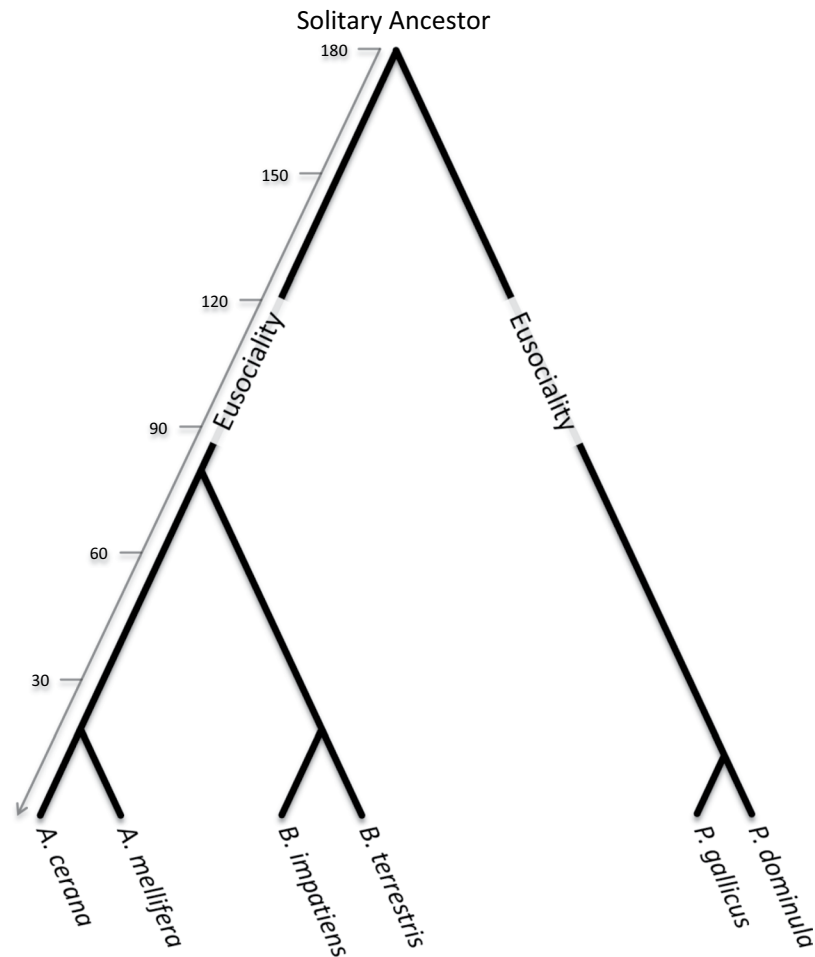
These above hypotheses have been integral for formulating the framework for eusocial evolution, but until recently, were challenging to address since it was difficult to estimate patterns of positive selection across the genomes of multiple non-model organisms. With diminishing sequencing costs and advancements in bioinformatic methods, it is possible now to sequence many individual genomes per species and use this data to estimate patterns of adaptive evolution<sup>19,20</sup>. Previous phylogenetic comparisons of several social and solitary bees and ants have provided a wealth of knowledge on the types of genes that show accelerated patterns of protein sequence evolution in eusocial insects<sup>14,21</sup>. However, it is difficult to directly quantify the selection coefficient on amino-acid changing mutations from such studies<sup>20,22,23</sup>. Population genomic studies provide robust methods for directly quantifying patterns of selection on genes and regulatory sequences<sup>12,16,24,25</sup>. To date, such studies have focused on either highly eusocial species<sup>16,25</sup> or multiple species with non-independent origins of sociality<sup>24</sup>.

Here, we present a comparative population genomic study involving *Apis* (honey bees), *Bombus* (bumblebees), and *Polistes* (paper wasps). *Polistes* of the family Vespidae display an intermediate level of social behaviour referred to as primitive eusociality. Both *Polistes* and *Bombus* are considered to be less socially complex relative to *Apis*, and share similar life history traits including; a solitary founding phase, small colony sizes, gyne (pre-queen) overwintering, establishment of dominance hierarchies, and less distinct caste divergence<sup>3,5,26–28</sup> (Table 1). Despite the similarities between *Polistes* and *Bombus*, *Bombus* is more socially complex. For example, *Bombus* workers can only lay haploid eggs<sup>29</sup>, while *Polistes* workers retain the potential to mate and lay fertilized diploid eggs<sup>26</sup>. In addition, *Bombus* possess morphologically distinct queen and worker castes, whereas *Polistes* queens and workers show no morphological differences and females retain caste totipotency into adulthood<sup>30</sup>. *Apis*, comparatively, is a highly advanced eusocial bee with large colony sizes, swarm founding, and distinct reproductive division of labour<sup>3</sup> (Table 1). Like *Bombus*, *Apis* workers can activate their ovaries in the absence of a queen to produce haploid drone eggs. *Apis* and *Bombus* are both corbiculate bees of the subfamily Apinae where eusociality has evolved once<sup>1,31,32</sup>. Vespidae is thought to have diverged from the family Apidae 180–150 million years ago, which has resulted in an equal divergence time between *Polistes* and *Bombus*, and *Polistes* and *Apis*<sup>1</sup> (Fig. 1). The species triad studied herein allows us to compare patterns of adaptive molecular evolution in two eusocial lineages with either similar (*Polistes* vs. *Bombus*) or different (*Polistes* vs. *Apis*) life histories and social complexity while controlling for divergence time.

We first performed a population genomic analysis of newly sequenced *P. dominula* (n = 10), and *P. gallicus* (n = 2) diploid workers to estimate patterns of positive selection in paper wasps, and then compared the dataset to published genomic studies on *Apis*<sup>16</sup> and *Bombus*<sup>24</sup>. The social ladder framework predicts greater similarities in the functions of genes under positive selection and enrichment of shared genes with signs of adaptive evolution between species with similar levels of social complexity (*Polistes* and *Bombus*) compared to those with different levels of social complexity (*Apis*). Qualitatively and quantitatively, we found greater similarities in patterns of adaptive molecular evolution between *Polistes* and *Bombus*. We also addressed the relationship between caste biased gene expression and patterns of adaptive evolution in the species triad. Interestingly, we found evidence that queen-biased genes show more adaptive evolution than worker-biased genes in *Bombus* and *Polistes*, the opposite pattern from previous studies of highly eusocial species.

## Results and Discussion

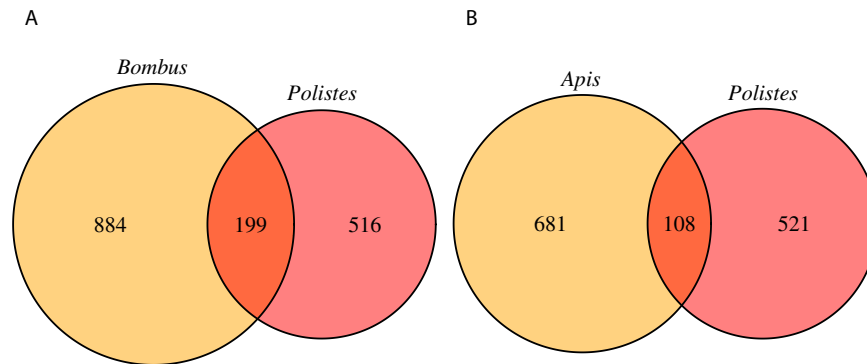
**Functions of genes experiencing adaptive evolution differ between eusocial species.** We used a Bayesian implementation of the McDonald-Kreitman (MK) test to estimate selection across the *Polistes* genome, and compared these levels to published estimates for *Apis* and *Bombus*<sup>16,24</sup>. The MK test quantifies adaptive evolution over the time-scale of species divergence by comparing patterns of synonymous and non-synonymous polymorphisms and divergence between two closely related species to estimate  $\gamma$ , the average selection coefficient on non-synonymous mutations scaled by the effective population size. We calculated  $\gamma$  for 9668 (81.8%) out of 11815 annotated genes in *Polistes dominula* (PdomGDB r1.2)<sup>33</sup>. We found 988 genes (10.45%) with signs of positive selection ( $\gamma > 1$ ) (Supplementary Fig. S1). Genes that showed signs of positive selection ( $\gamma > 1$ ) in *Polistes* were enriched for functional clusters associated with transcription and gene expression, particularly proteins containing zinc finger domains and RNA polymerases. Additionally, there were clusters of genes associated with fatty acid synthesis and metabolism, post-transcriptional regulation, and metal binding. The most significant term following false discovery rate correction was RNA 3'-end processing (GO:0031123).



**Figure 1.** Evolutionary relationships between *Polistes*, *Apis*, and *Bombus*. The divergence of *Apis* and *Bombus* is approximately 80–50 MYA while Apidae diverged from Vespidae ~150–180 MYA, suggesting a total divergence of 300–360 MY. The figure highlights two independent origins of eusociality; the first before the divergence of *Apis* and *Bombus* within the Apinae lineage and the second occurred in the Vespidae lineages, after the divergence from Apidae. Divergence time estimates in MYA. Diagram adapted from estimates in<sup>1,39,58,59</sup>.

The gene functions of adaptively evolving loci in *Polistes* are largely distinctive from those previously discovered for *Bombus* and *Apis*. In *Bombus* enriched GO terms are primarily associated with metabolic functions<sup>24</sup>, while *Apis* genes are enriched for terms associated with behaviour and sensory perception<sup>16</sup>. Though functional enrichment patterns of adaptively evolving genes are divergent between species, they are consistent with the social ladder framework. The social ladder framework postulates that for early to intermediate stages of social evolution, changes in patterns of gene regulation should be prominent<sup>17,18</sup>. This includes adaptive changes to genes such as transcription factors, but can also involve changes in non-coding sequences such as cis-regulatory regions, transcription factor binding sites, and microRNAs<sup>12,21,34</sup>. Our study lends support to this inference as many of the genes found to be under positive selection in primitively eusocial *Polistes* have functions associated with the regulation of gene expression. We found a significantly higher proportion of zinc finger proteins with signs of positive selection, which regulate gene expression by binding to DNA and controlling transcription of target genes<sup>35</sup>, in *Polistes* ( $n = 62$ ), relative to *Bombus* ( $n = 53$ ) ( $\chi^2 = 14.60$ ,  $p = 1.33e^{-4}$ ) and *Apis* ( $n = 15$ ) ( $\chi^2 = 25.13$ ,  $p = 5.37e^{-7}$ ), and we found a significantly higher proportion of genes associated with RNA polymerase under positive selection in *Polistes* ( $n = 54$ ), relative to *Bombus* ( $n = 49$ ) ( $\chi^2 = 10.68$ ,  $p = 1.09e^{-3}$ ) and *Apis* ( $n = 23$ ) ( $\chi^2 = 9.75$ ,  $p = 1.79e^{-3}$ ). We also found evidence for positive selection on a greater proportion of genes related to polyadenylation in *Polistes* ( $n = 11$ ), relative to *Bombus* ( $n = 4$ ) ( $\chi^2 = 7.12$ ,  $p = 7.62e^{-3}$ ) and *Apis* ( $n = 2$ ) ( $\chi^2 = 4.27$ ,  $p = 0.039$ ), and 3' end processing in *Polistes* ( $n = 20$ ) relative to *Bombus* ( $n = 10$ ) ( $\chi^2 = 10.36$ ,  $p = 1.29e^{-3}$ ) and *Apis* ( $n = 5$ ) ( $\chi^2 = 6.76$ ,  $p = 9.33e^{-3}$ ). Both processes are important for mRNA processing.

Of the three species studied herein, *Polistes* exhibit the most primitively eusocial societies with the most flexible castes; foundresses perform foraging and reproductive tasks during early colony stages, and workers maintain the ability to transition to reproductive queens into adulthood<sup>5,26</sup>. Flexibility in gene regulation contribute to the phenotypic plasticity of the species, allowing individuals to enhance their fitness during critical events during the colony lifetime, such as queen death<sup>36</sup>. Our data lends support to the importance of gene regulatory evolution in this primitively eusocial group of wasps, an idea which was also suggested from *de novo* genome sequencing of another *Polistes* species<sup>37</sup>.



**Figure 2.** Venn diagram of overlapping genes between *Polistes*, *Bombus*, and *Apis*. (A) Overlap of genes with positive selection ( $\gamma > 1$ ) between *Bombus* and *Polistes*, (B) Overlap of genes with positive selection ( $\gamma > 1$ ) between *Apis* and *Polistes*. The proportion of genes that overlap between *Bombus* and *Polistes* is significantly higher relative to the proportion of genes that overlap between *Polistes* and *Apis* (Fisher Exact Test  $p = 2.79e^{-10}$ ).

### Less socially complex species share greater overlap of orthologous genes undergoing adaptive evolution.

Eusociality is a model example of convergent evolution as it has evolved independently over multiple evolutionary lineages<sup>1,4,38</sup>. It may be that positive selection on a few key genes gave rise to eusociality in different lineages. Thus far, studies comparing genes under positive selection across multiple eusocial lineages have not found a high degree of overlap between genes experiencing positive selection in different eusocial species<sup>21,24</sup>.

We compared the proportion of overlapping genes under positive selection in *Polistes*, *Apis*, and *Bombus* (Supplementary Dataset 1). We hypothesized that if there is a common set of genes influencing the evolution of eusociality, we expect greater similarity of genes under positive selection between *Polistes* and *Bombus*, relative to *Polistes* and *Apis* due to similarities in life history and colony traits. We found the greatest overlap of genes under positive selection ( $n = 299$ ) between *Apis* and *Bombus* (Fisher Exact Test (one tail)  $p = 2.2e^{-16}$ ). However, this is not surprising given that *Bombus* and *Apis* are closely related, and shared a common ancestor ca. 80 MYA<sup>39</sup>. We found a significant overlap of 199 genes (12.4%) with signs of positive selection between *Polistes* and *Bombus* (Fisher Exact Test (one tail)  $p = 1.01e^{-15}$ ) (Fig. 2A), and a significant overlap of 108 genes (8.2%) (Fisher Exact Test (one tail)  $p = 2.28e^{-07}$ ) between *Polistes* and *Apis* (Fig. 2B). Overall, there were far more common genes between *Polistes* and *Bombus* relative to *Polistes* and *Apis* (Fisher Exact Test  $p = 2.79e^{-10}$ ). Finally, the three-way comparison between *Polistes*, *Bombus*, and *Apis* revealed a higher than expected overlap of genes under positive selection across all three species ( $n = 31$ ) ( $\chi^2 = 84.83$ ,  $p < 0.0001$ ).

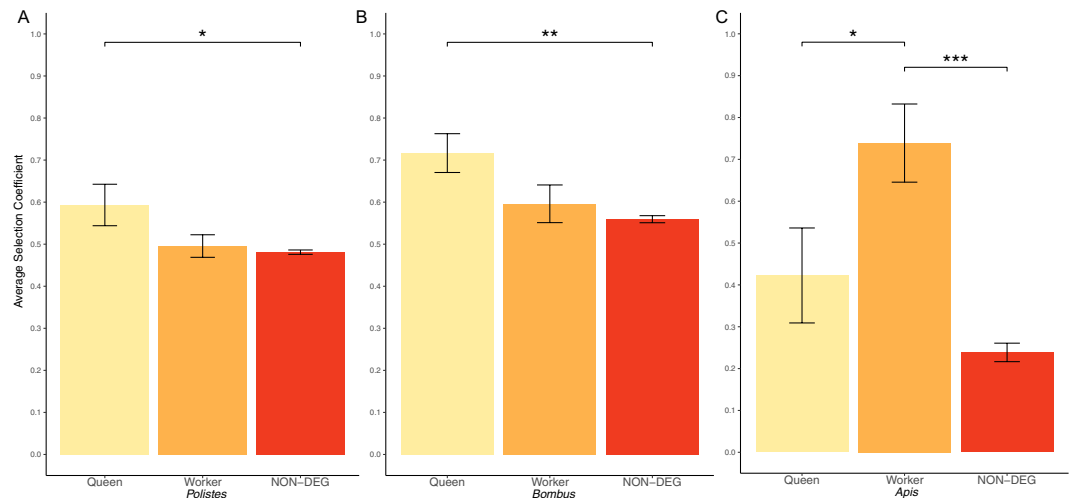
Our results highlight that there is greater overlap of genes undergoing adaptive evolution between *Polistes* and *Bombus* compared to *Polistes* and *Apis*. This is a compelling result given that Vespidae diverged from Apidae 180–150 MYA, and the evolutionary distance between *Polistes* and *Bombus* is similar to that of *Polistes* and *Apis*<sup>1</sup> (Fig. 1). The finding that species with similar life history traits and social complexity (Table 1) have greater similarity in patterns of adaptive evolution suggests that earlier stages of social evolution may be driven by positive selection on a few common genes, or that being eusocial can drive patterns of adaptive evolution for a few orthologous genes. Both scenarios highlight a much greater association between social complexity and patterns of molecular evolution than previously anticipated.

### Queen traits have greater influence on adaptive evolution in less socially complex species.

In addition to studying if similar genes and gene functions experience adaptive evolution in our eusocial species triad, we also quantified the extent of positive selection on queen-biased and worker-biased genes in *Polistes* and compared the patterns to *Bombus*<sup>24</sup> and *Apis*<sup>16</sup>. Caste-specific phenotypes are likely orchestrated via caste-biased gene expression<sup>9,40</sup> and thus, our population genetic analysis of caste-biased genes provides insights on the relative contribution of queen and worker traits to fitness and adaptation in different eusocial lineages.

We quantified expression patterns of six *Polistes dominula* queens and five *Polistes dominula* workers<sup>33</sup> to identify queen-biased (upregulated in queens) and worker-biased (upregulated in workers) genes. Using our population genomic data we were able to estimate the selection coefficient for 7612 out of 8946 genes with transcriptomic data. We observed a trend that queen-biased genes had a higher average  $\gamma$  ( $0.593 \pm \text{SEM } 0.0493$ ,  $n = 114$ ) compared to worker-biased genes ( $0.496 \pm \text{SEM } 0.0268$ ,  $n = 287$ ) and non-differentially expressed genes ( $0.481 \pm \text{SEM } 0.00517$ ,  $n = 7211$ ) (Fig. 3A). Average  $\gamma$  significantly differed among groups ( $F_{2,7609} = 3.735$ ,  $p = 0.024$ ), with queen-biased genes being significantly higher than non-differentially expressed genes (Tukey  $p = 0.020$ ), but not worker-biased genes (Tukey  $p = 0.113$ ). Worker-biased genes were not significantly different than non-differentially expressed genes (Tukey  $p = 0.858$ ). Furthermore, the proportion of genes with evidence of positive selection ( $\gamma > 1$ ) was significantly higher in queen-biased genes (19.3%,  $n = 22$ ) relative to worker-biased genes (11.1%,  $n = 32$ ) ( $\chi^2 = 4.02$ ,  $df = 1$ ,  $p = 0.045$ ).

The higher incidence of positive selection on queen-biased genes in *Polistes* is similar to patterns observed in *Bombus*<sup>24</sup>. Harpur, *et al.*<sup>24</sup> found higher levels of average  $\gamma$ , and a high proportion of genes with  $\gamma > 1$ , in genes that are significantly upregulated in reproductively-active (queens and foundresses) females relative to non-reproductively active females (gynes and workers) in bumblebees. To make the *Polistes* and *Bombus* dataset



**Figure 3.** Selection on caste-biased genes in (A) *Polistes*, (B) *Bombus*, and (C) *Apis*. Queen-biased genes have the highest  $\gamma$  and the highest proportion of genes with  $\gamma > 1$  in both *Polistes* and *Bombus*. *Apis* exhibits an opposite pattern where worker-biased genes have the highest  $\gamma$  and the highest proportion of genes with  $\gamma > 1$  (see Results for details). The error bars represent the standard error of the mean. NON-DEG = Non-differentially expressed genes. \* < 0.05, \*\* < 0.01, \*\*\* < 0.001.

comparable, we reanalyzed the *Bombus* dataset focusing only on queen-biased ( $n = 185$ ) and worker-biased genes ( $n = 192$ )<sup>24,41</sup> (Fig. 3B). Similar to *Polistes*, we found a trend in which queen-biased genes have a higher average  $\gamma$  ( $0.717 \pm \text{SEM } 0.046$ ) relative to worker-biased ( $0.596 \pm \text{SEM } 0.045$ ) and non-differentially expressed genes ( $0.559 \pm \text{SEM } 0.008$ ). Average  $\gamma$  significantly differed among groups ( $F_{2,5640} = 6.148$ ,  $p = 0.00215$ ), with queen-biased genes significantly higher than non-differentially expressed genes (Tukey  $p = 0.0017$ ), but not significantly different from worker-biased genes (Tukey  $p = 0.135$ ), and worker-biased genes were not significantly different than non-differentially expressed genes (Tukey  $p = 0.693$ ). Additionally, the proportion of genes with signatures of positive selection ( $\gamma > 1$ ) was significantly higher in queen-biased genes (31.4%,  $n = 58$ ) relative to worker-biased genes (17.7%,  $n = 34$ ) ( $\chi^2 = 7.19$ ,  $df = 1$ ,  $p = 0.007$ ) in *Bombus*, similar to *Polistes* (see above).

The enrichment of queen-biased genes among adaptively evolving genes in *Polistes* and *Bombus* are in stark contrast to *Apis* where worker-biased genes show the greatest levels of positive selection. This includes a greater average selection coefficient ( $F_{2,1688} = 15.8$ ,  $p = 1.6e^{-07}$ ) relative to queens (Tukey  $p = 0.045$ ) and non-differentially expressed genes (Tukey  $p = 2.0e^{-07}$ ) (Fig. 3C), and a greater proportion of genes undergoing adaptive evolution ( $\gamma > 1$ ) in worker-biased genes relative to queen-biased genes ( $\chi^2 = 12.81$ ,  $df = 1$ ,  $p = 0.0003$ )<sup>16</sup>.

There are two possible explanations for the observed similarities of adaptive evolution in caste-biased genes between *Polistes* and *Bombus*, in contrast to *Apis*. Both *Polistes* and *Bombus* are independently founding; nests are initiated by foundresses, who – if successful in establishing a nest – become the future queen. In contrast, *Apis* is swarm-founding; colonies are established by swarms of bees that include a mated queen and several thousand workers. For independent founding species, there are likely strong selective pressures on future queens to find and establish a nest, forage, and lay eggs<sup>42,43</sup>. There is often very high mortality rates at this stage of colony development<sup>44</sup>. These selection pressures may lead to higher levels of positive selection on genes expressed by queens. Alternatively, the relative contribution between worker-biased and queen-biased genes to adaptive evolution may be influenced by the number of workers within a colony, or the proportion of time that workers are present during a eusocial species' life cycle<sup>24</sup>.

Recent population genomic work on the advanced eusocial pharaoh ant, *Monomorium pharaonis* revealed similar patterns to *Bombus* and *Polistes* where queen-biased genes experienced the greatest rates of adaptive evolution<sup>25</sup>. Pharaoh ants, unlike *Bombus* and *Polistes*, found nests by budding (colony fission, Schmidt *et al.*<sup>45</sup>). However, like *Bombus* and *Polistes*, they have smaller nests of approximately 1200 individuals<sup>45</sup>; significantly smaller than *Apis* colonies that typically contain 40,000 to 60,000 workers<sup>46</sup>. Additionally, pharaoh ant colonies are polygynous, maintaining multiple queens per colony, and queen production increases when colony size is reduced<sup>45</sup>. The emphasis on queen numbers and their input to colony growth may influence the patterns of adaptive evolution seen in pharaoh ants. More comparative work using a greater number of species with varying social complexity is needed to better understand how colony founding, colony size, and other life history and eusocial traits influence patterns of adaptive evolution of caste-biased genes. This will enable us to fully understand the molecular processes involved in the evolution and elaboration of eusociality, and study the relationship between eusocial evolution and molecular evolution.

## Conclusion

Our comparative analyses have demonstrated that there are convergent patterns of adaptive molecular evolution associated with different levels of social complexity in three species of Hymenoptera. We utilized a species triad to compare two species that were either similar (*Polistes* or *Bombus*) or different (*Polistes* and *Apis*), while controlling for phylogenetic distance. We found that *Polistes* and *Bombus*, two less socially complex species that

evolved eusociality independently, shared greater patterns of adaptive molecular evolution relative to *Polistes* and *Apis*. This includes sharing more genes under positive selection, and greater rates of positive selection acting on queen-biased genes. Our study also highlighted the importance of adaptive changes in gene regulation during the earliest stages of social evolution; genes involved in regulating transcription were highly enriched among adaptively evolving genes in *Polistes*.

## Methods

**Genome Alignment, Variant Calling, and Filtration.** Methods for genome alignment, SNP detection, and filtering are described in detail in the supplementary material. In brief, paired-end Illumina genome sequencing was performed on ten *P. dominula* and two *P. gallicus* female worker samples. Reads were aligned to the unmasked *P. dominula* reference genome (PdomGDB r1.2)<sup>33</sup> using BWA-MEM<sup>47</sup>. Duplicates were marked using Picard (<http://broadinstitute.github.io/picard/>) and indel realignment was performed using the Genome Analysis Toolkit (GATK)<sup>48,49</sup>. Variants were detected with GATK's HaplotypeCaller using all species-specific alignments in unison. Variants were subsequently filtered of poor quality SNPs based on GATK's hard filter recommendations, upper and lower depth limits, missing data, and regions of high sequence homology. In addition, genes were filtered on a percent threshold for low to no coverage across coding regions. Genes that had poor coverage (based on lower depth limit) for >0.1 (10%) of the coding sequence were removed from further analysis.

**Variant Annotation and Filtration.** Variants that passed all of the filtering criteria were annotated for predicted effects on genes using SnpEff<sup>50</sup>. Genes were removed from the analysis if they contained warnings for an incomplete transcript, multiple stop codons, and no start codon. Additionally, genes were removed if they possessed annotations for lost stop codons, gain of stop codon, loss of start codon, or non-synonymous start variants. Tri-allelic variants were also discarded to limit the chances of retaining SNPs with potential sequencing error and to simplify downstream analysis for the large dataset<sup>51</sup>.

**Quantifying Selection.** A Bayesian implementation of the McDonald-Kreitman test<sup>23</sup> was used to estimate the prevalence of selection acting on genes in *Polistes*. Synonymous and non-synonymous substitutions were determined using the predicted gene annotations from SnpEff. Divergence data was based on fixed mutations between *P. dominula* and *P. gallicus* gene sequences, while polymorphisms were based on variable mutation sites in both species. Sites were removed if the comparison resulted in a triallelic variant and in cases where both species were fixed for the same allele. The Bayesian implementation of the McDonald-Kreitman test, SnPIRE, makes use of genome wide information and doesn't require a priori knowledge of species divergence parameters. The selection coefficient, gamma ( $\gamma$ ), is calculated for each gene, and represents the average selection coefficient on non-synonymous mutations in a gene scaled by the effective population size ( $\gamma = 2N_e s$ ) (Eliertson *et al.* 2012). Typically,  $\gamma > 1$  indicates positive selection<sup>23</sup>. Selection coefficients for *Apis* and *Bombus* were collected from previously published datasets<sup>16,24</sup>.

**Gene ontology.** Gene ontology (GO) was performed using DAVID 6.8 (2013–2017)<sup>52</sup> with default annotation databases including KEGG pathways and Interpro protein domains. We performed the analysis using *Drosophila melanogaster* fly base gene IDs. *Polistes* sequences orthologous to *Drosophila* were identified using reciprocal blastp matches with an E-value threshold of  $1e^{-10}$ . *Drosophila* orthologs were found for 6534 genes (55%) of the 11815 annotated genes in *Polistes* (PdomGDB r1.2)<sup>33</sup>. Reciprocal blastp matches were also found for *Apis* (n = 6986) and *Bombus* (n = 12393) using the same method. Gene ontology functional annotation clusters with an enrichment score  $\geq 1.3$ , and gene ontology terms with  $p < 0.05$  after FDR correction were determined to be of significance.

**Differential Gene Expression.** Difference in gene expression between *Polistes* queens and workers was determined from Illumina paired-end RNA-Seq libraries of six *P. dominula* queen (SRX1124061, SRX1124060, SRX1124059, SRX1124057, SRX1124056, SRX1124054) and five *P. dominula* worker (SRX1124053, SRX1124052, SRX1124051, SRX1124049, SRX1122234) whole heads<sup>33</sup>. Sequences were downloaded from GenBank and aligned to the unmasked *P. dominula* reference genome (PdomGDB r1.2)<sup>33</sup> using the default parameters of STAR's 2-pass method<sup>53</sup>. The STAR 2-pass method performs two alignments in which the splice junctions detected in the first alignment are used to guide the second alignment. Using the resulting BAM files, transcriptomes were assembled using Cufflinks<sup>54</sup>. Transcriptomes were produced with bias detection and correction algorithms using the *P. dominula* reference sequence (PdomGDB r1.2)<sup>33</sup>, and were quantified using the annotated reference genome. Assembled transcriptomes were merged using Cuffmerge to concatenate overlapping regions that agree in splice and orientation position. Cuffdiff was then used to find significant changes in gene level expression between worker and queen castes. Loci were removed from the analysis if they could not be identified after the merge step due to: the direct overlap of genes located on opposite strands, genes on the same strand that were merged due to shared transcripts, or clustering of three or more genes. Remaining genes were determined to be significantly differentially expressed if they passed the following thresholds: FDR < 0.05, and FPKM > 1. Differential expression data for *Apis mellifera* (i.e. the honey bee protein atlas which lists proteins with consistent patterns of differential expression across 26 tissues of queens and workers) and *Bombus impatiens* (i.e. microarray study that examined brain gene expression of queens, workers, foundresses, and gynes) were collected from previously published datasets<sup>41,55</sup>. We followed the same procedure used to previously identify caste biased genes<sup>16,24</sup>. Our study makes use of the best transcriptomic data sets currently available for *Apis*, *Bombus*, and *Polistes*, but clearly, the honey bee data set is much larger in scope relative to *Bombus* and *Polistes*. Studies of positive selection on caste biased genes will surely improve as richer transcriptomic data sets become available for *Polistes* and *Bombus*.

**Ortholog analysis between *Polistes*, *Bombus*, and *Apis*.** We conducted a cross-species orthology analysis involving *Polistes*<sup>33</sup>, *Bombus*<sup>56</sup>, and *Apis*<sup>57</sup>. Orthologs between species pairs were identified using reciprocal blastp matches with an E-value threshold of  $1e^{-10}$ . Multiple best blastp matches of the same sequence segment were considered. Reciprocal matches with corresponding gamma values for *Polistes* and *Apis*<sup>16</sup>, *Polistes* and *Bombus*<sup>24</sup>, and *Bombus* and *Apis*, revealed 7338, 6496, and 8204 ortholog matches respectively. A three-way reciprocal blastp between all three species produced 5943 gene orthologs.

**Data availability.** All sequence data used in this paper are available from NCBI's Sequence Read Archive (SRA) under BioProject PRJNA477544.

## References

- Peters, R. S. *et al.* Evolutionary history of the Hymenoptera. *Current Biology* **27**, 1013–1018 (2017).
- Batra, S. Nests and social behavior of halictine bees of India (Hymenoptera: Halictidae). *Indian J Entomol* **28**, 375–393 (1966).
- Michener, C. Comparative social behavior of bees. *Annu Rev Entomol* **14**, 299–342, <https://doi.org/10.1146/annurev.en.14.010169.001503> (1969).
- Wilson, E. O. The insect societies. (Belknap Press of Harvard University Press, 1971).
- West-Eberhard, M. The Social Biology of Polistine Wasps. *Misc Pub Mus Zool Univ Mich* **140**, 1–101 (1969).
- West-Eberhard, M. J. & Turillazzi, S. Natural history and evolution of paper wasps. (Oxford University Press, 1996).
- West-Eberhard, M. J. Phenotypic plasticity and the origins of diversity. *Annu Rev Ecol Syst* **20**, 249–278, <https://doi.org/10.1146/annurev.es.20.110189.001341> (1989).
- Linksvayer, T. A. & Wade, M. J. The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. *Q Rev Biol* **80**, 317–336, <https://doi.org/10.1086/432266> (2005).
- Sumner, S., Pereboom, J. J. M. & Jordan, W. C. Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, *Polistes canadensis*. *Proceedings. Biological sciences / The Royal Society* **273**, 19–26, <https://doi.org/10.1098/rspb.2005.3291> (2006).
- Berens, A. J., Hunt, J. H. & Toth, A. L. Comparative Transcriptomics of Convergent Evolution: Different Genes but Conserved Pathways Underlie Caste Phenotypes across Lineages of Eusocial Insects. *Mol Biol Evol* **32**, 690–703, <https://doi.org/10.1093/molbev/msu330> (2015).
- Toth, A. L. *et al.* Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* **318**, 441–444, <https://doi.org/10.1126/science.1146647> (2007).
- Molodtsova, D., Harpur, B. A., Kent, C. F., Seevananthan, K. & Zayed, A. Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors. *Front Genet* **5**, 1–7, <https://doi.org/10.3389/fgene.2014.00431> (2014).
- Johnson, B. R. & Tsutsui, N. D. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* **12**, 164, <https://doi.org/10.1186/1471-2164-12-164> (2011).
- Woodard, S. H. *et al.* Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci USA* **108**, 7472–7477, <https://doi.org/10.1073/pnas.1103457108> (2011).
- Jasper, W. C. *et al.* Large-Scale Coding Sequence Change Underlies the Evolution of Postdevelopmental Novelty in Honey Bees. *Mol Biol Evol* **32**, 334–346, <https://doi.org/10.1093/molbev/msu292> (2014).
- Harpur, B. A. *et al.* Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci USA* **111**, 2614–2619, <https://doi.org/10.1073/pnas.1315506111> (2014).
- Rehan, S. M. & Toth, A. L. Climbing the social ladder: the molecular evolution of sociality. *Trends Ecol Evol* **30**, 426–433, <https://doi.org/10.1016/j.tree.2015.05.004> (2015).
- Toth, A. L. & Rehan, S. M. Molecular evolution of insect sociality: An Eco-Evo-Devo perspective. *Annu Rev Entomol* **62**, 419–442 (2017).
- Hasselmann, M., Ferretti, L. & Zayed, A. Beyond fruit-flies: population genomic advances in non-Drosophila arthropods. *Brief Funct Genomics* **14**, 424–431, <https://doi.org/10.1093/bfgp/evl010> (2015).
- Kent, C. F. & Zayed, A. Population genomic and phylogenomic insights into the evolution of physiology and behaviour in social insects. *Adv Insect Physiol* **48**, 293–324 (2015).
- Kapheim, K. M. *et al.* Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1144 (2015).
- McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**, 652–654 (1991).
- Eilertson, K. E., Booth, J. G. & Bustamante, C. D. SnpPRE: selection inference using a Poisson random effects model. *PLoS Comput Biol* **8**, e1002806, <https://doi.org/10.1371/journal.pcbi.1002806> (2012).
- Harpur, B. A. *et al.* Queens and workers contribute differently to adaptive evolution in bumble bees and honey bees. *Genome Biol Evol* **9**, 2395–2402 (2017).
- Warner, M. R., Mikheyev, A. S. & Linksvayer, T. A. Genomic signature of kin selection in an ant with obligately sterile workers. *Mol Biol Evol* **34**, 1780–1787 (2017).
- Jandt, J. M., Tibbetts, E. A. & Toth, A. L. *Polistes* paper wasps: a model genus for the study of social dominance hierarchies. *Insectes Soc* **61**, 11–27, <https://doi.org/10.1007/s00040-013-0328-0> (2014).
- Pardi, L. 'La 'Dominazione'è il Ciclo Ovario Annuale in *Polistes gallicus* (L.). Ricerche sui Polistini VII. *Bollettino Istituto di Entomologia Università di Bologna* **15**, 25–84 (1946).
- Pardi, L. Dominance order in *Polistes* wasps. *Physiol Zool* **21**, 1–13 (1948).
- Bourke, A. & Ratnie, F. Kin-selected conflict in the bumble-bee *Bombus terrestris* (Hymenoptera: Apidae). *Proceedings of the Royal Society of London B: Biological Sciences* **268**, 347–355 (2001).
- Hunt, J. H. The evolution of social wasps. (Oxford University Press, 2007).
- Cardinal, S. & Danforth, B. N. The antiquity and evolutionary history of social behavior in bees. *PLOS one* **6**, e21086 (2011).
- Romiguier, J. *et al.* Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. *Mol Biol Evol* **33**, 670–678 (2015).
- Standage, D. S. *et al.* Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Mol Ecol* **25**, 1769–1784, <https://doi.org/10.1111/mec.13578> (2016).
- Simola, D. F. *et al.* Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res* **23**, 1235–1247, <https://doi.org/10.1101/gr.155408.113> (2013).
- Iuchi, S. Three classes of C2H2 zinc finger proteins. *Cell Mol Life Sci* **58**, 625–635, <https://doi.org/10.1007/PL00000885> (2001).
- Strassmann, J. E. *et al.* The Cost of Queen Loss in the Social Wasp *Polistes dominulus* (Hymenoptera: Vespidae). *J Kans Entomol Soc* **77**, 343–355, <https://doi.org/10.2317/E-15.1> (2004).
- Patalano, S. *et al.* Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc Natl Acad Sci USA* **112**, 13970–13975, <https://doi.org/10.1073/pnas.1515937112> (2015).

38. Branstetter, M. G. *et al.* Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Curr Biol* **27**, 1019–1025 (2017).
39. Cardinal, S. & Danforth, B. Bees diversified in the age of eudicots. *Proc R Soc B* **280**, 20122686, <https://doi.org/10.1098/rspb.2012.2686> (2013).
40. Ferreira, P. G. *et al.* Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome biology* **14**, R20, <https://doi.org/10.1186/gb-2013-14-2-r20> (2013).
41. Woodard, S. H., Bloch, G. M., Band, M. R. & Robinson, G. E. Molecular heterochrony and the evolution of sociality in bumblebees (*Bombus terrestris*). *Proc R Soc Lond B Biol Sci* **281**, 20132419, <https://doi.org/10.1098/rspb.2013.2419> (2014).
42. Woodard, S. H., Bloch, G., Band, M. R. & Robinson, G. E. Social regulation of maternal traits in nestfounding bumble bee (*Bombus terrestris*) queens. *J Exp Biol* **216**, 3474–3482 (2013).
43. Tibbetts, E. A. & Reeve, H. K. Benefits of foundress associations in the paper wasp *Polistes dominulus*: Increased productivity and survival, but no assurance of fitness returns. *Behav Ecol* **14**, 510–514, <https://doi.org/10.1093/beheco/arg037> (2003).
44. Zanette, L. R. & Field, J. Founders versus joiners: group formation in the paper wasp *Polistes dominulus*. *Anim Behav* **82**, 699–705 (2011).
45. Schmidt, A. M., Linksvayer, T. A., Boomsma, J. J. & Pedersen, J. S. Queen–worker caste ratio depends on colony size in the pharaoh ant (*Monomorium pharaonis*). *Insectes Soc* **58**, 139–144 (2011).
46. Winston, M. L. The biology of the honey bee. (Harvard University Press, 1991).
47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760, <https://doi.org/10.1093/bioinformatics/btp324> (2009).
48. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491–498, <https://doi.org/10.1038/ng.806> (2011).
49. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
50. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
51. Wang, Y., Lu, J., Yu, J., Gibbs, R. A. & Yu, F. An integrative variant analysis pipeline for accurate genotype/haplotype inference in population NGS data. *Genome Res* **23**, 833–842, <https://doi.org/10.1101/gr.146084.112> (2013).
52. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
53. Engström, P. G. *et al.* Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Methods* **10**, 1185–1191, <https://doi.org/10.1038/nmeth.2722> (2013).
54. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562–578, <https://doi.org/10.1038/nprot.2012.016> (2012).
55. Chan, Q. W. *et al.* Honey bee protein atlas at organ-level resolution. *Genome Res* **23**, 1951–1960 (2013).
56. Sadd, B. M. *et al.* The genomes of two key bumblebee species with primitive eusocial organization. *Genome biology* **16**, 76, <https://doi.org/10.1186/s13059-015-0623-3> (2015).
57. Elsik, C. G. *et al.* Finding the missing honey bee genes: lessons learned from a genome upgrade. *Bmc Genomics* **15**, 1 (2014).
58. Biewer, M., Schlesinger, F. & Hasselmann, M. The evolutionary dynamics of major regulators for sexual development among Hymenoptera species. *Front Genet* **6**, 1–11, <https://doi.org/10.3389/fgene.2015.00124> (2015).
59. Arévalo, E., Zhu, Y., Carpenter, J. M. & Strassmann, J. E. The phylogeny of the social wasp subfamily Polistinae: evidence from microsatellite flanking sequences, mitochondrial COI sequence, and morphological characters. *BMC Evol Biol* **4**, 8 (2004).

## Acknowledgements

The research was supported by a Discovery grant from the Natural Sciences and Engineering Council of Canada (NSERC) and York University Research Chair in Genomics (A.Z.), an NSERC CGS-M scholarship (K.A.D.), a NSF-IOB Behavioral Systems Grant (1456283 to A.L.T) and the São Paulo Research Foundation (FAPESP proc. 2015/05302-0 to A.R.).

## Author Contributions

K.A.D. and A.Z. wrote the manuscript. K.A.D. carried out the analyses, prepared all figures. L.B., A.R. and A.L.T. provided *Polistes* samples. B.A.H. provided the *Bombus* and *Apis* datasets. All authors reviewed and edited the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-28489-5>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018