Ancient Duplications Have Led to Functional Divergence of Vitellogenin-Like Genes Potentially Involved in Inflammation and Oxidative Stress in Honey Bees

Heli Salmela[1], Taina Stark[1,2], Dimitri Stucki[1], Siiri Fuchs[1], Dalial Freitak[1], Alivia Dey[3], Clement F. Kent[3], Amro Zayed[3], Kishor Dhaygude[1], Heikki Hokkanen[2], and Liselotte Sundström[1,*]

[1]Department of Biosciences, Centre of Excellence in Biological Interactions, University of Helsinki, Helsinki, Finland

[2]Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland

[3]Department of Biology, York University, Toronto, ON, Canada

*Corresponding author: E-mail: liselotte.sundstrom@helsinki.fi.

Accepted: January 28, 2016

Data deposition: This project has been deposited at at Dryad under the accession address http://dx.doi.org/10.5061/dryad.hq669.

Abstract

Protection against inflammation and oxidative stress is key in slowing down aging processes. The honey bee (*Apis mellifera*) shows flexible aging patterns linked to the social role of individual bees. One molecular factor associated with honey bee aging regulation is vitellogenin, a lipoglycophosphoprotein with anti-inflammatory and antioxidant properties. Recently, we identified three genes in Hymenopteran genomes arisen from ancient insect *vitellogenin* duplications, named *vg-like-A*, *-B*, and *-C*. The function of these *vitellogenin* homologs is unclear. We hypothesize that some of them might share gene- and protein-level similarities and a longevity-supporting role with *vitellogenin*. Here, we show how the structure and modifications of the *vg-like* genes and proteins have diverged from *vitellogenin*. Furthermore, all three *vg-like* genes show signs of positive selection, but the spatial location of the selected protein sites differ from those found in *vitellogenin*. We show that all these genes are expressed in both long-lived winter worker bees and in summer nurse bees with intermediate life expectancy, yet only *vg-like-A* shows elevated expression in winter bees as found in *vitellogenin*. Finally, we show that *vg-like-A* responds more strongly than *vitellogenin* to inflammatory and oxidative conditions in summer nurse bees, and that also *vg-like-B* responds to oxidative stress. We associate *vg-like-A* and, to lesser extent, *vg-like-B* to the antiaging roles of *vitellogenin*, but that *vg-like-C* probably is involved in some other function. Our analysis indicates that an ancient duplication event facilitated the adaptive and functional divergence of *vitellogenin* and its paralogs in the honey bee.

Key words: *Apis mellifera*, vitellogenin, protein structure, functional divergence, seasonality, gene expression.

Introduction

Chronic inflammation and increased oxidative stress appear to be major aging factors (Harman 1956; McGeer PL and McGeer EG 2004). Animals employ a broad anti-inflammatory and antioxidant machinery to decelerate aging. This machinery includes proteins, such as lipoproteins, that can suppress inflammation by covering hydrophobic surfaces brought about by inflammation. Inflammation can result from, for example, wounding or oxidative damage (Seong and Matzinger 2004; Cho and Seong 2009). Lipoproteins, in general, have roles beyond lipid transportation, and are strongly connected to health and longevity in humans (Atzmon et al. 2002; McKay et al. 2011; Hausman et al. 2012).

Honey bees are used as a model species in aging studies due to their remarkable aging plasticity (reviewed by Munch and Amdam 2010). Typically, honey bee workers—the nonreproductive females in a colony—spend the first weeks of their approximately 6-week-long life inside the hive mainly nursing brood (nurse bees). Around the age of 21 days, workers born in summer start foraging, an activity characterized by rapid physiological aging (Winston 1987). In temperate climates, honey bees do not hibernate. Instead, the workers born in autumn form a long-lived phenotype, and these winter workers (winter bees) can survive for several months or even over a year (Fluri et al. 1977; Munch and Amdam 2010; Munch et al. 2013). These two honey bee worker cohorts differ in their

capability to mitigate oxidative stress, the winter bees being most stress resistant (Munch and Amdam 2010; Munch et al. 2013).

An important part of the molecular machinery that regulates honey bee division of labor and life span is a protein called vitellogenin (Vg). This large lipid-carrying protein is produced in the fat body—a tissue with some analogy to the vertebrate liver—and stored there or secreted to the hemolymph, where it is a major protein component. Vg consists of several structural domains, and is modified in a complex manner by addition of glycans and phosphogroups (Tufail and Takeda 2008; Havukainen et al. 2012) that, typically, guide proteins in their biochemical tasks (Cao et al. 2009). Vg is best known for its role in the formation of egg yolk in egg-laying animals (Tufail and Takeda 2008); however, it also fulfills roles unrelated to reproduction in worker bees. Elevated Vg levels in worker bees provide protection against oxidative stress (Seehuus et al. 2006). This is due to the antioxidant property of the protein; Vg can neutralize free radicals (Seehuus et al. 2006; Havukainen et al. 2013). Another shielding effect of Vg is the recognition of dead and damaged cells in an anti-inflammatory manner (Havukainen et al. 2013). Furthermore, Vg has a positive effect on cellular immunity in the honey bee (Amdam et al. 2005). These properties link Vg to stress resistance and longevity in honey bee workers (Munch and Amdam 2010; Munch et al. 2013) as, especially, the long-lived winter bees have elevated *vg* expression levels in autumn and in early winter (Fluri et al. 1977; Aurori et al. 2014).

The *vg* gene appears to have experienced major past and ongoing bouts of adaptive evolution in the honey bee, with a high number of amino acid changes fixed by positive selection (Kent et al. 2011). There are several evolutionarily recent *vg* copies in the genomes of many insect species, for example, the ant *Solenopsis invicta* has four *vg* copies (Wurm et al. 2010), which are differentially expressed in the worker, queen, and male castes (Wurm et al. 2010; Corona et al. 2013; Feldmeyer et al. 2014). These *vg* copies are absent in the honey bee, but we have recently identified three genes homologous to *vg* in the Hymenopteran genomes sequenced so far: *vg-like-A*, *-B*, and *-C* (Morandin et al. 2014). It appears that these *vg* homologs have arisen from gene duplication events older than the species-specific *vg* duplications identified before (Wurm et al. 2010). The ancestral *vg* gene has been duplicated early in insect evolution, with one of the duplicated genes giving rise to *vg-like-A* and *-B* (found also in a species as distant as mosquito *Anopheles gambiae*), and the other yielded *vg* and *vg-like-C*. *Vg-like-C* is only found in Hymenoptera (Morandin et al. 2014). So far, expression of these insect *vg-like* genes has only been verified in ants (Morandin et al. 2014). The function of the *vg-like* genes has remained unclear.

Our objective was to examine if the *vg-like* genes have anti-inflammatory and antioxidative qualities similar to *vg* in the honey bee. We hypothesized that *vg-like* genes with the most structural, posttranslational, and evolutionary similarities to *vg* are likely also functionally closer to *vg*. Alternatively, *vg-like* genes of limited similarity to *vg* might target a subset of the *vg* tasks, or have a function of their own. Based on the known *vg* expression patterns, we hypothesized that a *vg-like* gene with antiaging properties would be more expressed in the long-lived winter bees than in summer nurse bees and the expression should get upregulated also in the summer nurse bees when exposed to stressors. To test these ideas, we conducted a suite of analyses, including database searching, sequence-based predictions, and homology modeling of the *vg-like* genes in the honey bee, comparing them with *vg*. We also studied the population genetics of *vg-like* genes to determine if they are subject to positive selection like *vg*, and if patterns of positive selection of the *vg-like* genes occur in the same or different protein domains. Finally, we measured *vg-like* gene expression in winter bees and summer nurse bees, and tested if injecting summer nurses with bee saline or an oxidative agent can provoke *vg-like* gene expression.

Materials and Methods

Vg-Like Gene/Protein Characterization

We extracted genomic information on the *vg-like* genes based on the honey bee genome (Honeybee Genome Sequencing Consortium 2006). Then, we ran sequence-based searches using different servers for prediction of posttranslational modifications and identification of structural protein domains. The following servers were used: SUPERFAMILY for protein domains, NetNGlyc 1.0 for glycosylation sites, NetPhos 2.0 for phosphorylation sites, SignalP 4.1 for a signal peptide, and the Honey Bee Peptide Atlas for protein localization. The latter is a mass-spectrometric database for peptides found in all three honey bee castes (workers, queens, and drones) in various tissues (Chan et al. 2011).

We built homology models of Vg and the Vg-like proteins using the MODELER software in the Discovery Studio 4.0 Modeling Environment (Accelrys Software Inc., San Diego). The X-ray structure of lamprey lipovitellin (PDB-ID: 1LSH) was used as the template as before for Vg-related proteins (Mann et al. 1999; Havukainen, Halskau, Skjaerven, et al. 2011). We modeled the amino acid residues 22–839 of Vg; 83–787 of Vg-like-A; 36–700 of Vg-like-B, and 27–321 of Vg-like-C. The carboxyl-terminus sequence of Vg, Vg-like-A, and -B was considered unreliable for modeling owing to gaps in the template structure. In addition, there is no structural data available for the carboxyl-terminal vWFD domain. Thus, these areas were not modeled. The modeling was based on multiple sequence alignment of Vg and Vg-like proteins as in Morandin et al. (2014). Multiple Vg sequence alignment (Avarre et al. 2007) was also used as a reference. Five models per protein were built, with the optimization level High and with loop refinement. The lowest energy model

was accepted as representative for each protein. The models were verified with the three-dimensional (3D) profile option in the software that compares the compatibility of the 3D structure of the protein and its sequence. Pymol software (The PyMOL Molecular Graphics System, Version 1.5.0.4; Schrödinger, LLC) was used for visualization.

Population Genetics Analysis

We first studied the rate of molecular evolution of *vg-like* genes using d*N*/d*S*, the ratio of the number of nonsynonymous substitutions per nonsynonymous site (d*N*) to the number of synonymous substitutions per synonymous site (d*S*). We used *Apis mellifera*'s Vg-like protein sequences to find orthologs in other *Apis* as well as *Bombus* using BLASTP (Johnson et al. 2008). The resulting protein sequences were first aligned with Clustal Omega (Sievers et al. 2011), then a refined coding sequence and protein alignment was produced using PAL2NAL 14 (Suyama et al. 2006). PAML 4 (Yang 2007) was then used to calculate d*N*/d*S* using published phylogenetic trees for *Apis* and *Bombus* (Arias and Sheppard 1996; Cameron et al. 2007). To test for patterns of accelerated molecular evolution, we used PAML with Model = 0, NS = 7 and 8, with omega and kappa kept variable, to compare model fit without and with positive selection. Separate runs were done using different codon frequency models, but on inspection this produced only small changes in calculated d*N*/d*S*, so results are reported for CodonFreq = 3. We report the average d*N*/d*S* for all sister species pairs in the tree. The final analysis included *A. mellifera*, *Apis cerana* (for all but vg-like-B), *Apis florea*, *Apis dorsata*, *Bombus impatiens*, and *Bombus terrestris*.

The d*N*/d*S* approach is useful for studying the rate of amino acid evolution, but some nonadaptive evolutionary process can mimic natural selection (reviewed by Kent and Zayed 2015). As such, we also carried out more robust tests of selection using population genetic data from *Apis*. We used sequences of *vg-like* genes from 11 *A. mellifera* and 1 *A. cerana* workers from a recently published large-scale population genomic data set (Harpur et al. 2014). The *A. mellifera* samples were derived from Africa representing a large stable population that is minimally impacted by human management (Harpur et al. 2014). Diploid worker genotypes were phased using PHASE v2.1.1 (Stephens et al. 2001) with input files and output files processed through SeqPHASE (Flot 2010). Population genetics summary statistics—nucleotide diversity (π), synonymous site divergence (K_s), neutrality tests, Tajima's *D*, and Fu & Li's *D*—were calculated using DNAsp v5.10 (Librado and Rozas 2009). DNAsp was also used to carry out McDonald–Kreitman (MK) tests that compare the ratio of nonsynonymous to synonymous fixed changes between species with polymorphisms within species (McDonald and Kreitman 1991). We then used mkprf to estimate mean and standard deviation of the parameter gamma (γ) that is used to represent the product of effective population size and the selection coefficient, using a burn-in of 10,000 and 100,000 runs (Bustamante et al. 2002). Significance was evaluated based on the posterior distributions; γ was deemed significant if no more than 2.5% of the posterior distribution overlapped with zero. The population genetic data and methods can be found in the digital repository http://dx.doi.org/10.5061/dryad.hq669.

The Honey Bee Worker Samples

We measured *vg* and *vg-like* gene expression in two sets of worker bees (*A. mellifera ligustica*): Winter bees and summer nurses. Winter bees were collected from three hives in the Paulig mansion and the Toivola garden in Helsinki in mid-November in 2014 (ten individuals per hive). Summer bee nurses were collected from five hives located next to each other at the Viikki campus in Helsinki in May–June (26–30 nurses per hive). Nurses were identified based on their behavior, that is, workers in the process of tending brood. This behavior is performed by 4- to 12-day-old nurse bees (Johnson 2010). Summer nurses were further exposed to three treatments (see below), whereas the winter bees received the same treatment as the Control group of the summer nurses.

Experimental Design of the Inflammation/Oxidation Test

Summer nurses from all five hives were exposed to three treatments (Control, Sham, and Paraquat), to test for expression responses of the *vg-like* genes. The Control group (N = 49, 9–10 individuals per hive) received no injection. The Sham group received an injection of 2 µl sterile honey bee saline (Dietemann et al. 2013) (N = 44, 7–10 individuals per hive). The purpose of this group is to reveal gene expression changes provoked by the inflammation caused by a needle prick. Vg protein reacts to tissue damage (Havukainen et al. 2013); hence, this gene would be expected to react to a sham injection. Oxidative stress was induced by injection of 2 µl of 50 µg/g Paraquat (Sigma-Aldrich) diluted in bee saline (N = 49, 9–11 individuals per hive). The Paraquat dosage that deemed to cause a level of moderate oxidative stress was based on an earlier study in the honey bee (Seehuus et al. 2006). Prior to injections, all bees, including the Control, were anesthetized by a standard cold treatment of 3 min at −20 °C (Dietemann et al. 2013), and were kept on ice during the injections. The injections were made between abdominal segments using a Hamilton G 30 micro syringe (Becton Dickinson) as described in Amdam et al. (2003). All bees, including the Control, were kept in plastic cages—five individuals in each—with 50% sugar solution for ad libitum feeding at 30 °C in a humidified incubator for 18 h, followed by tissue sampling.

Quantitative Real-Time PCR

For tissue sampling, the winter bees and summer nurses were anesthetized by cold treatment as above, and the gut and the stinger were removed on ice. The abdomen containing whole fat body tissue was detached, placed in 1 ml TRIsure (Bioline), and frozen at −80 °C until further processing. Before RNA extraction, the abdomens were homogenized with Tissue Lyser. The standard TRIsure procedure was used for RNA extraction (see manufacturer's instructions), and the isolated RNA was treated with DNase I (Thermo Scientific) to avoid amplification of genomic DNA.

The primers for the *vg-like* genes (supplementary table S1, Supplementary Material online) were designed on exon–intron boundaries to ensure that genomic DNA was not amplified, using NCBI Primer-BLAST, and synthesized by OligomerHelsinki, Finland. We analyzed the melt curve output to confirm that no primer dimers or DNA amplification was produced during the cycles. The reference genes were chosen based on previous Real-time PCR (qPCR) work in the honey bee (de Miranda and Fries 2008; Wang et al. 2012). The efficiencies of the primers were tested in their appropriate annealing temperatures, and the correlation coefficients (R^2) of the chosen primers all ranged from 0.963 to 0.998 (supplementary table S1, Supplementary Material online).

Real-time PCR (qPCR) was performed on single individuals as follows: Initial denaturation at 95 °C for 3 min; 39 cycles with denaturation at 95 °C for 15 s; and primer-pair-specific annealing temperature (supplementary table S1, Supplementary Material online) for 60 s. Reactions were done with BioRad iQ qPCR SYBR Green Master Mix using the BioRad CFX 384 qPCR Cycler in triplicates. We used 5 µl of 100 ng/µl RNA. By monitoring negative control samples (without reverse transcriptase) and melting curves, we verified that the qPCR assay was not confounded by DNA contamination or primer dimers.

Gene expression was analyzed based on standardized *Ct* values (ΔCt) (Livak and Schmittgen 2001), where *RP49* (de Miranda and Fries 2008) was used as the reference gene for *vg*, *vg-like-A*, and *vg-like-B* and *actin* (Wang et al. 2012) for *vg-like-C*. The reference genes (*RP49* and *actin*) were chosen based on their optimal primer behavior in the temperatures that differed for *vg-like-C* and the other genes. Because the raw *Ct* values of the two reference genes did not differ substantially across samples and individuals (visual comparison), all genes were analyzed in the same model. To compare the effects of treatment on gene expression among the genes, we performed a mixed effects model with treatment (Control, Sham, Paraquat), gene (*vg*, *vg-like-A*, *vg-like-B*, *vg-like-C*), and the interaction of treatment and gene as fixed factors. Colony of origin was included as random effect. For all statistical tests we used a threshold level of significance of $\alpha = 0.05$. Statistical analysis was carried out in R version 3.1.1 (R Core Team 2014), using the package nlme (Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team 2015; nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-119, http://CRAN.R-project.org/package=nlme, last accessed February 18, 2016).

Results

We used computational and in vivo analysis to investigate the role *vg-like* genes/proteins may play in honey bee longevity. The computational results include protein characterization via database searches, sequence-based predictions, and a population genetic analysis. The in vivo results are based on gene expression analysis on honey bee workers of two different cohorts, and subjected to different treatments.

vg-Like Gene/Protein Characterization

The results of database searches show that *vg* and the *vg-like* genes differ in their genomic organization; *vg* is located on chromosome LG4, *vg-like-A* on chromosome LG13, and *vg-like-B* and *-C* on the same chromosome LG2 ~4,480 kbp apart. Furthermore, *vg-like-C* coding sequence is approximately half the length of the other genes. The intron percentage also differed, being higher in *vg-like-C* (52.68%) and *vg-like-B* (33.40%), compared with *vg-like-A* (20.60%) and *vg* (11.92%). An in silico predicted splice variant was found only for *vg-like-B*; this splice variant lacks the last exon. The protein encoded by each of the genes belongs to the protein family pfam01347, which is also named as Vitellogenin_N. All Vg-like amino acid sequences contained at least one predicted glycosylation site, a secretion signal peptide, and a few predicted phosphorylation sites (table 1).

The localization of the Vg-like proteins was inspected using the honey bee mass-spectrometric data base (Chan et al. 2011) (table 1). The protein corresponding to *vg-like-A* was more abundant in workers than in queens. It was found largely in the fat body with surrounding tissues (the heart and the terga), and in the third pair of legs. In lesser quantities, it was also found in the scape (the antennal base), the second pair of legs, and the sterna (the abdominal cuticula). The protein corresponding to *vg-like-B* was not identified in the database. Finally, the protein corresponding to *vg-like-C* was found in all three castes (workers, queens, and males) in one location: The scape (antennal base). *Vg-like-C* more abundant in the drone and worker castes than in the queen caste.

The comparison of protein structure by homology modeling shows that Vg-like-A shares nearly all of the features of Vg (fig. 1*A* and *B* and supplementary table S2, Supplementary Material online). Also, the Vg-like-B exhibited most Vg features. In contrast, Vg-like-C is a truncated protein, and so differs considerably from Vg. A protein domain search using the SUPERFAMILY server shows that all the Vg-like proteins contain the N-terminal domain called N-sheet (fig. 1*A* and *B*). However, the small Vg-like-C has no other domains but the

Table 1

Sequence-Based Predictions of Glycosylation and Phosphorylation Sites and Signal Peptide in Vg and the Vg-Like Proteins, and Protein Localization Based on a Mass-Spectrometric Database

Protein Name	Glycosylation Sites	Signal Peptide	Phosphorylation Sites (Ser/Thr/Tyr)	Localization (Q = queen, W = worker, D = Drone)
Vg	3	16 aa	74/28/17	Q, W: Most tissues D: Many tissues Abundancy: Q>W≫D
Vg-like-A	6	18 aa	39/15/14	Q: Fat body, heart, leg W, D: Fat body, heart, leg, antenna Abundancy: W>Q≫D
Vg-like-B	11	35 aa	32/12/27	—
Vg-like-C	1	11 aa	13/4/3	Q, W, D: Antenna base Abundancy: D>W>Q

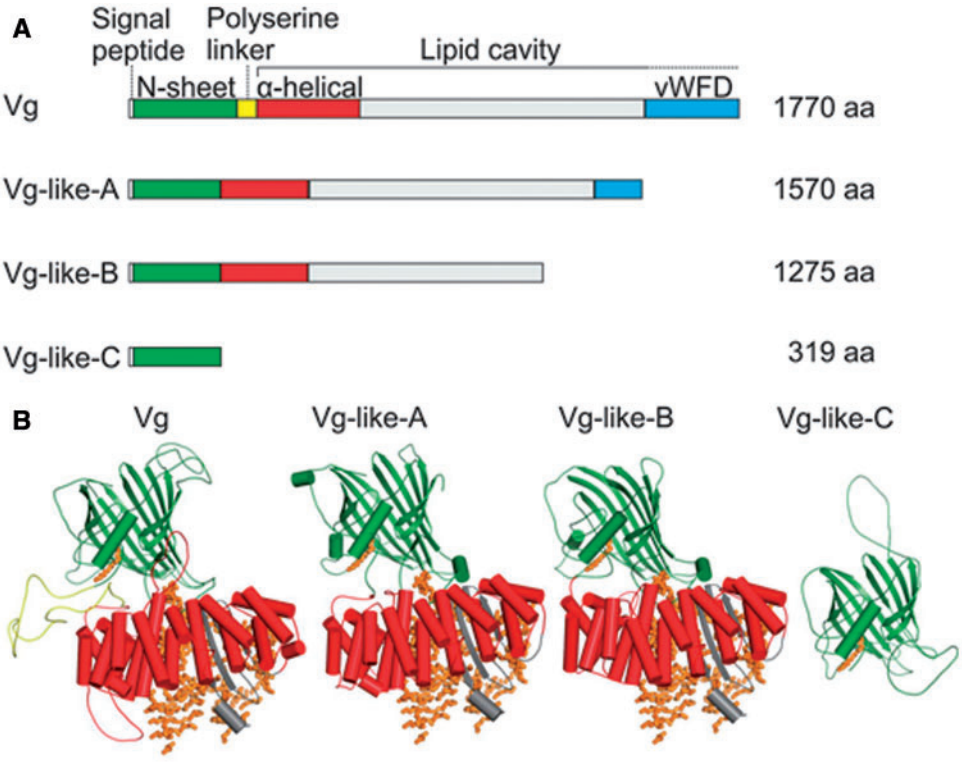

Fig. 1.—The domain architecture of Vg and Vg-like proteins in the honey bee. (*A*) A bar presentation of the full amino acid sequences. The domains N-sheet, α-helical, and vWFD are highlighted, as well as the polyserine linker. The approximate area that carries lipid molecules is indicated as Lipid cavity. (*B*) Partial homology models of the proteins. The lipids (orange) are captured from the template structure (PDB-ID = 1LSH). The colors in *A* correspond to the domain colors in *B*. Most of the gray areas in the lipid cavity and the vWFD are missing in the models (see Materials and Methods).

N-sheet. All Vg-like proteins lack the polyserine linker region following the N-sheet in Vg (fig. 1*A* and *B*). Vg-like-A and -B possess the N-terminal α-helical domain and most large β-sheets forming the lipid-binding cavity of Vg. In the Vg C-terminus, there is a vWFD domain containing a signature amino acid sequence followed by nine cysteine residues (Tufail and Takeda 2008); in Vg-like-A, the cysteine residues were missing. Vg-like-B fully lacked vWFD.

Population Genetic Analysis on vg-Like Genes

We first studied general patterns of amino acid evolution of Vg-like proteins using d*N*/d*S* (table 2). We did not find evidence of accelerated amino acid evolution, which may indicate positive selection, across the entire sequence of the Vg-like genes, although Vg-like-B's N-terminal domain did show significant patterns of accelerated evolution. However, both adaptive and neutral process can influence d*N*/d*S* ratios (Harpur and Zayed 2013; reviewed by Kent and Zayed 2015), and several characteristics of social insects may particularly distort d*N*/d*S* ratios (Kapheim et al. 2015), including biased gene conversion which is especially strong in *Apis* (Kent et al. 2012). We thus carried out more extensive population genetic tests of adaptive evolution in *Apis*.

We found evidence for positive selection acting on all or some parts of Vg-like proteins using a Bayesian implementation of the MK, which estimates the average population-size scaled selection coefficient on replacement mutations (gamma). Significant positive selection (i.e., large and positive gamma estimates) was found for the N-terminal domains (the N-sheet domain and the α-helical domain, the latter is present in Vg-like-A and -B only) of all three proteins, Vg-like-A, -B, and -C (table 2). Additionally, other parts of Vg-like-A are also under positive selection given the overall high gamma estimated across its entire amino acid sequence. Evidence of positive selection on the N-terminal domains of the Vg-like proteins using the MK test is corroborated by population genetic summary statistics, including negative values of Tajima's *D* and Fu & Li's *D*.

We mapped the fixed amino acid differences between *A. mellifera* and *A. cerana* in the N-terminal areas, for which we have reliable homology models of the Vg-like proteins (fig. 2). There were 40 fixed changes in Vg-like-A, 6 in Vg-like-B, and 20 in Vg-like-C. Certain functionally interesting areas appear to be enriched in amino acid changes in the Vg-like-A and -C protein structures. One such area is a bundle of loops that connect β-sheets in the N-sheet domain; 9/40 of fixed changes are located here in Vg-like-A and 5/20 in Vg-like-C (fig. 2). Another area is close to the hypothetical lipophilic ligand in Vg-like-C with 5/20 fixed changes (fig. 2). In Vg-like-A α-helical domain, the changes also appear to be enriched in the N-terminal beginning of the α-helical domain (7/40 changes). In Vg-like-B, the fixed changes appear to be evenly distributed.

Gene Expression and Test for Anti-Inflammatory and Anti-Oxidative Function

To compare seasonal changes in gene expression, we verified the *vg* and *vg-like* gene expression in two sets of honey bee workers—one set collected in winter (winter bees) and another collected in summer (summer nurse bees) (fig. 3). *Vg-like-A* shows a 13-fold and *vg* an11-fold increase in expression in winter bees, compared with summer bees, whereas the corresponding ratios for *vg-like-B* and *vg-like-C* are 1.4, and 0.4, respectively. Within seasonal cohorts the expression of *vg* was 16-fold compared with *vg-like-A* in winter bees, and 20-fold in summer nurse bees. In comparison with *vg-like-B*, the expression of *vg* was143-fold in winter bees, but only 18-fold in summer nurse bees. Compared with *vg-like-C*, the expression of *vg* was over 10,000 times higher in winter bees and 358 times higher in summer nurse bees. Thus, *vg* and *vg-like-A* both showed considerably increased expression levels in winter bees in conjunction with little seasonal changes in the *vg* to *vg-like-A* expression ratio. In contrast, winter bees showed negligible upregulation in *vg-like-B*, whereas the expression of *vg-like-C* in winter bees was 75% of that in summer nurse bees, hence, the difference in expression between these two genes and *vg* was markedly larger in winter bees than in summer nurse bees (fig. 3).

We tested whether expression changes could be provoked by Paraquat injection (oxidative stress) or by the inflammatory effect of a sham injection alone in summer nurse bees (fig. 4 and tables 3 and 4), and found a significant upregulation of *vg* and *vg-like-A* in the injected nurses, in both the Sham and the Paraquat treatment groups, compared with the noninjected Control group. The upregulation of *vg-like-A* was significantly higher than that of *vg*, both in the Sham treatment ($t = -3.3342$, DF = 538, $P = 0.0009$; table 4) and in the

Table 2
Summary Statistics for Molecular Evolution in the Vg-Like Genes

	VgA-whole	VgA-N[a]	VgB-whole	VgB-N	VgC-whole	VgC-N
Selection coeff (γ)	1.163*	2.699*	−0.687	2.660*	0.908	1.885*
Tajima's *D*	−1.158	−1.043	−1.761	−2.28005*	−0.653	−1.514
Fu & Li's *D*	−0.969	−1.018	−1.761	−3.47505*	−0.661	−1.640
π (*Apis. mellifera*)	0.008	0.008	0.000	0.000	0.024	0.021
K_s (divergence)	0.184	0.252	0.050	0.054	0.224	0.245
d*N*/d*S*[b]	0.202	0.186	0.160	0.121*	0.155	0.151

[a]*N* represents the N-terminal area visualized in figure 1; the N-sheet domain (Vg-like-A, -B, and -C) together with the α-helical domain (Vg-like-A and -B).
[b]Average d*N*/d*S* over sister species.
*Significant values.

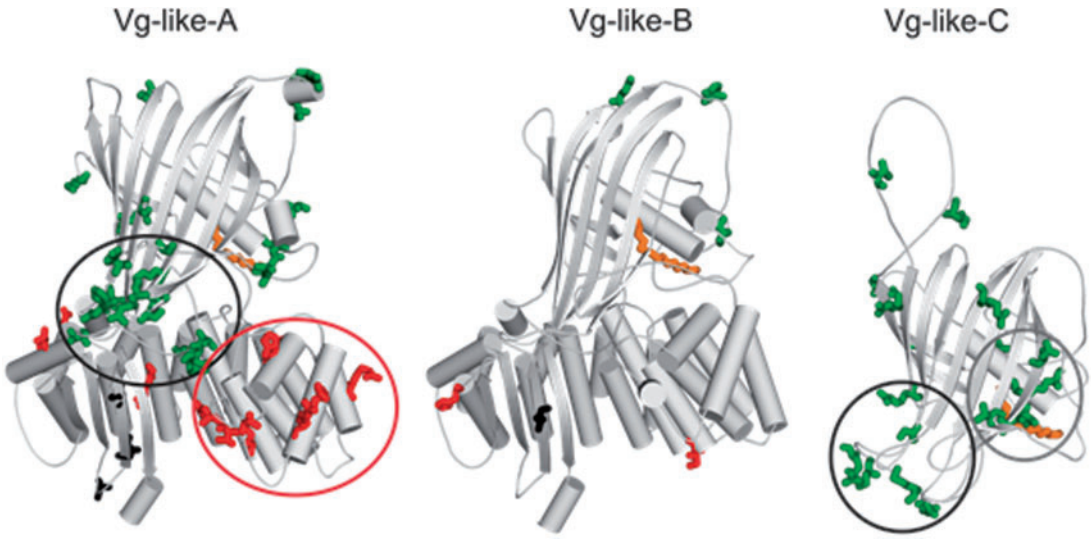

Fig. 2.—Location of the fixed amino acid differences between *Apis mellifera* and *Apis cerana* in Vg-like-A, -B, and -C proteins. The protein models of *A. mellifera* are shown in light gray. The amino acid residues with fixed changes are shown as sticks colored green in the N-sheet domain, red in the α-helical domain, and black in the remaining area (partial lipid cavity). For clarity, only the single lipid putatively located in the N-sheet domain is shown (orange). Some structural elements with known or putative functional importance have several fixed changes that are encircled: A loop area in the N-sheet (black circle), an area near the putative lipophilic ligand (gray circle in Vg-like-C), and an area in the α-helical domain (red circle in Vg-like-A).

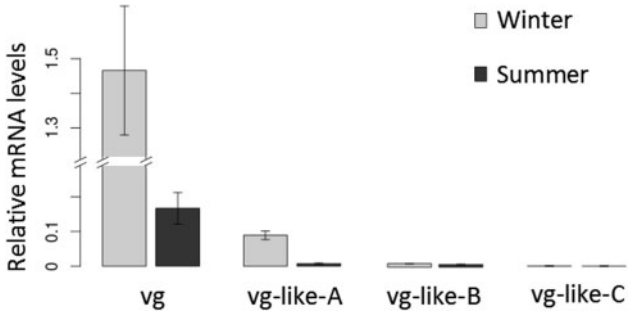

Fig. 3.—The expression of *vg* and the *vg-like*-genes in winter bees (W) and summer nurses (S) (mean ± standard error of the mean, $N = 30$ winter bees from 3 hives and $N = 49$ summer nurses from 5 hives, with 9–10 individuals per hive). The expression was measured by qPCR and normalized to a reference gene (*RP49* or *actin*). The summer nurses shown here are the same as the Control group in figure 4.

Paraquat treatment ($t = -3.3033$, DF = 538, $P = 0.0010$; Table 4). There was no significant difference between the Sham and the Paraquat treatment, which suggests that injection trauma alone can upregulate *vg* and *vg-like-A*. In the case of *vg-like-B*, the Paraquat exposure induced a significant gene upregulation, compared with the noninjected Control. However, neither the difference between the Control and Sham treatments, nor that between the Sham and Paraquat treatments was significant. There were no significant expression differences between the treatment groups for *vg-like-C*.

Discussion

Honey bee Vg is a multifunctional hemolymph protein that promotes longevity (reviewed by Munch et al. 2013). Here, we show that in the honey bee the homologous genes, *vg-like-A, vg-like-B*, and *vg-like-C*, have diverged both from *vg* and from each other, not only by their gene and protein architecture, but also by their evolutionary pattern and their function. We identify *vg-like-A* as the closest structural and functional relative to *vg*, with a likely role in honey bee aging.

We found that the genomic organization, such as chromosomal location, size, and the exon and intron content, differs between *vg* and the *vg-like* genes. This probably reflects their ancient origin in insects (Morandin et al. 2014). Only *vg-like-B* and *-C* were found to be located in the same chromosome in the honey bee, yet they are not the evolutionarily closest homologs. *Vg-like-A* and *-B* are more closely related, and *vg-like-C* has probably evolved from a more recent *vg* duplication event, according to the phylogeny by Morandin et al. (2014).

Our predictions based on amino acid sequence identified a secretion signal peptide and several glycosylation sites typical for secreted proteins (Cao et al. 2009) in all three Vg-like proteins. In addition, we found peptide hits for Vg-like-A and Vg-like-C outside the fat body. These findings suggest that Vg-like proteins apparently exit their production site and enter the hemolymph. The main production site is most likely the fat body, as Vg and most secreted proteins are produced in this tissue (Arrese and Soulages 2010). The presence of the N-sheet domain in all three Vg-like proteins

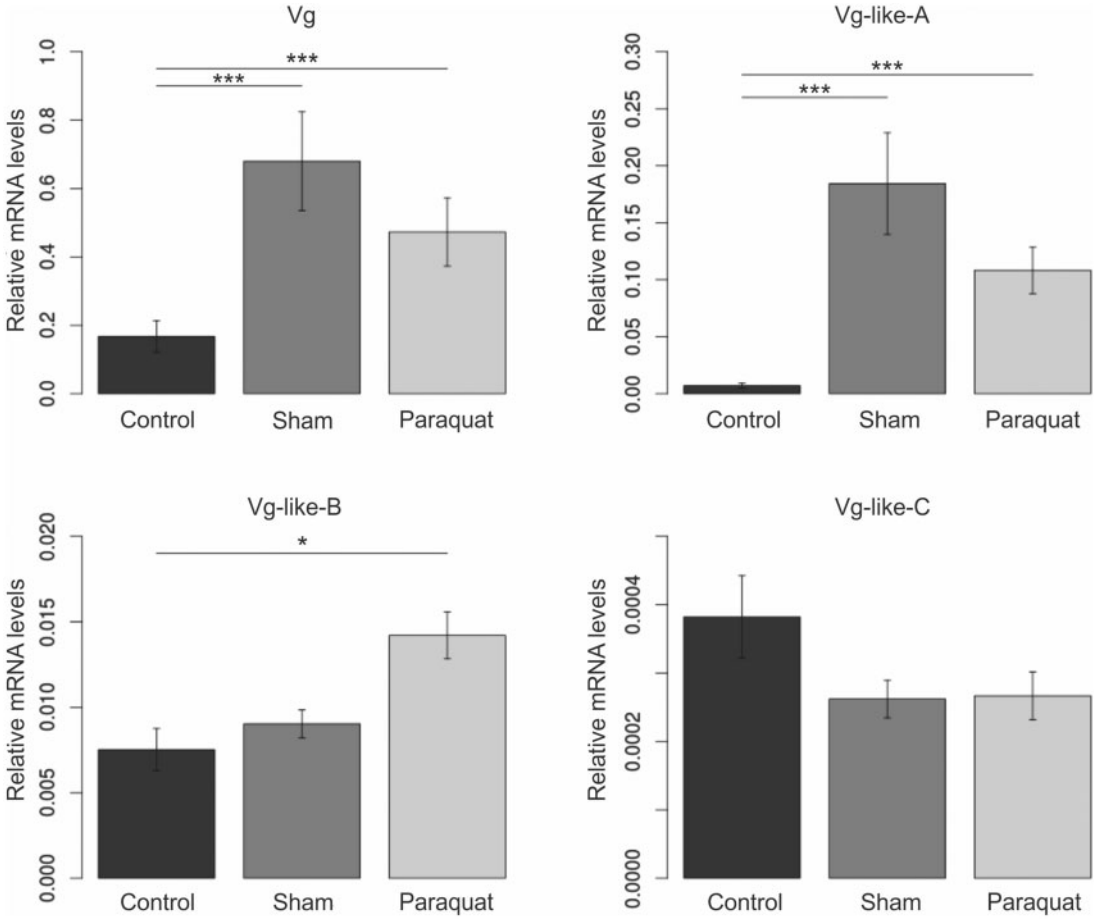

Fig. 4.—The expression of *vg* and the *vg-like* genes in summer nurse bees (mean ± standard error of the mean, $N = 49$, 44, and 49 for Control, Sham, and Paraquat, respectively, with 7–11 individuals from five hives). The three treatment groups were: Control (noninjected), Sham (injected with bee saline), and Paraquat (injected with oxidative poison Paraquat in bee saline). Significant differences are indicated (0.05 is denoted by *; <0.01 is denoted by ***). The expression was measured by qPCR and normalized to a reference gene.

Table 3

Pairwise Comparisons of Gene Expression Differences between the Three Treatment Groups for Each Gene in Summer Nurse Bees

Gene	Treatment Comparison	Estimate	Standard Error	*Z* Value	*P* Value
vg	Control–Sham	−1.77020	0.49000	−3.613	0.000546***
	Control–Paraquat	−2.01643	0.47383	−4.256	4.69×10^{-5}***
	Sham–Paraquat	0.24622	0.48764	0.505	0.690318
vg-like-A	Control–Sham	−4.09851	0.49776	−8.234	7.99×10^{-16}***
	Control–Paraquat	−4.25438	0.48423	−8.786	$<2.00 \times 10^{-16}$***
	Sham–Paraquat	0.15587	0.48999	0.318	0.794542
vg-like-B	Control–Sham	−0.53782	0.49481	−1.087	0.383639
	Control–Paraquat	−1.13756	0.48178	−2.361	0.029812*
	Sham–Paraquat	0.59974	0.48475	1.237	0.324015
vg-like-C	Control–Sham	0.47177	0.48721	0.968	0.401082
	Control–Paraquat	0.46004	0.47642	0.966	0.401082
	Sham–Paraquat	0.01174	0.48710	0.024	0.980779

*is denoted by 0.05
***is denoted by <0.01

Table 4

Effect of Treatments on the Expression of *Vitellogenin* and Differences in the Expression of *vg-Like* Genes Compared with *Vitellogenin*

Effect	Estimate (± standard error)	*t* value[a]	*P* value
Intercept	4.77 ± 0.44	t_{538} = 10.906	<0.0001
Sham	−1.77 ± 0.49	t_{538} = −3.613	0.0003
Paraquat	−2.02 ± 0.47	t_{538} = −4.256	<0.0001
vg-like-A	4.74 ± 0.48	t_{538} = 9.782	<0.0001
vg-like-B	2.89 ± 0.48	t_{538} = 5.962	<0.0001
vg-like-C	7.06 ± 0.48	t_{538} = 14.815	<0.0001
Sham × *vg-like-A*	−2.33 ± 0.7	t_{538} = −3.334	0.0009
Paraquat × *vg-like-A*	−2.24 ± 0.68	t_{538} = −3.303	0.001
Sham × *vg-like-B*	1.23 ± 0.7	t_{538} = 1.77	0.077
Paraquat × *vg-like-B*	0.88 ± 0.68	t_{538} = 1.301	0.194
Sham × *vg-like-C*	2.24 ± 0.69	t_{538} = 3.246	0.001
Paraquat × *vg-like-C*	2.48 ± 0.67	t_{538} = 3.686	0.0003

NOTE.—Estimated from a linear mixed model with d*Ct* values as dependent variable and colony as random effect.

[a]Subscripts indicate the degrees of freedom.

suggests that they could be taken up from the hemolymph by some tissues, given that the N-sheet contains the Vg receptor binding area (Roth et al. 2013). The Vg receptor transfers Vg to tissues located in the head, the ovaries, and the hypopharyngeal glands, among others (Guidugli-Lazzarini et al. 2008). However, further studies are needed to assess whether the same receptor facilitates the uptake of both Vg and Vg-like proteins, and to pinpoint the precise location of the Vg-like proteins.

Our peptide search suggests that Vg-like-A and -C are more abundant in the sterile workers than in the queen. Consequently, these proteins are unlikely candidates for egg yolk formation, which is one of the best described roles of Vg (Tufail and Takeda 2008). Furthermore, heavy phosphorylation is associated with the uptake of Vg to the eggs (Miller et al. 1982; Raikhel and Dhadialla 1992), but we show here that the number of predicted phosphorylation sites in all three honey bee Vg-like proteins is low compared with Vg, and they lack the highly phosphorylated polyserine linker structure. Previously, we have shown that the *vg-like* genes have relatively low expression levels in ant queens compared with ant workers (Morandin et al. 2014), which also suggests that Vg-like proteins are specialized to tasks other than egg yolk formation in social insects.

Based on homology modeling, Vg-like-A appears to resemble Vg more than the other Vg-like proteins by its protein structure. The presence of all structural elements with the exception of the polyserine linker might enable Vg-like-A to perform many tasks similar to those of Vg. Vg-like-B, on the other hand, has lost the polyserine linker, a vWFD domain of unknown function, and has a slightly truncated lipid cavity. These structural differences suggest that Vg-like-B can only perform a restricted subset of Vg's functional repertoire. Vg-like-C is strikingly different from the other homologs by containing a single domain, the N-sheet. The absence of a large lipid cavity suggests that this protein is unlikely to be involved in lipid metabolism, although the N-sheet in a vertebrate Vg is known to carry one lipid molecule (Anderson et al. 1998). Vg-like-C also lacks the α-helical domain, which mediates the anti-inflammatory actions of Vg (Havukainen et al. 2013). The Vg-like-C structure suggests a strict specialization to one or more limited tasks has taken place. This is further corroborated by the fact that Vg-like-C is only found in the scape, suggesting that it has possibly acquired a new neurobiological function.

Our population genetic analysis revealed significant positive selection in the entire *vg-like-A* gene, and in the area corresponding to the protein N-terminus in all the *vg-like* genes. In contrast, in *vg*, positive selection has been shown to affect the lipid cavity, but not the N-terminus (Kent et al. 2011). The fixed nonsynonymous differences that we mapped on the Vg-like protein models were concentrated to certain hotspot areas in Vg-like-A and Vg-like-C, such as one of the loop regions in the N-sheet domain. The role of these loops is unclear, but they may be involved in interactions with neighboring macromolecules (Havukainen, Halskau, Skjaerven, et al. 2011). In Vg-like-A, there was another hotspot for fixed amino acid differences in the α-helical domain, linked to Vg's anti-inflammatory actions (Havukainen et al. 2013). Changes in the α-helical domain could, possibly, modify a response to inflammation, but this remains speculative. In Vg-like-C, there was a hotspot surrounding a putative lipid ligand, which could be interpreted as an adaptation to different ligand molecules. We identified few fixed differences in Vg-like-B, which is in accordance with our previous phylogenetic analysis of insect *vg-like*-genes, where *vg-like-B* was found in the only branch in the phylogenetic tree that has not experienced a period of accelerated amino acid evolution after *vg* duplications (Morandin et al. 2014).

Our gene expression analysis provides the first evidence that all three *vg-like* genes are expressed in the honey bee, both in winter bees and summer nurses. The levels of gene expression are similar to previous findings in ants (Morandin et al. 2014); *vg* is expressed most heavily, followed by *vg-like-A* and *-B*, whereas the expression of *vg-like-C* is very modest. We also found that the expression of *vg-like-A* is higher in winter bees than in nurses—a pattern similar to *vg* in the honey bee (Dainat et al. 2012). This change is consistent with *vg-like-A* being linked to the longevity of winter bees, as is the case for *vg* (Munch and Amdam 2010; Munch et al. 2013). These results notwithstanding, we found that an increase in *vg-like-A* expression can also be triggered in summer nurse bees subjected to an injection. Vg is an anti-inflammatory protein that attaches to damaged cells (Havukainen et al. 2013), hence, it was expected that *vg* gene would react to a sham injection. This was observed also in the case of *vg-like-A*.

Indeed, both injections of saline and Paraquat induce a higher level of *vg-like-A* gene upregulation than for *vg*. Possibly, *vg-like-A* could be linked to the melanization reaction that is induced by injection. A Vg-related protein is involved in melanin synthesis in the mealworm *Tenebrio molitor* (Lee et al. 2000).

In addition to a protective role under inflammation and oxidative stress, Vg-like-A may also have other roles. The study by Shipman et al. (1987) suggests that Vg-like-A is a lipid transporter during metamorphosis, as this protein accumulated in hemolymph at the end of the larval stages (larval-specific very high density lipoprotein in that study). Also *vg* is transcribed in young developmental stages (Piulachs et al. 2003). Possibly, both Vg and Vg-like-A have a dual role in basic lipid trafficking during various life stages and, on the other hand, in protection against inflammation and oxidative stress. Lipoproteins of animals, including human, are known to have such dual roles in lipid metabolism and defense (Seong and Matzinger 2004; Cho and Seong 2009).

In this study, the effect of injection of bee saline alone did not significantly upregulate *vg-like-B* or *vg-like-C* expression. However, the joint effect of injection and oxidative stress (i.e., injection of Paraquat) leads to *vg-like-B* upregulation. A speculative interpretation is that *vg-like-B* has a specific role during oxidative stress, but is not needed to modify moderate inflammation, unlike *vg* and *vg-like-A* that respond to any inflammatory trauma caused by injection. *Vg-like-C* differed from the other genes analyzed here by its unresponsiveness to any injection treatments. The localization in the antenna as suggested by our peptide search, the dwarf size, and the low expression level all suggest that *vg-like-C* might have subfunctionalized to one of the roles of ancient *vg*, and this role is likely not related to aging. In subfunctionalization, a gene copy adopts a function of the ancestral pleiotropic gene after duplication (Hughes 1994). Hypothetically, *vg-like-C* could be involved in carrying a small hydrophobic ligand in the antenna. This speculation, however, calls for more experiments.

In summary, we show that Vg-like genes/proteins share structural similarities with Vg, but each of them also has unique gene/protein architecture and a distinct pattern of positive selection, which are indicative of a divergence of functions. Based on our gene expression studies, we suggest that *vg-like-A* has a similar longevity supporting role as *vg*, and that *vg-like-C* is unrelated to antiaging actions, but likely highly specialized to an unknown role. This subfunctionalization of the gene copies can help to understand the rapid historical and ongoing evolution of such a pleiotropic gene as *vg* (Havukainen, Halskau, Amdam 2011; Kent et al. 2011; Morandin et al. 2014). We also show that all copies of *vg* and *vg-like* genes are subject to positive selection, but that the precise target sites of the proteins are different from those identified in Vg. Indeed, it has been suggested that genes can evolve rapidly after duplication (Chen et al. 2013), and acquire novel functions or specialize on one function within a set of many. The ant-specific *vg* copies are differentially expressed in the queen and the workers (Wurm et al. 2011; Corona et al. 2013; Feldmeyer et al. 2014), which indicates that *vg* can acquire new functions via duplication even in an evolutionarily short time. Our population genetic evidence further supports this hypothesis.

Supplementary Material

Supplementary tables S1 and S2 are available at *Genome Biology and Evolution online* (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Academy of Finland (decision numbers 265971 [to H.S.], and 252411 and 284666 [to L.S.]); the University of Helsinki; the Finnish Beekeepers'Association (to T.S.); a Discovery Grant from the Natural Sciences and Engineering Council of Canada (to A.Z.), and by an Early Researcher Award from the Ontario Ministry of Research and Innovation (to A.Z.). We thank the Stadin Tarhaajat Beekeepers' Association and, in particular, Merja-Riitta Laurila for the winter bee samples. Special thanks to the research coordinator Tuula Lehtonen and the executive director Heikki Vartiainen at the Finnish Beekeepers' Association for support. We thank Dr Ying Wang at Arizona State University, Dr Eva Hystad at Norwegian University of Life Sciences, and the technicians Leila Miettinen, Minttu Ahjos, Leena Laaksonen, Heini Ali-Kovero, and Annukka Ruokolainen at University of Helsinki for expert advice.

Literature Cited

Amdam GV, Simoes ZL, Guidugli KR, Norberg K, Omholt SW. 2003. Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. BMC Biotechnol. 3:1–8.

Amdam GV, et al. 2005. Social reversal of immunosenescence in honey bee workers. Exp Gerontol. 40:939–947.

Anderson TA, Levitt DG, Banaszak LJ. 1998. The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. Structure 6:895–909.

Arias MC, Sheppard WS 1996. Molecular phylogenetics of honey bee subspecies (*Apis mellifera L.*) inferred from mitochondrial DNA sequence. Mol Phylogenet Evol. 5(3):557–66.

Arrese EL, Soulages JL. 2010. Insect fat body: energy, metabolism, and regulation. Annu Rev Entomol. 55:207–225. doi: DOI 10.1146/annurev-ento-112408-085356

Atzmon G, et al. 2002. Plasma HDL levels highly correlate with cognitive function in exceptional longevity. J Gerontol A Biol Sci Med Sci. 57:M712–M715.

Aurori C, et al. 2014. What is the main driver of ageing in long-lived winter honeybees: antioxidant enzymes, innate immunity, or vitellogenin? J Gerontol A Biol Sci Med Sci. 69(6):633–639.

Avarre JC, Lubzens E, Babin PJ. 2007. Apolipocrustacein, formerly vitellogenin, is the major egg yolk precursor protein in decapod crustaceans and is homologous to insect apolipophorin II/I and vertebrate apolipoprotein B. BMC Evol Biol. 7:3.

Bustamante CD, et al. 2002. The cost of inbreeding in *Arabidopsis*. Nature 416:531–534.

Camero S. Hines H. Williams P. 2007. A comprehensive phylogeny of the bumble bees (Bombus). Biological J Linnean Soc 91(1):161–188.

Cao J, et al. 2009. Identification of N-glycosylation sites on secreted proteins of human hepatocellular carcinoma cells with a complementary proteomics approach. J Proteome Res. 8:662–672.
Chan QW, Parker R, Sun Z, Deutsch EW, Foster LJ. 2011. A honey bee (*Apis mellifera L.*) PeptideAtlas crossing castes and tissues. BMC Genomics 12:290.
Chen S, Krinsky BH, Long M. 2013. New genes as drivers of phenotypic evolution. Nat Rev Genet. 14:645–660.
Cho NH, Seong SY. 2009. Apolipoproteins inhibit the innate immunity activated by necrotic cells or bacterial endotoxin. Immunology 128:e479–e486.
Corona M, et al. 2013. Vitellogenin underwent subfunctionalization to acquire caste and behavioral specific expression in the harvester ant *Pogonomyrmex barbatus*. PLoS Genet. 9(8):e1003730.
Dainat B, Evans JD, Chen YP, Gauthier L, Neumann P. 2012. Predictive markers of honey bee colony collapse. PLoS One 7:e32151.
de Miranda JR, Fries I. 2008. Venereal and vertical transmission of deformed wing virus in honeybees (*Apis mellifera* L.). J Invertebr Pathol. 98:184–189.
Dietemann V, Ellis JD, Neumann P. 2013. The COLOSS BEEBOOK, Volume I: standard methods for *Apis mellifera* research. Cardiff (UK): International Bee Research Association.
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.
Flot JF. 2010. seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. Mol Ecol Resour. 10:162–166.
Fluri P, Wille H, Gerig L, Lüscher M. 1977. Juvenile hormone, vitellogenin and haemocyte composition in winter worker honeybees (*Apis mellifera*). Experientia 33:1240–1241.
Guidugli-Lazzarini KR, et al. 2008. Expression analysis of putative vitellogenin and lipophorin receptors in honey bee (*Apis mellifera* L.) queens and workers. J Insect Physiol. 54:1138–1147.
Harman D. 1956. Aging: a theory based on free radical and radiation chemistry. J Gerontol. 11:298–300.
Harpur BA, Zayed A. 2013. Accelerated evolution of innate immunity proteins in social insects: adaptive evolution or relaxed constraint? Mol Biol Evol. 30:1665–1674.
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. 111:2614–2619.
Hausman DB, Fischer JG, Johnson MA. 2012. Protein, lipid, and hematological biomarkers in centenarians: definitions, interpretation and relationships with health. Maturitas 71:205–212.
Havukainen H, Halskau Ø, Amdam GV. 2011. Social pleiotropy and the molecular evolution of honey bee vitellogenin. Mol Ecol. 20:5111–5113.
Havukainen H, Halskau Ø, Skjaerven L, Smedal B, Amdam GV. 2011. Deconstructing honeybee vitellogenin: novel 40 kDa fragment assigned to its N terminus. J Exp Biol. 214:582–592.
Havukainen H, Underhaug J, Wolschin F, Amdam G, Halskau O. 2012. A vitellogenin polyserine cleavage site: highly disordered conformation protected from proteolysis by phosphorylation. J Exp Biol. 215:1837–1846.
Havukainen H, et al. 2013. Vitellogenin recognizes cell damage through membrane binding and shields living cells from reactive oxygen species. J Biol Chem. 288:28369–28381.
Honeybee Genome Sequencing Consortium. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443:931–949.
Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. Proc Biol Sci. 256:119–124.
Johnson BR. 2010. Division of labor in honeybees: form, function, and proximate mechanisms. Behav Ecol Sociobiol. 64:305–316.
Johnson M, et al. 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36:W5–W9.
Kapheim KM, et al. 2015. Genomic signatures of evolutionary transitions from solitary to group living. Science 348:1139–1143.
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. 20:5226–5235.
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 U S A. 109:18012–18017.
Kent CF, Zayed A. 2015. Population genomic and phylogenomic insights into the evolution of physiology and behaviour in social insects. Adv Insect Physiol. 48:293–324.
Lee K, et al. 2000. Activated phenoloxidase from *Tenebrio molitor* larvae enhances the synthesis of melanin by using a vitellogenin-like protein in the presence of dopamine. Eur J Biochem. 267(12):3695–3703.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.
Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402–408.
Mann CJ, et al. 1999. The structure of vitellogenin provides a molecular model for the assembly and secretion of atherogenic lipoproteins. J Mol Biol. 285:391–408.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351:652–654.
McGeer PL, McGeer EG. 2004. Inflammation and the degenerative diseases of aging. Ann N Y Acad Sci. 1035:104–116.
McKay GJ, et al. 2011. Variations in apolipoprotein E frequency with age in a pooled analysis of a large group of older people. Am J Epidemiol. 173:1357–1364.
Miller MS, Benore-Parsons M, White HB 3rd. 1982. Dephosphorylation of chicken riboflavin-binding protein and phosvitin decreases their uptake by oocytes. J Biol Chem. 257:6818–6824.
Morandin C, et al. 2014. Not only for egg yolk—functional and evolutionary insights from expression, selection, and structural analyses of *Formica* ant vitellogenins. Mol Biol Evol. 31:2181–2193.
Munch D, Amdam GV. 2010. The curious case of aging plasticity in honey bees. FEBS Lett. 584:2496–2503.
Munch D, Kreibich CD, Amdam GV. 2013. Aging and its modulation in a long-lived worker caste of the honey bee. J Exp Biol. 216:1638–1649.
Piulachs MD, et al. 2003. The vitellogenin of the honey bee, *Apis mellifera*: structural analysis of the cDNA and expression studies. Insect Biochem Mol Biol. 33:459–465.
R Core Team 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/
R Core Team 2015. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/
Raikhel AS, Dhadialla TS. 1992. Accumulation of yolk proteins in insect oocytes. Annu Rev Entomol. 37:217–251.
Roth Z, et al. 2013. Identification of receptor-interacting regions of vitellogenin within evolutionarily conserved beta-sheet structures by using a peptide array. Chembiochem 14:1116–1122.
Seehuus SC, Norberg K, Gimsa U, Krekling T, Amdam GV. 2006. Reproductive protein protects functionally sterile honey bee workers from oxidative stress. Proc Natl Acad Sci U S A. 103:962–967.
Seong SY, Matzinger P. 2004. Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses. Nat Rev Immunol. 4:469–478.

Shipman BA, Ryan RO, Schmidt JO, Law JH. 1987. Purification and properties of a very high density lipoprotein from the hemolymph of the honeybee *Apis mellifera.* Biochemistry 26:1885–1889.
Sievers F, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7:539.
Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 68:978–989.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612.
Tufail M, Takeda M. 2008. Molecular characteristics of insect vitellogenins. J Insect Phys. 54:1447–1458.
Wang Y, Brent CS, Fennern E, Amdam GV. 2012. Gustatory perception and fat body energy metabolism are jointly affected by vitellogenin and juvenile hormone in honey bees. PLoS Genet. 8:e1002779.
Winston ML, 1987. Biology of the Honey Bee. Cambridge (MA): Harvard University Press.
Wurm Y, Wang J, Keller L. 2010. Changes in reproductive roles are associated with changes in gene expression in fire ant queens. Mol Ecol. 19(6):1200–11.
Wurm Y, et al. 2011. The genome of the fire ant *Solenopsis invicta*. Proc Natl Acad Sci U S A. 108(14):5679–5684.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Associate editor: Dorothee Huchon