



**HAL**  
open science

## Quantitative traits loci (QTL) involved in body colour, wing morphometry, cuticular hydrocarbons and venom components in honeybee

Florence Mougel, Michel Solignac, Dominique Vautrin, Emmanuelle Baudry, John Ogden, Alain Tchapla, Hugues Schweitz, H el ene Gilbert

### ► To cite this version:

Florence Mougel, Michel Solignac, Dominique Vautrin, Emmanuelle Baudry, John Ogden, et al.. Quantitative traits loci (QTL) involved in body colour, wing morphometry, cuticular hydrocarbons and venom components in honeybee. *Apidologie*, 2012, 43 (2), pp.162-181. 10.1007/s13592-011-0090-0 . hal-01003629

**HAL Id: hal-01003629**

**<https://hal.science/hal-01003629>**

Submitted on 11 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

Copyright

# Quantitative traits loci (QTL) involved in body colour, wing morphometry, cuticular hydrocarbons and venom components in honeybee

Florence MOUGEL<sup>1,2</sup>, Michel SOLIGNAC<sup>1,2</sup>, Dominique VAUTRIN<sup>1</sup>,  
Emmanuelle BAUDRY<sup>3</sup>, John OGDEN<sup>4</sup>, Alain TCHAPLA<sup>4</sup>, Hugues SCHWEITZ<sup>5</sup>,  
Hélène GILBERT<sup>6,7</sup>

<sup>1</sup>Laboratoire Evolution Génome Spéciation, CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette Cedex, France

<sup>2</sup>Université Paris Sud, 91405 Orsay Cedex, France

<sup>3</sup>Laboratoire Ecologie Systématique et Evolution, Université Paris Sud, 91405 Orsay Cedex, France

<sup>4</sup>Groupe de Chimie Analytique de Paris Sud EA 4041, Laboratoire d'Etudes des Techniques et Instruments d'Analyse Moléculaire, I.U.T. Orsay, Université Paris Sud, Plateau de Moulon, 91400 Orsay Cedex, France

<sup>5</sup>Institut de Pharmacologie Moléculaire et Cellulaire, CNRS, 660 route des Lucioles, Sophia Antipolis, 06560 Valbonne, France

<sup>6</sup>UMR1313 Génétique Animale et Biologie Intégrative, INRA Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France

<sup>7</sup>UMR444 Laboratoire de Génétique Cellulaire, INRA Chemin de Borde Rouge, 31326 Castanet-Tolosan Cedex, France

Received 3 February 2011 – Revised 27 July 2011 – Accepted 11 August 2011

**Abstract** – In order to detect quantitative trait loci (QTL) on body composition and venom characteristics of honeybees, two F2 backcross progenies issued from crosses between subspecies *Apis mellifera mellifera* and *Apis mellifera ligustica* were genotyped for microsatellite markers and characterized for 31 quantitative traits. Three traits were recorded for body colour, 11 for wing morphometry, 12 for cuticular hydrocarbons and five for venom components. We detected 50 QTL responsible for quantitative variations in these traits, seven of which showed putative pleiotropy. Most of the QTL were specific to one progeny, revealing allelic variation of these QTL within the parental subspecies. The position of some of these QTL was accurately determined, with the minimum confidence interval reaching 76 kb and 1.5 cM, and containing a single gene.

honeybee / genetics / QTL / morphometry / cuticular hydrocarbons / venom

## 1. INTRODUCTION

The first genetic map built from RAPD markers (Hunt and Page 1995) initiated honeybee genetics and search for quantitative trait loci (QTL). A second step in genetic studies was achieved with the publication of Expressed

Sequence Tag (EST) database available for microarray expression studies (Whitfield et al. 2002). Finally, the complete genome was sequenced in 2006 (Honeybee Genome Sequencing Consortium 2006) and organized into oriented chromosomes thanks to a microsatellite-based genetic map (Solignac et al. 2007). Honeybees are progressively turning into a model species.

Since 1995, genetic and genomic publications focused however on specific properties of *Apis mellifera*, meaning agronomic interest and

Corresponding author: F. Mougél,

mougél@legs.cnrs-gif.fr

Manuscript editor: Klaus Hartfelder

eusocial behaviour. Description of QTL concerned foraging behaviour, such as, quantity of pollen stored by a colony (Hunt et al. 1995), quantity and quality of pollen and nectar taken and age of first foraging (Rueppell et al. 2004, 2006, 2011), aggressiveness behaviour (Arechavaleta-Velasco et al. 2003; Hunt et al. 1998, 1999), learning ability (Chandra et al. 2001), hygienic behaviour (Lapidge et al. 2002) or ovary size development linked to the reproductive status of a female (Linksvayer et al. 2009; Oxley et al. 2008; Rueppell et al. 2011). However, the present knowledge of honeybee genetics would allow approaching numerous other traits, for now mainly studied in *Drosophila melanogaster*.

We choose to focus in this study on four categories of traits which are expressed at the individual level instead of the colony level: cuticular hydrocarbons (12 traits), venom components (five traits), abdominal colouration (three traits) and wing venation (11 traits). Cuticular hydrocarbons are known to serve as ecological, behavioural and physiological signals in insect and other arthropod species (for a review, see Howard and Blomquist 2005). Venom components are of interest both for their impact on human health (Dotimas and Hider 1987; Hider 1988; Lima and Brochetto-Braga 2003) and for their pharmacological properties (Rajendra et al. 2004). Morphometric traits (abdominal colouration and wing venation) are part of the “standard morphometry” described by Ruttner et al. (1978) and Ruttner (1988). They have been used for the definition of the numerous subspecies of honeybee that originate from a wide range of geographic locations and are still widely studied (Diniz-Filho et al. 2000; Radloff et al. 2003; Ruttner et al. 2000).

All these traits were previously shown to be at least partially genetically determined justifying the search for determining QTL. Cuticular hydrocarbon profiles of worker bees taken from the same subfamily differed less than those between subfamilies present in a colony indicating a genetic control of the profile (Arnold et al. 1996, 2000). Venom quantity and composition was shown to be variable between species of the

genus *Apis* and subspecies of *A. mellifera* (Schmidt 1995; Schumacher et al. 1992). Finally, heritabilities of morphometric traits were estimated to be moderate to high with little effect of dominance (Oldroyd et al. 1991; Poklukar and Kezic 1994). The evidence of genetic variation for the four groups of traits indicates that QTL detection is promising.

In this paper, we describe the QTL characterisation for 31 traits from these four categories in honeybee. Analyses were applied to two F2 backcross progenies from crosses between two subspecies of honeybee divergent for most of the traits recorded: *Apis mellifera mellifera* and *Apis mellifera ligustica*. Details on QTL analyses, potential pleiotropy and fixation of alleles in the subspecies are described, and putative candidate genes are discussed when available.

## 2. MATERIALS AND METHODS

### 2.1 Crosses

The crosses are described in detail in Solignac et al. (2004). Workers from two F2 backcross progenies were genotyped and phenotyped for specific quantitative traits. Two F1 queens were obtained from crosses between virgin *A. mellifera ligustica* queens and *A. mellifera mellifera* drones. These subspecies were retained as most of the traits chosen for this analysis had previously been reported to show divergent levels or characteristics in the subspecies (Ruttner 1988; Schumacher et al. 1992), justifying the assumption that different alleles for major QTL affecting these traits will segregate in the progeny of F1 queens. These F1 hybrid queens (named B and V) were then double backcrossed with one drone from each of the two parental subspecies. The two resulting F2 backcross progenies comprised 92 and 95 workers for queen B and queen V, respectively, composed of two subfamilies each, from the two backcrosses: B1 (parental drone *A. mellifera mellifera*, 54 workers) and B2 (parental drone *A. mellifera ligustica*, 38 workers), for the progeny of queen B, and V1 (parental drone *A. m. mellifera*, 50 workers) and V2 (parental drone *A. mellifera ligustica*, 45 workers) for the progeny of queen V.

## 2.2 Genetic markers

The honeybee genetic map was built with 1,880 markers for the progeny of queen B and 662 for the progeny of queen V, leading to a total of 2,008 markers (Solignac et al. 2007). The same set of markers was used for QTL screening. Some markers were eliminated for various reasons: (1) they were at null distance from other markers with less missing genotypes; (2) they showed null alleles and genotypes could be obtained for one subfamily only; or (3) genotyping quality was too low, leading to a large number of missing data. Finally, 1,282 and 585 markers were selected for a first QTL genome scan for queens B and V, respectively. The mean distance between consecutive markers was 3.6 and 7.1 cM for queens B and V, respectively, with maximum distances of 15.9 and 32.9 cM, respectively.

After a first QTL detection, the density of markers was increased in the vicinity of the detected QTL in order to confirm their presence and improve their localisation. Markers that were heterozygous for queen V, but which had not been used initially were then genotyped. Moreover, using the genome sequence of the honeybee, 85 new markers were designed to complement the genotyping effort. Among these, 20 and 24 were genotyped in queen B and queen V, respectively, for the QTL regions of interest. The final density was around 5 cM between markers in each candidate region, except in three cases where it was not possible to define new markers closer than 7–9 cM due to a high recombination rate or a gap in the sequence assembly.

For markers of one specific QTL of interest (Col1) the number of genotyped progeny was enlarged to 277 workers for queen V to improve the detection accuracy.

Origin of the marker alleles could be identified systematically. For further analyses, alleles were noted 1 when coming from the subspecies *A. mellifera mellifera* and 2 when coming from the subspecies *A. mellifera ligustica*. As a consequence, the two queens had genotypes 12, and within each subfamily progeny were either 11 and 12 (if drone was 1) or 12 and 22 (if drone was 2). This notation ignores allelic variability within subspecies even when it exists.

## 2.3 Quantitative measures

The abdomen of each individual from the two progenies was photographed. Yellow surface area on tergites 3 and 4 and total abdomen surface area were measured using NIH image and ImageJ software (Abramoff et al. 2004). Three quantitative traits were obtained from the ratio between yellow and total surface area: yellow area on tergite 3 (Cou1), yellow area on tergite 4 (Cou2) and their sum (Cou3), all of these being expressed as a proportion of total abdomen area. Obviously, Cou3 was highly correlated with Cou1 and Cou2. Nevertheless, all three traits were used in QTL detection to maximise the chance of detecting low QTL effects.

Additional morphometric traits used were obtained from the right forewing of individuals. Nineteen points were picked out at venation intersections, and ten angular measures were calculated from them. These measures are part of the morphometric characters defined by Friedrich Ruttner (1988), to compare geographic races of honeybee. The correspondence with Ruttner's characters is as follows: Angle1=31, Angle2=28, Angle3=25, Angle4=21, Angle5=22, Angle6=30, Angle7=23, Angle8=26, Angle9=27 and Angle10=24. The length of the forewing was also measured for the queen B progeny (distance between the two most extreme points of the wing venation: d1–8). For all morphometric traits (abdominal colouration and angles of wing venation), measures were made twice and the final quantitative measure was the average of the two values.

To measure cuticular hydrocarbons, young bees were isolated for the first 15 days after emergence to prevent any admixture with other colony members. Fifteen-day-old individuals were frozen and immersed for 5 min in 1 ml of pentane solvent to dissolve cuticular hydrocarbons. The solvent was then kept at  $-20^{\circ}\text{C}$  to avoid evaporation and placed at room temperature for 30 min before analysis. Identification of cuticular hydrocarbons was performed using capillary gas chromatography–mass spectrometry (Incos 50, Finnigan). Their relative quantities were determined from the peak surface areas using both on-column and automate phase solid interface high-temperature capillary gas chromatography (Ogden et al. 1998). Measures were performed with a retention gap (2 m $\times$ 0.53 mm) followed by a

HT5 column (25 m×0.32 mm×0.10 μm), helium as gas vector using Hewlett Packard 5850 series II chromatograph. Component detection was carried out with Hewlett Packard flame ionisation detector. Twelve components were selected for QTL mapping. They comprised from 23 to 33 carbons. Six of these were alkane components (Cir1, C23; Cir2, C25; Cir3, C26; Cir4, C27; Cir7, C29; and Cir12, C33), five were alkene components (Cir6, C27:1; Cir8, C31:1a; Cir9, C31:1b; Cir10, C33:2; and Cir11, C33:1) and the last (Cir5) could not be determined.

For venom components, the poison sac was taken from each individual and placed in a 5-μl sterile water. The sac was squeezed two to three times and then removed from the tube. Samples were conserved at -20°C. They were kept 30 min at room temperature and homogenized before use. The relative quantity of the different venom components was determined by a modified high-performance liquid chromatography method (Ameratunga et al. 1995). Detection was performed at 280 nm. Five components could be quantified: noradrenaline (Nor), dopamine (Dop), apamin (Apa), phospholipase A2 (Pho) and melittin (Mel). Their relative quantities were estimated from the surface area of the peaks.

## 2.4 Statistical analysis

Quantitative variation within the whole sample, each family and each subfamily was studied using R 2.10.1 (R Development Core Team 2005). Normal distribution was tested using the Shapiro–Wilk test, applied to the four subfamilies. Medians and variances of each quantitative trait were compared between B and V progenies using Kruskal–Wallis tests and Fligner–Killeen tests, respectively. The advantage of these methods is that they rely on very few assumptions. In particular, normal distribution of the quantitative measure is not assumed for either tests, and variance equality is not required for the Kruskal–Wallis test. Overall tests for variance equality were also conducted for the four subfamilies using Fligner–Killeen test, and pairwise comparison were performed when overall tests were significant. Median values were also compared between each of the two subfamilies belonging to the same progeny (B1 was compared

with B2 and V1 was compared with V2) using Kruskal–Wallis tests. As most of the relations between quantitative measures were not linear, Spearman’s correlation coefficient calculation was made and tests applied. When multiple and non-independent tests were conducted (pairwise comparison of variance between subfamilies, correlation coefficient), a Holm correction of the *p* values for multiple testing was applied.

## 2.5 QTL detection

QTL detection was performed with QTLMap (Elsen et al. 1999). The two F2 backcross progenies were analysed separately, in order to detect specific QTL allelic combinations segregating in the putative heterozygous queens families. Interval mapping analyses were applied along all the linkage groups from the position of the first marker to the last (position  $x = \text{InitPos}$ ,  $\text{MaxPos}$ ). The genetic model for the traits was  $y_{ijk} = \mu_i + e_{ijk}$  under the null hypothesis  $H_0 = \text{“no QTL segregating”}$ , and  $y_{ijk} = \mu_i + \sum_j p_{jk}^x g_j + e_{ijk}$  under the alternative hypothesis  $H_1 = \text{“a QTL segregating at position } x\text{”}$ , where  $y_{ijk}$  is the phenotype for individual  $k$  from drone subfamily  $i$  having genotype  $j$ ,  $j = [11, 12, 22]$ ,  $\mu_i$  is the mean of the progeny from drone  $i$ ,  $p_{jk}^x$  is the probability for individual  $k$  to be of genotype  $j$  at the tested position  $x$ , computed according to Elsen et al. (1999), from the genealogy, genetic map and observed genotypes,  $g_j$  is the fixed effect for genotype  $j$ ,  $e_{ijk}$  is the residual normally distributed with mean 0 and standard deviation  $\sigma$ . The distribution of the residual assumes homoscedasticity between the two drone subfamilies. This hypothesis was shown to be robust but potentially lowering the power to detect QTL when heteroscedasticity was actually present (Goffinet et al. 1999). To avoid overparameterization, a constraint on genotypic effects was applied:  $\sum_j g_j = 0$ . A Newton–Raphson algorithm was used to maximise the likelihoods  $L_0$ , and  $L_1^x$  for each tested position  $x$ , and thus obtain estimates of genotypic effects. Likelihood ratio tests (LRT) $_x$  were computed as  $-\log(L_0/L_1^x)$ , with  $\max(\text{LRT}_x)_{x=\text{InitPos}, \text{MaxPos}}$  indicating the most likely position of a QTL on the linkage group considered.

From  $\hat{g}_{11}$ ,  $\hat{g}_{12}$  and  $\hat{g}_{22}$ , at the position of  $\max(\text{LRT}_x)$ , the trait variance explained by the QTL was computed as  $V_q = 0.25\hat{g}_{11} + 0.5\hat{g}_{12} + 0.25\hat{g}_{22}$ , approximating

the distribution of the genotypes in the queens progeny with the proportions 1:4, 1:2 and 1:4, for genotypes 11, 12 and 22, respectively. The proportion of variance explained by the QTL was further computed dividing  $V_q$  by the phenotypic variance of the trait  $V_p$  (estimated from centred quantitative measures, i.e. by subtraction of the subfamily mean). A test was applied to determine whether the QTL was additive, (partially) dominant or overdominant. The additive effect  $a$  of the QTL was estimated as  $(\hat{g}_{22} - \hat{g}_{11})/2$ , and the dominance effect  $d$  was estimated as  $\hat{g}_{12} - (\hat{g}_{22} + \hat{g}_{11})/2$ , with the ratio  $r = \frac{a}{d}$  quantifying the level of dominance for the QTL: if  $|r| \in [0;0.5]$ , the QTL was said to be additive or codominant  $Co$ , if  $|r| \in [0.5;0.8]$ , the QTL was said to be partially dominant  $d$ , if  $|r| \in [0.8;1.2]$ , the QTL was said to be dominant  $D$ , and if  $|r| > 1.2$ , the QTL was said to be overdominant  $SD$  in the following results.

Sixteen genetic maps, one for each chromosome, were successively scanned for the 31 quantitative traits. The first step consisted of screening the genome by interval mapping, taking 5 cM steps. When QTL were detected, new genotyping in the vicinity of the primo-detected QTL were performed and a second interval mapping was performed in the region of interest, using 1 cM steps. A large number of QTL detection analyses was performed during this study: 31 traits, corresponding to about 18 independent factors (>90% variation explained by the 18 highest eigenvalues of the correlation matrix of the traits), were tested on 16 independent linkage groups, corresponding to a total of 288 independent tests. To account for multiple testing, the threshold for the estimated LRT was defined in order to achieve a proportion of 5% of false-positive results at the level of the genome-wide analysis for each queen family (5% threshold=14.1). It corresponded to an approximate Bonferroni correction of the type I error of 5% for the test on each chromosome for each trait using the factor 288, leading to a  $1.7 \cdot 10^{-4}$  type I error for the chromosome-wide  $\chi^2$  test with 1 degree of freedom. Results at a 10% type I error at the genome-wide level are reported as a tendency for QTL detection. Confidence intervals (CI) were computed at the 90% level applying the log-drop-off technique, as proposed by Lander and Botstein 1989. Physical boundaries of the CI are from the v4 assembly of the honeybee genome.

## 2.6 Localisation of genes of interest

We searched for localisation of well-known genes in the CI described for the QTL. These genes are involved in the molecular pathways known to influence quantitative traits studied in model species such as *D. melanogaster* or had been previously described specifically in honeybee (venom components). We first looked for annotated genes in the genome (on the NCBI website: <http://www.ncbi.nlm.nih.gov/>) and when genes were not annotated, we blasted homologous sequences from model organisms against the RefSeq protein dataset of honeybee on the NCBI website.

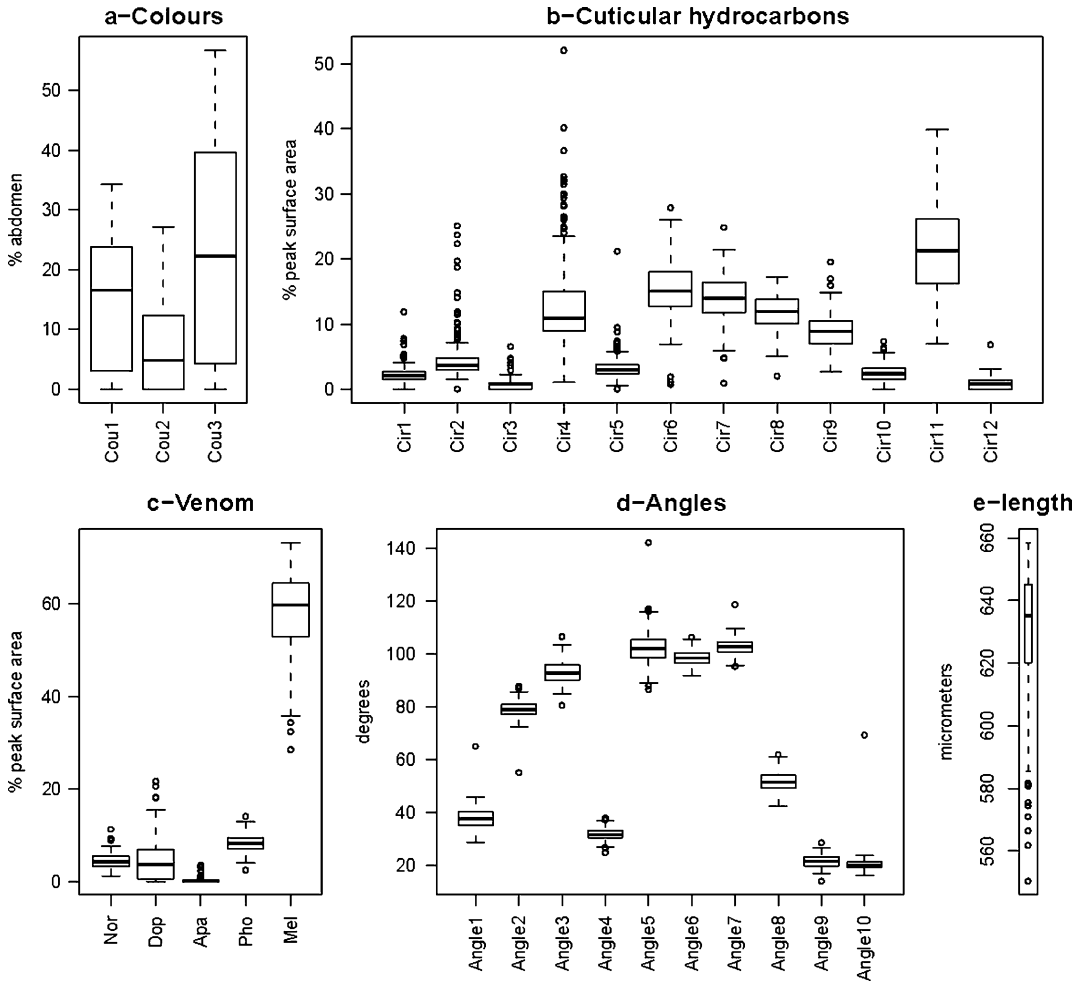
For body colour QTL, we localised pigment production genes and patterning genes (Wittkopp et al. 2002; Wittkopp and Beldade 2009). We took advantage of the genes described in Dearden et al. (2006) and Drapeau et al. (2006). For venom component QTL, we sought the localisation of genes coding for mellitin, apamin (Gmachl and Kreil 1995), dipeptidyl peptidase and phospholipase A2 (Hoffman 2006), and genes involved in catecholamine synthesis, such as dopamine and noradrenaline (Pendleton et al. 1998; Squire et al. 2003). Study was made on two main categories of enzymes involved in the biochemical pathway of cuticular hydrocarbons: elongase (Chertemps et al. 2007; Howard and Blomquist 2005) and desaturase (Dallerac et al. 2000; Eigenheer et al. 2002; Legendre et al. 2008), which had previously been shown to involve quantitative changes. Finally, genes described as influencing wing vein pattern (Blair 2007; De Celis and Diaz-Benjumea 2003; Molnar et al. 2007; Sander et al. 2010; Yan et al. 2009) were compared with morphometric QTL. For this last category, we also used some of the genes described in Dearden et al. (2006).

## 3. RESULTS

### 3.1 Description of quantitative variation

Body colour is a very distinctive character of the two parental subspecies studied here, *A. mellifera mellifera* and *A. mellifera ligustica*, between which it differs markedly. The resulting variability observed in the progeny was high for the three traits measured (Figure 1). Percentage





**Figure 1.** Quantitative variability of the different sets of traits (the progenies of queens B and V are pooled). Variability is represented by the boxplot function of R 2.10.1. The *thick line* indicates the median position. The *two walls of the box* correspond to first and third quartiles. *Whiskers* are separated from the box by a 1.5 interquartile range (3rd quartile minus 1st quartile). *Circles* represent individual measures outside the *whiskers*.

of yellow surface ranged from 0 to 34 for Cou1, 0 to 27 for Cou2 and 0 to 57 for Cou3. It also clearly appeared that the yellow surface was greater on the third abdomen segment (Cou1) than on the fourth (Cou2). As a result, it was mainly the third segment that contributed to the total yellow surface of the abdomen (Cou3).

The most abundant cuticular hydrocarbons had between 27 and 33 carbons, with a predominance of the compound Cir11 (mean, 21.5%), an alkene with 33 carbons and one double bond. The average composition ob-

served in the two progenies was divergent from that described in Chaline et al. (2005). These previous authors indicated that the major component was Cir4 (19.5% of the total hydrocarbons present on the cuticle) followed by Cir11 (15.8%). Moreover, the same authors described Cir6 as a low proportion compound (2.7%) while it was the second most abundant compound in our study (mean, 15.3%). However, the proportions cited in Chaline et al. (2005), were obtained from individual foragers while our results were obtained from isolated

emerging bees. Maturation with age, as well as compound exchange between individuals could explain such divergences. Large variability was observed for the twelve compounds. Distributions seemed to be highly spread for traits Cir1 to Cir5, for which numerous measurements were found outside the whisker bounds.

The relative abundance of venom components melittin (mean, 57.8%) and phospholipase A2 (mean, 8.3%) was similar to that described in the literature (Dotimas and Hider 1987). However, the relative abundances of the three other components were the opposite to those found in previous studies: we observed 4.4% noradrenaline and 4.6% dopamine, while these had previously been described to represent less than 1% of the venom, and we observed only 0.2% apamin, which had previously been quantified at between 1% and 3% of the venom (Dotimas and Hider 1987). However, these previously published data were obtained from a mixture of venom from numerous bees. Furthermore, the authors did not specify the subspecies from which this venom was collected. Differences in venom composition between individual bees and subspecies could explain these discrepancies. Indeed, in our study, the proportion of each component showed clear variability within the progeny.

The ten venation angles measured had all about the same magnitude of variability. Angle5 had the largest range (30° between the smallest and the largest value) while Angle10 had the smallest (7° between the smallest and the largest value), and all other angles had a range around 20°. This was true when one point, which clearly diverged from the distribution, had been discarded. The deviant measures came from a single individual of the queen V progeny and were discarded from subsequent analyses. Finally, large variability was also observed in the wing size trait.

The distribution of all traits was compared with the normal distribution using Shapiro–Wilk tests. The comparison was conducted for each trait within the subfamilies, this level being the most relevant for QTL mapping (drone effect

removed for interval mapping). Normality was respected in most cases, except when distribution was highly spread (Cir1 to Cir4) or comprised frequent null values (colour traits, Dop and Apa). However, no transformation was applied to these traits as previous tests (unpublished data) showed that the Gaussian model applied here for QTL detection was robust to similar deviations from normality.

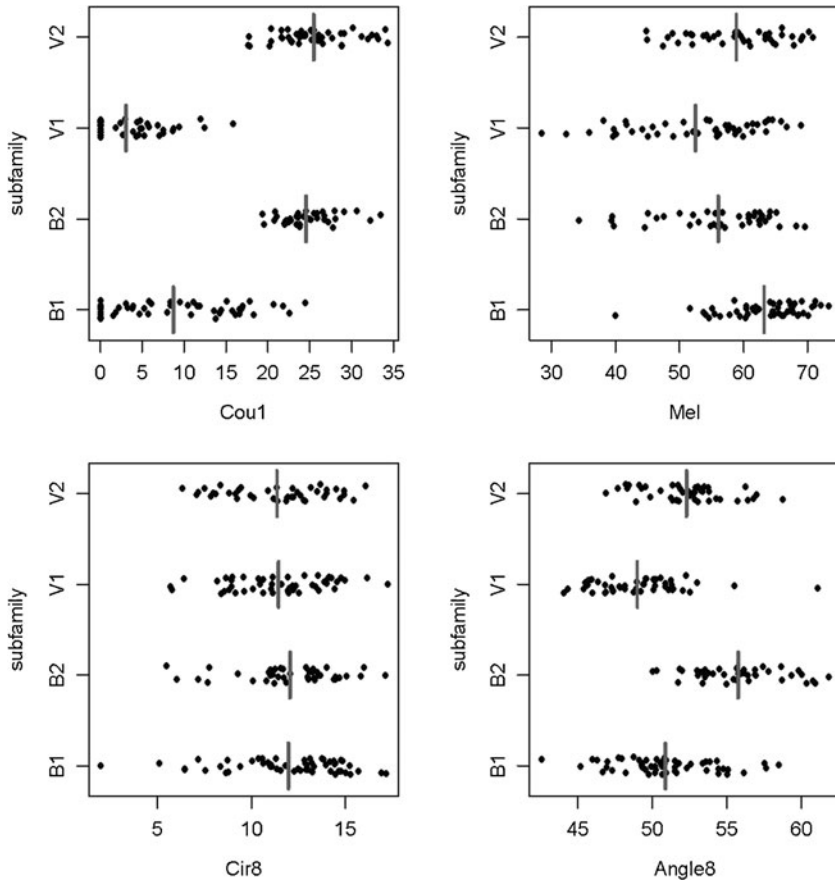
Distribution of the traits within each subfamily was studied. Examples for Cou1, Mel, Cir8 and Angle8 are shown in Figure 2. Paternal contribution to the mean of the group was high in most cases. Consequently, it was taken into account in the model for QTL detection. Furthermore, similar differences between the two subfamilies were found in B and V progenies for 17 out of 30 traits (for example, see Cou1 and Angle8 in Figure 2). This suggested the existence of subspecies effects. In two cases (Mel and Angle5 in Figure 2), the opposite effect of paternal contribution were observed between the two progenies. For the 11 remaining traits, no significant difference was observed between subfamilies in at least one progeny (for example, see Cir8 in Figure 2). Despite the subspecies effect, traits were not fixed within subspecies: medians and/or variances were significantly different between the two progenies in 24 out of 30 traits (not shown).

Correlations were tested in each family on centred quantitative values (subfamily mean subtracted). Table I shows only significant correlations after correction for multiple testing; these were only found between traits from the same category (colours, cuticular hydrocarbons, venom components or angles) and mostly concerned cuticular hydrocarbons.

### 3.2 QTL localization

As a first step, interval mapping was performed with steps of 5 cM for each chromosome, each progeny and each trait. Seventy-two significant LRT peaks were detected from the first screening at the suggestive level of 10% at the genome-wide level. The second interval mapping was applied





**Figure 2.** Quantitative variation within the four different subfamilies (B1, B2, V1 and V2) for Cou1, Mel, Cir8 and Angle8. Each point is an individual measure. Units are the same as in Figure 1. Grey lines indicate the mean value of the trait in each subfamily.

with 1 cM steps on these specific regions, covering at least the confidence interval and no less than 40 cM. This second screening resulted in the confirmation of 65 LRT peaks out of the 72 primo-detected. The higher marker density resulted in a better definition of the LRT peaks: (1) CI being reduced by 5 to 24 cM for 20 peaks and (2) maximum LRT position being better defined (shifts up to 21.2 cM in the most dramatic case). However, in few cases, CI were enlarged in the second finer screening. Increases were generally low (less than 5 cM), with one notable exception for the QTL affecting Mel on chromosome 7 (progeny V). In this case, the CI was enlarged by 14.5 cM. This larger peak was

probably composed of two different peaks only seeable when marker density increased. Specific effort was applied to a peak on chromosome 1, where 192 supplementary individuals from progeny V had records for Cou1, Cou2 and Cou3 traits. Figure 3 shows the results of screenings with the two density maps for trait Cou1 in progenies B and V (where the tested progeny in the 1 cM screening was enlarged to 277 individuals). New analyses allowed a dramatic increase of the test statistic significance for progeny V, from a  $p$  value of 0.018 to a  $p$  value lower than 0.0001 at the genome-wide level.

Some of the LRT peaks were observed in the same regions for different quantitative traits. In

**Table I.** Test for correlations within families for centred values (Spearman's method adjusted for multiple testing, see text).

Trait 1	Trait 2	Family	
		B	V
Cou1	Cou2	*	NS
	Cou3	**	**
Cou2	Cou3	**	**
Nor	Dop	NS	***
Cir1	Cir2	**	NS
	Cir7	***	NS
Cir2	Cir3	*	*
	Cir4	**	**
	Cir7	**	**
	Cir8	**	**
	Cir9	NS	*
Cir3	Cir11	**	**
	Cir4	**	*
	Cir7	**	***
	Cir8	**	**
	Cir9	NS	*
Cir4	Cir10	***	**
	Cir11	**	**
	Cir7	**	**
	Cir8	**	**
	Cir9	NS	*
Cir5	Cir10	***	**
	Cir11	**	**
Cir6	Cir9	NS	***
	Cir11	NS	*
Cir7	Cir9	**	NS
	Cir11	**	**
Cir8	Cir8	**	**
	Cir10	NS	*
	Cir11	**	**
Cir9	Cir10	NS	***
	Cir11	***	***
Cir10	Cir12	**	***
	Cir11	*	*
Angle4	Cir11	*	*
	Angle5	**	**
Angle5	Angle10	**	**
	Angle7	NS	NS
	Angle9	***	*
	Angle10	*	NS
		**	NS

Only pairs of traits for which significant correlation was found in at least one family are mentioned

NS non-significant

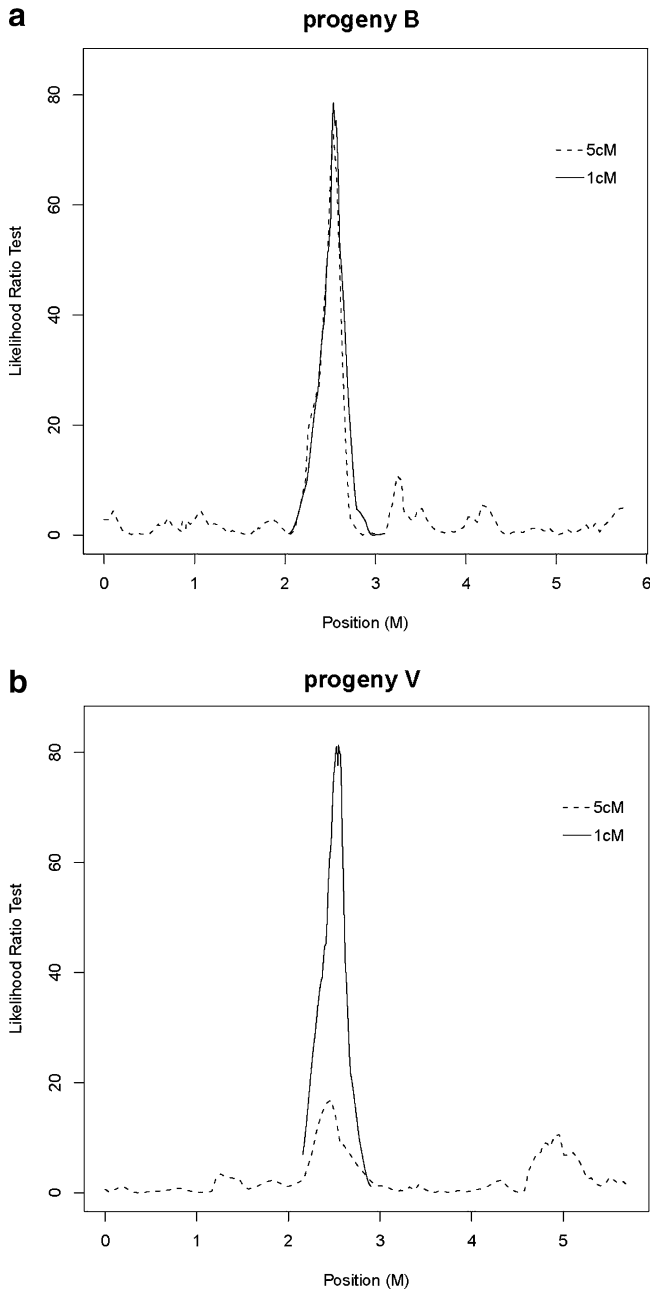
\*Significant at the 5% level

\*\*Significant at the 1% level

\*\*\*Significant at the 1% level

most cases, the traits involved were from the same category. Such cases suggested the existence of pleiotropic QTL, or at least pleiotropic chromosomal regions. By counting each pleiotropic region as one QTL, and adding QTL detected for single traits, the analysis resulted in 50 different QTL for the 31 studied quantitative traits: seven QTL for colouration (named Col1 to Col6b), 9 for venom components (Venom1 to Venom9), 12 for cuticular hydrocarbons (HC1 to HC12) and 22 for wing morphometry (Morpho1 to Morpho22) (Table II). As expected, the pleiotropic QTL were partly in agreement with the phenotypic correlations observed. For example, Col1 and Col2 jointly influenced the correlated traits Cou1 and Cou3 and Cou2 (for Col1 only). Inversely, no significant phenotypic correlations were found between traits influenced by HC2, HC3 or Morpho2. An eighth pleiotropic QTL could have been identified for Cou1 and Cou3 on chromosome 16, but as the CI did not overlap, we preferred to describe two distinct and genetically linked QTL (Col6a and Col6b). Finally, three pairs of LRT peaks concerning traits from different categories had almost identical CI (Venom7 and Morpho7, Venom9 and Morpho8, HC5 and Morpo21). They were described as single trait QTL in this first stage. Information about the LRT peaks is summarized in Figure 4. This clearly shows a striking result: the QTL, except Col1 and HC3, were only found in one progeny. HC3 was a special case because it influenced different traits in progenies B (Cir7 and Cir12) and V (Cir5). Moreover, in many cases, several different QTL were found for one trait in one progeny when no QTL were found for the same trait in the other progeny. Finally, no QTL was found (neither in B or V progeny) for Pho, Cir1, Cir3 and Cir4.

Another striking result is the high number of QTL showing dominance: only 18 QTL showed additive patterns, and in 19 cases the test concluded to overdominance. The ratio of dominance over additive effect was higher than 5 in four cases, reaching a maximum value of 16 (Angle8 on chromosome 6). Moreover, for



**Figure 3.** Results of interval mapping along chromosome 1 for Cou1 (pleiotropic QTL Col1). Results from first screening (5 cM, dotted line) and second screening (1 cM, continuous line) are shown on the same graph. **a** Interval mapping in the queen B progeny; **b** interval mapping in the queen V progeny; the second screening was performed with an enlarged sample size of 277 individuals. The genetic position is slightly different between the two progenies due to variations in the genetic maps. Two sub-peaks were detected at the same physical position in each case. The second peak was discarded from the CI in queen B progeny due to a high difference in significance.

**Table II.** Synthesis of all detected QTL.

QTL name	$K^a$	Queen	Trait	LRT	Confidence interval			Direction <sup>d</sup>	Dominance <sup>e</sup>	Vq/Vp <sup>f</sup>	
					Range (kb) <sup>b</sup>	Start (kb) <sup>b</sup>	End (kb) <sup>b</sup>				
Col1	1	V	Cou1*	81.3	1,511	12,007	13,518	5.5	82	SD	31%
	1	V	Cou3*	71.3	1,250	12,007	13,257	5.0	62	SD	25%
	1	B	Cou3	65.8	347	13,171	13,518	2.5	20	d	85%
	1	B	Cou1	78.5	261	13,257	13,518	1.5	20	D	83%
	1	B	Cou2	19.0	356	13,518	13,874	5.5	9	Co	40%
Col2	15	B	Cou1	25.8	550	5,728	6,278	10.5	28	d	37%
	15	B	Cou3	21.1	550	5,728	6,278	14.0	28	D	30%
HC1	4	B	Cir2	15.4	1,409	3,069	4,478	18.5	61	d	36%
	4	B	Cir8	15.4	1,409	3,069	4,478	18.5	61	Co	33%
HC2	4	B	Cir11	13.8	509	3,342	3,851	8.0	19	Co	32%
	4	B	Cir7	14.3	412	3,342	3,754	5.0	19	Co	32%
	13	B	Cir12	19.9	658	133	791	10.0	15	d	44%
	13	B	Cir9	15.4	712	133	845	16.5	16	d	34%
	13	B	Cir2	13.3	651	663	1,314	20.5	34	SD	23%
HC3	16	V	Cir5	19.3	2,422	172	2,594	9.5	91	d	32%
	16	B	Cir7	14.8	426	1,164	1,590	3.0	9	d	24%
Morpho1	16	B	Cir12	19.7	403	2,062	2,465	3.5	23	Co	44%
	1	V	Angle7	16.5	602	17,561	18,163	10.0	11	Co	41%
	1	V	Angle4	15.9	924	18,162	19,086	6.5	12	d	29%
Morpho2	1	V	Angle5	21.8	893	18,162	19,055	5.5	12	Co	45%
	2	V	Angle7	26.4	452	5,662	6,114	4.0	17	d	42%
	2	V	Angle6	16.7	721	5,987	6,708	8.0	3	Co	34%
Col3	1	B	Cou2	14.5	635	21,691	22,326	4.0	27	SD	21%
	8	V	Cou2	15.4	1,374	7,771	9,145	10.5	75	SD	21%
	8	V	Cou3	14.2	402	3,933	4,335	6.8	8	Co	27%
	16	B	Cou1	18.4	125	3,132	3,257	5.0	0	d	27%
	16	B	Cou3	15.5	166	3,359	3,525	4.0	1	SD	18%

Table II (continued).

QTL name	K <sup>a</sup>	Queen	Trait	LRT	Confidence interval			Direction <sup>d</sup>	Dominance <sup>e</sup>	V <sub>q</sub> /V <sub>p</sub> <sup>f</sup>		
					Range (kb) <sup>b</sup>	Start (kb) <sup>b</sup>	End (kb) <sup>b</sup>				Range (cM)	Nb genes <sup>c</sup>
Venom1	16	B	Nor	14.5	208	2,924	3,132	9.0	1	-	SD	20%
Venom2	3	V	Dop	17.7	551	4,075	4,626	11.0	21	-	D	27%
Venom3	3	V	Dop	15.5	796	6,830	7,626	11.0	33	-	d	28%
Venom4	1	B	Apa	16.2	1,020	1,669	2,689	11.0	26	-	D	34%
Venom5	1	V	Apa	15.8	829	16,996	17,825	12.5	18	+	D	42%
Venom6	5	B	Apa	17.0	489	5,055	5,544	14.0	29	+	D	39%
Venom7	4	V	Mel	21.4	128	1,462	1,590	5.0	5	-	d	35%
Venom8	4	V	Mel	13.9	79	12,153	12,232	1.5	1	+	Co	26%
Venom9	7	V	Mel	13.1	1,921	7,386	9,307	21.5	43	-	D	21%
HC4	6	B	Cir2	14.6	1,164	14,015	15,179	16.5	26	+	D	32%
HC5	12	V	Cir5	13.1	547	3,821	4,368	3.0	3	+	d	22%
HC6	1	V	Cir6	13.3	983	14,693	15,676	7.0	15	+	Co	28%
HC7	2	V	Cir6	16.5	1,165	8,024	9,189	11.0	32	+	SD	20%
HC8	10	B	Cir8	15.7	747	1,273	2,020	9.0	4	-	SD	19%
HC9	8	B	Cir9	17.5	77	1,009	1,086	3.5	1	+	SD	26%
HC10	14	B	Cir9	27.5	258	4,331	4,589	4.0	8	-	D	54%
HC11	5	V	Cir10	14.9	501	11,586	12,087	12.0	25	+	D	29%
HC12	11	B	Cir11	13.1	240	5,160	5,400	6.2	0	+	SD	22%
Morpho3	10	B	Angle1	14.1	295	11,650	11,945	3.5	10	+	d	29%
Morpho4	5	V	Angle2	15.5	683	6,181	6,864	12.5	30	+	D	26%
Morpho5	4	B	Angle3	16.2	224	2,084	2,308	8.0	12	+	Co	28%
Morpho6	3	B	Angle4	18.3	423	12,036	12,459	4.0	19	+	d	35%
Morpho7	4	V	Angle4	18.1	128	1,462	1,590	8.0	5	+	d	32%
Morpho8	7	V	Angle4	16.3	706	7,386	8,092	11.5	19	+	SD	26%
Morpho9	16	V	Angle6	13.9	309	3,402	3,711	10.4	4	-	d	25%
Morpho10	4	B	Angle7	13.5	81	71	152	2.0	6	-	Co	27%
Morpho11	8	V	Angle7	17.6	522	10,743	11,265	5.0	12	-	SD	23%
Morpho12	10	B	Angle7	15.1	387	5,719	6,106	9.0	22	-	Co	36%
Morpho13	1	B	Angle8	17.1	608	3,250	3,858	13.5	11	+	SD	18%

**Table II** (continued).

QTL name	K <sup>a</sup>	Queen	Trait	LRT	Confidence interval			Direction <sup>d</sup>	Dominance <sup>e</sup>	V <sub>q</sub> /V <sub>p</sub> <sup>f</sup>		
					Range (kb) <sup>b</sup>	Start (kb) <sup>b</sup>	End (kb) <sup>b</sup>				Range (cM)	Nb genes <sup>c</sup>
Morpho14	1	B	Angle8	13.6	569	7,519	8,088	6.0	31	+	SD	17%
Morpho15	6	B	Angle8	14.8	140	9,018	9,158	6.3	0	-	d	21%
Morpho16	6	B	Angle8	13.9	963	15,109	16,072	6.6	32	+	SD	15%
Morpho17	9	B	Angle8	15.2	520	9,962	10,482	12.0	34	-	D	30%
Morpho18	5	B	Angle9	13.4	624	10,715	11,339	9.0	27	+	SD	24%
Morpho19	7	V	Angle9	14.8	149	12,528	12,677	10.0	5	+	Co	30%
Morpho20	13	B	Angle9	16.6	444	1,102	1,546	6.0	15	-	Co	35%
Morpho21	12	B	Angle10	13.8	547	3,821	4,368	6.0	3	+	Co	29%
Morpho22	4	B	d1-8	14.3	294	7,708	8,002	4.0	3	-	SD	18%

Pleiotropic QTL influencing at least two traits from the same category are listed first

LRT likelihood ratio test

<sup>a</sup>Chromosome location of the QTL

<sup>b</sup>The QTL localization was based on an enlarged sample size of 277 individuals

<sup>c</sup>Range, start and end in kb are from the v4 assembly of honeybee genome

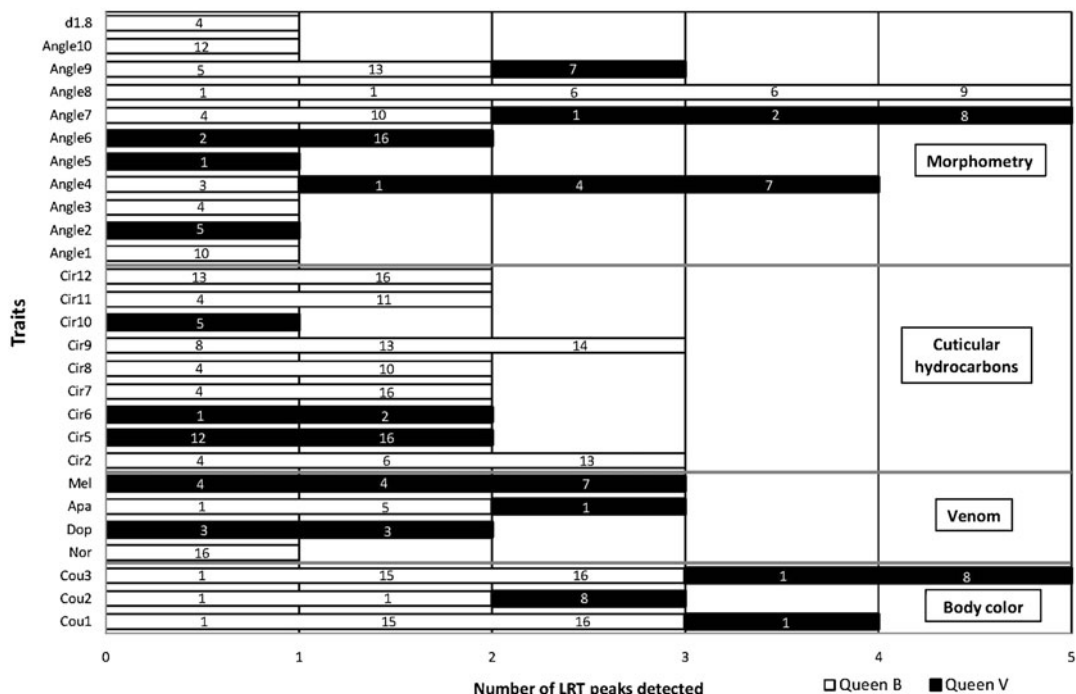
<sup>d</sup>Number of annotated genes in the confidence interval

<sup>e</sup>Direction of the effect with minus (-) indicating a decrease of the measured character with the *A. m. mellifera* allele at the QTL

<sup>f</sup>Co (codominance,  $|d/a| < 0.5$ ), d (partial dominance,  $0.5 < |d/a| < 0.8$ ), D (complete dominance,  $0.8 < |d/a| < 1.2$ ), SD (overdominance,  $|d/a| > 1.2$ )

<sup>f</sup>Percentage of phenotypic variance explained by the QTL





**Figure 4.** Number of LRT peaks detected for each trait in each progeny (progeny B in *white* and progeny V in *black*). The number on each bar corresponds to the number of the chromosome bearing the LRT peak.

five traits in progeny B (Cou1, Cou3, Cir9 and Angle8) and one in progeny V (Angle7), the addition of the QTL effects explained more than 100% of the phenotypic variance. More than 70% of the phenotypic variance is explained for five other traits within progeny (Apa, Cir2 and Cir12 in progeny B, Mel and Angle4 in progeny V).

The range of the CI was highly variable: between 76 and 2,422 kb, corresponding to 1 and 22 cM, respectively. This variability, and the large values for some CI, could not result from discrepancies in the density of the genetic maps between these regions, as density was increased to similar values in all segments, but could indicate the presence of more than one gene of interest in the same region.

### 3.3 Genes of interest within QTL CI

Table III lists genes known to be involved in molecular processes of the quantitative variation studied here and localized in or close to the CI

of the detected QTL. These genes were either known from model species such as *D. melanogaster* and previously identified as homologous genes in honeybee, or specific genes mapped to honeybee (venom components). We searched for genes involved in pigment production as well as in patterning. Two well-known genes (*Tan* and *Ebony*) are present in or very close to the CI of Col1: the first clearly distinct peak in queen V (Figure 3b) was very close to *Tan*, and the second peak, shared between the two queens (Figure 3a, b), included *Ebony*. One of the patterning genes could be located in the CI of Col1 (*wnt4*). This pleiotropic regulatory protein influences pigmentation, among other things, and belongs to the developmental genes described in Dearden et al. (2006). However, no particular gene could be identified in or close to the CI of the six other QTL for body colour.

Cuticular hydrocarbons are derived from fatty acids. Among the 12 and six putative

**Table III.** Genes of interest found in the confidence intervals of the QTL detected.

QTL	Gene/function	ID	<i>K</i>	Start	Stop	<i>d</i> from CI
Col1	Tan	GB13643	LG1	11,991,507	11,996,967	10,630
	Wnt4 Signal transduction	GB10450	LG1	12,597,050	12,616,120	Within
	Ebony	GB19941	LG1	13,297,018	13,311,847	Within
Venom3	Dopamine beta hydroxylase	GB16337	LG3	7,514,385		Within
HC3	Long-chain fatty acid elongation	GB12176	LG16	1,762,074	1,764,525	Within
	Long-chain fatty acid elongation	GB19345	LG16	1,909,141	1,933,372	Within
	Long-chain fatty acid elongation	GB19268	LG16	1,819,639	1,830,897	Within
	Long-chain fatty acid elongation	GB14916	LG16	1,938,602	1,984,821	Within
	Long-chain fatty acid elongation	GB16107	Un.655=LG16	1,767,245	1,817,244	Within
	Long-chain fatty acid elongation	GB13826	Un.655=LG16	1,767,245	1,817,244	Within
HC5	Desaturase	GB18070	LG12	4,460,820	4,462,763	92,819
Morpho1	Engrailed	GB15566	LG1	19,145,586	19,191,405	59,815
Morpho3	Yellow (Yellow-y)	GB19464	LG10	11,741,768	11,747,281	Within
Morpho18	Wnt11 Signal transduction	GB16243	LG5	11,040,606	11,046,486	Within
Morpho11	Hephaestus, Notch signalling	GB12758	LG8	10,988,503	11,037,222	Within

*K* chromosome number (or linkage group (LG)), *d* from CI distance from confidence interval of the QTL (in bp)

genes annotated for homology with elongase and desaturase, respectively, four of the elongase genes were located close to each other in the CI of HC3. Furthermore, several annotated genes were located in unknown scaffold in the published assembly (not attributed to a linkage group). However, Robertson et al. (2007) manually incorporated some of these unknown scaffolds into the smallest chromosomes. Among them, scaffold Un.655, comprising two genes homologous to elongase, was placed between scaffolds 16.7 and 16.8, within the CI of HC3. It seems that several genes coding for elongase are clustered in this region, as observed in *D. melanogaster*. One gene homologous to desaturase was also observed in the vicinity of HC5, but the distance from the CI was quite large (93 kb).

Dopamine and noradrenaline are well known as neurotransmitters from the catecholamine

group (Squire et al. 2003). One of the well-described enzymes involved in the synthesis of catecholamine could be found in the CI of Venom3: dopamine beta hydroxylase catalyses the transformation of dopamine to noradrenaline. However, no candidate gene could be located in or close to the CI of QTL Venom4 to Venom9 influencing mellitin and apamin traits.

We compared the location of the identified genes involved in the wing vein pattern of honeybee (Dearden et al. 2006) with the CI found for angles. Despite the large number of QTL and the large number of candidate genes, few co-locations were found. Furthermore, we signalled the presence of a yellow gene in the CI of Morpho3, although this gene is not described as influencing wing vein characteristics. In *D. melanogaster*, this gene is required for the deposition of melanin, a pigment

differentially expressed between vein and inter-vein cells.

#### 4. DISCUSSION

Quantitative variation is complex and in most cases only partly explained by known genetic variants. Genetic dissection of quantitative traits implies the estimation of the number of loci involved, of the allelic effect of each locus and of the interactions between alleles at the same locus and at different loci.

Honeybee genetics is not highly developed despite the agronomic interest of the species. One reason is honeybee social organization and the fact that most agronomic traits are expressed at the colony level. Furthermore, inbreeding is highly unfavourable due to single-locus sex determination (Hasselmann et al. 2001), which precludes genetic studies on homozygous strains. On the other hand, effects of environmental variation are somewhat attenuated, although not eliminated, by nest breeding of young bees. We intended to bypass the lack of divergent homozygous strains by crossing divergent subspecies, as frequently done in livestock species. The assumption is that allelic variation at the locus of interest will be greater between subspecies than within subspecies. Indeed, subspecies effect was detected for most of the traits. However, within subspecies variation appeared to be high for most, if not all of the traits. As a consequence, over the 50 QTL detected, only two of them were shared between the two progenies, and up to five QTL were detected for one progeny when no QTL at all were detected in the other. Comparing the variances estimated for each progeny for the different traits (not shown) confirmed both the hypothesis of simple allelic variation within subspecies and the hypothesis of existence of dominance and/or epistasis effects on the traits studied. As a consequence, our strategy of searching for QTL in two different progenies was successful in maximising the number of QTL detected, given the number of genotypes and phenotypes recorded.

Another striking result is the high number of QTL showing overdominance. This is probably the result of the between-subspecies cross where association of very distinct alleles can generate new phenotype variability. This result would correspond to the heterosis effect described in genetic processing of crops or livestock species (Gallais 1988; Hochholdinger and Hoecker 2007). Finally, it should be noted that part of the detected dominance could also be a consequence of the deviations from normality in the distribution of some of the traits (for Cir2 for example) and needs confirmation.

Besides the discrepancies in QTL detected between the progenies, the proportion of phenotypic variance explained by the QTL was also variable (Table II). It should be noted that the sum of  $V_q/V_p$  for a single trait in a single progeny may exceed 100%, which is an overestimation of individual QTL effects and a known consequence of the models that maximise the variance at one QTL position to detect QTL. Adding effects estimated independently at the different significant positions is straightforward to obtain a general view of the genetic architecture of the traits, but it is not statistically correct. Only models jointly accounting for all loci, and potentially their interactions, would allow precise QTL variance estimations, with only marginal improvement of location accuracy. However, high number of progeny is required to achieve accurate estimation of all genotypic effects when multiple positions are tested jointly, and no software is available that accounts for double backcross progeny. In most cases two or three QTL were sufficient to explain more than 70% of the phenotypic variance. This suggests the existence of a low number of major genes involved in the quantitative variation of the studied traits. The lowest proportion of variance explained by one QTL was 15%, probably corresponding to the lower limit for detection of one QTL in our protocol.

No QTL was detected in either the B or V progenies for the four traits Phos, Cir1, Cir3 and Cir4. No single reason can explain this observation for all the traits. Their phenotypic variance was of the same magnitude as the

variance of other traits for which QTL were detected. Three of the four traits with no QTL did not fit a normal distribution (Cir1, Cir3 and Cir4), but the fourth one did (Pho); there were other cases where traits for which QTL were detected did not fit this assumption either (Cir2). One explanation could be the larger environmental variances and difficulties in measuring these traits, resulting in lower phenotyping accuracies. However, it is unlikely that environmental variance diverges strongly between traits from the same category. Finally, these traits could be under the influence of numerous genes, each contributing only a small part of the overall variance, e.g. <15% of the phenotypic variance, and were thus undetectable with the sample and genetic maps used in this study.

Seven pleiotropic QTL were detected in this study, as well as three cases of co-localization (Venom7 and Morpho7, Venom9 and Morpho8, HC5 and Morpo21). One would expect, these genetic linkages to be observed when correlations between traits were significant, as for Col1, Col2, HC1 and Morpho1, but they were less clear for other pleiotropic QTL and co-localizations. Even without the Holm correction to account for multiple testing, correlations supporting HC2, HC3 and Morpho2 remained undetectable. One possibility is that the interval comprised two different genes influencing the traits. These genes would be close enough to generate low genetic correlations between the traits, but the genetic correlations could not be estimated in our populations. Among other explanations for the lack of correlations in agreement with pleiotropic QTL, opposite correlations resulting from environment and/or other QTL might mask correlations due to pleiotropic QTL or segments.

Some genes known to influence the trait studied in model organisms could be located in the CI of the QTL found. For example, the presence of Tan and Ebony precisely in the two peaks detected in the Col1 QTL is a strong validation of the region proposed. However, despite the large size of the gene space sought (about 200 genes over the four pathways), only

a few genes could be assigned to the QTL found. This result might be linked to the unfinished status of honeybee genome assembly. The published version of the genome was still fragmented, with 186 Mb positioned on 626 mapped scaffolds (located in a specific chromosome) and 40 Mb in unmapped scaffolds (Honeybee Genome Sequencing Consortium 2006). Some superscaffolding efforts have been made to reduce the gaps in the sequence (Robertson et al. 2007), but numerous scaffolds are still referred to as “unknown”, as their position in the genome is not determined. In fact some genes of putative influence on the traits studied here were located in these “unknown” scaffolds and might be relocated in our CI in the future. The recent release of a new version of the genome (4.5) improved considerably this picture with 203 Mb assembled into chromosomes but the sequence is still dispersed over 340 mapped scaffolds and 5,304 unmapped scaffolds representing 31 Mb. Furthermore, besides the gaps in the sequence assembly, the number of annotated genes in *A. mellifera* is small compared with *D. melanogaster*. The RefSeq protein set used in Blast requests on the NCBI website comprises only 10,409 proteins for honeybee, but 18,556 for fruit fly. This difference is likely to be due to better knowledge of *Drosophila* genetics. Furthermore, the large evolutionary distance between honeybee and model organisms is unfavourable for annotation of its gene set. In comparison, the RefSeq protein set comprises 15,995 genes for *D. pseudoobscura*, a close relative of *D. melanogaster*, and 12,902 genes for *Anopheles gambiae*, which is more distant, but also in the order Diptera. The publication of the *A. mellifera* genome assembly also suggests that annotation was very stringent and maybe missed some known genes. This is confirmed by a recent study of the completeness of genome sequencing projects (Parra et al. 2009), which showed that the coverage of the honeybee genome sequence is good but that the annotation is not complete even for genes highly conserved between eukaryotes. Another annotation difficulty may result from honeybee

genes having more and longer introns than those of fruit fly. All together, gaps in the sequence assembly and incomplete annotation are probably responsible for the absence of genes in the CI of two QTL (Col6a and Morpho15). It is also likely that the number of genes was underestimated in many of the CI. Whatever the gaps in genome and annotation, the lack of identified genes in the CI of the QTL may also be related to genes and/or pathways specific to honeybee. The strength of our approach is the opportunity it offers to describe new genes involved in observed quantitative variation, adding to our understanding of these traits even in the era of complete genome description.

## 5. CONCLUSIONS

Fifty different QTL could be described in this study, and some of them with highly significant LRT at the genome-wide level. The generally large proportion of variance explained by the detected QTL suggests the existence of few major genes responsible for the genetic architecture of traits in the four categories studied. The traits revealed great variability within the two parental subspecies, resulting in the detection of very different sets of QTL in the two queen families analysed. Few candidate genes could be proposed based on the knowledge of molecular pathways and confidence intervals described. However, this number could increase with improvement of the annotation of the honeybee genome and completeness of the genome assembly.

## ACKNOWLEDGEMENTS

We would like to thank Gérard Arnold and Aurore Arnold for their technical support.

**Locus de caractères quantitatifs (QTL) impliqués dans la couleur du corps, la morphométrie alaire, les hydrocarbures cuticulaires et les composants du venin chez l'abeille.**

**Abeille / génétique / QTL / morphométrie / venin / hydrocarbures cuticulaires**

**Quantitative Traits Loci (QTL) für Körperfärbung, Flügelmorphometrie, kutikuläre Kohlenwasserstoffe und Giftzusammensetzung der Honigbiene**

**Honigbiene / Genetik / qtl / Morphometrie / kutikuläre Kohlenwasserstoffe / Gift**

## REFERENCES

- Abramoff, M.D., Magalhaes, P.J., Ram, S.J. (2004) Image processing with ImageJ. *Biophotonics Int.* **11**, 36–42
- Ameratunga, R.V., Hawkins, R., Prestidge, R., Marbrook, J. (1995) A high efficiency method for purification and assay of bee venom phospholipase A2. *Pathology* **27**, 157–160
- Arechavaleta-Velasco, M.E., Hunt, G.J., Emore, C. (2003) Quantitative trait loci that influence the expression of guarding and stinging behaviors of individual honey bees. *Behav. Genet.* **33**, 357–364
- Arnold, G., Quenet, B., Cornuet, J.M., Masson, C., De Schepper, B., Estoup, A., Gasqui, P. (1996) Kin recognition in honeybees. *Nature* **379**, 498
- Arnold, G., Quenet, B., Masson, C. (2000) Influence of social environment on genetically based subfamily signature in the honeybee. *J. Chem. Ecol.* **26**, 2321–2333
- Blair, S.S. (2007) Wing vein patterning in *Drosophila* and the analysis of intercellular signaling. *Annu. Rev. Cell Dev. Biol.* **23**, 293–319
- Chaline, N., Sandoz, J.C., Martin, S.J., Ratnieks, F.L., Jones, G.R. (2005) Learning and discrimination of individual cuticular hydrocarbons by honeybees (*Apis mellifera*). *Chem Senses* **30**, 327–335
- Chandra, S.B., Hunt, G.J., Cobey, S., Smith, B.H. (2001) Quantitative trait loci associated with reversal learning and latent inhibition in honeybees (*Apis mellifera*). *Behav. Genet.* **31**, 275–285
- Chertemps, T., Duportets, L., Labeur, C., Ueda, R., Takahashi, K., Saigo, K., Wicker-Thomas, C. (2007) A female-biased expressed elongase involved in long-chain hydrocarbon biosynthesis and courtship behavior in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **104**, 4273–4278
- Dallerac, R., Labeur, C., Jallon, J.M., Knipple, D.C., Roelofs, W.L., Wicker-Thomas, C. (2000) A delta 9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **97**, 9449–9454
- De Celis, J.F., Diaz-Benjumea, F.J. (2003) Developmental basis for vein pattern variations in insect wings. *Int. J. Dev. Biol.* **47**, 653–663
- Dearden, P.K., Wilson, M.J., Sablan, L., Osborne, P.W., Havler, M., McNaughton, E., Kimura, K., Milshina, N.V., Hasselmann, M., Gempe, T., Schioett, M., Brown, S.J., Elsik, C.G., Holland, P.W., Kadowaki, T., Beye, M. (2006) Patterns of conservation and change in honey bee developmental genes. *Genome Res.* **16**, 1376–1384



- R Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Diniz-Filho, J.A., Hepburn, R., Radloff, S.E., Fuchs, S. (2000) Spatial analysis of morphological variation in African honeybees (*Apis mellifera* L.) on a continental scale. *Apidologie* **31**, 191–204
- Dotimas, R.E., Hider, R.C. (1987) Honeybee venom. *Bee World* **68**, 51–71
- Drapeau, M.D., Albert, S., Kucharski, R., Prusko, C., Maleszka, R. (2006) Evolution of the yellow/major royal jelly protein family and the emergence of social behavior in honey bees. *Genome Res.* **16**, 1385–1394
- Eigenheer, A.L., Young, S., Blomquist, G.J., Borgeson, C.E., Tillman, J.A., Tittiger, C. (2002) Isolation and molecular characterization of *Musca domestica* delta-9 desaturase sequences. *Insect Mol. Biol.* **11**, 533–542
- Elsen, J.M., Mangin, B., Goffinet, B., Boichard, D., Le Roy, P. (1999) Alternative models for QTL Detection in livestock. I. General introduction. *Genet. Sel. Evol.* **31**, 213–224
- Gallais, A. (1988) Heterosis: its genetic basis and its utilization in plant breeding. *Euphytica* **39**, 95–104
- Gmachl, M., Kreil, G. (1995) The precursors of the bee venom constituents apamin and MCD peptide are encoded by two genes in tandem which share the same 3'-exon. *J. Biol. Chem.* **270**, 12704–12708
- Goffinet, B., Le Roy, P., Boichard, D., Elsen, J.M., Mangin, B. (1999) Alternative models for QTL detection in livestock. III. Heteroskedastic model and models corresponding to several distributions of the QTL effect. *Genet. Sel. Evol.* **31**, 341–350
- Hasselmann, M., Fondrk, M.K., Page, R.E., Beye, M. (2001) Fine scale mapping in the sex locus region of the honey bee (*Apis mellifera*). *Insect Mol. Biol.* **10**, 605–608
- Hider, R.C. (1988) Honeybee venom: a rich source of pharmacologically active peptides. *Endeavour* **12**, 60–65
- Hochholdinger, F., Hoecker, N. (2007) Towards the molecular basis of heterosis. *Trends Plant Sci.* **12**, 427–432
- Hoffman, D.R. (2006) Hymenoptera venom allergens. *Clin. Rev. Allergy Immunol.* **30**, 109–128
- Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**, 931–949
- Howard, R.W., Blomquist, G.J. (2005) Ecological, behavioral, and biochemical aspects of insect hydrocarbons. *Annu. Rev. Entomol.* **50**, 371–393
- Hunt, G.J., Page Jr., R.E. (1995) Linkage map of the honey bee, *Apis mellifera*, based on RAPD markers. *Genetics* **139**, 1371–1382
- Hunt, G.J., Page Jr., R.E., Fondrk, M.K., Dullum, C.J. (1995) Major quantitative trait loci affecting honey bee foraging behavior. *Genetics* **141**, 1537–1545
- Hunt, G.J., Guzman-Novoa, E., Fondrk, M.K., Page Jr., R.E. (1998) Quantitative trait loci for honey bee stinging behavior and body size. *Genetics* **148**, 1203–1213
- Hunt, G.J., Collins, A.M., Rivera, R., Page Jr., R.E., Guzman-Novoa, E. (1999) Quantitative trait loci influencing honeybee alarm pheromone levels. *J. Hered.* **90**, 585–589
- Lander, E.S., Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199
- Lapidge, K.L., Oldroyd, B.P., Spivak, M. (2002) Seven suggestive quantitative trait loci influence hygienic behavior of honey bees. *Naturwissenschaften* **89**, 565–568
- Legendre, A., Miao, X.X., Da Lage, J.L., Wicker-Thomas, C. (2008) Evolution of a desaturase involved in female pheromonal cuticular hydrocarbon biosynthesis and courtship behavior in *Drosophila*. *Insect Biochem. Mol. Biol.* **38**, 244–255
- Lima, P.R.D., Brochetto-Braga, M.R. (2003) Hymenoptera venom review focusing on *Apis mellifera*. *J. Venom Anim. Toxins* **9**, 149–162
- Linksvayer, T.A., Rueppell, O., Siegel, A., Kaftanoglu, O., Page Jr., R.E., Amdam, G.V. (2009) The genetic basis of transgressive ovary size in honeybee workers. *Genetics* **183**, 693–707
- Molnar, C., Holguin, H., Mayor Jr., F., Ruiz-Gomez, A., De Celis, J.F. (2007) The G protein-coupled receptor regulatory kinase GPRK2 participates in hedgehog signaling in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **104**, 7963–7968
- Ogden, J., Pardo, C., Tchaplaj, A. (1998) Method development for the analysis of bee cuticular and esters by gas chromatography. *J. Chromatogr. Sci.* **36**, 287–292
- Oldroyd, B., Rinderer, T., Buco, S. (1991) Heritability of morphological characters used to distinguish European and Africanized honeybees. *Theor. Appl. Genet.* **82**, 499–504
- Oxley, P.R., Thompson, G.J., Oldroyd, B.P. (2008) Four quantitative trait loci that influence worker sterility in the honeybee (*Apis mellifera*). *Genetics* **179**, 1337–1343
- Parra, G., Bradnam, K., Ning, Z., Keane, T., Korf, I. (2009) Assessing the gene space in draft genomes. *Nucleic Acids Res.* **37**, 289–297
- Pendleton, R.G., Rasheed, A., Roychowdhury, R., Hillman, R. (1998) A new role for catecholamines: ontogenesis. *Trends Pharmacol. Sci.* **19**, 248–251
- Poklucar, J., Kezic, N. (1994) Estimation of some characteristics of hind legs and wings of honeybee workers (*Apis mellifera carnica* Polm) using the half-sibs method. *Apidologie* **25**, 3–11
- Radloff, S.E., Hepburn, R., Bangay, L.J. (2003) Quantitative analysis of intracolony and intercolony morphometric variance in honeybees, *Apis mellifera* and *Apis cerana*. *Apidologie* **34**, 339–351



- Rajendra, W., Armugam, A., Jeyaseelan, K. (2004) Toxins in anti-nociception and anti-inflammation. *Toxicon* **44**, 1–17
- Robertson, H.M., Reese, J.T., Milshina, N.V., Agarwala, R., Solignac, M., Walden, K.K., Elsie, C.G. (2007) Manual superscaffolding of honey bee (*Apis mellifera*) chromosomes 12–16: implications for the draft genome assembly version 4, gene annotation, and chromosome structure. *Insect Mol Biol.* **16**, 401–410
- Rueppell, O., Pankiw, T., Nielsen, D.I., Fondrk, M.K., Beye, M., Page Jr., R.E. (2004) The genetic architecture of the behavioral ontogeny of foraging in honeybee workers. *Genetics* **167**, 1767–1779
- Rueppell, O., Chandra, S.B., Pankiw, T., Fondrk, M.K., Beye, M., Hunt, G., Page, R.E. (2006) The genetic architecture of sucrose responsiveness in the honeybee (*Apis mellifera* L.). *Genetics* **172**, 243–251
- Rueppell, O., Metheny, J.D., Linksvayer, T., Fondrk, M. K., Page Jr., R.E., Amdam, G.V. (2011) Genetic architecture of ovary size and asymmetry in European honeybee workers. *Heredity* **106**, 894–903
- Ruttner, F. (1988) Biogeography and taxonomy of honeybees. Springer, Berlin
- Ruttner, F., Tassencourt, L., Louveaux, J. (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie* **9**, 363–381
- Ruttner, F., Pour, E.M., Fuchs, S. (2000) Ecoclines in the near east along 36° N latitude in *Apis mellifera* L. *Apidologie* **31**, 157–165
- Sander, V., Eivers, E., Choi, R.H., De Robertis, E.M. (2010) *Drosophila* Smad2 opposes Mad signaling during wing vein development. *Plos One* **5**, E10383
- Schmidt, J.O. (1995) Toxinology of venoms from the honeybee genus *Apis*. *Toxicon* **33**, 917–927
- Schumacher, M.J., Schmidt, J.O., Egen, N.B., Dillon, K.A. (1992) Biochemical variability of venoms from individual European and Africanized honeybees (*Apis mellifera*). *J. Allergy Clin. Immunol.* **90**, 59–65
- Solignac, M., Vautrin, D., Baudry, E., Mougel, F., Loiseau, A., Cornuet, J.M. (2004) A microsatellite-based linkage map of the honeybee, *Apis mellifera* L. *Genetics* **167**, 253–262
- Solignac, M., Mougel, F., Vautrin, D., Monnerot, M., Cornuet, J.M. (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
- Squire, L.R., Bloom, F.E., McConnell, S.K., Roberts, J. L., Spitzer, N.C., Zigmond, M.J. (2003) Neurotransmitters. In: *Fundamental neuroscience*, pp. 163–196. Academic, San Diego
- Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C. G., Liu, L., Pardinas, J.R., Robertson, H.M., Soares, M.B., Robinson, G.E. (2002) Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res* **12**, 555–566
- Wittkopp, P.J., Beldade, P. (2009) Development and evolution of insect pigmentation: genetic mechanisms and the potential consequences of pleiotropy. *Semin. Cell Dev. Biol.* **20**, 65–71
- Wittkopp, P.J., True, J.R., Carroll, S.B. (2002) Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* **129**, 1849–1858
- Yan, S.J., Zartman, J.J., Zhang, M., Scott, A., Shvartsman, S.Y., Li, W.X. (2009) Bistability coordinates activation of the EGFR and DPP pathways in *Drosophila* vein differentiation. *Mol. Syst. Biol.* **5**, 278