Population genomics of the honey bee reveals strong signatures of positive selection on worker traits

Brock A. Harpur[a,1], Clement F. Kent[a,1,2], Daria Molodtsova[a], Jonathan M. D. Lebon[b], Abdulaziz S. Alqarni[c], Ayman A. Owayss[c], and Amro Zayed[a,3]

[a]Department of Biology and [b]Department of Computer Science and Engineering, York University, Toronto, ON, Canada M3J 1P3; and [c]Department of Plant Protection, College of Food and Agriculture Sciences, King Saud University, Riyadh 11451, Kingdom of Saudi Arabia

Edited by Joan E. Strassmann, Washington University in St. Louis, St. Louis, MO, and approved December 10, 2013 (received for review August 15, 2013)

Most theories used to explain the evolution of eusociality rest upon two key assumptions: mutations affecting the phenotype of sterile workers evolve by positive selection if the resulting traits benefit fertile kin, and that worker traits provide the primary mechanism allowing social insects to adapt to their environment. Despite the common view that positive selection drives phenotypic evolution of workers, we know very little about the prevalence of positive selection acting on the genomes of eusocial insects. We mapped the footprints of positive selection in *Apis mellifera* through analysis of 40 individual genomes, allowing us to identify thousands of genes and regulatory sequences with signatures of adaptive evolution over multiple timescales. We found Apoidea- and *Apis*-specific genes to be enriched for signatures of positive selection, indicating that novel genes play a disproportionately large role in adaptive evolution of eusocial insects. Worker-biased proteins have higher signatures of adaptive evolution relative to queen-biased proteins, supporting the view that worker traits are key to adaptation. We also found genes regulating worker division of labor to be enriched for signs of positive selection. Finally, genes associated with worker behavior based on analysis of brain gene expression were highly enriched for adaptive protein and *cis*-regulatory evolution. Our study highlights the significant contribution of worker phenotypes to adaptive evolution in social insects, and provides a wealth of knowledge on the loci that influence fitness in honey bees.

natural selection | kin selection | social evolution | taxonomically restricted genes

Eusocial behavior evolved multiple times in insects and is characterized in part by extreme asymmetries in the reproductive potential of individuals (1). This asymmetry is most pronounced in advanced eusocial insects, with their fertile queen and sterile worker castes. Darwin first recognized that natural selection cannot directly optimize worker phenotypes because workers are usually sterile (2). Hamilton (3, 4) developed kin-selection theory to describe the conditions that allow natural selection to indirectly optimize worker phenotypes if such phenotypes benefit their fertile kin. It is commonly believed that worker traits, such as sib-care, foraging, and colony defense, play important roles in allowing colonies to adapt to their environment (5–7). However, despite the central role of kin-selection and inclusive fitness theory in the field of Sociobiology (8, 9), we lack knowledge on the pattern and prevalence of positive selection acting on the genomes of eusocial insects.

Population genomic studies provide unprecedented opportunities to detect signatures of selection on DNA sequences over different timescales (10). There are several tests of selection that can be applied to genome-wide datasets. The McDonald–Kreitman (MK) test is arguably the best method for detecting selection on protein coding sequences because of its robustness to changes in a species' demography, which often confounds other tests of selection (10, 11). A recent Bayesian implementation of this classic test uses genome-wide estimates of polymorphism and divergence to improve statistical power (11). Outlier tests of selection are also less sensitive to population demography, which affect all loci within a genome; loci under selection thereby appear as outliers in the empirical distribution of genome-wide data (12–14). In spatially structured populations, outlier tests of genetic differentiation are especially useful in identifying loci underlying local adaptation (10, 15).

The honey bee, *Apis mellifera*, provides an ideal system for applying population genomics to understand the evolutionary forces shaping eusocial insect genomes. The honey bee is arguably the most well-known social insect at the level of behavior, physiology, and genetics, and there are many rich datasets that detail caste-specific transcriptomic and proteomic phenotypes (16, 17). The bee genome is relatively small (236 Mb) and lacks many repetitive elements (18), making assembly via short-read sequencing highly feasible. Finally, the honey bee's genetically and phenotypically distinct population groups in Africa, Asia, and Europe (19, 20) provide an opportunity to examine how the honey bee genome adaptively diverged in response to the different selective pressures experienced across its large and diverse native range (21, 22).

To this end, we undertook a comprehensive population genomic study of the honey bee by sequencing the genomes of 40 individual bees from different geographic regions, including a closely related species. Our goals were to first identify genomic regions with signs of positive selection and then examine the

Significance

Most hypotheses explaining the evolution of sociality in insects assume that positive selection drives the evolution of worker traits. Yet we know little about the extent of natural selection acting on social insects. We produced a map of positive selection for the honey bee through analysis of 40 individual genomes. We found strong evidence of positive selection acting on genes and regulatory sequences, and we discovered that mutations in worker-biased proteins tend to have greater fitness effects than mutations in queen-biased proteins. We also found many instances of positive selection acting on genes that influence worker traits, suggesting that worker phenotypes represent a major vector for adaptation in social insects.

Author contributions: B.A.H., C.F.K., and A.Z. designed research; B.A.H., C.F.K., D.M., J.M.D.L., A.S.A., and A.A.O. performed research; B.A.H., C.F.K., and D.M. analyzed data; and B.A.H., C.F.K., and A.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in NCBI's short read archive (accession no. SRP029219). The discovered SNPs can be viewed as a track on www.BeeBase.org.

[1]B.A.H. and C.F.K. contributed equally to this work.

[2]Present address: HHMI Janelia Farm Research Campus, Ashburn, VA 20147.

[3]To whom correspondence should be addressed. E-mail: zayed@yorku.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1315506111/-/DCSupplemental.

degree to which genes associated with worker traits contribute to adaptive evolution. Our study provides unparalleled insights on the genes and traits underlying adaptation in social insects.

Results

Genomic Diversity in *Apis mellifera*. We sequenced the diploid genomes of *A. mellifera* workers sampled from the four genetically distinct honey bee lineages (19, 20) in Africa (n = 11 workers), Asia (n = 10), East Europe (n = 9), and West/Northern Europe (n = 9) at an average coverage of 38× (Table S1). We also sequenced a single *Apis cerana* worker as an outgroup. We conducted preliminary Sanger sequencing of several randomly chosen exons to ensure that our collected specimens were not admixed (23). We discovered 12,041,303 SNPs in the 39 sequenced *A. mellifera* genomes, many of which were validated using independent datasets. We used the identified SNPs to confirm the population structure of the sampled bees. As expected, the 39 *A. mellifera* workers were assigned to four distinct populations and our sampled bees had very low levels of admixture (Fig. S1 and Table S2). Given that human management increases admixture levels in honey bees, the nonadmixed bees studied herein provide the best approximation of the four *A. mellifera* evolutionary lineages before human management (23).

Signatures of Positive Selection over Intermediate Timescales. We used a Bayesian implementation of the MK test (11) to estimate the strength and direction of selection on 12,303 genes since divergence between *A. mellifera* and *A. cerana* ∼5–25 Mya (24, 25). The MK test requires polymorphism data from at least one species (i.e., *A. mellifera*) and divergence data from at least one outgroup sequence (i.e., *A. cerana*) (10, 26, 27), and the Bayesian implementation of the MK test allows for the estimation of the population size-scaled selection coefficient γ on nonsynonymous mutations (11). Although the MK test is very robust to changes in population demography (11), we conservatively implemented this test using the polymorphism data from African bees only, which represent a large stable population that is minimally impacted by human management (23, 28). We found that most genes in the bee genome (approximately 90%) have γ between −1 and 1 (Fig. 1*A*); 0.9% of genes have $\gamma < -1$, consistent with strong purifying selection, whereas 9.3% of genes have $\gamma > 1$, consistent with strong positive selection (Dataset S1).

Signatures of Positive Selection over Short Timescales. Positive selection facilitating local adaptation creates loci with outlier levels of genetic differentiation (F_{ST}) relative to the rest of the genome (12, 13). We used outlier levels of F_{ST} to identify loci that have likely experienced geographically restricted positive selection since divergence of *A. mellifera*'s four evolutionary lineages ∼1 Mya to 11,000 y ago (24, 25). We used two approaches to detect genomic windows (≥5 kb) (Dataset S2) and SNPs with outlier levels of F_{ST} (Fig. 1*B* and Dataset S3) in the six pairwise population comparisons between the four bee lineages. The two approaches were highly concordant: outlier SNPs were significantly enriched within outlier windows (Fisher's exact test; $P < 2.2 \times 10^{-16}$ for all pairwise comparisons) and, on average, 55.5% of SNPs within outlier windows were themselves outlier SNPs. We detected an average of 5,715 outlier windows with extreme levels of genetic differentiation in the six pairwise population comparisons. Outlier SNPs contained alleles that were either nearly or completely fixed in pairwise population comparisons (F_{ST} ranged from 0.89 to 1). We found that SNPs with outlier F_{ST} in *A. mellifera* occur mostly in putative *cis*-regulatory regions: 18.5% of SNPs found 500 bp upstream of genes are outliers relative to 12.3% in exons, 8.5% in introns, and 8.6% in intergenic regions. However, there is still a considerable amount of positive selection acting on protein sequences: 11% of nonsynonymous mutations were outlier SNPs and outliers SNPs were enriched for nonsynonymous SNPs (Fisher's exact test, $P < 2.2 \times 10^{-15}$).

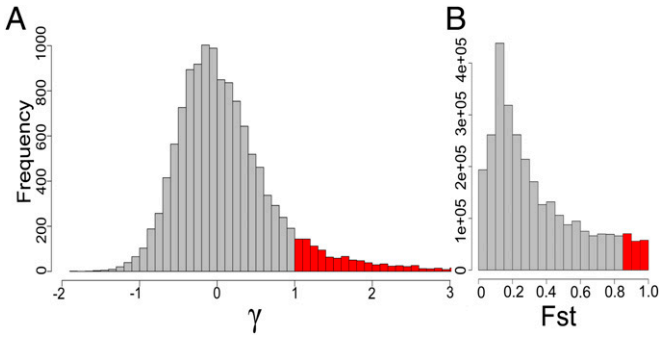

Fig. 1. Loci underpinning adaptive evolution in honey bees. Histogram of (*A*) the population-size scaled selection coefficient (γ) for 12,303 genes; γ ranges from −1.64 to 11.8, but we truncated the histogram at γ = 3 for readability. There are 88 genes with γ > 3. (*B*) Histogram of pairwise genetic differentiation (F_{ST}) between African and West European honey bees for 3,392,632 SNPs. F_{ST} histograms for the other five pairwise comparisons are found in Fig. S2. Areas in red represent outlier loci with signatures of adaptive evolution.

Biological Significance of Loci Underlying Positive Selection. We used Gene Ontology (GO) tools (29) to investigate the possible function of adaptively evolving loci. Genes associated with G protein-coupled receptors (GPCRs) and GPCR-signaling were enriched among adaptively evolving protein and regulatory loci over intermediate and short timescales (Dataset S4). GPCRs translate sensory inputs into cellular responses and are thus crucial for tuning an organism's physiology and behavior in response to the environment; this is particularly intriguing given the degree to which pheromones within a colony affect the biology of the different honey bee castes. We also found many annotation clusters enriched among adaptively evolving loci, including genes associated with adult behavior, cognition, nervous system development, metabolism, and steroid hormones (Dataset S4).

Selection on Taxonomically Restricted Genes. The gene content of genomes is dynamic over evolutionary time, and genomes contain both "old" genes and "new" genes. Old genes originated in an evolutionary-distant common ancestor and orthologous copies are found across many distant taxa, whereas new genes originated recently and are found only in specific taxonomic groups. Taxonomically restricted genes (TRGs) have been the subject of recent attention because they are predicted to be drivers of phenotypic evolution (30). The genomes of social insects harbor many TRGs, which are hypothesized to play an important role in the elaboration of sociality (31). TRGs in ants (32), bees (33), and wasps (34) tend to show, on average, worker-biased expression, which suggests that they play an important role in the evolution of worker phenotypes. We used the hierarchical catalog of orthologs in OrthoDB v.6 (35) to classify honey bee genes to four mutually exclusive groups: *Apis*-restricted, Apoidea-restricted, and Hymenoptera-restricted genes, as well as genes found in honey bees and at least one other insect order (Dataset S5). We then asked if TRGs exhibit differences in adaptive protein evolution over intermediate timescales.

We found a significantly higher proportion of *Apis*-restricted, Apoidea-restricted, and Hymenoptera-restricted genes with signs of strong positive selection ($\gamma > 1$) relative to genes found in other insects; 20.4% of *Apis* genes (n = 88), 21.8% of Apoidea genes (n = 215), and 15% of Hymenoptera genes (n = 1,321) have $\gamma > 1$ relative to 9% of genes found in other insects (n =

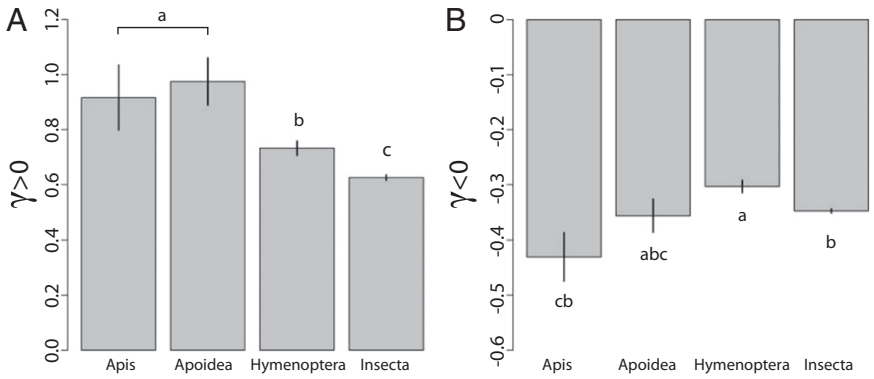

Fig. 2. (*A*) Taxonomically restricted genes have higher rates of adaptive evolution. For genes with signs of positive selection ($\gamma > 0$), γ is significantly higher in *Apis*-restricted and Apoidea-restricted genes, intermediate in Hymenoptera-restricted genes, and lowest for genes found in other insect orders. (*B*) For genes with signs of negative selection ($\gamma < 0$), *Apis*-restricted genes have the highest levels of negative selection. Error bars denote SEM.

8,686; χ^2, $P < 0.0003$ for all tests comparing *Apis*, Apoidea, and Hymenoptera genes relative to genes found in other insects). Furthermore, Apoidea-restricted genes have a significantly higher proportion of genes with $\gamma > 1$ relative to Hymenoptera-restricted genes (χ^2, $P = 0.025$). We also found that among *A. mellifera* genes with signs of positive selection ($\gamma > 0$), those found in all insects had the lowest average γ, those found in the Hymenoptera had intermediate average γ, and those found in the Apoidea had the highest average γ. *Apis*- and Apoidea-specific genes did not differ with respect to γ, but the differences between these two groups (i.e., *Apis* + Apoidea) and Hymenoptera- and insect genes were highly significant (Wilcoxon test, $P < 10^{-10}$). Average γ for Apoidea was more than three times higher than γ for genes found in all insects (Fig. 2*A*). We also observed differences in the prevalence of negative selection ($\gamma < 0$) among TRGs, with *Apis*-specific genes having significantly stronger purifying selection relative to Hymenoptera-specific genes (Fig. 2*B*) (Wilcoxon test, $P < 0.01$).

Adaptive Evolution of Queen-Biased and Worker-Biased Proteins. We investigated the degree to which worker and queen phenotypes contribute to colony fitness by examining if proteins with caste-biased expression show differences in the prevalence of positive selection. We used a list of caste-biased proteins from the *Honey Bee's Protein Atlas* (16), which provides quantitative proteomic data for 26 tissues assayed in queens and workers. Most honey bee proteins are expressed in both queen and worker tissues. We obtained γ estimates for 90 and 79 proteins that had higher expression in workers relative to queens (i.e., worker-biased) or vice versa (i.e., queen-biased), based on average whole-body expression (16; also see Dataset S5 herein); these proteins consistently exhibited higher expression in workers relative to queens, or vice versa, in most of the 26 tissues in the *Honey Bee Protein Atlas*. Although few in number, caste-biased proteins provide an objective way to identify sets of genes that are relevant to caste-biased phenotypes. We make the reasonable assumption that the evolution of worker-biased proteins is mostly shaped by forces acting on worker phenotypes, and that the evolution of queen-biased proteins is mostly shaped by forces acting on queen phenotypes. We found that worker-biased proteins had a significantly higher γ relative to queen-biased proteins (workers: average $\gamma = 0.77$; queens: average $\gamma = 0.42$; Wilcoxon test; $P < 0.0016$). Proteins that were not differentially expressed between queens and workers ($n = 1{,}095$) are expected to have the greatest levels of pleiotropy and constraint (36), and indeed they have significantly lower γ relative to worker-biased and queen-biased proteins (Wilcoxon test; $P < 0.013$) (Fig. 3*A*). We also found that worker-biased proteins were enriched for signatures of local adaptation. When benchmarked against non-differentially expressed proteins, we found that worker-biased proteins showed a greater enrichment of nonsynonymous outlier SNPs in more tissues and over a larger number of pairwise lineage comparisons relative to queen-biased proteins (Fisher's exact test: $P = 0.0005$).

Worker Traits and Colony Fitness. We investigated if genes that are a priori known to influence worker phenotypes showed signatures of positive selection.

Worker brain gene expression and behavior. There is strong evidence that shifts in brain gene expression mediate shifts in behavior in workers (17, 37, 38). Given the considerable and possibly adaptive differences in worker behavior between the honey bee's four evolutionary lineages (39), we predicted that differentially expressed genes (DEGs) associated with worker behavior would be enriched for signs of adaptive divergence. We queried 27 microarray experiments from the BeeSpace project that assayed the brain transcriptomes of nearly 1,000 workers across several natural or experimentally induced behavioral states (reviewed by refs. 17 and 37). We found that DEGs associated with 23 of 27 behavioral states have regulatory regions with significantly more outlier SNPs than expected by chance in at least one pairwise population comparison after correcting for multiple tests [false-discovery rate (FDR), $\alpha = 0.001$; $P < 2.2 \times 10^{-4}$] (Fig. S3). Eighteen of 27 behavioral states were enriched for coding sequences with significantly more nonsynonymous outlier SNPs than expected by chance (FDR, $\alpha = 0.001$; $P < 2 \times 10^{-6}$) (Fig. S3).

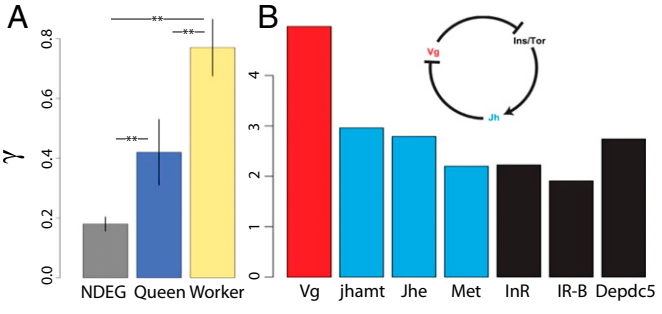

Fig. 3. Genes associated with worker phenotypes show signs of adaptive evolution in honey bees. (*A*) Worker-biased proteins have significantly higher selection coefficients relative to queen-biased proteins, and non-differentially expressed proteins (NDEG). Error bars denote SEM; $^{**}P < 0.01$. (*B*) Genes causally associated with worker division of labor have very high selection coefficients in the honey bee.

The enrichment of outlier loci in differentially expressed genes across most BeeSpace experiments indicates that genes associated with worker behavior are enriched for signatures of positive selection underlying local adaptation.

Worker division of labor. Worker honey bees undergo an age-related division of labor that allows them to transition from in-hive tasks to foraging and colony defense over time. This division of labor is regulated through an unusual interaction between the egg yolk protein Vitellogenin (Vg), juvenile hormone (JH) and JH-signaling, and insulin-like/TOR signaling (40–46) (Fig. 3*B*). The mutually repressive relationship between Vg and JH is unique to worker honey bees, prompting researchers to hypothesize that these conserved genes and signaling pathways were co-opted via natural selection to regulate worker division of labor in *Apis* (47, 48). Vg was previously shown to be under positive selection based on analysis of several exons (28), and our complete analysis shows its selection coefficient to be even higher than previously reported ($\gamma = 4.97$ vs. 1.88). Vg in turn regulates the central insulin/Tor growth pathway (49) and both of the bee's insulin receptors and the Depdc5 gene—part of a complex which sensitizes Tor signaling to cellular amino acid levels (50)—are under positive selection. Juvenile hormone acid methyltransferase (*Jhamt*) and juvenile hormone esterase (*Jhe*) are the proximal biosynthetic and catabolic enzymes for JH (51), and Met/Gce2 is the key cofactor in JH receptor complexes (52, 53); all of these are under significant and strong positive selection. We also investigated if *foraging* (54) and *malvolio* (55)—both implicated in worker division of labor—experience positive selection. We had previously estimated that *foraging* experiences nearly neutral evolution based on analysis of four exons (28), but our complete analysis herein indicated that *foraging* experiences positive selection ($\gamma = 0.99$). On the other hand, *malvolio* appears to be constrained ($\gamma = -0.33$). Given their causal involvement in regulating worker division of labor, signatures of selection on the above-mentioned genes (Fig. 3*B*) supports the hypothesis that worker division of labor has major influence on colony fitness.

Major royal jelly proteins. Workers have specialized hypopharyngeal glands that are used to synthesize royal jelly for feeding nestmates (39). The honey bee genome contains several genes that encode Major Royal Jelly Protein (Mrjp), and most of these genes are highly expressed in the hypopharyngeal glands of workers (56). The eight Mrjp genes studied herein had significantly higher gamma relative to other genes (Wilcoxon test, $P = 0.0015$) and three of eight genes had $\gamma > 2$ (binomial $P = 0.00003$), indicating high levels of positive selection. This list included Mrjp1 (*royalactin*), which is essential for inducing queen–worker differentiation (57). We also detected significant signs of positive selection on Mrjp4 and Mrjp7, which are known to be expressed only in workers and not in any other caste or developmental stage (56).

Discussion

The honey bee is a model eusocial organism and our analyses provide novel insights on the process of adaptive evolution in social insects. We found strong evidence of positive selection acting on protein coding sequences in the honey bee. The highest levels of selection were observed in genes that were taxonomically restricted to bees, whereas Hymenoptera-specific genes had intermediate levels of selection. The fact that Apoidea-specific genes had similar selection coefficients relative to *Apis*-specific genes suggests that adaptive evolution in the social honey bee is partially fueled by novel genes that were found in solitary ancestors. Although there is evidence that sociality evolved by co-opting conserved genetic toolkits (58), our results suggest that taxonomically restricted genes play an important and disproportionately large role in the adaptive evolution of social insects. Additionally, we uncovered a substantial amount of adaptive regulatory sequence evolution when contrasting differences in allele frequency between the four honey bee lineages studied herein. Our results, along with recent findings of rapid evolution of transcription-factor binding sites in social insects (31), suggests that *cis*-regulatory changes play an important role in the evolution of insect societies.

The fitness of a colony is determined by the traits of fertile members who monopolize reproduction, and by the traits of sterile workers who build and maintain the colony, feed the queen and the brood, collect food and resin, maintain temperature homeostasis, and sacrificially defend the colony against intruders (39). It is often thought that worker behavior and phenotypic plasticity provide the primary mechanism that allows insect colonies to adapt to their environment (5–7), and our population genomic data support this view. We showed that proteins with worker-biased expression have significantly higher selection coefficients relative to queen-biased proteins. We also showed that genes with known effects on worker division of labor and genes associated with nursing brood tend to be under strong positive selection in honey bees. Furthermore, we showed that genes associated with worker behavior and behavioral plasticity, based on extensive studies of brain gene expression, were enriched for signatures of adaptive *cis*-regulatory and protein evolution.

It was previously shown that genes with worker-biased brain expression have lower rates of protein evolution relative to queen-biased genes based on analysis of *Apis* and *Nasonia vitripennis* alignments (59), a result that is apparently inconsistent with our finding of higher rates of adaptive evolution of worker-biased proteins. However, our study used a more comprehensive database of caste-biased proteins (i.e., proteomic differences assayed in 26 tissues versus transcriptomic differences assayed in one tissue), included TRGs that we have shown to experience higher rates of positive selection (i.e., *Apis–Nasonia* alignments would have excluded *Apis*- and Apoidea-specific genes), and directly quantified adaptive evolution (11) [i.e., general measures of protein evolution (59) are affected by both adaptive, neutral, and nonadaptive causes (60)]. Our population genomics study strongly indicates that worker transcriptomic and proteomic phenotypes are enriched for signatures of positive selection.

Worker honey bees are effectively sterile but they can produce haploid sons in queenless colonies. Given the rarity of worker reproduction under queen-right conditions (61), the lower number of drones produced by queenless colonies relative to queen-right colonies (62), and lack of evidence showing that worker-laid drones have similar fitness as queen-laid drones, it is reasonable to assume that indirect kin-selection is mostly responsible for the adaptive evolution of worker traits. Recent theory suggests that, all other factors being equal, indirect selection on workers will be effectively weaker than direct selection on queens (63), especially when queens are polyandrous, as in *A. mellifera* (64). However, our work shows that indirect selection does not necessarily impede adaptive evolution of the worker caste, possibly because mutations in worker-biased genes tend to—on average—have higher colony-level fitness effects.

The field of genomics has greatly enriched research in sociobiology by providing knowledge on the molecular basis underlying caste differentiation and caste-specific phenotypes. Our population genomics approach allowed us to identify loci that affect fitness in honey bees, "the alleles that matter!" (65). We have shed some light on the biological and social relevance of such loci but more studies are needed to understand the molecular and phenotypic basis of adaptation in honey bees (66). We believe that the rich genomic resources provided herein will be instrumental in developing and testing mechanistic and evolutionary-explicit models of how and why social behavior evolved.

Materials and Methods

Sequencing, Alignment, and SNP Calling. Genomic DNA was extracted from each bee using a DNeasy Blood & Tissue Kit from Qiagen, and sent for Illumina Hi-Seq sequencing (50-bp reads) at Génome Québec Innovation Centre at McGill University. Each bee was sequenced in a single Hi-Seq lane. We implemented a bioinformatics pipeline to align sequencing reads, detect SNPs, and filter out highly repetitive regions of the bee genome from analysis (*SI Materials and Methods*). Overall, we were able to study genetic diversity in 227.6 Mb (~96%) of *A. mellifera*'s genome, and the sequenced *A. mellifera* workers had an average coverage depth of 38×. Five researchers manually examined over 100 kb of sequence to ensure the accuracy of our alignment and SNP calls.

Validation of SNPs. We used three datasets to validate the SNPs identified herein (NGS SNPs). (*i*) Some of the bees analyzed by us were previously used to sequence several nuclear genes using Sanger technology (23, 28, 60, 67). We compared 270 different Sanger sequences covering 169,791 bp to our NGS dataset: 97% of sequences had identical numbers of SNPs. (*ii*) We compared 1,088,415 SNPs from the reference *A. mellifera* genome (18) to NGS SNPs: 88% of the SNPs were present in our dataset, either as SNPs (82.2%) or as indel polymorphisms (5.8%). (*iii*) We also validated 85% of SNPs derived from sequencing Africanized honey bees (18). Given the large level of genetic diversity in honey bees, we do not expect to find a high (>95%) correspondence between NGS SNPs and those found in *ii* and *iii*. The large level of validation reported herein, especially when comparing NGS and Sanger sequences derived from the same bees (97% validation), indicates that the vast majority of SNP calls are accurate.

Population Structure. We used the program ADMIXTURE (68) to estimate the population origin and admixture levels of the sequenced bees. We tested K = 1–6 populations (100 times per K) assuming no prior knowledge of population origin. We randomly selected 25,000 SNPs separated by at least 5 Kb from across the genome; singleton SNPs (i.e., derived allele present in a single bee) were excluded from this analysis. We repeated this analysis with three sets of 25,000 randomly chosen SNPs to test the robustness of ADMIXTURE results.

MK Analysis. We used a Bayesian implementation of the MK test (11) to estimate the prevalence of selection acting on genes. We used perl scripts to determine if SNPs were nonsynonymous or synonymous using predictions from the bee's official gene set (OGSv3.2) (69). Divergence data were based on fixed mutations between *A. cerana* and *A. mellifera* sequences. We restricted our MK analysis to genes with sequence coverage in all African bees. We used the following measures to guard against spurious alignment of coding sequences in *A. mellifera* and noncoding sequences in *A. cerana*: (*i*) We used expression data derived from RNA sequencing of *Apis cerana* worker brains (70) to mask portions of *A. mellifera* exons that have no evidence of expression in *A. cerana*; (*ii*) we checked all coding-sequence alignments for the presence of frame-shifting indels. When we discovered a frame-shifting indel in an exon, we excluded the downstream sequence of that exon. Genes with no SNPs were excluded from analyses.

Outlier SNPs and Windows. F_{ST} was estimated for all six pairwise comparisons involving the four sampled *A. mellifera* populations following Weir and Cockerham (71) as implemented in GENEPOP v4.2 (72). Weir and Cockerham's (71) method provides accurate estimates of F_{ST} given uneven or small sample sizes (73). In each population comparison, SNPs with a minimum allele frequency <0.025 and SNPs not meeting our masking criteria were excluded from analysis. We used two independent methods to identify loci and regions with outlier levels of F_{ST}. First, we classified any SNP in the top 5% of the empirical distribution of F_{ST} as an outlier. Across our dataset, outlier SNPs were significantly differentiated based on exact G-tests (74) ($q << 10^{-8}$ after FDR correction). Second, we used a creeping-window algorithm (14) that estimates mean F_{ST} for overlapping 5-kb windows containing at least 30 SNPs. Analyses were also performed with 7- and 10-kb windows and results remained consistent across the different window sizes. To avoid estimating F_{ST} across sequence gaps, windows with SNPs spaced greater than 5-kb apart were skipped (14). For the creeping-window approach, outlier windows were statistically identified using simulation as follows: (*i*) we rescanned the genome 10 million times and randomly sampled new F_{ST} values for every SNP in a given window (14); (*ii*) windows were deemed outliers if observed average F_{ST} in a window was above the 95th percentile of the empirical distribution of expected F_{ST}, following stringent FDR correction ($q < 0.025$) (75). Within a range of overlapping windows, only the most significant window was considered an outlier. Because the two methods of detecting outlier loci were highly concordant (see text above), we used the first method for most analyses because it allowed us to precisely determine the genomic context (i.e., coding vs. noncoding) of outlier loci. All F_{ST}-based analyses were performed on each pairwise population comparison ($n = 6$) and corrected for multiple testing using FDR (76).

GO Analysis. We used the program DAVID 6.7 (29) to examine if adaptively evolving loci are enriched for specific functional annotation clusters using default parameters. We first identified the *Drosophila* homologs of positively selected bee genes using blastp match (evalue threshold 1e-10). We were able to find fly homologs for 54.3% of genes in OGSv3.2.

Bee Protein Atlas. The *Honey Bee Protein Atlas* (16) provides protein expression data in 26 tissues in queens and workers for 1,728 proteins in OGSv3.2. We examined if significantly worker-biased proteins, averaged across the different tissues (16), have different γ relative to significantly queen-biased proteins using a Wilcoxon nonparametric test. We also counted the number of cases where worker-biased proteins were enriched for nonsynonymous outlier SNPs relative to all proteins found in a given tissue; this analysis was repeated for 26 tissues and for each of the six pairwise population comparisons (a total of 156 tests). We performed a similar analysis for queen-biased genes. After first ensuring that queen-biased and worker-biased proteins did not significantly differ in length, we compared the number of significant and nonsignificant (FDR, $\alpha < 0.05$) tests of enrichment in worker-biased and queen-biased proteins using a Fisher's exact test. The *Honey Bee Protein Atlas* also provided a proteomic contrast of drones and workers. Worker-biased proteins had higher selection coefficients relative to drone-biased proteins but the number of drone-biased proteins was too small to warrant a statistical analysis.

BeeSpace Project. We obtained lists of differentially expressed genes in the brains of worker honey bees from 27 microarray experiments targeting several aspects of worker behavior associated with behavioral maturation, foraging, and aggression (17; reviewed by ref. 37). We compared the number of outlier SNPs in putative *cis*-regulatory sequences (i.e., 500-bp upstream of start codon), and the number of outlier nonsynonymous SNPs in exons, in DEGs, and non-DEGs for each of the 27 experiments. Across the experiments, DEGs were not significantly longer than non-DEGs, and thus enrichment of outlier SNPs in the exons of DEGs was not caused by differences in gene length.

Statistical Analyses and Power. All statistical analyses were performed in R (77). All comparisons were performed with nonparametric tests unless otherwise stated. FDR corrections were based on the methods of either Benjamini–Hochberg (α values reported) (76) or Storey (q values reported) (75); the latter was used when the number of statistical tests was large. We used appropriate samples sizes for estimating γ (20 haploid chromosomes from Africa) (78) and F_{ST} (18–20 haploid chromosomes per population) (73).

ACKNOWLEDGMENTS. We thank L. Packer, C. Eardley, P. De la Rua, A. Oleksa, P. Kozmus, and M. Hasselmann for providing bee samples, the staff at McGill University/Génome Québec Innovation Centre's for sequencing, K. Eilertson for help with SnIPRE (Selection inference using a Poisson random-effects model), T. Linksvayer for stimulating discussions, C. Elsik and C. Childers for facilitating submission to BeeBase, and the Honey Bee Genome Sequencing Consortium for their efforts on upgrading the genome and for providing access to gene predications. This study was supported in part by financial and logistical support from King Saud University (A.S.A. and A.A.O.); an Elia Scholarship from York University (to B.A.H.); and a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada and an Early Researcher Award from the Ontario Ministry of Research and Innovation (to A.Z.).

1. Wilson EO, Hölldobler B (2005) Eusociality: Origin and consequences. *Proc Natl Acad Sci USA* 102(38):13367–13371.
2. Darwin C (1909) *The Origin of Species* (P. F. Collier & Son, New York).
3. Hamilton WD (1964) The genetical evolution of social behaviour. I. *J Theor Biol* 7(1): 1–16.
4. Hamilton WD (1964) The genetical evolution of social behaviour. II. *J Theor Biol* 7(1): 17–52.
5. Wilson EO (1985) The sociogenesis of insect colonies. *Science* 228(4707):1489–1495.
6. Sagili RR, Pankiw T, Metz BN (2011) Division of labor associated with brood rearing in the honey bee: How does it translate to colony fitness? *PLoS ONE* 6(2):e16785.

7. Wray MK, Mattila HR, Seeley TD (2011) Collective personalities in honeybee colonies are linked to colony fitness. *Anim Behav* 81(3):559–568.
8. Strassmann JE, Page RE, Jr., Robinson GE, Seeley TD (2011) Kin selection and eusociality. *Nature* 471(7339):E5–E6, author reply E9–E10.
9. Abbot P, et al. (2011) Inclusive fitness theory and eusociality. *Nature* 471(7339):E1–E4, and author reply E9–E10.
10. Begun DJ, et al. (2007) Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* 5(11):e310.
11. Eilertson KE, Booth JG, Bustamante CD (2012) SnIPRE: Selection inference using a Poisson random effects model. *PLOS Comput Biol* 8(12):e1002806.
12. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 12(12):1805–1814.
13. Nielsen R (2005) Molecular signatures of natural selection. *Annu Rev Genet* 39: 197–218.
14. Qanbari S, et al. (2012) A high resolution genome-wide scan for significant selective sweeps: An application to pooled sequence data in laying chickens. *PLoS ONE* 7(11): e49525.
15. Hohenlohe PA, et al. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* 6(2):e1000862.
16. Chan QW, et al. (2013) Honey bee protein atlas at organ-level resolution. *Genome Res* 23(11):1951–1960.
17. Chandrasekaran S, et al. (2011) Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. *Proc Natl Acad Sci USA* 108(44): 18020–18025.
18. Weinstock GM, et al.; Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443(7114): 931–949.
19. Ruttner F (1988) *Biogeography and Taxonomy of Honeybees* (Springer, New York).
20. Whitfield CW, et al. (2006) Thrice out of Africa: Ancient and recent expansions of the honey bee, *Apis mellifera*. *Science* 314(5799):642–645.
21. Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proc Natl Acad Sci USA* 105(9):3421–3426.
22. Chávez-Galarza J, et al. (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Mol Ecol* 22(23):5890–5907.
23. Harpur BA, Minaei S, Kent CF, Zayed A (2012) Management increases genetic diversity of honey bees via admixture. *Mol Ecol* 21(18):4414–4421.
24. Arias MC, Sheppard WS (2005) Phylogenetic relationships of honey bees (Hymenoptera:Apinae:Apini) inferred from nuclear and mitochondrial DNA sequence data. *Mol Phylogenet Evol* 37(1):25–35.
25. Kotthoff U, Wappler T, Engel MS (2013) Greater past disparity and diversity hints at ancient migrations of European honey bee lineages into Africa and Asia. *J Biogeogr* 40(10):1832–1838.
26. Hartl DL, Clark AG (2007) *Principles of Population Genetics* (Sinauer Associates, Sunderland, MA), 4th Ed.
27. Bustamante CD, et al. (2005) Natural selection on protein-coding genes in the human genome. *Nature* 437(7062):1153–1157.
28. Kent CF, Issa A, Bunting AC, Zayed A (2011) Adaptive evolution of a key gene affecting queen and worker traits in the honey bee, *Apis mellifera*. *Mol Ecol* 20(24): 5226–5235.
29. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
30. Chen SD, Krinsky BH, Long MY (2013) New genes as drivers of phenotypic evolution. *Nat Rev Genet* 14(9):645–660.
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(8):1235–1247.
32. Feldmeyer B, Elsner D, Foitzik S (2014) Gene expression patterns associated with caste and reproductive status in ants: Worker-specific genes are more derived than queen-specific ones. *Mol Ecol* 23(1):151–161.
33. Johnson BR, Tsutsui ND (2011) Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* 12:164.
34. Ferreira PG, et al. (2013) Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol* 14(2):R20.
35. Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV (2013) OrthoDB: A hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res* 41 (Database issue):D358–D365.
36. Fisher RA (1930) *The Genetic Theory of Natural Selection* (Dover, New York).
37. Zayed A, Robinson GE (2012) Understanding the relationship between brain gene expression and social behavior: Lessons from the honey bee. *Annu Rev Genet* 46: 591–615.
38. Whitfield CW, et al. (2006) Genomic dissection of behavioral maturation in the honey bee. *Proc Natl Acad Sci USA* 103(44):16068–16075.
39. Winston ML (1987) *The Biology of the Honey Bee* (Harvard Univ Press, Cambridge, MA).
40. Wang Y, et al. (2010) Down-regulation of honey bee IRS gene biases behavior toward food rich in protein. *PLoS Genet* 6(4):e1000896.
41. Nelson CM, Ihle KE, Fondrk MK, Page REJ, Amdam GV (2007) The gene *vitellogenin* has multiple coordinating effects on social organization. *PLoS Biol* 5(3):e62.
42. Ament SA, et al. (2012) The transcription factor *ultraspiracle* influences honey bee social behavior and behavior-related gene expression. *PLoS Genet* 8(3):e1002596.
43. Ament SA, Wang Y, Robinson GE (2010) Nutritional regulation of division of labor in honey bees: Toward a systems biology perspective. *Wiley Interdiscip Rev Syst Biol Med* 2(5):566–576.
44. Ament SA, Corona M, Pollock HS, Robinson GE (2008) Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc Natl Acad Sci USA* 105(11):4226–4231.
45. Sullivan JP, Fahrbach SE, Robinson GE (2000) Juvenile hormone paces behavioral development in the adult worker honey bee. *Horm Behav* 37(1):1–14.
46. Amdam GV, Omholt SW (2003) The hive bee to forager transition in honeybee colonies: The double repressor hypothesis. *J Theor Biol* 223(4):451–464.
47. Amdam GV, Norberg K, Hagen A, Omholt SW (2003) Social exploitation of vitellogenin. *Proc Natl Acad Sci USA* 100(4):1799–1802.
48. Amdam GV, Norberg K, Fondrk MK, Page RE, Jr. (2004) Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proc Natl Acad Sci USA* 101(31):11350–11355.
49. Corona M, et al. (2007) Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proc Natl Acad Sci USA* 104(17):7128–7133.
50. Bar-Peled L, et al. (2013) A tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1. *Science* 340(6136):1100–1106.
51. Jindra M, Palli SR, Riddiford LM (2013) The juvenile hormone signaling pathway in insect development. *Annu Rev Entomol* 58:181–204.
52. Li M, Mead EA, Zhu J (2011) Heterodimer of two bHLH-PAS proteins mediates juvenile hormone-induced gene expression. *Proc Natl Acad Sci USA* 108(2):638–643.
53. Bernardo TJ, Dubrovsky EB (2012) The *Drosophila* juvenile hormone receptor candidates methoprene-tolerant (MET) and germ cell-expressed (GCE) utilize a conserved LIXXL motif to bind the FTZ-F1 nuclear receptor. *J Biol Chem* 287(10):7821–7833.
54. Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE (2002) Influence of gene action across different time scales on behavior. *Science* 296(5568):741–744.
55. Ben-Shahar Y, Dudek NL, Robinson GE (2004) Phenotypic deconstruction reveals involvement of manganese transporter *malvolio* in honey bee division of labor. *J Exp Biol* 207(Pt 19):3281–3288.
56. Drapeau MD, Albert S, Kucharski R, Prusko C, Maleszka R (2006) Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome Res* 16(11):1385–1394.
57. Kamakura M (2011) Royalactin induces queen differentiation in honeybees. *Nature* 473(7348):478–483.
58. Toth AL, Robinson GE (2007) Evo-devo and the evolution of social behavior. *Trends Genet* 23(7):334–341.
59. Hunt BG, et al. (2010) Sociality is linked to rates of protein evolution in a highly social insect. *Mol Biol Evol* 27(3):497–500.
60. Harpur BA, Zayed A (2013) Accelerated evolution of innate immunity proteins in social insects: Adaptive evolution or relaxed constraint? *Mol Biol Evol* 30(7):1665–1674.
61. Visscher PK (1989) A quantitative study of worker reproduction in honey bee colonies. *Behav Ecol Sociobiol* 25(4):247–254.
62. Page RE, Erickson EH (1988) Reproduction by worker honey bees (*Apis mellifera* L.). *Behav Ecol Sociobiol* 23(2):117–126.
63. Linksvayer TA, Wade MJ (2009) Genes with social effects are expected to harbor more sequence variation within and between species. *Evolution* 63(7):1685–1696.
64. Hall DW, Goodisman MAD (2012) The effects of kin selection on rates of molecular evolution in social insects. *Evolution* 66(7):2080–2093.
65. Rockman MV (2012) The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution* 66(1):1–17.
66. Barrett RDH, Hoekstra HE (2011) Molecular spandrels: Tests of adaptation at the genetic level. *Nat Rev Genet* 12(11):767–780.
67. Kent CF, Minaei S, Harpur BA, Zayed A (2012) Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *Proc Natl Acad Sci USA* 109(44):18012–18017.
68. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.
69. Elsik CG, et al. (2014) Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* 15:86.
70. Wang ZL, et al. (2012) Transcriptome analysis of the Asian honey bee *Apis cerana cerana*. *PLoS ONE* 7(10):e47954.
71. Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370.
72. Raymond M, Rousset F (1995) Genepop (v. 1.2)—Population genetics software for exact tests and ecumenicism. *J Hered* 86(3):248–249.
73. Willing EM, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE* 7(8):e42649.
74. Goudet J, Raymond M, de Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics* 144(4):1933–1940.
75. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100(16):9440–9445.
76. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—A practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57(1):289–300.
77. Team RDC (2011) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
78. Andolfatto P (2008) Controlling type-I error of the McDonald-Kreitman test in genomewide scans for selection on noncoding DNA. *Genetics* 180(3):1767–1771.