

Adaptive evolution of a key gene affecting queen and worker traits in the honey bee, *Apis mellifera*

CLEMENT F. KENT, AMER ISSA, ALEXANDRA C. BUNTING and AMRO ZAYED

Department of Biology, York University, 4700 Keele Street, Toronto, Ontario, Canada

Abstract

The vitellogenin egg yolk precursor protein represents a well-studied case of social pleiotropy in the model organism *Apis mellifera*. Vitellogenin is associated with fecundity in queens and plays a major role in controlling division of labour in workers, thereby affecting both individual and colony-level fitness. We studied the molecular evolution of vitellogenin and seven other genes sequenced in a large population panel of *Apis mellifera* and several closely related species to investigate the role of social pleiotropy on adaptive protein evolution. We found a significant excess of nonsynonymous fixed differences between *A. mellifera*, *A. cerana* and *A. florea* relative to synonymous sites indicating high rates of adaptive evolution at vitellogenin. Indeed, 88% of amino acid changes were fixed by selection in some portions of the gene. Further, vitellogenin exhibited hallmark signatures of selective sweeps in *A. mellifera*, including a significant skew in the allele frequency spectrum, extreme levels of genetic differentiation and linkage disequilibrium. Finally, replacement polymorphisms in vitellogenin were significantly enriched in parts of the protein involved in binding lipid, establishing a link between the gene's structure, function and effects on fitness. Our case study provides unequivocal evidence of historical and ongoing bouts of adaptive evolution acting on a key socially pleiotropic gene in the honey bee.

Keywords: evolution, pleiotropy, positive selection, social behaviour, vitellogenin

Received 6 July 2011; revision received 8 August 2011; accepted 16 August 2011

Introduction

Because of its association with fecundity and reproduction, the study of vitellogenin has special significance for understanding the evolution of division of labour in social insects. Vitellogenin encodes the major egg yolk precursor proteins found in most oviparous animals (Chen *et al.* 1997; Sappington & Raikhel 1998; Tufail & Takeda 2008). In insects, vitellogenin proteins are synthesized in the fat body, secreted into the haemolymph and sequestered by oocytes, thereby providing the future embryo with protein, lipid and carbohydrates (reviewed by Tufail & Takeda 2008). Vitellogenin (Vg) has been extensively studied in the model eusocial organism the honey bee (*Apis mellifera*), where it has multiple phenotypic effects in both queens and workers (Amdam *et al.* 2003). Vg is linked to fecundity in the

honey bee: Vg expression is higher in the reproductive queens in comparison with the effectively sterile workers (Barchuk *et al.* 2002; Amdam *et al.* 2003; Corona *et al.* 2007), and circulating Vg titres match queen egg-laying rate (Engels 1974). Vg also affects lifespan and oxidative stress resistance in both queens and workers (Amdam *et al.* 2005a; Corona *et al.* 2007). Further, Vg affects several important traits in worker bees, either directly or through its repressive effect on juvenile hormone (Amdam & Omholt 2003). For example, nurse bees use Vg to make 'brood food'—secretions fed to larvae (Amdam *et al.* 2003). Additionally, Vg regulates worker division of labour—a major component of colony fitness (Sagili *et al.* 2011)—by pacing workers' transition from nursing to foraging (Nelson *et al.* 2007). The age at onset of foraging varies between different behaviourally distinct honey bee subspecies (Brillet *et al.* 2002) in a presumably adaptive way (Amdam *et al.* 2005b). Vg, therefore, contributes to colony fitness

Correspondence: Amro Zayed, Fax: 416-736-5698; E-mail: zayed@yorku.ca

through its effects on queen fecundity and lifespan, as well as through worker sib-care and division of labour.

Vitellogenin has pleiotropic effects on queen and worker traits in the honey bee. It is unclear whether protein-coding genes can adaptively evolve in social insects if they affect different traits in different castes, such as vitellogenin. It has been suggested that the evolution of eusociality involved the co-optation of pleiotropic gene networks regulating reproduction and maternal care in solitary insects (West-Eberhard 1996; Amdam *et al.* 2004), presumably through natural selection. Yet, genes with pleiotropic effects are expected to be evolutionarily constrained if mutations that increase fitness for one trait can decrease overall fitness via antagonistic effects on other traits (Williams 1957; Fisher 1958; Otto 2004). Genes with socially pleiotropic effects can be similarly constrained if they have antagonistic effects on multiple castes, resulting in sub-optimum queen or worker phenotypes (Linksvayer & Wade 2005). Indeed, several authors have suggested that gene duplication and caste-specific expression are needed to escape the constraining effects of social pleiotropy and propel queen/worker phenotypic divergence (West-Eberhard 1996; Gadagkar 1997). The fire ant genome contains four copies of vitellogenin that are expressed in a caste-specific manner (Wurm *et al.* 2011), suggesting that vitellogenin may have been constrained because of social pleiotropy in the fire ant. However, positive selection can still act on a mutation if its net effects on fitness across all castes are positive (Linksvayer & Wade 2005). Clearly, empirical data sets are needed to investigate the effect of social pleiotropy on molecular evolution, as a prelude to understanding phenotypic specialization and the evolution of division of labour in social insects.

We sequenced most of Vg's coding region (4.05 of 5.31 kb) in 41 *A. mellifera* workers, and in *A. cerana*, *A. florea* and *A. dorsata*. We also partially sequenced seven single-copy nuclear genes for comparative analysis (*Erk7*, *For* and five randomly chosen genes; Table S1, Supporting information). *Erk7* and *For* are also expressed in queens and workers and are involved in the regulation of worker division of labour (Ben-Shahar *et al.* 2003; Whitfield *et al.* 2003). We used the data set to study molecular evolution of vitellogenin in honey bees and to better understand the effect of social pleiotropy on molecular evolution in social organisms.

Methods

Population samples

We sampled three major population groups of *A. mellifera* (Whitfield *et al.* 2006) from Africa ($N = 19$ workers,

38 chromosomes per locus), East Europe ($N = 10$ workers, 20 chromosomes per locus) and West Europe ($N = 12$ workers, 20 chromosomes per locus) (Fig. S1, Supporting information). We also used one worker of each of *A. cerana*, *A. florea* and *A. dorsata* sampled from Thailand (two chromosomes per locus per species).

PCR and sequencing

DNA was extracted from thoracic tissue using a Qiagen DNeasy blood and tissue kit. PCR primers were developed to target exons using the honey bee genome as a template (Table S2, Supporting information). PCR was conducted in 25 μ L reactions containing 1 u Tsg DNA polymerase (BioBasic, Canada) 2–4 mM of $MgSO_4$, 200 μ M of each dNTP, 0.5 μ M of each primer and 20–100 ng of DNA. PCR purification and Sanger sequencing using both forward and reverse primers were carried out at McGill University and The G  n  me Qu  bec Innovation Centre (Qu  bec, Canada).

Sequence analysis and alignment

Sequencing reads were imported into Sequencher (v4.10.1) and trimmed of low-quality bases. Heterozygous sites were detected using Sequencher's Call Secondary Peaks command with a 34% threshold (i.e. intensity threshold for the fluorescently weaker allele at a heterozygous site), and all heterozygous sites were manually checked for errors. All sequences were aligned using Sequencher using the published *Apis mellifera* genome as a reference. The single nucleotide polymorphisms (SNPs) analysed herein were discovered from the alignments as allelic differences within or between *A. mellifera*, *A. cerana*, *A. florea* and *A. dorsata*.

Population genetic analysis

Diploid worker genotypes were phased using Phase (Stephens & Scheet 2005); the software estimates the most likely allelic phase for individuals with heterozygous genotypes, given the observed association between alleles in individuals with homozygous genotypes. We used Structure 2.3.3 (Pritchard *et al.* 2000; Hubisz *et al.* 2009) to examine the population structure in *A. mellifera* samples and to test for the presence of individuals of mixed ancestry. Our analysis is consistent with three distinct *A. mellifera* population groups in Africa, West Europe and East Europe (Groups A, M, and C following Ruttner 1988) corroborating recent studies (Whitfield *et al.* 2006; Zayed & Whitfield 2008). We detected several admixed individuals (Fig. S1, Supporting information); such bees were included in the global analysis

of *Apis mellifera*, but were excluded when conducting intraspecific analysis of the African, West European and East European *Apis mellifera* subspecies groups.

We used DNAsp 5.10 (Librado & Rozas 2009) to estimate several population genetic parameters and summary statistics. Probability values for tests were determined through 100 000 coalescent simulation runs using estimated recombination rates. Linkage disequilibrium measures calculated by DNAsp were exported to Gold (Abecasis & Cookson 2000) and plotted using the R^2 measure. We used DNAsp to carry out McDonald–Kreitman (MK) tests (McDonald & Kreitman 1991), which compare the number of nonsynonymous (NS) to synonymous (SS) sites that are polymorphic within species to those that are fixed between species; a higher ratio of fixed-to-polymorphic NS sites vs. SS sites indicates positive selection. DNAsp also provides an estimate of α , the proportion of amino acid changes fixed by positive selection (Stoletzki & Eyre-Walker 2011). We used mkprf (Bustamante *et al.* 2002, 2005) to estimate the mean and standard deviation (SD) of γ , representing the product of the effective population size and the selection coefficient ($\gamma = 2N_e s$), using default parameters and 100 000 runs.

We used Lositan (Antao *et al.* 2008) to examine whether some SNPs had outlier estimates of genetic differentiation, as quantified using F_{st} . Only SNPs with a minor allele frequency of 5% or higher were included in this analysis, which was carried out using 100 000 runs and applying a false discovery rate (FDR) correction (Benjamini & Hochberg 1995). We ran Lositan three times, each involving a pairwise comparison between our three sampled population groups. To examine whether genetic differentiation in Vg was elevated over genome wide levels, we compared F_{st} at polymorphic sites in Vg to those found in the seven genes studied herein, as well as a previously published data set (Zayed & Whitfield 2008) that estimated F_{st} in the same subspecies groups. Zayed & Whitfield (2008) previously showed that F_{st} in exon SNPs was negatively correlated with GC content. Combining data from Zayed & Whitfield (2008) with the current data set strengthened the negative correlation of F_{st} with GC ($P = 0.0007$, $R^2 = 0.13$). We compared the residual F_{st} estimates (i.e. after removing the effect of GC) in Vg vs. the rest of the data set using a *t*-test. General linear model analysis was carried out in R 2.8.1 using the stats package (R Development Core Team 2011).

Protein structure

We translated the Vg coding sequences obtained herein and aligned them to a partial protein model (Thompson & Banaszak 2002) using ClustalW2 (Larkin *et al.* 2007)

and cross-checked using Muscle (Edgar 2004) (Fig. S2, Supporting information). Polymorphic sites and linkage disequilibrium were visualized on the model using Chimera 1.5.2 (Pettersen *et al.* 2004; Meng *et al.* 2006). Atom positions from the protein model were imported into R 2.8.1, and the minimum Euclidean distance between each amino acid residue atom and any lipid atom (D_{min}) was calculated for all residues. We examined whether NS SNPs in a region of the protein associated with lipid binding were nearer to lipids by comparing D_{min} for NS sites in this region to D_{min} for SS sites; sites were classified as in contact with lipid when D_{min} was 4 Å or less.

Results

Vg exhibits high levels of adaptive evolution in the genus *Apis*

Vitellogenin had the highest levels of nucleotide diversity in *A. mellifera* ($\pi = 0.0086$) when compared to the other genes ($\pi = 0.0019 \pm 0.0011$ SD, $n = 7$). Vg also exhibited the highest level of divergence K (Jukes–Cantor), with 6.3% sequence divergence between *A. mellifera* and *A. cerana*, vs. an average of 2.9% sequence divergence in the other genes. The MK test comparing Vg in *A. mellifera* and *A. cerana* was highly significant ($\alpha = 0.81$; 53:98, polymorphic-to-fixed NS sites; and 206:71 SS sites, respectively; G-test with Williams correction, $G = 63.0$; $P < 10^{-10}$). MK tests using *A. florea* or *A. dorsata* gave similar results (*A. florea*: $\alpha = 0.84$, $P = 0.00004$; *A. dorsata*: $\alpha = 0.81$, $P = 0.003$). The MK test is largely insensitive to population demography (McDonald & Kreitman 1991; but see Stoletzki & Eyre-Walker 2011). Our study included 12 *A. mellifera scutellata* workers sampled from Africa and representing a large stable (Table S3, Supporting information) and lightly managed honey bee subspecies (Winston 1987; Ruttner 1988). A MK test utilizing only *A. m. scutellata* and *A. cerana* was highly significant ($P < 0.000005$) demonstrating that selection, and not demographic history, is driving the observed adaptive signatures in Vg. We also used a Bayesian method to estimate γ acting on our studied genes (Bustamante *et al.* 2002, 2005); positive values of γ indicate positive selection. γ was significantly positive for Vg (mean $\gamma = 1.88$) and for *Erk7* ($\gamma = 2.89$), but not significant for the other six genes (Fig. 1a). Our analysis also detected heterogeneous patterns of molecular evolution in both Vg and *Erk7*, with some domains and exons experiencing positive selection while others evolved neutrally or were constrained (Fig. 1b, Table S4, Supporting information). Interestingly, the N-terminal barrel region of vitellogenin has been identified as a conserved region

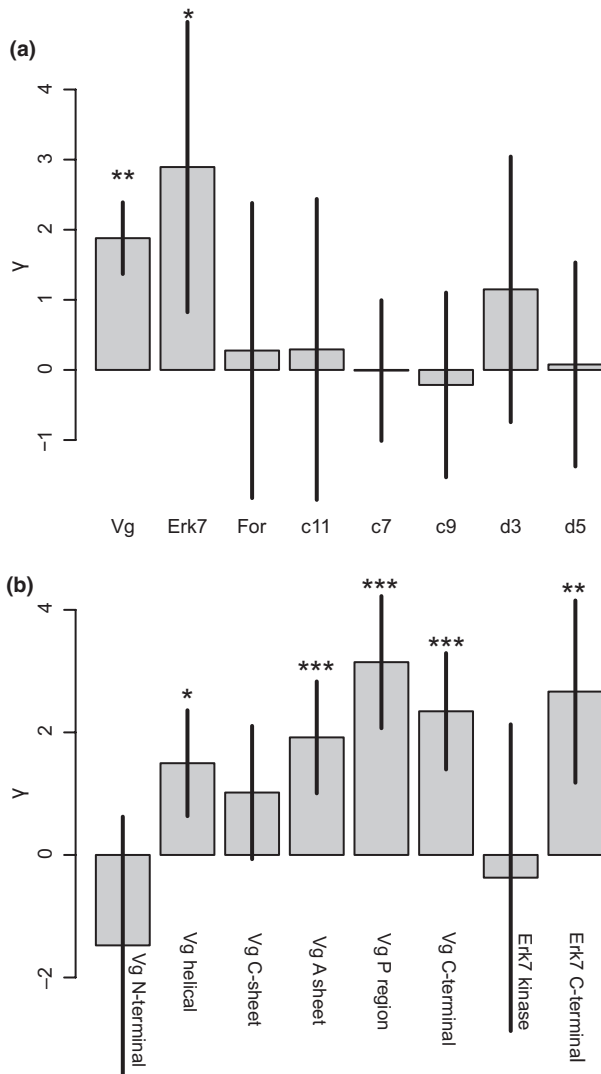


Fig. 1 Evidence for positive selection on vitellogenin in honey bees. Mean $\gamma = 2N_e s$ ($\pm$ SD; positive values indicate positive selection) for (a) all studied genes, and (b) for distinct functional regions in vitellogenin and Erk7 (See Table S4, Supporting information, for region definitions). Both vitellogenin (Vg) and Erk7 show significant gene-wide signatures of positive selection (a), although both Vg's N-terminal and Erk7's Kinase domain lack signs of positive selection (b).

in insects (Havukainen *et al.* 2011) and is the only region in this gene with a negative selection coefficient in our data (Fig. 1b).

Vg exhibits high levels of recent and ongoing adaptive evolution in Apis mellifera

Apis mellifera originated in Africa and colonized East Europe and West Europe via two independent expansions (Whitfield *et al.* 2006; Zayed & Whitfield 2008). We used population genetic summary statistics to infer

recent and ongoing bouts of adaptive evolution associated with *A. mellifera*'s out-of-Africa expansions. Positive selection is expected to increase linkage disequilibrium and skew the allele frequency spectrum towards rare-variants in the vicinity of a selected site (Nielsen 2005). After correction for multiple testing, we found many sites with significant linkage disequilibrium in Vg in both East European and West European honey bee subspecies (Fig. 2a,b; Table S3, Supporting information). For example, linkage disequilibrium statistics *B* (Wall 1999) and *Za* (Rozas *et al.* 2001) are both significant for Vg in both East Europe and West Europe subspecies groups but not for African subspecies groups. Further, these statistics were not significant at Erk7 and the five randomly chosen genes in East Europe, West Europe and Africa (Table S3, Supporting information). We also detected a significant skew in the allele frequency spectrum, as quantified by the Tajima's *D* statistic and the Fu & Li *F** test, at Vg but not in the other genes when all *Apis mellifera* samples were considered (Table S3, Supporting information). Although both changes in demography and drift can mimic selection by causing a skew in the allele frequency spectrum and linkage disequilibrium, these changes are expected to affect the entire genome and not a specific locus. Therefore, significant values of *D*, *F** and linkage disequilibrium at Vg but not the random genes suggest that the selection is operating on the former.

Positive selection underlying local adaptation is also expected to increase genetic differentiation between populations at the selected sites relative to the rest of the genome (Lewontin & Krakauer 1973; Beaumont 2005; Kelley *et al.* 2006). Overall, polymorphisms in Vg had significantly higher levels of genetic differentiation quantified using *F_{st}* when compared to other genes ($t = 3.65$, d.f. = 96, $P = 0.0004$). We also detected 64 SNPs with outlier values of *F_{st}* in Vg after correcting for differences in allele frequencies between sites. The 64 outlier SNPs were spatially clustered at approximately five distinct regions within Vg: of the 64 sites, 50% are within 10 bp of another outlier, and 94% are within 100 bp of another outlier. Additionally, the majority of the outlier SNPs was in or close to areas of high linkage disequilibrium. Finally, 27% of the outlier loci in Vg were replacement polymorphisms in regions of high linkage disequilibrium with alleles that were fixed in one or both of the European honey bee groups (Fig. 2b). The above patterns are indicative of recent or ongoing positive selection acting on functional variation, which in turn skews the allele frequency spectrum and increases genetic differentiation at nearby neutral sites.

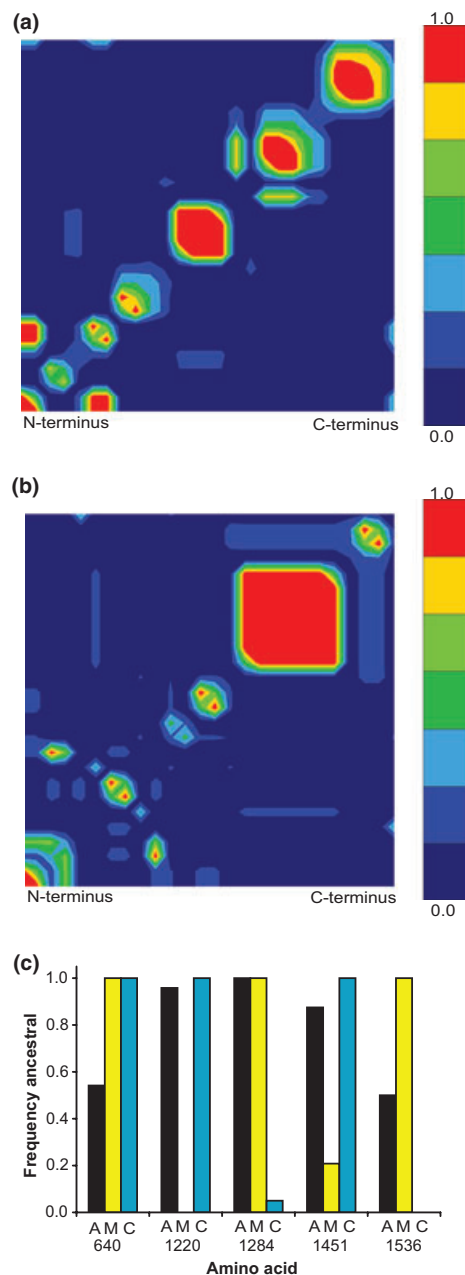


Fig. 2 Evidence for positive selection on vitellogenin within *Apis mellifera*'s geographic subspecies. Both Western European (a) and Eastern European bees (b) have regions of highly significant linkage disequilibrium (red, $R^2 > 85\%$) concentrated in C-terminal portions of the gene. (c) Candidate functional polymorphisms experiencing selective sweeps in vitellogenin (Vg)—all sites are nonsynonymous polymorphisms with outlier levels of genetic differentiation and high levels of linkage disequilibrium. Ancestral alleles were determined by comparison to *Apis cerana* and *Apis florea* sequences. Sites are labelled by amino acid residue number; A, African; M, Western European; C, Eastern European. Note fixation of derived allele for site 1220 in Western European bees and for site 1536 in Eastern European bees.

Lipid-binding regions of the Vg protein are enriched for adaptive amino acid changes

How do signatures of selection in Vg relate to the protein's structure and function? Vitellogenin's lipovitellin lipid-binding structure (Raag *et al.* 1988; Anderson *et al.* 1998) consists of a highly conserved N-terminal region that forms a beta-barrel with receptor binding sites (Li *et al.* 2003; Havukainen *et al.* 2011), followed by a region of alpha helices that define and partially enclose a lipid-binding cavity (Fig. S2, Supporting information). We distinguish a 'ball and socket' region of the protein (Anderson *et al.* 1998), which represents the lipid-binding cavity's insertion into the N-terminal beta-barrel; this ball-and-socket region encloses the area of highest resolved phospholipid densities in vitellogenin (Thompson & Banaszak 2002).

We aligned Vg's amino acid sequences with the sequence used for crystallography (Larkin *et al.* 2007) and displayed sites of significant linkage disequilibrium, genetic differentiation and amino acid changes (i.e. fixed differences in the genus, or polymorphic within *Apis mellifera*) (Pettersen *et al.* 2004; Meng *et al.* 2006). We quantified how close each mapped amino acid residue in the ball-and-socket region was to the phospholipid cargo by estimating the smallest distance of any non-hydrogen atom in the residue from any phospholipid. We found that residues close to phospholipids were significantly enriched for likely adaptive

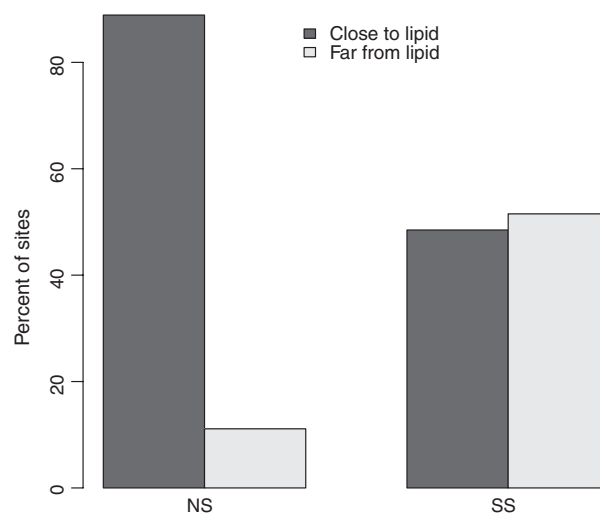


Fig. 3 Nonsynonymous mutations in vitellogenin are significantly enriched in parts of the protein associated with lipid binding. As control, synonymous mutations, believed to be neutral, do not show an association with lipid binding. Most nonsynonymous mutations in vitellogenin are estimated to be adaptive, suggesting that vitellogenin's lipid binding capacity is of both functional and adaptive relevance to the bee.

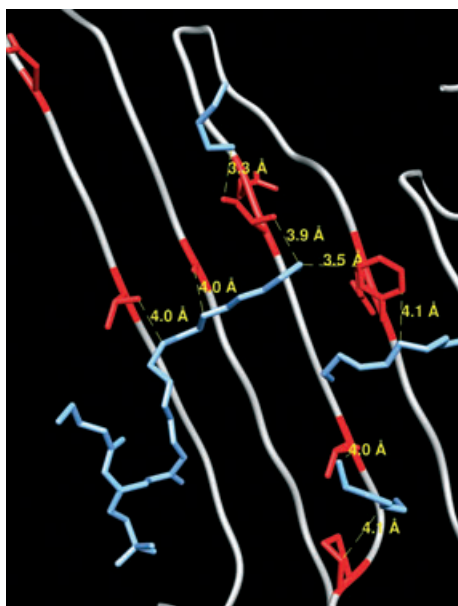


Fig. 4 A portion of vitellogenin where points of contact between bound lipids and the protein's lipid-carrying cavity occur at sites with nonsynonymous (NS) mutations in the bee. Part of the vitellogenin's (Vg) ball-and-socket region is shown in ribbon format along with lipid molecules (blue chains). Sites in contact with lipids (minimum carbon-carbon distances are depicted in yellow) have nonsynonymous mutations (red).

replacement mutations than residues farther from phospholipids (Fig. 3, Fig. S3, Supporting information).

Two putative polymorphisms experiencing recent or ongoing positive selection (Fig. 2c, Table S5, Supporting information) were located in the ball-and-socket region of the protein (Fig. 4). Residues 640 and 1451 are replacement polymorphisms with outlier F_{st} values and with side chains as little as 4.02 Å away from a bound phospholipid (i.e. they are in contact). This protein region contains a remarkable number of replacement changes in residues whose side chains contact phospholipids (residues 594, 608, 627—Fig. 4). There are also several regions of high linkage disequilibrium (not shown) in the ball-and-socket region. The striking juxtaposition of sites of potential adaptive change with residues in contact with the phospholipid cargo suggests selection on Vg's capacity to carry phospholipids in *Apis mellifera*.

Discussion

Adaptive evolution on Vg

Our data point to a clear signature of historical, recent and/or ongoing positive selection acting on the honey bee's vitellogenin. MK tests clearly showed that nonsynonymous mutations in Vg contribute significantly more to divergence vs. polymorphism relative to synonymous

and presumably neutral mutations. This is a hallmark signature of positive selection acting rapidly to fix beneficial nonsynonymous mutations resulting in reduced within species polymorphism and high between species divergence at such sites. We also detected several signatures of positive selection acting on Vg within the honey bee's geographic subspecies groups, as evident by significant deviations in Vg's allele frequency spectrum, outlier values of genetic differentiation and high levels of linkage disequilibrium relative to other genes in our data set. The above three observations are classic signatures of recent and ongoing selective sweeps.

Interestingly, Vg had high rates of molecular evolution in the honey bee, as evident by high rates of within *Apis mellifera* polymorphisms and high rates of divergence in the genus. The genomic location of Vg partially explains our observations: Vg is located on *Apis mellifera*'s fourth chromosome, which has the highest recombination rate (27.6 cM/Mb) among the bee's 16 chromosomes (Beye *et al.* 2006). Our estimates of recombination at Vg are also 5.9 times higher when compared to the average recombination rate at the other genes studied herein. There is a positive trend between recombination and polymorphism rates in *Apis mellifera* and other organisms (Beye *et al.* 2006), which can either be caused by adaptive (e.g. reduced neutral diversity in low recombination regions because of positive/negative selection on linked sites) or neutral processes (e.g. recombination increases mutation rates). Further, if local adaptation involves disruptive selection acting to fix alternative alleles in the different honey bee subspecies groups, as our data on Vg indicate, then we would expect a high level of polymorphism maintained within *Apis mellifera* at that gene.

Our evidence for positive selection acting on Vg demonstrates that the gene affects fitness in honey bees. Because queens in eusocial insects are specialized for egg laying, it is possible that the observed patterns of adaptive evolution on Vg reflect selection on queen reproductive traits and fecundity. Vg hemolymph titre is proportional to the queen's egg-laying rate (Engels 1974; Engels *et al.* 1990; Amdam *et al.* 2003) and egg-laying rate correlates with number of swarms (number of new daughter queens)—a direct barometer of fitness (McLellan & Rowland 1986). It is also conceivable that adaptive evolution on Vg can result through positive selection acting on worker traits through their net benefit on related reproductive kin. Vitellogenin is a source for the royal jelly produced by workers and fed to larvae (Amdam *et al.* 2003), so just as it affects queen egg laying, it may also affect the rearing of larvae hatching from those eggs. Additionally, Vg plays an important role in regulating the nurse-to-forager transition, which is a key aspect of worker division of labour that can

impact colony fitness by affecting colony foraging rates, foraging efficiency and growth rates (Sagili *et al.* 2011). Although it is currently difficult to determine whether positive selection on Vg is primarily affecting queen and/or worker traits, our work has highlighted many putative mutations that are adaptively evolving that can be assayed for phenotypic effects on queens and workers in the future.

In addition to demonstrating an overall signature of positive selection on Vg, our work also showed that nonsynonymous mutations in vitellogenin are significantly enriched in parts of the protein associated with lipid binding. Our finding suggests that differences in Vg's lipid-binding capacity are affecting queen and/or worker fitness. A recent analysis of the N-terminal, conserved barrel region of Vg, suggested its increased hydrophobicity in bees might be important in lipid binding (Havukainen *et al.* 2011). We agree with this suggestion, and our data more explicitly show that larger regions show evidence of selection for lipid binding. Vg has been shown to bind steroid hormones such as JH and ecdysteroids in the hemolymph and carry them to the ovaries in some insects (Lagueux *et al.* 1981; Yamada *et al.* 2005; Tufail & Takeda 2008), and binding of JH to lipid carrier molecules has been shown to protect it from degradation (Engelmann 2000). Vg's antagonistic association with JH in *Apis mellifera*, combined with our finding that Vg's lipid-binding capacity is adaptively evolving in the bee, suggests the intriguing possibility that Vg titres may directly affect steroid hormone action via binding of lipophilic hormone carriers.

Social pleiotropy and adaptive evolution on Vg

Our study provides the first evidence of positive selection acting on a socially pleiotropic gene in social insects. In addition to Vg, we also studied two genes that are presumably socially pleiotropic based on expression in queens and workers: Erk7 (1.25 kb sequenced) and For (1.09 kb sequenced). Similar to Vg, both Erk7 and For have been previously implicated in regulating worker division of labour (Grozinger *et al.* 2007; Nelson *et al.* 2007; Elango *et al.* 2009). We found very robust patterns of positive selection acting on Erk7, but For was extremely constrained; we found no replacement changes in For within or between *Apis mellifera* and *Apis cerana*, compared with 151 replacement changes in Vg. Our results indicate that social pleiotropy, on its own, does not necessarily restrict adaptive protein evolution. Both Vg and Erk7 are expressed in a limited number of tissues (Lyko *et al.* 2010), while For is ubiquitously expressed (Dawson-Scully *et al.* 2007; Kent *et al.* 2009), and this finding suggests that traditional forms of pleiotropy (i.e. expression of a gene in

multiple tissues and developmental stages, e.g. Wright *et al.* 2004) are more constraining than social pleiotropy.

Why does social pleiotropy appear to not constrain adaptive evolution on Vg and Erk7? Social pleiotropy is expected to constrain protein evolution only if pleiotropic alleles have opposing fitness effects on different castes. It is possible that alleles with socially pleiotropic effects are rarely antagonistic. Alternatively, the conditions under which antagonistic social pleiotropy can occur seem narrow. Positive selection acts indirectly on the effectively sterile worker caste, making it less efficient compared to selection acting directly on queens: selection on an allele because of worker fitness effects is half as strong as selection on queen fitness effects (Linksvayer & Wade 2009). Therefore, the conditions under which antagonistic social pleiotropy stifles protein evolution are restricted; an allele must have opposing fitness effects in queens and workers, and the strength of positive (or negative) selection on worker traits must balance out the strength of negative (or positive) selection on queen traits. Under such conditions, antagonistic alleles can remain segregating in the population (Linksvayer & Wade 2005). However, if the last condition is not satisfied, then selection can still fix mutations with opposing fitness effects in queens and workers.

Our data lend support to the above-mentioned hypothesis. The large proportion of nonsynonymous polymorphisms fixed by selection in Vg suggests that functional mutations in the gene do not often have antagonistic effects or that their net effect on fitness across queens and workers are positive. We also found 53 segregating nonsynonymous mutations in *Apis mellifera*'s vitellogenin. These mutations likely have a spectrum of fitness effects including: no effect on fitness (i.e. neutral), antagonistic pleiotropic effects (i.e. appear to be neutral if the sum of fitness effects across castes are balanced) or positive effects on one or both castes (i.e. on their way to fixation). Our population genetic data indicate that many of these mutations are indeed experiencing ongoing positive selection and are thereby not likely to have strong antagonistic effects. If our case study of Vg can be generalized, our work indicates that social pleiotropy does not necessarily limit adaptive protein evolution in social insects.

Acknowledgements

We thank L. Packer, N. De Silva and members of the Thailand Inventory Group for Entomological Research (TIGER) project (NSF DEB-0542864), C. Eardley, P. De la Rua, A. Oleksa, P. Kozmus and M. Hasselmann for providing bee samples, K. Sofier, S. Munir, S. El Khazendar, F. Zaidi, and M. Aryaee for laboratory assistance, and S. Minaei for providing data. We also thank B. Harpur, J. Gixti, J. Shore and L. Packer for

providing helpful comments on this manuscript. This study was funded through a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC), and an Early Researcher Award from the Ontario Ministry of Research and Innovation to AZ. CK is supported by an NSERC Postdoctoral Fellowship.

References

- Abecasis GR, Cookson WOC (2000) GOLD—graphical overview of linkage disequilibrium. *Bioinformatics*, **16**, 182–183.
- Amdam GV, Omholt SW (2003) The hive bee to forager transition in honeybee colonies: the double repressor hypothesis. *Journal of Theoretical Biology*, **223**, 451–464.
- Amdam GV, Norberg K, Hagen A, Omholt SW (2003) Social exploitation of vitellogenin. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 1799–1802.
- 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. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 11350–11355.
- Amdam GV, Aase ANTO, Seehuus SC *et al.* (2005a) Social reversal of immunosenescence in honey bee workers. *Experimental Gerontology*, **40**, 939–947.
- Amdam GV, Norberg K, Omholt SW *et al.* (2005b) Higher vitellogenin concentrations in honey bee workers may be an adaptation to life in temperate climates. *Insectes Sociaux*, **52**, 316–319.
- Anderson TA, Levitt DG, Banaszak LJ (1998) The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure*, **6**, 895–909.
- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.
- Barchuk AR, Bitondi MG, Simoes ZLP (2002) Effects of juvenile hormone and ecdysone on the timing of vitellogenin appearance in hemolymph of queen and worker pupae of *Apis mellifera*. *Journal of Insect Science*, **2**, 1–8.
- Beaumont MA (2005) Adaptation and speciation: what can FST tell us? *Trends in Ecology and Evolution*, **20**, 435–440.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **57**, 289–300.
- Ben-Shahar Y, Leung H-T, Pak WL, Sokolowski MB, Robinson GE (2003) cGMP-dependent changes in phototaxis: a possible role for the foraging gene in honey bee division of labor. *Journal of Experimental Biology*, **206**, 2507–2515.
- Beye M, Gattermeier I, Hasselmann M *et al.* (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Research*, **16**, 1339–1344.
- Brillet C, Robinson GE, Bues R, LeConte Y (2002) Racial differences in division of labor in colonies of the honey bee. *Ethology*, **108**, 115–126.
- Bustamante CD, Nielsen R, Sawyer SA *et al.* (2002) The cost of inbreeding in Arabidopsis. *Nature*, **416**, 531–534.
- Bustamante CD, Fledel-Alon A, Williamson S *et al.* (2005) Natural selection on protein-coding genes in the human genome. *Nature*, **437**, 1153–1157.
- Chen JS, Sappington TW, Raikhel AS (1997) Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry. *Journal of Molecular Evolution*, **44**, 440–451.
- Corona M, Velarde RA, Remolina S *et al.* (2007) Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 7128–7133.
- Dawson-Scully K, Armstrong GA, Kent C, Robertson RM, Sokolowski MB (2007) Natural variation in the thermotolerance of neural function and behavior due to a cGMP-dependent protein kinase. *PLoS ONE*, **2**, e773.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.
- Elango N, Hunt BG, Goodisman MAD, Yi SV (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 11206–11211.
- Engelmann F (2000) The interactions between juvenile hormone (JH), lipophorin, vitellogenin, and JH esterases in two cockroach species. *Insect Biochemistry and Molecular Biology*, **30**, 793–803.
- Engels W (1974) Occurrence and significance of vitellogenins in female castes of social Hymenoptera. *Integrative and Comparative Biology*, **14**, 1229–1237.
- Engels W, Kaatz H, Zillikens A *et al.* (1990) Honey bee reproduction: vitellogenin and caste-specific regulation of fertility. In: *Advances in Invertebrate Reproduction 5* (eds Hoshi M, Yamashita O), pp. 495–502. Elsevier, Amsterdam.
- Fisher RA (1958) *The Genetical Theory of Natural Selection*. Dover, New York.
- Gadagkar R (1997) The evolution of caste polymorphism in social insects: Genetic release followed by diversifying evolution. *Journal of Genetics*, **76**, 167–179.
- Grozinger CM, Fan Y, Hoover SE, Winston ML (2007) Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Molecular Ecology*, **16**, 4837–4848.
- Havukainen H, Halskau O, Skjaerven L, Smedal B, Amdam GV (2011) Deconstructing honeybee vitellogenin: novel 40 kDa fragment assigned to its N terminus. *Journal of Experimental Biology*, **214**, 582–592.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.
- Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM (2006) Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Research*, **16**, 980–989.
- Kent CF, Daskalchuk T, Cook L, Sokolowski MB, Greenspan RJ (2009) The *Drosophila* foraging gene mediates adult plasticity and gene-environment interactions in behaviour, metabolites, and gene expression in response to food deprivation. *PLoS Genetics*, **5**, e1000609.
- Lagueux M, Harry P, Hoffmann JA (1981) Ecdysteroids are bound to vitellin in newly laid eggs of locusts. *Molecular and Cellular Endocrinology*, **24**, 325–338.
- Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphism. *Genetics*, **74**, 175–195.
- Li A, Sadasivam M, Ding JL (2003) Receptor–ligand interaction between vitellogenin receptor (VtGR) and vitellogenin (Vtg), implications on low density lipoprotein receptor and apolipoprotein B/E. *Journal of Biological Chemistry*, **278**, 2799–2806.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Linksvayer TA, Wade MJ (2005) The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. *Quarterly Review of Biology*, **80**, 317–336.
- Linksvayer TA, Wade MJ (2009) Genes with social effects are expected to harbor more sequence variation within and between species. *Evolution*, **63**, 1685–1696.
- Lyko F, Foret S, Kucharski R *et al.* (2010) The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biology*, **8**, e1000506.
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, **351**, 652–654.
- McLellan AR, Rowland CM (1986) A honeybee colony swarming model. *Ecological Modelling*, **33**, 137–148.
- Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics*, **7**, 339.
- Nelson CM, Ihle KE, Fondrk MK, Page REJ, Amdam GV (2007) The gene *vitellogenin* has multiple coordinating effects on social organization. *PLoS Biology*, **5**, e62.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.
- Otto SP (2004) Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proceedings of the Royal Society B-Biological Sciences*, **271**, 705–714.
- Pettersen EF, Goddard TD, Huang CC *et al.* (2004) UCSF Chimera – a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**, 1605–1612.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Raag R, Appelt K, Xuong NH, Banaszak L (1988) Structure of the lamprey yolk lipid-protein complex lipovitellin-phosvitin at 2.8 Å resolution. *Journal of Molecular Biology*, **200**, 553–569.
- Rozas J, Gullaud M, Blandin G, Aguade M (2001) DNA variation at the *rp49* gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. *Genetics*, **158**, 1147–1155.
- Ruttner F (1988) *Biogeography and Taxonomy of Honeybees*. Springer, New York.
- 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**, e16785.
- Sappington TW, Raikhel AS (1998) Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochemistry and Molecular Biology*, **28**, 277–300.
- Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, **76**, 449–462.
- Stoletzki N, Eyre-Walker A (2011) Estimation of the neutrality index. *Molecular Biology and Evolution*, **28**, 63–70.
- Thompson JR, Banaszak LJ (2002) Lipid-protein interactions in lipovitellin. *Biochemistry*, **41**, 9398–9409.
- Tufail M, Takeda M (2008) Molecular characteristics of insect vitellogenins. *Journal of Insect Physiology*, **54**, 1447–1458.
- Wall JD (1999) Recombination and the power of statistical tests of neutrality. *Genetical Research*, **74**, 65–79.
- West-Eberhard MJ (1996) Wasp societies as microcosms for the study of development and evolution. In: *Natural History and Evolution of Paper-Wasps* (eds Turillazzi S, West-Eberhard MJ), pp. 290–317. Oxford University Press, Oxford.
- Whitfield CW, Cziko AM, Robinson GE (2003) Gene expression profiles in the brain predict behavior in individual honey bees. *Science*, **302**, 296–299.
- Whitfield CW, Behura SK, Berlocher SH *et al.* (2006) Thrive out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, **314**, 642–645.
- Williams GC (1957) Pleiotropy, natural selection, and the evolution of senescence. *Evolution*, **11**, 398–411.
- Winston ML (1987) *The Biology of the Honey Bee*. Harvard University Press, Cambridge, MA.
- Wright SI, Yau CB, Looseley M, Meyers BC (2004) Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Molecular Biology and Evolution*, **21**, 1719–1726.
- Wurm Y, Wang J, Riba-Grognuz O *et al.* (2011) The genome of the fire ant *Solenopsis invicta*. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 5679–5684.
- Yamada R, Yamahama Y, Sonobe H (2005) Release of ecdysteroid-phosphates from egg yolk granules and their dephosphorylation during early embryonic development in silkworm, *Bombyx mori*. *Zoological Science*, **22**, 187–198.
- Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 3421–3426.

C.K.'s research focuses on the genetics of social behaviour. A.I. and A.C.B. are Research at York undergraduate students interested in genetics and evolution. A.Z.'s research focuses on understanding the genetic basis of worker phenotypes, the evolution of genes underlying worker phenotypes, and how worker phenotypes contribute to colony fitness in social insects.

Data accessibility

Individual sequence data and DNA alignments have been deposited to Genbank (accessions JN556047–JN557798). The accessions also include relative ancestry of each of our sampled individuals to the honey bee's African, E. and W. European subspecies groups, as presented in Supporting Fig. S1.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Population structure of sampled *Apis mellifera* workers.

Fig. S2 Vitellogenin protein with a visual representation of population genetic summary statistics.

Fig. S3 Nonsynonymous residues in the Vg ball-and-socket region.

Table S1 List of genes sequenced for this study.

Table S2 Sequencing primers.

Table S3 Population genetic summary statistics for *Apis mellifera*.

Table S4 Protein Regions of Vg and Erk7.

Table S5 Nonsynonymous changes at sites with LD and high F_{st} .

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.