Defense Response in Brazilian Honey Bees (*Apis mellifera scutellata* × spp.) Is Underpinned by Complex Patterns of Admixture

Brock A. Harpur1,*, Samir M. Kadi2, Ricardo O. Orsi2, Charles W. Whitfield3, and Amro Zayed4

1Department of Entomology, Purdue University
2Departamento de Produção Animal, Faculdade de Medicina Veterinária e Zootecnia de Botucatu, Universidade Estadual Paulista, UNESP, Botucatu, São Paulo, Brazil
3Department of Entomology, University of Illinois at Urbana–Champaign
4Department of Biology, Faculty of Sciences, York University, Toronto, Canada

*Corresponding author: E-mail: bharpur@purdue.edu

Accepted: 13 June 2020

Abstract

In 1957, an invasive and highly defensive honey bee began to spread across Brazil. In the previous year, Brazilian researchers hoped to produce a subtropical-adapted honey bee by crossing local commercial honey bees (of European origin) with a South African honey bee subspecies (*Apis mellifera scutellata*; an A-lineage honey bee subspecies). The resulting cross—African hybrid honey bees (AHBs)—escaped from their enclosure and spread through the Americas. Today, AHB is the most common honey bee from Northern Argentina to the Southern United States. AHBs are much more likely to sting nest intruders than managed European-derived honey bee colonies. Previous studies have explored how genetic variation contributes to differences in defense response between European-derived honey bee and AHB. Although this work demonstrated very strong genetic effects on defense response, they have yet to pinpoint which genes influence variation in defense response within AHBs, specifically. We quantified defense response for 116 colonies in Brazil and performed pooled sequencing on the most phenotypically divergent samples. We identified 65 loci containing 322 genes that were significantly associated with defense response. Loci were strongly associated with metabolic function, consistent with previous functional genomic analyses of this phenotype. Additionally, defense-associated loci had nonrandom and unexpected patterns of admixture. Defense response was not simply the product of more A-lineage honey bee ancestry as previously assumed, but rather an interaction between A-lineage and European alleles. Our results suggest that a combination of A-lineage and European alleles play roles in defensive behavior in AHBs.

Key words: honey bee, genomics, admixture, ancestry, behavior.

Significance

In 1957, an invasive and highly defensive honey bee began to spread across Brazil. In the previous year, Brazilian researchers hoped to produce a subtropical-adapted honey bee by crossing local commercial honey bees (of European origin) with a South African honey bee subspecies (*Apis mellifera scutellata*; an A-lineage honey bee subspecies). The resulting cross—African hybrid honey bees (AHBs)—escaped from their enclosure and spread through the Americas. AHBs are much more likely to sting nest intruders than managed European-derived honey bee colonies. We identified 65 loci containing 322 genes that were significantly associated with defense response. Loci were strongly associated with metabolic function, consistent with previous functional genomic analyses of this phenotype. Additionally, defense-associated loci had nonrandom and unexpected patterns of admixture. Defense response was not simply the product of more A-lineage honey bee ancestry as previously assumed, but rather an interaction between A-lineage and European alleles.
**Introduction**

In 1956, Brazilian researchers hoped to produce a subtropical-adapted honey bee by crossing commercial honey bees with a South African honey bee subspecies (*Apis mellifera scutellata*; A-lineage) (Kerr 1957; Nogueira-Neto 1964; Winston 1987, 1992). At the time, Brazil’s commercial stock was a mixed population that originated from at least two genetically distinct European sources: honey bee subspecies of the M-lineage (e.g., *Apis mellifera melissae* and *Apis mellifera iberiensis*) and C-lineage (e.g., *Apis mellifera ligustica* and *Apis mellifera carnica*) (Kerr 1957; Sheppard 1989a, 1989b; Crane 1999). Twenty-six *A. m. scutellata* queens were brought into Brazil from South Africa for breeding trials. It was shortly after this introduction that some of these colonies swarmed, hybridized with local stocks, and began spreading across Brazil (reviewed by Michener [1975], Spivak et al. [1991], and Winston [1992]). The escaped colonies rapidly reproduced and spread across South America, making their way to Central America by 1982, and to the United States by 1990. Since their escape and spread, they have replaced the European-derived honey bee (EHB) population in Brazil (De Jong 1996; Gonçalves 2004; Francoy et al. 2009). These invasive colonies have been called “killer honey bees” and are typically referred to as “Africanized honey bees.” The former for media attention and their extreme defensiveness and the latter because *A. m. scutellata* are an African honey bee subspecies. We have chosen to use the term African hybrid honey bee (AHB). We note that there are dozens of honey bee subspecies in Africa (Ruttner 1988) and research to date has demonstrated that *A. m. scutellata* is the primary contributor of A-lineage ancestry to AHB (Pinto et al. 2004; Francoy et al. 2009; Nelson et al. 2017).

AHBs are phenotypically and genetically distinct from other honey bees in the Americas and around the world (Collins et al. 1982; Lobo et al. 1989; Winston 1992; Martin and Medina 2004; Kadri et al. 2016; Nelson et al. 2017). The most readily observable phenotypic difference is their extreme defense response. In a standardized defense-response assay, AHBs respond faster and in greater numbers to perceived threats than European honey bees resulting in as many as eight times more stings to a perceived threat (Collins et al. 1982). High defensiveness persists across their range (Schneider et al. 2004) but see Rivera-Marchand et al. (2012). Genetically, AHBs are most similar to A-lineage honey bees (Clarke et al. 2002; Whitfield et al. 2006; Zayed and Whitfield 2008; Rivera-Marchand et al. 2012; Chapman et al. 2015; Kadri et al. 2016). However, portions of the genome still share ancestry with (i.e., have high genetic similarity to) European C- and M-lineage subspecies (Clarke et al. 2002; Whitfield et al. 2006; Zayed and Whitfield 2008; Rivera-Marchand et al. 2012; Chapman et al. 2015; Nelson et al. 2017). This remaining European ancestry is predicted to be associated with phenotypic variation within AHB (Avalos et al. 2017; Nelson et al. 2017).

Decades of research effort has been spent investigating honey bee defense response, especially in AHBs. Defense response—the suite of behaviors expressed by honey bees that ultimately lead to the release of their sting (Breed et al. 2004; Nouvian et al. 2016)—is a highly complex behavior with substantial interpopulation variation (Hunt et al. 1998; Breed et al. 2004; Uzunov et al. 2014). Some of this variation can be explained by environmental conditions inside and outside the colony (reviewed by Nouvian et al. [2016]) and see Rittschof and Robinson (2013) and Rittschof, Coombs, et al. (2015). Genetics also explains a portion of this variation, especially between AHB and European-derived populations in common apiaries (Kerr 1974; Stort 1975a, 1975b; Hunt et al. 1998, 1999, 2007). Narrow-sense heritability for defense response ranges from 0.14 to 0.43 (Brascamp et al. 2016).

The large fraction of phenotypic variation that can be explained by genetics has been the driving force behind quantitative trait loci (QTL) and functional genetic studies to date. Using crosses between AHB and EHB, there have been 15 broad QTL described for various components of the honey bee’s defense response and the production of honey bee alarm pheromone compounds (e.g., Hunt et al. 1998, 1999). Additional gene expression studies using AHB and EHB have found core sets of genes consistently up- or downregulated in defensive honey bees. For example, in defensive worker honey bees or those exposed to alarm pheromone, oxidative phosphorylation pathway transcripts are downregulated and glycolytic pathway transcripts are upregulated (Chandrasekaran et al. 2011; Li-Byarlay et al. 2014; Rittschof, Grozinger, et al. 2015). One particularly surprising finding from all the work to date has been the role of parent-of-origin effects on defense response. More aggressive colonies result from the cross of EHB queens mated to AHB drones than the inverse cross (Breed et al. 2004; Gibson et al. 2015). Recent observations have suggested this may be the result of parent-of-origin gene expression. Two gene clusters within major-effect QTLs for defense response and alarm pheromone production have strong maternal-expression bias in hybrid (EHB × AHB) worker brains (Gibson et al. 2015).

The functional studies to date have been invaluable to our understanding of the genetics of defense response; however, no study has yet explored how defense response varies specifically within AHB. With our detailed understanding of the proximate mechanisms of defense response, AHB provides a unique opportunity to explore how the introgression of long-separated populations (here, European and African honey bee lineages) and can influence phenotypic variation in an invasive population (Rius and Darling 2014). In the case of AHB, a gene flow event resulted in a hybridized population with a fitness advantage over local populations and with clear differences in phenotype (Winston 1992). We made use of a
recently curated population genomics data set for AHB (Kadri et al. 2016) to identify which loci are associated with defense response within AHB and quantify how variation in intraspecific genetic admixture at these loci could contribute to variation in defense response.

Materials and Methods

Defense-Response Assay and Sampling

We quantified the defense response of 116 Brazilian AHB colonies from four apiaries within São Paulo State (fig. 1A, supplementary table S1, Supplementary Material online, and Kadri et al. 2016). These apiaries were within approximately 15 km of each other. Each colony began as a wild-caught swarm early in the beekeeping season. Two weeks prior to phenotyping, we standardized colonies within Langstroth boxes to consist of seven brood frames and three nectar and pollen frames.

We measured the defense response of each colony using the Black Suede Ball test (Stort 1975a). The test is performed by gently swinging a suede ball in front of the colony entrance. The ball stimulates a defense response in bees at the colony entrance and their recruits additional bees to sting the ball. After 1 min of swinging, the ball is removed and sealed in a plastic bag. Defense response is scored as the number of stings found in the suede ball. We repeated this test three times over three days in April and May 2014 and averaged the number of stings across the tests. We tested all 116 colonies on each testing day. From each colony, we collected at least 15 workers from the brood chamber in 95% ethanol and stored them at −80°C until DNA extraction.

Genome Sequencing, Alignment, and Single-Nucleotide Polymorphism Calling

From the sample of 116 colonies, we extracted high-quality genomic DNA from the 15 most- and 15 least-defensive colonies (top and bottom 10% of the data; fig. 1). A single DNA extraction was performed for each colony by pooling together one-quarter of a worker’s thorax from each of 12 workers into a single 2-ml tube. We then used a Mag-Bind Blood DNA kit (Omega Biotek Store) to extract DNA. Each pool was sequenced with Illumina Hi-Seq 2500. The resulting data set and detailed computational methods are available as an open-access resource (Kadri et al. 2016) and the raw data are available on NCBI SRA (PRJNA324081). In brief, we aligned the reads for each colony individually to AMEL_v4.5 using BWA v0.7.5 (Li and Durbin 2010) and STAMPY v1.0.21 (Lunter and Goodson 2011) and jointly called single-nucleotide polymorphisms (SNPs) using VARSCAN v2.3.7 (Koboldt et al. 2009) and GATK UnifiedGenotyper (DePristo et al. 2011). All alignment and SNP calling were performed jointly on the high-defense and low-defense cohorts.

![Fig. 1](https://example.com/fig1.png)

Fig. 1.—(A) Histogram of average sting response (average number of stings per minute) for 116 Brazilian AHB colonies. We sequenced the genomes of the 15 most- and least-defensive colonies (fewer than 39 stings or >87 stings in a minute). We repeated this assay on ten non-AHB colonies within Canada and present the median response for comparison. (B) Distribution of \( F_{st} \) between highly and less-defensive colonies in Brazil. We observed an excess of sites with \( F_{st} \) exceeding values observed in simulation (red line).
Differentiated Sites
Sites associated with defense response are expected to have significant differences in allele frequency between the high- and low-defense cohorts. To identify such sites, we calculated the pairwise fixation index ($F_{st}$) between the two cohorts and the difference in allele frequency at each site using POOPOOLATION 2 (Kofler et al. 2011). We repeated this calculation with a permuted data set consisting of two cohorts that were built by sampling from all sequenced colonies without replacement.

To reduce noise in estimates of $F_{st}$ across the genome, we calculated the median $F_{st}$ of bins of 301 SNPs running along the genome. Any of these windows which had a median $F_{st} > 99.95\%$ of $F_{st}$ values (median $F_{st}$ window >0.0063) and contained at least three outlier $F_{st}$ SNPs ($F_{st} > 0.032$; greater than observed in our simulated data set) were deemed to be highly divergent and considered a candidate for defense response. None of these criteria was ever observed together in a single window of our permuted data set (see Results).

To further validate our cutoffs, we used the R package HardyWeinberg v1.6.1 to sample genotypes at 1,000,000 sites from a multinomial distribution with average allele frequencies of $P = 0.1–0.9$ independently from two populations each of 180 diploids (12 diploid workers from 15 colonies). We then estimated $F_{st}$ at each site between each population and compared these estimates with our observed single-site estimates of $F_{st}$ between highly and less-defensive colonies (see Results).

Estimating Local Ancestry
To estimate levels of A- or M-lineage across the genomes of each AHB sample, we used ANCESTRY_HMM (Corbett-Detig and Nielsen 2017). This method uses a hidden Markov model to estimate local ancestry within samples of arbitrary ploidy. We extracted all sites that were at least 5 kb apart and estimated the recombination rate between these in the ancestral genome using a recent honey bee recombination map (Liu et al. 2015). We repeated this calculation with a permuted data set consisting of two cohorts that were built by sampling from all sequenced colonies without replacement.

To reduce noise in estimates of $F_{st}$ across the genome, we calculated the median $F_{st}$ of bins of 301 SNPs running along the genome. Any of these windows which had a median $F_{st} > 99.95\%$ of $F_{st}$ values (median $F_{st}$ window >0.0063) and contained at least three outlier $F_{st}$ SNPs ($F_{st} > 0.032$; greater than observed in our simulated data set) were deemed to be highly divergent and considered a candidate for defense response. None of these criteria was ever observed together in a single window of our permuted data set (see Results).

Between cohorts, we estimated the difference in M-lineage ancestry at each site as $D_m = \text{mean}(\text{M-lineage ancestry in highly defensive colonies}) - \text{mean}(\text{M-lineage ancestry in less-defensive colonies})$. Where $D_m < 0$ there is more M-lineage ancestry at a site in the less-defensive colonies (i.e., more A-lineage ancestry in highly defensive colonies). To estimate significant differences in the proportions of A- and M-lineage ancestry between highly and less-defensive colonies, we used permutation. From the entire data set, we randomly sampled two sets of 15 colonies without replacement. For each site in the genome of this sample, we estimated the mean M-lineage ancestry and $D_m$. We repeated this protocol for 10,000 permutations. We deemed any significant $D_m$ to be one with a two-sided probability of $<0.001$. This corresponded to any site with $|D_m| \geq 0.08$, or an 8% difference in ancestral allele frequency between highly and less-defensive colonies.

Remapping QTL Regions and Genes with Reported Defense-Response Association
We remapped six previously reported defense-response QTL (Hunt et al. 1998, 1999, 2007; Guzman-Novoa et al. 2002; Lobo et al. 2003; Gibson et al. 2015; Harpur 2020). For each QTL, we took all reported marker sequences within the significant QTL regions and used BlastN against the honey bee genome (E-value $<1e^{-5}$). We extracted these regions and added 50 kb to either side of the QTL to account for potential errors in conversion.

Statistics and Gene Ontology Analyses
To determine if candidate genes were functionally enriched, we performed hypergeometric tests with DAVID 6.8 (Huang et al. 2009) against Gene Ontology and KEGG pathway terms using Apis mellifera gene calls against a background of all genes in the honey bee genome. We exported any result with $P < 0.05$. All tests were performed in R v 3.3.2 (R Core Team 2010) and were parametric unless otherwise stated.

Results
Identifying Defense-Response-Associated Sites
We observed no significant differences in the mean defense response among apiaries (ANOVA; $F_{3,112} = 0.13; P = 0.943$) and selected the 15 most- and least-defensive colonies for genome sequencing (see Materials and Methods and supplementary table S1, Supplementary Material online). Assuming defense response is heritable (Hunt et al. 1998, 1999), mutations that influence defensiveness should vary in frequency between the highly and less-defensive colonies. Conversely, regions of the genome that are not associated with defense response should have relatively low levels of genetic differentiation. We found that the highly and less-defensive cohorts exhibited very little genetic differentiation at most sites across the genome. The average $F_{st}$ between the two cohorts was $0.0065 \pm 0.0086$ SD (fig. 1B). Compared with both a permuted data set and a simulated data set, there was significantly more genetic differentiation between highly and less-defensive colonies than expected, particularly at the high end of the $F_{st}$ distribution. In a simulated data set of 1,000,000 independent sites, we never observed $F_{st} > 0.032$ (fig. 1B). Both data sets had significantly lower $F_{st}$ compared with the observed data set overall ($F_{st}$ permuted $= 0.0019; F_{st}$ simulated $= 0.0013; P < 0.00001$). There were more SNPs with high
differentiation in the observed set compared with the permuted: We only saw 108 SNPs with $F_{st} > 0.046$ compared with 7,237 in the observed data.

Across the genome, we identified 65 genomic loci containing 322 genes with high levels of genetic differentiation (see Materials and Methods) between the highly and less-defensive cohorts (fig. 2 and supplementary tables S2 and S3, Supplementary Material online; hereafter called defense-associated loci). On average, there was a 7.1% difference in allele frequency between the two cohorts in these regions with an average of 22% for the most differentiated SNPs within these regions. This difference is significantly higher than any other similar-sized set of SNPs chosen at random from the AHB genome (Permutation test $N = 1,000,000$; mean difference $= 0.052$; $P < 0.001$).

Defense-associated loci overlapped with two previously reported QTL for defense response: an alarm pheromone QTL on chromosome 12 and a defense-response QTL on chromosome 3 (Hunt et al. 1998, 1999; Guzman-Novoa et al. 2002; Lobo et al. 2003; Gibson et al. 2015). There were additional significant windows within at least 1 Mb of each of the five reported defense-response QTL. Taken together, our data provide evidence that the regions of the genome we have identified quantitatively contribute to defense response in AHB.

Defense-Associated Regions Are Differentially Admixed

Next, we asked if highly defensive colonies are more closely related to the M- or A-lineages at defense-associated loci relative to less-defensive colonies. To do this, we made use of a recently developed method to estimate ancestral proportions across the genome (admixture mapping) in pooled-sequencing data (Corbett-Detig and Nielsen 2017). This admixture mapping procedure confirmed that, on average, AHBs share 86% of their ancestry with A-lineage (supplementary fig. S1, Supplementary Material online) (Weinstock et al. 2006) with the remainder largely made up of M-lineage. We found only 1.5% C-lineage ancestry remaining within S~Paulo State’s AHBs, consistent with previous findings for this area (Clarke et al. 2002; Whitfield et al. 2006; Zayed and Whitfield 2008; Nelson et al. 2017). We also support previous findings of elevated M-lineage ancestry on a single regions of chromosome 11 (supplementary fig. S1, Supplementary Material online) (Nelson et al. 2017).
We found no significant evidence of a genome-wide difference in C-lineage ancestry between highly and less-defensive cohorts across all sites ($t = -0.68; P = 0.49$). Repeating this analysis within only defense-associated regions of the genome (as identified above), we again found no significant evidence of different levels of C-lineage ancestry between the two cohorts ($t = 1.85; P = 0.064$). When we compared the level of M-lineage ancestry between the two cohorts, we found that less-defensive colonies (mean $M = 14.1\%$) had a slightly, but a significantly higher level of M-lineage ancestry genome-wide when compared with highly defensive colonies (mean $M = 13.8\%; t = 3.37$; $P = 0.004$).

**Fig. 3.**—Average fixation index ($F_{st}$) between highly and less-defensive AHB colonies in Brazil within previously reported QTL. Red boxes are previously identified QTL (named above). Approximate location on each chromosome is depicted below as a percentage of the total chromosome. Significant regions are starred.
All defense-associated regions had evidence of differential admixture (fig. 4); however, only two were significant and were cases in which there was more M-lineage ancestry in more defensive colonies.

Defense-Associated Genes Have Metabolic Function

We next determined if the set of 322 candidate genes we identified have been previously associated with defense response in gene expression studies. Several recent studies identified genes and gene networks that are differentially expressed in the brains of defensive honey bees (e.g., guards, soldiers, or bees exposed to alarm pheromone) relative to corresponding controls (Alaux et al. 2009; Chandrasekaran et al. 2011) reviewed by Zayed and Robinson (2012). As well, a recent study identified a core set of genes whose expression is consistently associated defense response across Animalia (Rittschof et al. 2014). We compared our set of 322 genes with each of these sets. Twenty genes overlapped with one or more of the studies above (supplementary table S2, Supplementary Material online) and we find no significant evidence of overrepresentation of our candidate genes within these studies (Fisher Exact test, \( P > 0.1 \) for all comparisons).

We found evidence of common functional categories underpinning defense response in AHBs. Previous studies discovered that the brain metabolic activity of a defensive honey bee is shifted from oxidative phosphorylation to glycolysis (Alaux et al. 2009; Chandrasekaran et al. 2015). We found a slight but significant enrichment of our candidate genes within three metabolic pathways: “Ribosome biogenesis” (KEGG PATHWAY; ame03008), “Galactose metabolism” (KEGG PATHWAY; ame00052), and “Starch and sucrose metabolism” (ame00500; Hypergeometric test; \( P < 0.05 \); supplementary table S4, Supplementary Material online).

Genes involved in sugar metabolism have previously been shown to influence the expression of defense response (Alaux et al. 2009; Li-Byarlay et al. 2014; Chandrasekaran et al. 2015). We found that a major regulator of glycolysis, hexokinase2 (GB47079) was among the set of genes we identified with high \( F_{st} \) between the two cohorts (figs. 2 and 4 and supplementary table S2, Supplementary Material online). We found that this gene contained three SNPs with \( F_{st} > 99.9\% \) of all values in the genome between high and low-defense cohorts—an average of 21% difference in allele frequency. All of these SNPs fell within introns (fig. 3). This gene was within a region that had slightly but significantly (average of 3% different) more A-lineage ancestry in more highly defensive colonies (mean \( F_{st} = 0.25 \) relative to the rest of the genome (genomic mean = 0.21; fig. 5)(Harpur et al. 2014).

A Gene Cluster Is Associated with Defense Response and Differentially Admixed

Two gene clusters within QTLs on chromosomes 3 and 12 have previously been shown to be expressed in a maternally biased direction in crosses between AHB and EHB (Gibson et al. 2015). One of our putative defense-associated regions overlaps with the cluster on chromosome 3 (figs. 2 and 6). This cluster contained a region of high \( F_{st} \) between highly defensive and less-defensive cohorts and more defensive colonies had significantly more M-lineage ancestry (mean \( M = 0.17 \) relative to less-defensive colonies \( M = 0.11; P < 2.2\times10^{-16} \); fig. 6). In total, this accounted for nine genes with both high \( F_{st} \) and significant differences in ancestry within this cluster. We did not find the same patterns on the cluster in chromosome 12.
Discussion

Beekeepers have intentionally crossed long-separated and highly differentiated honey bee subspecies for centuries (Langstroth 1865; Ruttner 1988; Whitfield et al. 2006; Harpur et al. 2014; Rius and Darling 2014; Byatt et al. 2016). In doing so, they have created seminaturalized experimental genetic populations. We suggest that these introduced populations provide a useful opportunity to explore how admixture contributes to phenotypic variation and evolution (Rius and Darling 2014). Defense response, in particular, provides an exciting avenue to explore these questions as it has a relatively well-characterized genetic basis (Hunt et al. 1998, 1999) and varies among honey bee populations (Hunt et al. 1998; Breed et al. 2004; Uzunov et al. 2014) with AHBs being most defensive in North America (Collins et al. 1982; Breed et al. 2004; Gibson et al. 2015).

The elevated defense response in AHB relative to other genotypes in North and South America is a direct result of A-lineage introgression (Kerr 1957, 1974; Kerr and Bueno 1970; Stort 1975a, 1975b). However, we have provided some subtlety to this observation: The most defensive colonies in our study had both more A-lineage ancestry genome-wide and significantly higher levels of M-lineage ancestry within specific defense-response QTLs. The mechanism through which M-lineage alleles contribute to defense response has yet to be uncovered. Both the A- and M-lineages are noted as being defensive (Fletcher 1978; Pinto et al. 2014) and may vary in the behaviors ultimately leading to sting release (Breed et al. 2004). A combination of alleles from both genetic backgrounds at defense-associated loci may lead to a higher defense response than alleles strictly from one or the other genetic background. Such cases of transgressive introgression are frequently reported in hybrid plant populations (Rieseberg et al. 1999; Goulet et al. 2017).

How M- and A-lineage alleles interact to influence defense response is an important remaining question. The clearest example of how M-lineage alleles might contribute to defense response in AHBs can be found on chromosome 3. This region was highly differentiated between highly and less-defensive colonies, had higher M-lineage ancestry in more highly defensive colonies, overlapped with a previous QTL (sting-2) for defense response (Hunt et al. 1998, 1999), and a cluster of 12 genes whose expression is maternally biased in hybrid AHB × EHB worker brains (Gibson et al. 2015). On the latter point, hybrid workers resulting from a cross between an EHB queen and an AHB drone are more defensive than the inverse cross (Breed et al. 2004; Guzman-Novoa et al. 2005; Gibson et al. 2015). In the brains of adult hybrid workers from this cross,
the genes within *sting-2* have a strong expression bias toward maternally inherited EHB gene copies versus paternally inherited-AHB copies (Gibson et al. 2015). We found evidence that this locus is associated with defense response in AHB with the most defensive colonies having higher M-lineage ancestry at this locus. It is still unclear how genes expressed at this locus contribute variation to the expression of defense response. However, given the evidence of association in multiple data sets and the presence of at least two putative transcription factors (Gibson et al. 2015), it seems a worthwhile candidate for future functional analysis.

The role of sugar metabolism in defense response has been established for some time (Alaux et al. 2009; Li-Byarlay et al. 2014; Chandrasekaran et al. 2015). Our associated genes contained an enriched set involved in sugar metabolism (supplementary table S1, Supplementary Material online) including a major regulator of glycolysis, *hexokinase2* (GB47079). Highly defensive bees (A-lineage or AHB) and bees exposed to alarm pheromone have elevated whole-body metabolic rates—a measure of oxidative phosphorylation—and higher rates of glycolysis in the brain (Alaux et al. 2009; Chandrasekaran et al. 2015; reviewed by Harrison and Fewell [2002], Rittschof, Grozinger et al. [2015], and Southwick et al. [1990]). Events that lead to defense response in honey bees are associated with a hypoxia-like brain state (Rittschof and Robinson 2013). During hypoxia, glycolysis is initiated in part by *hexokinase2* (Semenza 2007; Wolf et al. 2011). At the level of the neuron, this shift has been suggested to result in changes in excitability which may enhance responsiveness to cues that induce defense response (Juge et al. 2010; Li-Byarlay et al. 2014; Chandrasekaran et al. 2015; Valdebenito et al. 2016). We suggest that functional analysis of AHB and EHB variants of GB47079 may prove fruitful to our understanding of the link between metabolism and defense response.

Our genomic approach allowed us to identify 65 loci within the AHB genome that have significant levels of genetic differentiation between the most-defensive and least-defensive colonies and have evidence of differential admixture. We have some confidence in the candidates we identified because 1) they fall within or near previously identified QTLs, 2) our candidates have molecular functions that are consistent with their involved in defensive behavior based on functional genomic studies of defensive bees, and 3) some of our candidates overlapped with regions of the honey bee genome that exhibit imprinting in association with defensive behavior and with gene whose expression had been previously associated with defensiveness. We also found some evidence of differential admixture contributing to extreme defensiveness in AHB. This work along with previous studies on AHB in Brazil and Puerto Rico are beginning to demonstrate the phenotypic and evolutionary consequences of intraspecific genetic admixture in honey bee populations (Avalos et al. 2017; Nelson et al. 2017). Future studies exploring variation in defense response, or any phenotype in North and South American honey bees, should consider the potentially important role that admixture plays in phenotypic diversity (Rius and Darling 2014). Honey bees in the Americas originate from at least three highly divergent ancestral lineages and relatedness to these lineages varies substantially across the honey bee’s introduced range (Whitfield et al. 2006; Harpur et al. 2015; Calfee et al. 2020). This variation likely contributes to observed phenotypic differences among honey bee populations (Avalos et al. 2017), could impact the power of genome-wide association studies (Hellwege et al. 2017), and cause variation in which variants are associated with phenotypes of interest (Sirugo et al. 2019).
Finally, the pooled-sequencing approach we used here is particularly useful for identifying genomic variation underpinning colony-level traits such as nest defense. As a superorganism, honey bee colonies are composed of thousands of individuals composed of up to 20 different patrilines. The interactions between individuals within a colony can have drastic influences on colony-level phenotypes such as aggression (Rittschof, Coombs et al. 2015). By creating a pooled “colony genome,” we could use within-colony allele frequencies to look at the ultimate expression of the phenotype across the colony as a whole. This procedure should be very useful in future iterations of association mapping within social insects.

**Supplementary Material**

**Supplementary data** are available at Genome Biology and Evolution online.

**Acknowledgments**

This study was funded by 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.). S.M.K. was supported by a scholarship from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP Process 2014/10150-2). B.A.H. would like to thank the members of the Purdue Bee Lab, Joshua Gibson, Greg Hunt, Ernesto Guzman, Gard Otis, and Arian Avalos for their informative discussions on defense response in honey bees.

**Literature Cited**

Alaux C, et al. 2009. Honey bee aggression supports a link between gene regulation and behavioral evolution. Proc Natl Acad Sci U S A. 106(36):15400–15405.

Avalos A, et al. 2017. A soft selective sweep during rapid evolution of gentle behaviour in an Africanized honeybee. Nat Commun. 8(1):1550.

Brascamp DW, William A, Boigenzahn C, Bijma P, Veerkamp RF. 2016. Heritabilities and genetic correlations for honey yield, gentleness, calmness and swarming behaviour in Austrian honey bees. Apidologie 47(6):739–748.

Breed MD, Guzman-Novoa E, Hunt GJ. 2004. Defensive behavior of honey bees: organization, genetics, and comparisons with other bees. Annu Rev Entomol. 49(1):271–288.

Byatt MA, Chapman NC, Latty T, Oldroyd BP. 2016. The genetic consequences of the anthropogenic movement of social bees. Insect Soc. 63(1):15–24.

Calfee E, Agra MN, Palacio MA, Ramírez SR, Coop G. 2020. Selection and hybridization shaped the Africanized honey bee invasion of the Americas. bioRxiv: 2020.03.17.994632. doi:10.1101/2020.03.17.994632.

Chandrasekaran S, et al. 2011. Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. Proc Natl Acad Sci U S A. 108(44):18020–18025.

Chandrasekaran S, et al. 2015. Aggression is associated with aerobic glycolysis in the honey bee brain. Genes Brain Behav. 14(2):158–166.

Chapman NC, et al. 2015. A SNP test to identify Africanized honeybees via proportion of “African” ancestry. Mol Ecol Resour. 15(6):1346–1355.

Clarke KE, Rinderer TE, Franck P, Quezada-Euán JG, Oldroyd BP. 2002. The Africanization of honeybees (Apis mellifera L.) of the Yucatan: a study of a massive hybridization event across time. Evolution 56:1462–1474.

Collins AM, Rinderer TE, Harbo JR, Bolten AB. 1982. Colony defense by Africanized and European honey bees. Science 218(4567):72–74.

Corbett-Detig R, Nielsen R. 2017. A hidden Markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. PLoS Genet. 13(1):e1006529.

Crane E. 1999. The world history of beekeeping and honey hunting. New York: Routledge.

De Jong D. 1996. Africanized honey bees in Brazil, forty years of adaptation and success. Bee World 77(2):67–70.

DePristo MA, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43(5):491–498.

Fletcher DJ. 1978. The African bee, Apis mellifera adansonii, in Africa. Annu Rev Entomol. 23(1):151–171.

Franczy TM, et al. 2009. Morphometric and genetic changes in a population of Apis mellifera after 34 years of Africanization. Genet Mol Res. 8(2):709–717.

Gibson JD, Arechavaleta-Velasco ME, Tsuruda JM, Hunt GJ. 2015. Biased allele expression and aggression in hybrid honeybees may be influenced by inappropriate nuclear-cytoplasmic signaling. Front Genet. 6:343.

Gonçalves LS, editor. 2004. Proceedings of the 8th IBRA International Conference on Tropical Bees and VI Encontro sobre Abelhas; Ribeirão Preto, Brasil.

Goulet BE, Roda F, Hopkins R. 2017. Hybridization in plants: old ideas, new techniques. Plant Physiol. 173(1):65–78.

Guzman-Novoa E, Hunt GJ, Uribe JL, Smith C, Arechavaleta-Velasco ME. 2002. Confirmation of QTL effects and evidence of genetic dominance of honeybee defensive behavior: results of colony and individual behavioral assays. Behav Genet. 32:95–102.

Guzman-Novoa E, et al. 2005. Paternal effects on the defensive behavior of honeybees. J Hered. 96(4):376–380.

Harpur BA. 2020. Hunt honey bee markers. In: Data Dryad: https://doi.org/10.5061/dryad.ns1rn8ppp.

Harpur BA, et al. 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. Proc Natl Acad Sci U S A. 111(7):2614–2619.

Harpur BA, et al. 2015. Assessing patterns of admixture and ancestry in Canadian honey bees. Insect Soc. 62(4):479–489.

Harrison JF, Fewell JH. 2002. Environmental and genetic influences on flight metabolic rate in the honey bee, Apis mellifera. Comp Biochem Physiol Part A Mol Integr Physiol. 133(2):323–333.

Hellwege JN, et al. 2017. Population stratification in genetic association studies. Curr Protoc Hum Genet. 95(1):1.22.1–1.22.23.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 4(1):44–57.

Hunt GJ, Collins AM, Rivera R, Page RE, Guzmán-Novoa E. 1999. Quantitative trait loci influencing honeybee alarm pheromone levels. J Hered. 90(5):585–589.

Hunt GJ, Guzman-Novoa E, Fondrk MK, Page RE. 1998. Quantitative trait loci for honey bee stinging behavior and body size. Genetics 148(3):1203–1213.

Hunt GJ, et al. 2007. Behavioral genomics of honeybee foraging and nest defense. Naturwissenschaften 94(4):247–267.

Juge N, et al. 2010. Metabolic control of vesicular glutamate transport and release. Neuron 68(1):99–112.

Kadri SM, Harpur BA, Orsi RO, Zayed A. 2016. A variant reference data set for the Africanized honeybee. Sci Data 3(1):160097.
