Characterization of Quantitative Trait Loci for the Age of First Foraging in Honey Bee Workers

By: Olav Rueppell

Rueppell O. (2009) Characterization of quantitative trait loci for the age of first foraging in honey bee workers. Behavior Genetics, *39(5), 541-553*

Abstract:
Identifying the basis of quantitative trait loci (QTL) remains challenging for the study of complex traits, such as behavior. The honey bee is a good model combining interesting social behavior with a high recombination rate that facilitates this identification. Several studies have focused on the pollen hoarding syndrome, identifying multiple QTL as the genetic basis of its behavioral components. One component, the age of first foraging, is central for colony organization and four QTL were previously described without identification of their genomic location. Enabled by the honey bee genome project, this study provides data from multiple experiments to scrutinize these QTL, including individual and pooled SNP mapping, sequencing of AFLP markers, and microsatellite genotyping. The combined evidence confirms and localizes two of the previous QTL on chromosome four and five, dismisses the other two, and suggests one novel genomic region on chromosome eleven to influence the age of first foraging. Among the positional candidates the *Ank2*, *PKC*, *Erk7*, and *amontillado* genes stand out due to corroborating functional evidence. This study thus demonstrates the power of combined, genome-based approaches to enable targeted studies of a manageable set of candidate genes for natural behavioral variation in the important, complex social trait "age of first foraging".
Keywords: Genetic architecture, Complex traits, Social insects, Life history, Division of labor, Behavioral ontogeny, AFLP, Foraging onset

Article:

Introduction

Most biologically-relevant traits, particularly behavior, are determined by a complex interplay between genes and the environment, which makes variation in both and their interactions influential for the observed phenotypes (Robinson et al. *2008*). This observation together with the development of genetic high-throughput techniques that allow genome-wide studies in a wide array of organisms have let to a recent surge of studies on complex traits. Their genetic architecture can be studied by quantitative trait loci (QTL) mapping either in defined crosses or natural populations (Vasemagi and Primmer *2005*). Numerous studies have elucidated the number and strength of genetic influences on many traits of interest, but a remaining challenge is to identify the underlying genes and characterize the specific genetic variation that is responsible for the complex traits (Phillips *1999*). With the completion of an increasing number of genome projects, this transition from QTL to a specific genetic element is becoming feasible in more systems but progress has been slow, partly because an identified QTL region corresponds to hundreds of potential candidate genes in most species (Mackay *2001*).

In contrast, QTL regions in the honey bee (*Apis mellifera* L.) are physically small and tend to contain relatively few genes, facilitating systematic evaluation of the potential candidates (Hunt et al. *2007*). The small physical size of QTL intervals is due to the exceptionally high recombination rate of the honey bee genome (Beye et al. *2006*; Solignac et al. *2007*). The honey bee is a long-established model in behavioral research, including behavioral genetics (Page et al. *2002*) and results of its genome project have been published (Honeybee Genome Consortium *2006*). The honey bee exhibits many complex behavioral phenotypes that result from the

species' ecology and sociobiology, and pronounced natural phenotypic variation for most traits in this species provides a good basis for forward genetic approaches, particularly QTL analyses.

QTL analyses have been performed for a broad range of behavioral traits in honey bees, including defensiveness (Hunt et al. *1998*), guarding (Arechavaleta-Velasco and Hunt *2004*), learning (Chandra et al. *2001*), hygienic behavior (Lapidge et al. *2002*), ethanol sensitivity (Ammons and Hunt *2008*), and worker sterility (Lattorff et al. *2007*; Oxley et al. *2008*). However, the most concerted QTL mapping effort has been directed toward understanding the genetics of the pollen hoarding syndrome. This suite of correlated behavioral and physiological traits was originally identified in artificially selected high- and low pollen-hoarding strains (Page and Fondrk *1995*) but is of general importance to understanding honey bee behavior and social evolution (Pankiw *2003*; Page and Amdam *2007*). The pollen hoarding syndrome includes behavioral choices of foragers, sucrose and light sensitivity, learning, locomotor activity, ovary size, and the age at which workers transition from in-hive tasks to foraging (=age of first foraging: AFF) (Page and Erber *2002*; Rueppell et al. *2004*, *2006a*; Humphries et al. *2005*; Page and Amdam *2007*).

Initially, two QTL (*pln1* and *pln2*) were mapped based on the colony-level pollen hoarding phenotype and they were reconfirmed by studying foraging choices in individual workers (Hunt et al. *1995*). Two more *pln* QTL were identified (*pln3* and *pln4*) in subsequent studies (Page et al. *2000*; Rüppell et al. *2004*) based on their effects on foraging choices in individual workers. Moreover, some of these QTL have shown pleiotropic effects on sucrose response and AFF (Rueppell et al. *2004*, *2006a*). The *pln* QTL regions have been located in the honey bee genome sequence and revealed some promising candidate genes that invoke the insulin/insulin-like signaling cascade and ecdysteroid signaling (Hunt et al. *2007*; Wang et al. *2009*), linking the pollen hoarding syndrome to female reproductive physiology (Page and Amdam *2007*).

In addition, the selected pollen hoarding strains have been used to study the genetic architecture of a 66% difference in the AFF between the high and low pollen hoarding strains (Rueppell et al. *2004*). This study revealed four significant QTL, two in a backcross to the high pollen hoarding line and two in the reciprocal backcross to the low pollen hoarding line (Rueppell et al. *2004*). The study showed that these four *aff* QTL have no overlap with the *pln* QTL regions, but their own genomic location could not be determined because the analyses were performed with Amplified Fragment Length Polymorphism (AFLP™) markers (Rueppell et al. *2004*). This has precluded follow-up studies on potential candidate genes, and hence the present study was designed to localize the *aff* QTL.

As a benchmark of the age-based division of labor among honey bee workers, the AFF is the central variable in the life history of honey bee workers, determining colony food intake and individual worker mortality (Rueppell et al. *2007*, *2008*). It is an exemplary complex trait determined by an interaction of environmental (Schulz et al. *1998*), social (Le Conte and Hefetz *2008*), physiological (Nelson et al. *2007*), and genetic (Robinson and Huang *1998*; Robinson *2002*) influences.

Gene expression analyses have identified thousands of differentially expressed genes between the in-hive nurse bee and forager states (Whitfield et al. *2003*). The gene expression patterns seem to be somewhat conserved in the genus *Apis* (Sen-Sarma et al. *2007*) but identification of consistently up- or down-regulated individual genes has proven problematic (Adams et al. *2008*). Moreover, it is clear that the transition between in-hive tasks and foraging needs to be decoupled from chronological age and early developmental processes for meaningful analyses (Whitfield et al. *2006b*). Comprehensive experimental studies of gene expression patterns, including physiological manipulations and comparisons between subspecies (Whitfield et al. *2006b*) have been able to narrow the list of expression candidates for the transition from nursing to foraging. These gene expression data and the functional annotation of honey bee genes (Honeybee Genome Consortium *2006*) now provide effective criteria for evaluating positional candidate genes in localized QTL regions for the AFF (Wayne and McIntyre *2002*). Therefore, I conducted multiple studies to re-evaluate and localize the reported AFF QTL (Rueppell et al. *2004*), search the corresponding genomic intervals for candidate genes, and assess their overlap with previously suggested functional candidates for the pollen hoarding syndrome and particularly AFF.

Methods and material

The first step in this study was to re-analyze the original mapping populations that led to the characterization of four AFF (age of first foraging) QTL (Rueppell et al. *2004*). In brief, these populations were reciprocal backcrosses derived from the "High Pollen Hoarding" and "Low Pollen Hoarding" strains (High and low strain, respectively) after 18 generations of bi-directional selection for the amount of pollen stored in the hive (Page and Fondrk *1995*), as described in Rueppell et al. (*2004*). After emergence in an incubator, workers were color marked, introduced into a common hive environment and monitored for their AFF. The individuals were collected when returning from their first foraging trip and stored for DNA extraction and genetic analyses (Rueppell et al. *2004*).

Selective individual SNP genotyping

Forty individual workers with extreme phenotypes from each of the original high and low backcross populations (HBC and LBC, respectively: Rueppell et al. *2004*) were chosen (Fig. 1) and genotyped at 280 SNPs with Sequenom's™ MALDI-TOF mass spectrometry technique (Ragoussis et al. *2006*). These SNPs were selected from 1,136 verified SNPs (Whitfield et al. *2006a*) to provide the best genomic coverage. Of the 280 SNP markers, 94 and 95 proved variable in the HBC and LBC, respectively. The differences in AFF between the genotypic SNP classes were evaluated by Mann–Whitney *U*-tests, followed by a simple Bonferroni correction for multiple comparisons because the sparse genomic coverage made these association tests essentially independent (Lander and Botstein *1989*).

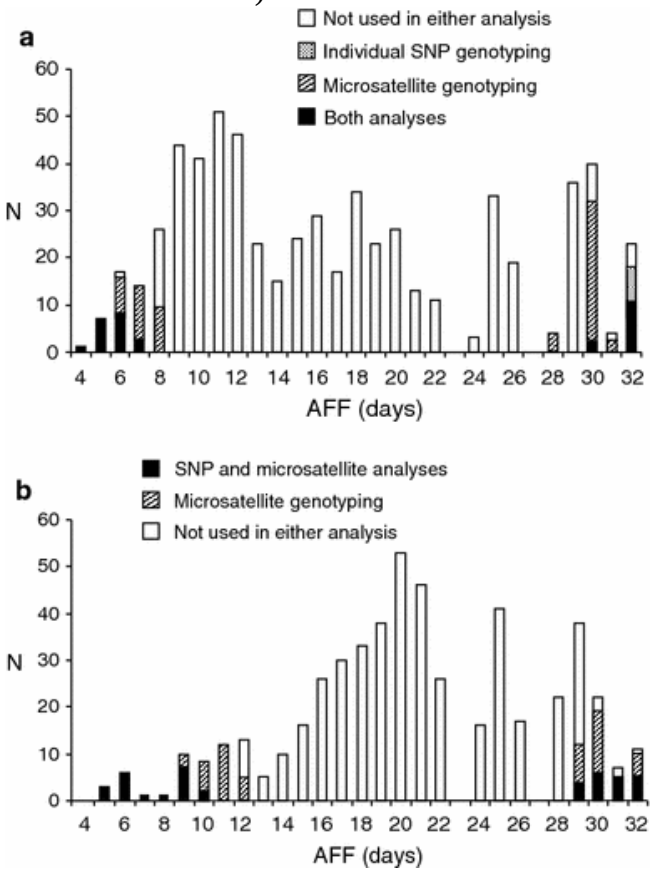

Fig. 1 ***Phenotypic distribution of the age of first foraging (AFF) in the high backcross (a) and the reciprocal low backcross (b) with the subsamples used for the different analyses***

Linkage map construction

Two new linkage maps were constructed with Mapmaker 3.0b (Lander and Botstein *1989*; Lincoln et al. *1993*). The new SNP markers were ordered based on their genomic location in the *Apis mellifera* genome 4.0 and then used as a linkage backbone to assign the previously genotyped AFLP markers (Rueppell et al. *2004*) to the 16 chromosomes. The initial default linkage criteria were a LOD > 4 and a linkage distance <37.5 cM (Kosambi mapping function was used throughout this study; Solignac et al. *2007*). The remaining markers were tested for

linkage (LOD > 2) to flanking markers of coverage gaps (>30 cM) and to markers at the ends of the linkage groups whenever comparison to the physical genome sequence indicated a lack of coverage by the genetic map. Throughout the process, linkage distances between SNP markers and the total recombination size of each chromosome was compared to existing estimates (Beye et al. *2006*; Solignac et al. *2007*). AFLP markers that led to significant (>5 cM and >10% of the respective linkage interval) map expansion were excluded.

Markers that had shown a significant effect in a previous analysis (Rueppell et al. *2004*) but were unlinked at this point of the analysis were placed to their most likely map location, if linkage to any mapped marker was suggested with LOD > 4. Finally, map order was checked by the ripple command (Lincoln et al. *1993*), locally permuting all potential map orders. Based on the constructed linkage maps, a new QTL analysis for AFF was performed on the combined SNP and AFLP data. I used the computer program MapQTL4.0 (Van Ooijen et al. *2002*) for interval mapping with empirically generated LOD significance thresholds by bootstrapping, as in previous analyses (Rueppell et al. *2004*).

Pooled SNP analysis

Additionally, I performed a SNP analysis on pooled samples derived from the HBC and LBC with all 1,136 verified SNPs (Whitfield et al. *2006a*). From each backcross, I generated three low AFF and three high AFF pools with 13–15 individuals from the phenotypic extremes (fractioned-pool design: Korol et al. *2007*). These individuals were overlapping to a large degree with the above-described samples that were individually genotyped. Before pooling, individual DNA concentrations were determined in duplicate on a Nanodrop™ spectrophotometer. The DNA pools were made up with equal DNA contributions from each individual and standardized to 100 ng/μl where possible. The pooled DNA was genotyped on an Illumina® BeadArray™ Reader as in previous studies of these SNPs (Whitfield et al. *2006a*).

The results were evaluated in two different ways. First, I used the automatically generated SNP genotypes to determine which SNPs were most differentiated between the low and high pools. The two alternative genotypes were coded as 1 and −1 and a differentiation index between the low and high tails was calculated as: , with H_n representing the genotypic value of the *n*th high pool and L_n representing the genotypic value of the *n*th low pool.

Second, I used the raw signal data to estimate allele frequencies in the pooled samples to test for marker-QTL linkage (Darvasi and Soller *1994*). The allele frequencies were estimated using a modified *k*-correction factor (Hoogendoorn et al. *2000*) that was based on the average signal ratios of both homozygous and heterozygote genotype classes, determined from independently genotyped individuals which were kindly provided by Charlie Whitfield (University of Illinois). To determine the minor allele frequency, P_a, for each locus in each sample pool, the difference in signal ratio between the heterozygote and the corresponding homozygote was determined as , where *A* and *B* are the two alternative signal values corresponding to the alternative alleles *a* and *b* from the homozygous (bb) or heterozygous (ab) genotype class, with . Then the allele frequency in each pool is computed as ; with P_a in [0–0.5]. In the backcross design this allele frequency can be directly translated to the number of heterozygous individuals in the pool by multiplication with 2*n*, with *n* representing the number of individuals in the DNA pool.

A simple χ^2 test combining the three pools from each phenotypic extreme was used to test for non-random genotypic composition between the two phenotypic extremes. Furthermore, the average phenotype of the alternative genotypic classes can be computed at each locus for a conservative estimate of single marker significance and allele substitution effect by assigning the average pool phenotype to each individual in each pool (Darvasi and Soller *1994*). Loci were excluded if their minor allele frequency estimates combined from all pools deviated significantly ($P < 0.01$) from the expected value of 0.25. With an overall sample size of 80 in the combined HBC pools and 90 in the LBC the acceptable range for P_a was [0.175–0.325] and [0.178–0.322], respectively. This resulted in the inclusion of 207 SNPs in the HBC analysis and 212 SNPs in the LBC analysis.

Sequencing of AFLP markers

AFLP bands that were linked to *aff*1–4 (for details on AFLP generation and scoring see Rueppell et al. *2004*) were cut out of the dried poly-acrylamide gel for cloning and sequencing to determine their genomic location. The bands were cut based on their autoradiograph and their excision was verified by re-exposure of the gel. For each of the four AFF QTL, bands from 4 to 6 different markers were cut and re-amplified using the specific AFLP primers (Rueppell et al. *2004*). The re-amplified product was run on a low melting point agarose gel and extracted from the gel with a Geneclean™ kit (Qbiogene, CA). The extracted DNA was cloned into a p-GemT™ vector (Promega, WI) and sequenced on an ABI Prism™ sequencer (Applied Biosystems, CA) after direct PCR sequencing attempts had failed due to multiple sequences present. The cloned sequences were located in the *Apis mellifera* genome 4.0 by BLASTN searches with default settings at the BeeBase server (http://racerx00.tamu.edu/).

Microsatellite reconfirmation

Based on the results of the individual and pooled SNP genotyping, the QTL mapping, and genomic location of the AFLP band sequences, I specifically tested nine genomic regions in the HBC and seven in the LBC for a genetic effect on AFF with microsatellite markers. Microsatellite loci (Solignac et al. *2007*) were selected based on their location, amplification, and variability in the respective cross. Select 96 individuals (Fig. 1) of extreme phenotypes in either the HBC or LBC were genotyped. I used a tailed-primer approach (Schuelke *2000*), labeling microsatellite alleles with a universal M13-primer coupled to IRD700 or IRD800, for detection with a DNA Analyzer 4300™ (Licor Inc., Lincoln, NB). Alleles were amplified with a touchdown PCR protocol, decreasing the annealing temperature from 68 to 48°C (Schug et al. *2004*). PCR reactions were carried out in 10 µl: 1 µl of 10 ng/µl of template DNA was added to a master mix that contained 1 µl of 2 mM dNTPs, 0.25 µl of 10 µM forward primer, 0.5 µl of 10 µM reverse primer, 0.5 µl of 1 µM of IRD-labeled M13 primer, 1 µl of 10× PCR buffer (20 mM $MgCl_2$, 100 mM Tris–HCl [pH 8.3], 500 mM KCl), 0.04 µl of 5 u/µl Taq DNA polymerase, and 5.71 µl dH_2O. PCR products of different size and IRD label were combined and analyzed on 25 cm long gels run at 1,000 V for 2–3 h. Genotypes were scored in duplicate. The significance of differences in AFF between the two genotypic classes was evaluated by Mann–Whitney *U*-tests.

Candidate gene evaluation

The evidence from a former study (Rueppell et al. *2004*), and current QTL mapping, single marker genotyping, and AFLP sequencing were combined to suggest three QTL with specified genomic locations. Candidate genes were evaluated in the 97% confidence interval of these QTL by comparison with the list of predicted genes in these genomic regions (Hunt et al. *2007*). The positional candidates were further evaluated functionally by comparison to their *Drosophila* homologs (http://www.flybase.org/) and other closest homologs (searching the NCBI and GeneOntology databases). The candidate gene list was also compared to candidate gene lists derived from gene expression studies on the transition of honey bee workers from in-hive duties to foraging (Whitfield et al. *2003*, *2006b*).

Results

Selective individual SNP genotyping

Stronger genetic effects were indicated by the SNP markers in the HBC than in the LBC, with one marker (*ahb2647*, group 11.37) effect being significant even after Bonferroni correction (Table 1). Only one marker was significant in the LBC before Bonferroni correction. All top SNP markers were incorporated into the joint SNP/AFLP genetic maps (below) as predicted by their physical location in the genome, except for the LBC marker *est3350*, which mapped to the fourth chromosome in the center of AFF3 (below) instead of the predicted location on chromosome three.

Table 1 ***SNP markers with the strongest genetic effects in the HBC and LBC, from approximately 100 variable SNPs genotyped in 40 individuals in each cross***

Backcross	SNP #	Genome location (scaffold in Amel4.0)	Effect	Uncorrected significance
HBC	ahb2647	11.37	$Z_{(17,22)} = 3.69$	$P < 0.0005$

Backcross	SNP #	Genome location (scaffold in Amel4.0)	Effect	Uncorrected significance
	est159	1.18	$Z_{(17,23)} = 3.28$	$P = 0.001$
	est8741	11.41	$Z_{(15,24)} = 3.01$	$P = 0.004$
LBC	est3350	3.10	$Z_{(16,24)} = 2.38$	$P = 0.017$

Joint AFLP/SNP map and QTL mapping

The initial, chromosome-building step involved 94 variable SNP markers in the HBC and 95 in the LBC, of which respectively 12 and 11 were excluded because they did not link to the adjacent markers, as predicted by their physical location. The total recombination size of these backbones were 2,817 cM and 2,739 cM, respectively (Table 2). The incorporation of 223 additional AFLP markers in the HBC and 230 in the LBC increased the overall map sizes (genome coverage) to 4,218 cM and 3,745 cM, respectively (Table 2), which is comparable to other recent map size estimates of the honey bee genome (Beye et al. *2006*; Solignac et al. *2007*).

Table 2 ***Genomic coverage from SNP and AFLP markers for QTL mapping of AFF***

Chrom.	Number of SNPs	Size (cM)	Scaffold range	Total number of markers	Total size (cM)	Number of gaps >30 cM
HBC						
1	10	400	1.08–1.75	29	521	6
2	7	216	2.14–2.43	30	439	4
3	6	172	3.18–3.31	31	296	1
4	5	202	4.08–4.27	24	320	4
5	8	239	5.03–5.33	20	283	2
6	7	256	6.22–6.51	23	324	3
7	4	174	7.10–7.37	14	233	3
8	5	167	8.07–8.35	21	277	1
9	4	185	9.02–9.25	14	271	3
10	3	112	10.19–10.45	5	106	0
11	5	149	11.14–11.41	12	218	2
12	5	175	12.20–12.30	26	279	2
13	5	232	13.05–13.17	15	237	3
14	4	93	14.13–14.24	16	184	1
15	3	45	15.26–15.33	22	188	2
16	1	N/A	16.19	5	42	0
Sum	82	2,817	N/A	307	4,218	37
LBC						
1	17	483	1.07–1.82	47	612	5
2	7	258	2.23–2.43	21	272	1
3	5	158	3.10–3.26	26	247	1
4	4	159	4.08–4.23	24	234	0
5	9	177	5.12–5.33	18	200	1
6	4	251	6.07–6.51	17	262	2

Chrom.	Number of SNPs	Size (cM)	Scaffold range	Total number of markers	Total size (cM)	Number of gaps >30 cM
7	4	181	7.10–7.37	18	278	2
8	4	160	8.07–8.35	20	226	2
9	4	142	9.02–9.26	16	217	2
10	2	72	10.19–10.33	16	253	2
11	5	148	11.14–11.41	13	196	2
12	4	100	12.20–12.30	13	125	1
13	8	221	13.01–13.17	15	228	3
14	5	137	14.13–14.24	20	217	1
15	2	92	15.26–15.33	10	129	0
16	1	N/A	16.19	6	47	0
Sum	85	2,739	N/A	300	3,743	25

In the HBC, the QTL analysis suggested two regions of major effect within wide linkage gaps on chromosomes six (66 cM gap between scaffolds 6.35 and 6.44, LOD = 47.4) and ten (62 cM gap between scaffolds 10.19 and 10.33, LOD = 26.8). The first region was associated with the marker that had the strongest single marker effect but was unlinked in the previous analysis: E5M3445 (Rueppell et al. *2004*), the second QTL was not linked to any previously reported marker. The end of chromosome eleven, containing two of the most significant single SNP markers (Table 1) showed only a modest LOD score of 1.4, similar to the genomic region around *pln1* (Rueppell et al. *2004*). The two previous QTL *aff1* and *aff2* were contained in two separate AFLP linkage groups that lacked linkage to any SNP marker. Thus, they were without genomic localization and therefore not considered further.

In the LBC, one significant QTL (LOD = 3.4) was displayed in the middle of a linkage gap on chromosome six (>70 cM gap between scaffolds 6.7 and 6.21). The previously identified *aff3* was recovered on chromosome four (scaffold 4.21) due to close linkage to one SNP marker (*ahb7032*). The incorporation of this SNP marker led to a reduced LOD score of 2.2 in this region (Fig. 2a), with the chromosome-wide significance threshold being 2.1. The inclusion or omission of the stray SNP *est3350* did not change the results. AFLP markers associated with *aff4* made up a small linkage group without initial linkage to any chromosome. However, this group could be inserted into a linkage gap between SNP markers on scaffold 5.12 and 5.21 (based on the AFLP sequence data: see below), and the resulting linkage group showed a LOD score peak of 3.1 (Fig. 2b), with a chromosome-wide significance threshold of 2.0.

Pooled SNP analysis

Based on the automatically generated genotypes, the most differentiated SNPs had a differentiation index of four (Table 3). Among these, four came from the HBC and only one from the LBC. The single marker analysis based on the raw signal ratios of the pooled samples resulted in four significant χ^2-values for separate SNP markers: In the HBC, allelic composition of the high and low AFF pools were significantly different at *est7677* (scaffold 10.37; $\chi^2 = 28.9$, $df = 1$, $P_{\text{uncorrected}} < 0.00001$) and at *ahb2647* (scaffold 11.37; $\chi^2 = 14.5$, $df = 1$, $P_{\text{uncorrected}} = 0.00012$), with corresponding allelic substitution effects of 14.8 (95% CI: 10.4–19.3) and 10.5 (5.4–15.5) days. In the LBC, SNPs *ahb1791* (scaffold 10.15) and *est5056* (scaffold 6.29) showed significant allelic effects (9.8 (5.4–14.1) days, $\chi^2 = 16.6$, $df = 1$, $P_{\text{uncorrected}} = 0.00005$; and 9.2 (4.8–13.6) days, $\chi^2 = 14.8$, $df = 1$, $P_{\text{uncorrected}} = 0.00012$, respectively).

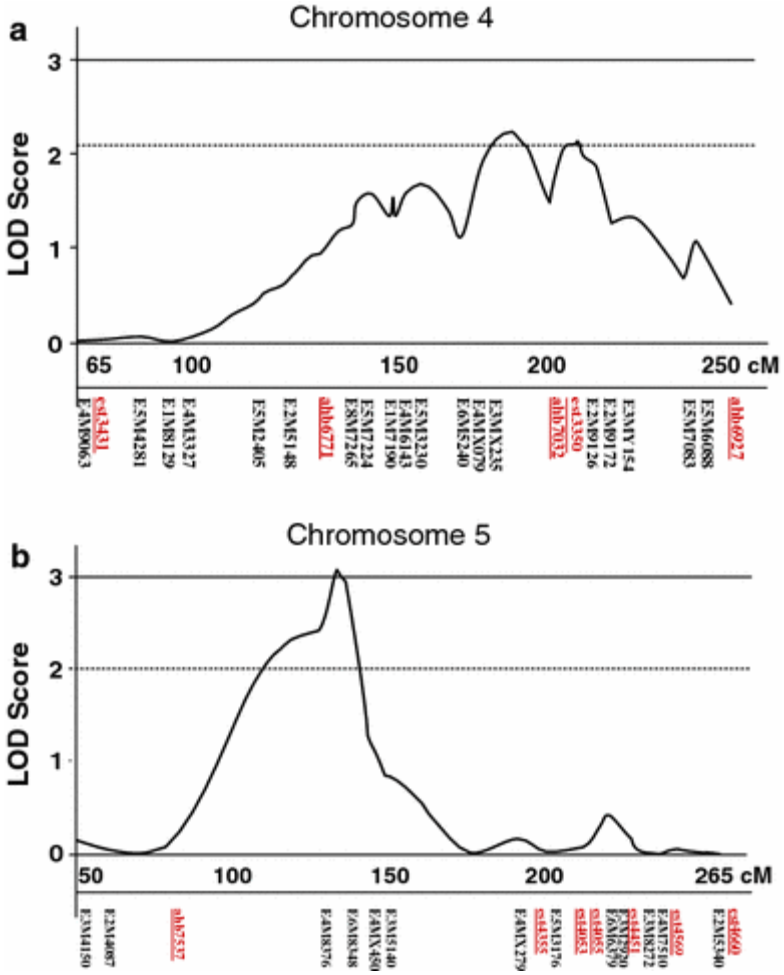

Fig. 2 ***Recovered QTL from the previous mapping of the age of first foraging (AFF) in honey bee workers. a AFF3 was located on chromosome four, albeit with a reduced LOD score, through joint SNP and AFLP mapping and sequencing of the original AFLP markers. b AFF4, located through sequencing of the original AFLP bands, was placed into a corresponding linkage gap in the combined SNP and AFLP map on chromosome five. Approximate genetic genomic positions (in centi-Morgans: cM) are indicated***

Table 3 ***Most differentiated SNPs between high and low AFF DNA pools, based on automatic genotyping***

SNP	Scaffold	Backcross	Low pool genotypes[a]	High pool genotypes[a]
ahb1573	Group01.08	HBC	AB, AB, AB	NC, NC, BB
ahb176	Group01.64	HBC	AB, AB, AB	BB, AB, BB
est6434	Group08.35	HBC	BB, BB, BB	AB, BB, AB
ahb2647	Group11.37	HBC	AB, AB, AB	BB, AB, BB
ahb2046	Group10.19	LBC	AB, AB, NC	BB, BB, NC

[a]***"NC" indicates an intermediate signal ratio that cannot be reliably scored as either genotype, indicating a relatively even mixture of genotypes***

Sequencing of AFLP markers

The cloning of AFLP bands associated with AFF1-4 yielded sequences that were associated with different genomic locations (Table 4). The most likely position of the QTL was assumed to be the genome region with most sequence matches to the AFLP bands. *aff1* and *aff2* had two matches each for scaffold 6.35 and 1.79, respectively. *aff3* was matched to 4.18–4.22 by four consecutive hits and *aff4* to 5.18–5.20 by three hits.

Table 4 *Genomic localization (Amel4.0 scaffold and E-value of BLASTN sequence match) of sequences of AFLP-bands linked to previous AFF QTL*

AFF1	AFF2	AFF3	AFF4
E4M4375 **Group6.35 (6e-70)**	E3M2201 Group1.11 (1e-38)	E7M7700 **Group4.18 (1e-59)**	E3M5140 GroupUn.1865 (9e-64) **Group5.20 (1e-31)**
E6M4177 **Group6.35 (5e-72)**	E6M5259 **Group1.79 (1e-57)**	E4MX079 Group9.10 (6e-20)	E4M10450 **Group5.19 (0.0)**
E1M4270 GroupUn.s (1e-70) Group1.66 (e-113) Group15.29 (1e-45)	E2M4093 Group3.1 (2e-29) Group3.36 (1e-30) Group5.12 (4e-24)	E3M10235 **Group4.18 (e-112)** Group7.35 (e-103)	
E3M3297 Group7.44 (e-110)	E3M2155 Group2.14 (1e-62) **Group1.79 (3e-54)**	E2M9126 Group13.14 (5e-78)	E4M8376 Group8.18 (e-108) **Group5.18 (e-157)**
		E2M9172 Group7.35 (e-103) **Group4.18 (3e-40)**	
		E3M11154 **Group4.22 (1e-56)**	

Several sequences were found multiple times and instances where genetic and physical maps coincide are bold-faced

Microsatellite reconfirmation

Many of the AFF candidate genomic regions that were suggested by the previous analyses could be ruled out based on the individual genotyping with microsatellites (Table 5). In the HBC, the QTL region comprising scaffolds 11.33–11.37 could be reconfirmed with a significant effect of the microsatellite marker *K1118*, located on scaffold 11.35. The marker showed an allele substitution effect of 6.7 days. In the LBC, *aff3* and *aff4* were reconfirmed with significant microsatellite effects on scaffold 4.21 (*K0411*, substitution effect: 5.8 (0.9–10.7) days) and 5.18 (new microsatellite OR5.18A, forward primer: CGTGTACACCGATGATCCTG, reverse primer: CGAACATCGCCCATAAAGAT, substitution effect: 4.5 (0.1–8.8) days), respectively (Table 5).

Table 5 *Genotypic effects on AFF of microsatellite markers in selected genome regions with their two-tailed probability*

Locus	Genome region	Reason for study	Back-cross	Mann–Whitney *U* test (significance)[a]
AP106	1.11	AFLP sequencing (AFF2), Pooled SNP analysis	HBC	$Z_{(32,56)} = -0.1$, $P = 0.944$
SV137	1.63	Pooled SNP analysis	HBC	$Z_{(50,44)} = -1.2$, $P = 0.224$
AT154	1.79	AFLP sequencing (AFF2)	HBC	$Z_{(51,45)} = -0.2$, $P = 0.833$
UN338	3.01	AFLP sequencing (AFF2)	HBC	$Z_{(42,54)} = -0.1$, $P = 0.904$
UN125T	3.36	AFLP sequencing (AFF2)	HBC	$Z_{(52,44)} = -0.3$, $P = 0.729$
OR6.38A	6.38	AFLP sequencing (AFF1) and QTL mapping	HBC	$Z_{(53,39)} = -0.5$, $P = 0.641$
K0756	7.44	AFLP sequencing (AFF1)	HBC	$Z_{(47,48)} = -1.0$, $P = 0.331$
UN392B	10.23	QTL mapping	HBC	$Z_{(50,45)} = -1.5$, $P = 0.141$
K1118	**11.35**	**Strongest individual SNP effect in HBC,**	**HBC**	**$Z_{(53,43)} = -2.6$, $P = 0.009$**

Locus	Genome region	Reason for study	Back-cross	Mann–Whitney *U* test (significance)[a]
		pooled SNP analysis		
AP106	1.11	Pooled SNP analysis	LBC	$Z_{(39,29)} = -0.0$, $P = 0.985$
K0358	3.10	Strongest individual SNP effect in LBC	LBC	$Z_{(45,48)} = -0.8$, $P = 0.404$
K0411	**4.21**	**AFLP sequencing (AFF3) and QTL mapping**	**LBC**	$\mathbf{Z_{(32,41)} = -2.0, P = 0.049}$
OR5.18A	**5.18**	**AFLP sequencing (AFF4) and QTL mapping**	**LBC**	$\mathbf{Z_{(50,41)} = -2.1, P = 0.038}$
AT153	6.13	QTL mapping	LBC	$Z_{(47,44)} = -0.7$, $P = 0.459$
AT202	6.29	Pooled SNP analysis	LBC	$Z_{(45,47)} = -0.1$, $P = 0.887$
AP033	10.15	Pooled SNP analysis	LBC	$Z_{(40,49)} = -1.6$, $P = 0.120$

[a]***Significant markers are printed in bold***

Candidate gene evaluation

The 97% confidence interval of *aff3* was estimated to contain the scaffolds 4.18–4.22 based on their original LOD-score traces (Rueppell et al. *2004*) combined with the positional information from sequencing the AFLP markers. For *aff4*, the 97% CI was similarly determined to include scaffolds 5.16–5.19. For the newly determined genetic effect in the HBC on the 11th chromosome, a meaningful LOD-score trace could not be calculated, and therefore, I evaluated genes within an approximated interval of 1.8 million base pairs surrounding the most significant marker effects (11.33–11.40).

For *aff3*, 56 putative genes were evaluated for their functional significance between positions 8.29 Mb and 9.92 Mb on chromosome four. This gene list contained three hypothetical loci without recognizable homologs, and five genes of unknown function. Among the functional candidates the following loci are noteworthy: genes involved in the nervous system, *Ank2* (LOC409051), *cheerio* (LOC409697), *chameau* (LOC408793); cell signaling components: *Rgk2* (LOC411209), *G-protein α-subunit* (LOC411011), *Sac1* (LOC411013), and the putative homolog of *ERK7*, a MAP (serine/threonine-protein) kinase (LOC408917); the immunity gene *cactus* (LOC411012), the putative pheromone receptor *GR68a* (LOC725297), and *dur-1* (LOC724697), which is up-regulated in *C. elegans* dauer stages, due to starvation.

For AFF4, the evaluated CI spanned from 8.0 to 9.3 Mb on chromosome five and included 34 genes as positional candidates, four of which were hypothetical loci without homologs and seven genes of unknown function. Potential functional candidates include *sas* (LOC412968), a receptor involved in larval molting, the CG9778 homolog (LOC412965), which is involved in neurotransmitter secretion, the peptide prohormone processing enzyme *amontillado* (LOC408835), and two serine/threonine-protein kinases similar to *widerborst* (LOC551258) and CG14305 (LOC411093), respectively.

In the QTL region on chromosome eleven, the selected 1.8 million base pairs (from 12.2 to 14 Mb) contained 143 predicted genes. Thirteen of these were hypothetical loci and 14 had no assignable molecular or biological function. Among the other genes, I identified 27 functional candidate genes, including components of signaling cascades, such as *1D-myo-inositol-trisphosphate 3-kinase* (Ip3 K), *protein kinase C* (Pkc), *Akt1* (LOC413430), *radish* (LOC409307), and a *Ras* related protein (LOC413339); genes involved in nervous system function, such as *acetylcholinesterase* (*ache*: LOC410270), *stoned* (LOC552808), and *ether a go-go* (*eag*: LOC552834); genes involved in gene regulation, such as *sec5* (LOC726595), *haywire* (LOC551734), and a histone acetyltransferase (LOC412705) and -methyltransferase (LOC411458); and genes involved in hormonal signaling, such as a signal peptide protease (LOC409870), *pro-hormone 4* (LOC409241), and the homolog of CG32432 (LOC724460). The selection of these subsets from the relatively long list of positional candidate genes based on their putative function is necessarily subjective and none of the positional candidate genes can be excluded based on this

study. Information on all positional candidate genes can be directly accessed through the NCBI Map viewer for the honey bee (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=7460).

Discussion

With several complementary approaches, enabled by the honey bee genome (Honeybee Genome Consortium *2006*), it was possible to localize three quantitative trait loci for the fundamental life-history transition in honey bee workers from in-hive tasks to foraging. In contrast to correlational gene expression differences which may be highly variable (Adams et al. *2008*), these QTL are causal for the observed behavioral variation. The transition from in-hive tasks to foraging, measured as the age of first foraging (AFF), is an exemplary complex behavioral trait, but the identified QTL explain a substantial amount of the total phenotypic variation. Maximum single marker estimates for the explained variation are 16.3% (AFF3) and 14.8% (AFF4) in the LBC and 29.5% (new QTL on chromosome 11), although these values are presumably overestimates of the QTL effects in the general population due to the selective genotyping of the samples.

The current study presents a significant advance over a previous QTL mapping study of the AFF (Rueppell et al. *2004*) because the genome-derived, sequence-specific markers have permitted a more stringent mapping approach, the genomic localization of the QTL, and their reconfirmation with additional markers. The results enable a direct comparison of positional candidate genes to candidate lists from gene expression studies and functional follow-up studies on the most promising candidates (Phillips *1999*; Mackay *2001*; Hunt et al. *2007*; Wang et al. *2009*). With *aff3* and *aff4*, two of four previously identified QTL were recovered and located on the genome, while the other two could not be confirmed. Instead, I identified another significant QTL on chromosome 11, between scaffolds 11.33 and 11.40.

Both recovered QTL (*aff3* and *aff4*) had been detected in the backcross to the low pollen hoarding strain with modest effect size (Rueppell et al. *2004*). Here, the combined data from AFLP-band sequencing, QTL mapping, and targeted microsatellite genotyping provide a high degree of confidence about their existence and genomic location. Support for the newly reported QTL on chromosome 11 is also strong due to the combined evidence of the individual SNP, pooled SNP, and microsatellite genotyping. The calculated allele substitution effects at the microsatellite markers in the range of 4–7 days are considerable, given an average AFF of 16 and 21 days in the HBC and LBC, respectively (Rueppell et al. *2004*). However, it is not clear how accurate these estimates of the substitution effects are. On the one hand, they represent overestimates due to selective genotyping (Darvasi and Soller *1994*). On the other hand, these markers are not necessarily representing the peak of the QTL effect and consequently have a smaller allele substitution effect than the QTL itself. Comparatively, the allelic effects calculated from the pooled DNA genotyping clearly represent overestimates due to their more selective genotyping.

The QTL mapping based on an incomplete joint AFLP-SNP linkage map, the AFLP band sequencing, and the SNP genotyping of selective, pooled DNA samples all implicated more genomic regions then the three final QTL deduced from the combination of the experiments. However, there was little overlap between these additional candidate regions and any single line of evidence coming from these experiments may be flawed for the following reasons. While a strict mapping protocol proved efficient to anchor a large number of anonymous AFLP markers to relatively few, positioned SNP markers with a restricted sample size (Table 2), over 1/3 of the previously mapped AFLP markers (Rueppell et al. *2004*) could not be anchored to any chromosome. These included markers that were associated with the original *aff1* and *aff2* QTL (Rueppell et al. *2004*), leading to the rejection of these two QTL as true effects. Furthermore, this study demonstrated the problem of AFLP band collision (Gort et al. *2006*) which is magnified when sequencing AFLP markers. Since only one primer is labeled, many unlabelled DNA fragments of the same length can co-migrate with the targeted marker during electrophoresis and be sequenced instead. These hidden bands explain the heterogeneity of sequences obtained from sequencing the AFLP bands that were found linked to AFF-QTL previously (Rueppell et al. *2004*).

Although containing over 300 markers each, the genomic HBC and LBC maps contained significant coverage gaps. In three of these gaps a significant QTL was indicated by interval mapping. However, all three of these

QTL were refuted by additional microsatellite genotyping, demonstrating the necessity for adequate genomic coverage for interval mapping and the benefits of comparing interval mapping results with single marker analyses.

The pooled SNP analysis was designed to provide a more complete coverage of the genome in a cost-effective way (Darvasi and Soller *1994*; Hoogendoorn et al. *2000*). After a conservative exclusion of a number of markers the remaining markers did not indicate any effect in the QTL regions that were mapped in linkage gaps. However, the pooled SNP analysis suggested four other regions to have significant effects. Among these, the region 11.33–11.40 also contained one of the most differentiated SNPs in the analysis of the automatically generated DNA pool genotypes, two of the most significant individually-scored SNPs, and was further corroborated by microsatellite evidence. In contrast, the remaining three regions that were suggested by the pooled SNP analysis could be refuted based on evidence from individually genotyped SNP or microsatellite markers.

The identified AFF-QTL regions do not coincide with any of the *pln*-QTL or QTL for stinging (Hunt et al. *2007*). The confidence intervals of the AFF-QTL also do not include the candidate genes *malvolio* (Ben-Shahar et al. *2004*), *PKG* (Ben-Shahar et al. *2002*), or *vitellogenin* (Nelson et al. *2007*), but all three regions contain several interesting candidates that warrant future study. While the gene content in AFF3 and AFF4 is moderate, the unexpectedly high number of positional candidate genes for the new AFF QTL on chromosome 11 has made the selection of the most likely candidate more challenging. The selection of candidates is further complicated by the complexity of the AFF trait. The transition from performing various in-hive tasks to foraging outside the hive involves multiple changes that combine social and environmental effects (Schulz et al. *1998*; Le Conte and Hefetz *2008*) with internal processes such as reproductive physiology (Nelson et al. *2007*; Page and Amdam *2007*) and brain reorganization (Fahrbach et al. *1998*), which makes numerous genes potentially influential. However, a few of the listed candidate genes in all three regions stand out due to independent, corroborating evidence.

In AFF3, the *Ank2* gene codes for a structural protein associated with the presynaptic membrane cytoskeleton. In *Drosophila* this gene is essential for synaptic stability (Koch et al. *2008*) and it could therefore play a crucial role in brain reorganization (Fahrbach et al. *1998*) influencing the transition from in-hive to foraging worker. Furthermore, the *Ank2* gene is up-regulated in developing honey bee workers relative to queens (Barchuk et al. *2007*) which suggests that it is a versatile gene that may have been co-opted during the social evolution of workers. The second, best candidate in AFF3 is the protein kinase LOC408917, which is the putative *ERK7* homolog based on best matches in reciprocal TBLASTN analyses. It is one of the top functional candidates due to its expression pattern (Whitfield et al. *2006b*). This gene is highly up-regulated in foragers compared to nurses, even when foraging experience is prevented. It consistently corresponds to the bees treatment with methoprene, a juvenile hormone analog that can influence the AFF (Sullivan et al. *2000*), and its expression is correlated with honey bee subspecies differences in AFF (Whitfield et al. *2006b*).

For AFF4, there is no strong support for any of the positional candidate genes but the most promising candidate is *amontillado*, a peptide hormone processing protease with subtilisin activity that has been found to affect larval hatching behavior (Wilson et al. *2008*). Even though its primary role is probably molting, *amontillado* in *Drosophila* leads to abandonment of larval wandering behavior (Rayburn et al. *2003*), similar to the *for* gene, which has been implicated in honey bee foraging behavior (Ben-Shahar et al. *2002*). Its mRNA is highly enriched in *Drosophila* adult nervous tissues (Chintapalli et al. *2007*) but is also expressed in larval neuroendocrine tissues (Tomancak et al. *2002*). Another possible candidate is LOC412965, whose *Drosophila* ortholog CG9778 is identified as calcium-dependent phospholipid binding protein involved in neurotransmitter secretion and/or vesicle-mediated transport (Wilson et al. *2008*). Its mRNA is 22×-fold enriched in the adult *Drosophila* brain (Chintapalli et al. *2007*). LOC412965 has the eighth most significant difference in gene expression between 17-day old foragers and 17-day old workers that were prevented from foraging and its expression is affected by manganese treatment (Whitfield et al. *2006b*).

The most likely candidate gene in the third region is *PKC*, a serine-threonine kinase that is involved in calcium- or diacylglycerol-dependent intracellular signaling and can be influenced by octopamine (Farooqui *2007*). Octopamine is linked to the onset of foraging in worker bees (Barron et al. *2002*) and there are consistent differences in PKC protein levels between the studied high- and low-pollen hoarding strains (Humphries et al. *2003*). If the different PKC titers represent the QTL effect, its basis could be either a *cis*-acting regulatory mutation or a mutation affecting protein stability. The involvement of PKC in learning and sensory perception further suggests its involvement in the determination of the age of first foraging by modulating internal or external stimuli that influence the onset of foraging (Humphries et al. *2003*).

There are two other good candidate genes in the same area, *acetylcholinesterase* and *eag*. The former shows expression differences in the brain between in-hive and foraging worker bees (Shapira et al. *2001*). It is essential for termination of the signal of cholinergic synapses by hydrolyzing acetylcholine. Behavioral phenotypes of different non-lethal alleles include altered motor activity and phototaxis (Hall et al. *1980*), both of which change during the transition from in-hive tasks to foraging. The *eag* protein displays voltage-gated potassium channel activity and the gene has been associated in *Drosophila* with learning and memory, perception of smell, and flight behavior (Homyk and Sheppard *1977*; Honjo and Furukubo-Tokunaga *2005*), but also transcriptional regulation (Wilson et al. *2008*) that could account for longer-term behavioral changes from in-hive worker to forager.

All QTL regions include a set of candidate genes among which a few stand out due to their phenotypic characterization in *Drosophila* and/or corroborating data from complementary studies in honey bees. However, each QTL region also contains a number of predicted genes with unknown function and negligible homology to other genes. If verified, these genes could prove the most interesting group of genes to explain the evolution of social behavior. In contrast to genes that have retained much of their structure and function across taxa (Robinson et al. *2008*), the reason for their pronounced molecular divergence could be their unique functions in social life.

Conclusion

This study has identified three genomic regions in two crosses between the selected high- and low-pollen hoarding honey bee strains that influence their differences in timing of the onset of foraging. This behavioral trait is part of the pollen hoarding syndrome, a suite of behavioral differences that influences behavioral variation among wild type bees (Amdam et al. *2006*) and subspecies differences (Pankiw *2003*). The pollen hoarding syndrome may therefore be fundamental for understanding social organization in honey bees (Page and Amdam *2007*). The AFF is a complex trait that is affected by many influences and it is to be expected that it shows only a partial genetic overlap with other traits of the pollen hoarding syndrome (Rueppell et al. *2004*). Specifically, the identified QTL do not have explicit ties to female reproductive physiology (Page and Amdam *2007*) and may therefore present other components of the genetic architecture of the pollen hoarding syndrome that are not female specific (Rueppell et al. *2006b*). The timing of the transition from in-hive to forager bees has critical effects on colony energy budget and individual mortality (Rueppell et al. *2007, 2008*). Thus, it is a central trait for colony fitness and social evolution and a similar age-based division of labor can be found in a variety of social insects. The identified QTL regions together with gene expression studies (Whitfield et al. *2006b*) provide a manageable number of candidate genes for detailed studies of the genetic basis of this important, complex social behavioral trait.

Acknowledgments I would like to thank Robert Page (ASU) for breeding and providing the original mapping populations, as well as allowing me to use the software MapQTL, Charlie Whitfield (UIUC) for practical assistance with the pooled SNP genotyping and providing his raw data to enable my analyses, and Greg Hunt (Purdue), Michael Munday (UNCG) and Patrick Nolan (UNCG) for advice and help regarding the analysis of candidate genes. Two careful reviewers significantly improved the quality of this manuscript. Practical help was provided by Kim Fondrk (UCD) and Meredith Humphries (UCD) with cloning the AFLP fragments, and Jackie Metheny (UNCG) and Emily Meznar (UNCG) with microsatellite genotyping. Financial support was provided by NSF (#0615502) and NIH (NIA PO1 AG22500).

References

Adams HA, Southey BR, Robinson GE, Rodriguez-Zas SL (2008) Meta-analysis of genome-wide expression patterns associated with behavioral maturation in honey bees. BMC Genomics 9:503. doi:10.1186/1471-2164-9-503

Amdam GV, Csondes A, Fondrk MK, Page RE Jr (2006) Complex social behaviour derived from maternal reproductive traits. Nature 439:76–78. doi:10.1038/nature04340

Ammons AD, Hunt GJ (2008) Identification of quantitative trait loci and candidate genes influencing ethanol sensitivity in honey bees. Behav Genet 38:531–553. doi:10.1007/s10519-008-9218-z

Arechavaleta-Velasco ME, Hunt GJ (2004) Binary trait loci that influence honey bee (Hymenoptera:Apidae) guarding behavior. Ann Entomol Soc Am 97:177–183. doi:10.1603/0013-8746(2004)097[0177:BTLTIH]2.0.CO;2

Barchuk AR, Cristino AS, Kucharski R, Costa LF, Simoes ZLP, Maleszka R (2007) Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. BMC Dev Biol 7:70

Barron AB, Schulz DJ, Robinson GE (2002) Octopamine modulates responsiveness to foraging-related stimuli in honey bees (*Apis mellifera*). J Comp Physiol [A] 188:603–610. doi:10.1007/s00359-002-0335-5

Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE (2002) Influence of gene action across different time scales on behavior. Science 296:741–744. doi:10.1126/science.1069911

Ben-Shahar Y, Dudek NL, Robinson GE (2004) Phenotypic deconstruction reveals involvement of manganese transporter *malvolio* in honey bee division of labor. J Exp Biol 207:3281–3288. doi:10.1242/jeb.01151

Beye M, Gattermeier I, Hasselmann M, Gempe T, Schioett M, Baines JF, Schlipalius D, Mougel F, Emore C, Rueppell O, Sirvio A, Guzman-Novoa E, Hunt G, Solignac M, Page RE (2006) Exceptionally high levels of recombination across the honey bee genome. Genome Res 16:1339–1344. doi:10.1101/gr.5680406

Chandra SBC, Hunt GJ, Cobey S, Smith BH (2001) Quantitative trait loci associated with reversal learning and latent inhibition in honeybees (*Apis mellifera*). Behav Genet 31:275–285. doi:10.1023/A:1012227308783

Chintapalli VR, Wang J, Dow JAT (2007) Using FlyAtlas to identify better *Drosophila* models of human disease. Nat Genet 39:715–720. doi:10.1038/ng2049

Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138:1365–1373

Fahrbach SE, Moore D, Capaldi EA, Farris SM, Robinson GE (1998) Experience-expectant plasticity in the mushroom bodies of the honeybee. Learn Mem 5:115–123

Farooqui T (2007) Octopamine-mediated neuromodulation of insect senses. Neurochem Res 32:1511–1529. doi:10.1007/s11064-007-9344-7

Gort G, Koopman WJM, Stein A (2006) Fragment length distributions and collision probabilities for AFLP markers. Biometrics 62:1107–1115. doi:10.1111/j.1541-0420.2006.00613.x

Hall JC, Alahiotis SN, Strumpf DA, White K (1980) Behavioral and biochemical defects in temperature sensitive acetylcholinesterase mutants of *Drosophila melanogaster*. Genetics 96:939–965

Homyk T, Sheppard DE (1977) Behavioral mutants of *Drosophila melanogaster*. 1. Isolation and mapping of mutations which decrease flight ability. Genetics 87(9):5–104

Honeybee Genome Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443:931–949. doi:10.1038/nature05260

Honjo K, Furukubo-Tokunaga K (2005) Induction of cAMP response element-binding protein-dependent medium-term memory by appetitive gustatory reinforcement in *Drosophila* larvae. J Neurosci 25:7905–7913. doi:10.1523/JNEUROSCI.2135-05.2005

Hoogendoorn B, Norton N, Kirov G, Williams N, Hamshere ML, Spurlock G, Austin J, Stephens MK, Buckland PR, Owen MJ, O'Donovan MC (2000) Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. Hum Genet 107:488–493.

doi:10.1007/s004390000397

Humphries MA, Muller U, Fondrk MK, Page RE Jr (2003) PKA and PKC content in the honey bee central brain differs in genotypic strains with distinct foraging behavior. J Comp Physiol [A] 189:555–562. doi:10.1007/s00359-003-0433-z

Humphries MA, Fondrk MK, Page RE Jr (2005) Locomotion and the pollen hoarding behavioral syndrome of the honeybee (*Apis mellifera* L.). J Comp Physiol [A] 191:669–674. doi:10.1007/s00359-005-0624-x

Hunt GJ, Page RE Jr, Fondrk MK, Dullum CJ (1995) Major quantitative trait loci affecting honey bee foraging behavior. Genetics 141:1537–1545

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

Hunt GJ, Amdam GV, Schlipalius D, Emore C, Sardesai N, Williams CE, Rueppell O, Guzman-Novoa E, Arechavaleta-Velasco M, Chandra S, Fondrk MK, Beye M, Page RE Jr (2007) Behavioral genomics of honeybee foraging and nest defense. Naturwissenschaften 94:247–267. doi:10.1007/s00114-006-0183-1

Koch I, Schwarz H, Beuchle D, Goellner B, Langegger M, Aberle H (2008) *Drosophila* ankyrin 2 is required for synaptic stability. Neuron 58:210–222. doi:10.1016/j.neuron.2008.03.019

Korol A, Frenkel Z, Cohen L, Lipkin E, Soller M (2007) Fractioned DNA pooling: a new cost-effective strategy for fine mapping of quantitative trait loci. Genetics 176:2611–2623. doi:10.1534/genetics.106.070011

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lapidge KL, Oldroyd BP, Spivak M (2002) Seven suggestive quantitative trait loci influence hygienic behavior of honey bees. Naturwissenschaften 89:565–568

Lattorff HMG, Moritz RFA, Crewe RM, Solignac M (2007) Control of reproductive dominance by the thelytoky gene in honeybees. Biol Lett 3:292–295. doi:10.1098/rsbl.2007.0083

Le Conte Y, Hefetz A (2008) Primer pheromones in social hymenoptera. Annu Rev Entomol 53:523–542. doi:10.1146/annurev.ento.52.110405.091434

Lincoln SE, Daly MJ, Lander ES (1993) Constructing genetic linkage maps with MAPMAKER/EXP version 3.0: a tutorial and reference manual. In: A whitehead institute for biomedical research technical report, Cambridge, MA

Mackay TFC (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35:303–339. doi:10.1146/annurev.genet.35.102401.090633

Nelson CM, Ihle KE, Fondrk MK, Page RE Jr, Amdam GV (2007) The gene vitellogenin has multiple coordinating effects on social organization. PLoS Biol 5:e62. doi:10.1371/journal.pbio.0050062

Page RE Jr, Amdam GV (2007) The making of a social insect: developmental architectures of social design. BioEssays 29:334–343. doi:10.1002/bies.20549

Page RE Jr, Erber J (2002) Levels of behavioral organization and the evolution of division of labor. Naturwissenschaften 89:91–106. doi:10.1007/s00114-002-0299-x

Page RE Jr, Fondrk MK (1995) The effects of colony level selection on the social organization of honey bee (*Apis mellifera* L.) colonies—colony level components of pollen hoarding. Behav Ecol Sociobiol 36:135–144. doi:10.1007/BF00170718

Page RE Jr, Fondrk MK, Hunt GJ, Guzman-Novoa E, Humphries MA, Nguyen K, Greene AS (2000) Genetic dissection of honeybee (*Apis mellifera* L.) foraging behavior. J Hered 91:474–479. doi:10.1093/jhered/91.6.474

Page RE Jr, Gadau J, Beye M (2002) The emergence of hymenopteran genetics. Genetics 160:375–379

Pankiw T (2003) Directional change in a suite of foraging behaviors in tropical and temperate evolved honey bees (*Apis mellifera* L.). Behav Ecol Sociobiol 54:458–464. doi:10.1007/s00265-003-0640-1

Phillips PC (1999) From complex traits to complex alleles. Trends Genet 15:6–8. doi:10.1016/S0168-

9525(98)01622-9

Ragoussis J, Elvidge GP, Kaur K, Colella S (2006) Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. PLOS Genet 2:920–929. doi:10.1371/journal.pgen.0020100

Rayburn LYM, Gooding HC, Choksi SP, Maloney D, Kidd AR, Siekhaus DE, Bender M (2003) *amontillado*, the *Drosophila* homolog of the prohormone processing protease PC2, is required during embryogenesis and early larval development. Genetics 163:227–237

Robinson GE (2002) Genomics and integrative analyses of division of labor in honeybee colonies. Am Nat 160:S160–S172. doi:10.1086/342901

Robinson GE, Huang Z-Y (1998) Colony integration in honey bees: genetic, endocrine and social control of division of labor. Apidology 29:159–170. doi:10.1051/apido:19980109

Robinson GE, Fernald RD, Clayton DF (2008) Genes and social behavior. Science 322:896–900. doi:10.1126/science.1159277

Rueppell O, Pankiw T, Nielson DI, Fondrk MK, Beye M, Page RE Jr (2004) The genetic architecture of the behavioral ontogeny of foraging in honey bee workers. Genetics 167:1767–1779. doi:10.1534/genetics.103.021949

Rueppell O, Chandra SBC, Pankiw T, Fondrk MK, Beye M, Hunt GJ, Page RE Jr (2006a) The genetic architecture of sucrose responsiveness in the honey bee (*Apis mellifera* L.). Genetics 172:243–251. doi:10.1534/genetics.105.046490

Rueppell O, Page RE Jr, Fondrk MK (2006b) Male behavioural maturation rate responds to selection on pollen hoarding in honeybees. Anim Behav 71:227–234. doi:10.1016/j.anbehav.2005.05.008

Rueppell O, Bachelier C, Fondrk MK, Page RE Jr (2007) Regulation of life history determines lifespan of worker honey bees (*Apis mellifera* L.). Exp Gerontol 42:1020–1032. doi:10.1016/j.exger.2007.06.002

Rueppell O, Linford R, Gardner P, Coleman J, Fine K (2008) Aging and demographic plasticity in response to experimental age structures in honeybees (*Apis mellifera* L.). Behav Ecol Sociobiol 62:1621–1631. doi:10.1007/s00265-008-0591-7

Rüppell O, Pankiw T, Page RE Jr (2004) Pleiotropy, epistasis and new QTL: the genetic architecture of honey bee foraging behavior. J Hered 95:481–491. doi:10.1093/jhered/esh072

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234. doi:10.1038/72708

Shapira M, Thompson CK, Soreq H, Robinson GE (2001) Changes in neuronal acetylcholinesterase gene expression and division of labor in honey bee colonies. J Mol Neurosci 17:1–12. doi:10.1385/JMN:17:1:1

Solignac M, Mougel F, Vautrin D, Monnerot M, Cornuet JM (2007) A third-generation microsatellite-based linkage map of the honey bee, *Apis mellifera*, and its comparison with the sequence-based physical map. Genome Biol 8:R66. doi:10.1186/gb-2007-8-4-r66

Sullivan JP, Jassim O, Fahrbach SE, Robinson GE (2000) Juvenile hormone paces behavioral development in the adult worker honey bee. Horm Behav 37:1–14. doi:10.1006/hbeh.1999.1552

Van Ooijen JW, Boer MP, Jansen RC, Maliepaard C (2002) MapQTL 4.0, software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, The Netherlands

Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. Mol Ecol 14:3623–3642. doi:10.1111/j.1365-294X.2005.02690.x

Wang Y, Amdam GV, Rueppell O, Wallrichs M, Fondrk MK, Kaftanoglu O, Page RE Jr (2009) PDK1 and HR46 gene homologs tie social behavior to ovary signals. PLoS ONE 4:e4899. doi:10.1371/journal.pone.0004899

Wayne ML, McIntyre LM (2002) Combining mapping and arraying: an approach to candidate gene identification. Proc Natl Acad Sci USA 99:14903–14906. doi:10.1073/pnas.222549199

Whitfield CW, Cziko A-M, Robinson GE (2003) Gene expression profiles in the brain predict behavior in

individual honey bees. Science 302:296–299. doi:10.1126/science.1086807

Whitfield CW, Behura SK, Berlocher SH, Clark AG, Johnston JS, Sheppard WS, Smith DR, Suarez AV, Weaver D, Tsutsui ND (2006a) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. Science 314:642–645. doi:10.1126/science.1132772

Whitfield CW, Ben-Shahar Y, Brillet C, Leoncini I, Crauser D, LeConte Y, Rodriguez-Zas S, Robinson GE (2006b) Genomic dissection of behavioral maturation in the honey bee. Proc Natl Acad Sci USA 103:16068–16075. doi:10.1073/pnas.0606909103

Wilson RJ, Goodman JL, Strelets VB, FlyBase Consortium (2008) FlyBase: integration and improvements to query tools. Nucleic Acids Res 36:D588–D593. doi:10.1093/nar/gkm930