GRAHAM, ALLIE MARIE, M.S. The Genetic Architecture of Reproductive Differences in Workers of Africanized and European Honey Bees, *Apis mellifera*. (2009) Directed by Dr. Olav Rueppell. 94 pp.

The Western Honeybee (*Apis mellifera*) displays a special form of social behavior called eusociality. The evolution of its reproductively specialized castes and social behavior from a solitary ancestor may be explained by the reproductive ground plan hypothesis. This hypothesis predicts a relationship between the variation of ovary size and -activity and social behavior. At the phenotypic level, ovary size has been associated with a whole set of behavioral phenotypes, known as the pollen hoarding syndrome. While many of these phenotypes are potentially influenced by regulatory pathways, involving juvenile hormone and vitellogenin, the exact genetic links between ovary size determination and social behavior are still unknown.

To test the generality of the hypothesized genetic linkage between reproductive and social behavior, I investigated the genetic architecture of ovary size differences between Africanized and European honey bees. Two backcrosses of a hybrid queen and Africanized drones that resulted in transgressive worker ovary phenotypes were studied for pleiotropic effects of existing behavioral QTL and potential new QTL with a combination of SNP and microsatellite markers. Analyses show small but significant effects on ovary size for some of the behavioral QTL, as predicted by the reproductive ground plan hypothesis. In addition, I detected two new QTL of major effect on ovary size. I describe potential candidate genes for the QTL and suggest that the detected major and minor effects could reflect genetic control of caste divergence and worker division of

labor, respectively, representing two distinct stages of honey bee social evolution that may be connected via female reproductive physiology.

THE GENETIC ARCHITECTURE OF REPRODUCTIVE

DIFFERENCES IN WORKERS OF AFRICANIZED AND

EUROPEAN HONEY BEES, *APIS MELLIFERA*.

by

Allie Marie Graham

A Thesis Submitted to
the Faculty of The Graduate School
at The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2009

Approved by

_____
Committee Chair

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of The

Graduate School at the University of the North Carolina at Greensboro.

Committee Chair_____

Committee Members_____

_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Social Evolution

The Western Honeybee (*Apis mellifera)* is an important organism not only due to

its role in agriculture as a primary pollinator, but also because of its rare form of social

behavior: eusociality. In general, social behavior is common throughout the animal

kingdom but its sophistication varies across a wide range (Sherman et al., 1995). A social

group is characterized by the presence of multiple generations and a communal system of

caring for the group, mainly the young. Eusocial organisms have a caste system based on

reproductive status, in addition to overlapping of generations and cooperative care for the

young (Crespi et al., 1995). Eusociality is fairly uncommon except in insects, such as

bees, ants, wasps and termites. There are, however, a few exceptions in crustaceans

(Duffy et al., 2000) and mammals (Jarvis, 1981).

The ultimate explanation for the evolution of eusociality is kin selection. This

general principle of natural selection predicts that altruistic behavior will be favored

when the benefit of an altruistic action (b), weighted by the relatedness between donor

and the recipient (r), outweighs the cost to the donor (c): $b*r > c$ (Hamilton 1964). Under

this circumstance, genes may be better passed on through relatives than through direct

reproduction (Wilson, 1971; Foster et al., 2005). Thus, in order for the altruistic behavior

to evolve by kin selection,

individuals must be related. In haplo-diploid species, such as the social Hymenoptera, sex is determined by fertilization with males generally being haploid and females diploid. This makes female colony members highly related (Figure 1) and could explain the predominance of eusociality in Hymenoptera (Hamilton, 1965; Foster et al. 2005).The advantage could have continued long enough to lead to the evolution of synergistic social systems, which then allowed the maintenance of social behavior (Trivers and Hare, 1976).

The typical Hymenopteran social system is exemplified by the honey bee (*Apis mellifera* L): Since the queen bee is the only reproducing individual, all the daughters are sisters. Drones (male bees) pass on their entire genome and full sisters are therefore more related to each other than to their own offspring, if they were to produce any. Full sisters share seventy-five percent of their genes, on average (Hamilton, 1965). If the worker sisters were to reproduce, their own offspring would only share fifty precent of their genes. Therefore, the production of full sisters rather than daughters is advantageous (Queller et al, 1998). The principle of kin selection is not only true for eusocial organisms, but also for some vertebrates that cooperatively breed, and are not subject to sex determination via haplodiploidy (Griffin et al., 2003). Even though kin selection provides an ultimate explanation for why eusociality may have evolved, it does not give any proximate explanation that would suggest a mechanism for how this particular social form evolved from solitary ancestors.

Honeybee Biology

The social structure of *A. mellifera* is composed of several castes within a typical colony: hundreds of male drones, thousands of female workers, and one female queen (Lindauer, 1953; Rösch, 1925). The drones' only function is to mate with queens of other colonies to perpetuate their genes. Otherwise, they provide no communal service to the hive. After mating, the queen specializes exclusively on egg production and is groomed and fed for this purpose by the workers. The workers are in charge of all remaining tasks, and consequently the overall health and well-being of the hive, by providing food, caring for the brood, and defending the hive. The duties performed by the workers occur relative to the particular age of the individual (age polyethism), creating a temporal caste ontogeny (Free, 1965; Lindauer, 1953; Rösch, 1925). The individual emerges ready for in-hive tasks, but in the later part of its life, there is a switch to tasks outside the hive. As the workers transition between in-hive tasks, there is considerable overlap, which can lead to the performance of multiple taks at any given age (Ribbands, 1953; Free 1965). Initially, young workers start tasks that are related to cleaning the nest and preparing the cells. These tasks are important because it rids the hive of unsanitary debris, and also synchronizes the emergence of another group of workers that are laid in the newly cleaned cells. Next, the worker tends to the brood and the queen by regular inspection and feeding (nurse bee), after which, the individual focuses on building the combs, and handling and storing food. During the final part of the workers' life, tasks such as ventilating, guarding, and ultimately foraging, are performed. The lifespan of a foraging bee can be as short as 4 to 5 days due to extensive flight trips (Rueppell et al. 2007), and

3

is accompanied by the degeneration of food and wax glands, and the fraying of wings. Generally, there is a sharp transition from in-hive to outside tasks, and foragers do not revert to in-hive tasks under normal circumstances, but the age at which the individuals transition is highly variable (Lindauer, 1953; Ribbands, 1952; Rösch, 1925).

The eggs laid by the queen have the potential to develop into any one of the three castes. Sex is determined by fertilization, and specifically the sex determination locus (Beye et al., 2003). Normally, the unfertilized eggs develop into drones, while a fertilized egg can develop either into a worker or a queen. Female differentiation depends on the quantity and quality of food that is fed to the developing larva (Haydak 1943; Hrassnigg et al., 2005; Johnson 2003). The food fed to potential queens is classified as royal jelly, which is particularly rich in proteinaceous mandibular gland secretions and sugars. The worker larvae are fed more hypopharyngeal gland secretions and also more honey and pollen during the latter part of larval development (Jung-Hoffmann, 1966). These nutritional factors are then translated into gene expression differences (Evans and Wheeler 1999, 2000; Wheeler et al., 2006; Amdam et al., 2007), which translate into differences in behavior, physiology, and morphology, that differentiate queens from workers and thus help the specialization to their particular social roles.

The most important of these differences is the female reproductive system between queens and workers. The female reproductive system of a honeybee consists of two ovaries, oviducts, spermatheca and the vulva (Figure 2). Each ovary consists of several ovarioles, in which several eggs may mature simultaneously when the ovary is active. Queen ovaries can contain up to 160 ovarioles (Winston 1987). As the egg

develops, it moves down the ovariole and exits via the oviducts. Queens are able to selectively fertilize eggs which are passing into the oviduct by releasing a small amount of sperm which is stored within the spermatheca. Workers are not able to selectively fertilize eggs because they do not engage in mating flights, and therefore do not have any sperm. Workers also have drastically reduced ovaries with few ovarioles and are facultatively sterile. However, in a queenless situation, workers will activate their ovaries and lay eggs destined to be drones (Velthuis, 1970).

Based on the reproductive biology of the honeybee, there should be specific associations between aspects of the ovary. For the queen and worker, the more ovarioles there are within an ovary, the more eggs can be produced and ovary size is thus quantitatively classified as reproductive potential. Because the left and right ovaries are functionally equivalent, genes are expected to affect them equally. Thus, the two ovaries are likely correlated with each other. Nevertheless, the two ovaries may contain differing numbers of ovarioles, creating asymmetry. The relative asymmetry of ovariole number between the two ovaries should be unrelated to ovary size and is not expected to show genetic overlap with ovary size measurements. Previous studies have confirmed the prediction that more ovarioles per ovary bias workers towards ovary activation in the absence of reproductive suppression by the queen (Makert et al., 2006). Thus, the size of the ovary should be correlated with ovary activation under queenless conditions.

Honeybee Worker Life History

There is not only reproductive inquality between the two female castes, but a continuum of reproductive inequality within the worker population (Makert et al., 2006). Ovary sizes, measured as ovariole number, are variable in the workers, which leads to a different tendency to activate their ovaries to reproduce in the absence of a queen. Typically, the level of oocyte maturation, or ovary activity, is inhibited by the presence of a queen through pheromone control (Butler, 1957; Queller et al, 1998). Both ovary size and activity are used as a direct measure of reproductive physiology and can predict reproductive output when no queen is present and workers compete for direct reproduction (Makert et al. 2006).

Even under normal social circumstances, the worker ovary influences, or is at least correlated with, major hormone and protein titers that govern worker division of labor and thus the ovary could potentially be a central player in the evolution of worker division of labor, including foraging regulation (Amdam et al., 2006, 2007; Hepperle and Hartfelder 2001; Johnson 2003; Nelson et al., 2007; Page and Erber 2002; Robinson and Vargo 1997). The major regulators implied so far are juvenile hormone, vitellogenin and the insulin-like signaling pathway. They have been experimentally connected to growth, longevity and reproduction in a variety of organisms, including the honeybee (Amdam et al., 2004; Corona et al, 2007). In honeybees, they have also been implicated in the division of labor (Ament et al., 2008).

Juvenile hormone (JH) is thought to have a managerial role in the trait associations seen between social behavior and reproductive physiology. In many insect

6

species, nutrition and JH levels are correlated in adult females ,with better nutrition resulting in higher JH levels and, as a result, egg production (Wheeler, 1996). A suppresion of JH production leads to reduced gonotropic function and conversely, when JH is not suppressed, it has a positive effect on oocyte maturation and reproductive behavior. JH also plays a key role in female caste determination in the honeybee. Queen larvae begin to show substantially higher levels of juvenile hormone during development relative to those larvae destined to be workers (Rachinsky et al., 1990). It has also been implicated in the inhibition of apoptosis during larval development in the ovaries (Schmidt-Capella and Hartfelder, 1998). During adulthood, the level of JH is low in nurses, and then increases once individuals become foragers (Robinson 2002). Thus, JH is clearly an important factor in the mediation of life history traits, reproduction, social behavior, and caste differentiation.

Vitellogenin is a yolk precursor protein which correlates with pollen foraging, large ovary size, as well as behavioral development within the caste system (Amdam et al, 2007). JH interacts with vitellogenin by actively suppressing vitellogenin in a mutual, negative feedback loop, causing workers to be fixed as foragers. The upregulation of vitellogenin, therefore, causes the fixation of nurse bees within their behavioral sub-caste, irrespective of age. Experimentally, it has been shown that high levels of vitellogenin are present in nurse bees, and inversely, low levels present in foragers. The variable expression in protein levels within the worker caste can be explained by their task-association (Amdam et al., 2003, 2007; Robinson et al. 1992; Robinson and Vargo 1997).

Vitellogenin is used as a source of protein fed to the larvae by the nurse bees, but is absent in foragers which do not tend to the brood.

In addition to affecting the expression of vitellogenin, JH also interacts with several genes in the insulin-like signaling pathway that regulate the female caste differentiation, worker physiology and developmental history (Amdam al, 2007; Sullivan et al, 2003; Wheeler et al., 2006). It is an evolutionarily conserved pathway that regulates growth and size in many organisms, through a series of downstream signals, which include JH and vitellogenin in insects (Nijhout 2003; Wu et al., 2006). Differential feeding is known to be crucial in caste determination and behavioral ontogeny. Therefore, the insulin-like signaling pathway that is involved in nutrient sensing has a suggestive role in *A. mellifera* behavioral ontogeny and caste determination (Page and Amdam, 2007).

Honeybee workers exhibit a pronounced, sophisticated temporal division of labor described above (Free, 1965; Lindauer, 1953; Rösch, 1925). In addition, workers also show genetic effects on behavioral specialization (Amdam et al. 2004; Hunt et al., 2007; Rüppell et al., 2004). The best documented case involves foraging preferences, which have been studied in detail, using strains that were selected for low and high pollen hoarding behavior (Page and Fondrk, 1995; Page and Erber 2002). Honey bee workers will forage for carbohydrates, in the form of nectar, and protein, in the form of pollen. Many individuals specialize in one over the other, and this preference is exhibited throughout life. Artificial selection has resulted in changed worker foraging preferences,

but also caused a set of correlated changes, known as the pollen hoarding syndrome (Rueppell et al. 2006a,b; Page and Amdam 2007).

Individuals exhibiting higher pollen hoarding behavior generally have larger ovaries, developed faster into egg-laying workers in the absence of a queen, and also exhibit an earlier shift from nest activites to foraging outside of the hive, than individuals exhibiting low pollen hoarding behavior (Amdam et al. 2006). Also, the high pollen hoarding individuals collect more dilute concentrations of nectar (Rueppell et al, 2006a). The age at which high pollen hoarding workers commence foraging is earlier than both unselected bees and the bees selected for low pollen hoarding behavior (Pankiw and Page, 2001). Other behavioral traits have been associated with pollen hoarding behavior, such as sensitivity to sucrose and learning ability. High pollen hoarding bees were more sensitive to sucrose than low pollen hoarding bees (Page et al., 1998). In high and low pollen hoarding bees, individuals with a high gustatory response to sucrose learned better than bees with a low gustatory response (Scheiner et al., 2001).

This behavioral distinction between nectar and pollen foragers and between artificially selected high and low pollen hoarding bee strains is also reflected in racial differences between European and Africanized honeybees that show similar trait associations (Pankiw, 2003). European honeybees typically forage more for nectar, forage later in life, are less sensitive to sucrose, have fewer ovarioles and, overall, have less reproductive potential. On the other hand, the Africanized honeybees show a foraging preference for pollen, forage earlier in life, are more responsive to sucrose, have more ovarioles, and are characterized by a higher reproductive potential (Pankiw, 2003).

Africanized honey bees originate from an African subspecies, *Apis mellifera scutellata* Lepeletier, which is native to the savannah of eastern and southern South Africa. In the 1950's, *A.m. scutellata* was introduced to tropical Brazil through both the inadvertent release of *scutellata* queens and deliberate distribution of African/European racial hybrids to improve Brazilian bee-keeping and honey production (Spivak et al., 1991). Neotropical *A.mellifera* populations were poorly adapted to tropical ecological and climatological conditions in Brazil, and primarily existed in managed apiaries (Sheppard et al., 1991). The density of European derived wild colonies during the 1950's was apparently low (Michener, 1975; Taylor, 1988), which allowed for the success of the African honey bee and resulted in the extensive and rapid range expansion that followed. Within 20 years the descendent "Africanized" honey bees reached their southern limits in Argentina (Kerr et al., 1982), while their northward expansion continues to this day in the southern United States (e.g. Sheppard et al., 1991). The Africanized bee is naturally more defensive than European subspecies such that it is more likely to respond to disturbances by stinging perceived predators, have faster stinging responses, and have greater sensitivity to stimuli (Hunt et al., 1998). This type of response exhibited by the Africanized honey bees has been tied to aspects the pollen hoarding syndrome, such as foraging behavior and sucrose responsiveness (Hunt et al. 1998; Pankiw, 2003).

Thus, the two types of foragers, the two selected strains, and the two races mirror the trait associations of the two phases of a hypothetical, solitary ancestor, non-reproductive and reproductive, as predicted by the RGPH of social evolution in honeybees (see below).

Behavioral Genetic Studies of the Pollen Hoarding Syndrome

The genetic architecture of the pollen hoarding syndrome has been investigated in several mapping studies. These studies have revealed four quantitative trait loci (QTL) regions that influence several aspects of foraging specialization (Hunt et al., 1995; Page et al., 2000; Rüppell et al., 2004). Though not necessarily genes themselves, QTL regions are stretches of DNA that likely contain the genes that underlie variation in a particular, quantitative phenotypic trait.

These QTL regions were named *pln1*, *pln2*, *pln3* and *pln4* and exhibited partial genetic overlap to other aspects of the pollen hoarding syndrome, in particular the age at onset of foraging (Rueppell et al. 2004b), and sucrose sensitivity (Rueppell et al. 2006a). Initially, three QTLs designated as *pln1*, *pln2*, and *pln3* were detected based on the quantity of pollen in colonies from a backcross population derived from artificially selected high and low pollen hoarding strains (Hunt et al, 1994; Page et al, 2000). The effects of these QTL regions were confirmed with the association of marker alleles near the QTLs with individual foraging traits within independent crosses (Page et al., 2000: Rüppell et al., 2004). Therefore, these QTLs are relatively robust. *Pln2* and *pln3* were shown to influence the discrimination for the sugar concentration of nectar collected (Hunt et al., 1995; Page et al., 2000). *Pln1* was shown to directly affect the age at which individuals initiated foraging behavior (Rueppell et al., 2004) and was revealed to interact with *pln2* and *pln3* to affect various foraging decisions of workers (Rüppell et al., 2004). Another study found an additional QTL, designated as *pln4*, directly influencing foraging specialization and gustatory response (Rüppell et al., 2004; Rueppell et al., 2006a) and

mapping to about 50 cM from *pln1*. Additional QTL could not be excluded and the genetic architecture is further complicated by QTL regions displaying epistatic effects with one another (Rüppell et al., 2004).

Additional QTL regions have been identified for the age of first foraging (*aff3*, *aff4*, *aff* new) (Rueppell et al., 2004; Rueppell, 2009). Initial experiments measuring the age of first foraging, using the same pollen hoarding lines, revealed four significant QTL: two in the high pollen hoarding line, and two in the reciprocal low pollen hoarding line (Rueppell et al., 2004). The QTLs, designated *aff1-4*, showed no overlap with the previously identified *pln* QTL regions. However, their location in the genome could not be accurately determined due to use of Amplified Fragment Length Polymorphisms (AFLP^{TM}) markers. A recent study intended on localizing the *aff* QTL, reconfirmed the presence of two of the four, as well as uncovering another QTL: *aff3*, *aff4*, new *aff* with defined genomic locations on chromosomes 4, 5, and 11, respectively (Rueppell, 2009).

Within the *pln* QTL regions on chromosome 1 and 13 are several specific genes of interest, which are involved in insulin-like signaling (Hunt et al. 2007). Insulin-like signaling has been suggest to underlie *Drosophila* QTL that determine ovary size (Orgogonzo et al. 2006). Recently, Wang et al. (2009) identified two positional candidate genes, *PDK1* and *HR46*, in two *pln* regions that are involved in insulin-like and hormonal signaling. These genes had a pleiotropic effect on ovary size and foraging behavior, and showed differential expression between the selected strains and between low and high ovary size groups (Wang et al. 2009).

12

The Reproductive Ground Plan Hypothesis in Honeybees

The evolution of a eusocial species with different castes and elaborate social behavior from a solitary ancestor is theorized by the reproductive ground plan hypothesis (Amdam et al, 2004; West-Eberhard 1987, 1996). This hypothesis suggests that reproductive cycle controls have been co-opted during social evolution to control social behavior and functions: Many solitary insects are characterized by a gonotropic cycling between a non-reproductive phase and a reproductive phase (Kelly et al., 1987). Reproductive activity of an individual is correlated with ovary activation (egg production) and a suite of other traits.

There are no consistent alternative proximate hypotheses to explain how eusocial behavior in insects has evolved. However, the primary characteristic that defines sociality is reproductive division of labor. This was hypothesized to provide the organizational mechanism which aided in the evolution of eusociality by M.J West-Eberland (1987), termed the ovarian ground plan hypothesis. This hypothesis was further elaborated on and adapted for *Apis* by Amdam et al. (2004) and called the Reproductive Ground Plan Hypothesis (RGPH; explained in detail below).

The non-reproductive phase of the gonotropic cycle of a putative solitary bee ancestor is characterized by reproductive inactivity and a tendency to forage for carbohydrates, in the form of nectar, to ensure survival. Depending on favorable environmental circumstances, the organism can then cycle into the reproductive phase, which is characterized by reproductive activation, foraging for pollen, and heightened sensitivity (Page and Amdam 2007). The original reproductive ground plan hypothesis

(RGPH) as proposed by West-Eberhart (1996) contends that eusocial insects evolved the queen and worker castes by the modification of the regulatory gene network initially controlling the reproductive and non-reproductive phases of the solitary ancestor, respectively. The queens have retained the reproductive phase, while the worker castes have retained the non-reproductive phase. With slight modifications, this evolutionary hypothesis could be applicable to many eusocial organisms in the order Hymenoptera (Hunt et al., 2007) but also different contexts within honeybees, such as worker sub-caste differentiation (Amdam and Omholt 2003; Amdam et al., 2003, 2004, 2006, 2007). Most empirical work (to be described below) has focused on the honeybee model, *Apis mellifera* (L.).

Oldroyd and Beekman (2008) break down the RGPH into three separate hypotheses which compare different "classes" of bees: the original RGPH, modified RGPH, and the forager RGPH. Their original RGPH focuses on primary caste differences, suggesting that honeybee workers express behaviors found in non-reproductive solitary insects, while the queens express phenotypes seen in solitary insects while in the reproductive phase. The modified RGPH postulates that within the worker caste, the observed temporal division of labor is explained by the expression of reproductive and foraging phases of solitary insects, where young workers perform nest-bound work (reproductive) and older workers forage (foraging phase). Finally, the forager RGPH proposes that more reproductively tuned individuals express pollen foraging behavior, while non-reproductive foragers specialize on nectar foraging (Oldroyd and Beekman 2008).

Phenotypic associations between social behavior and reproductive traits in honey bee workers have been proposed as empirical evidence to support the modified and forager RGPH (Amdam et al, 2004; Amdam et al, 2006; Amdam et al, 2007). Ovary size varies within workers, and this variation has been shown to correlate with foraging preferences and differing levels of vitellogenin (Amdam et al., 2004, 2006). These phenotypic associations are also predicted by the RGPH and in turn support it because the phenotypic correlations between strains and wildtype bees (Amdam et al. 2006) and within experimental crosses (Wang et al. 2009) indicate pleiotropic genetic control modules as suggested by the RGPH. In fact, vitellogenin dynamics have been suggested as a potential mechanistic explanation of these trait associations of the pollen hoarding syndrome (Nelson et al. 2007), particularly since it down-regulates JH (Guidugli et al. 2005).

Oldroyd and Beekman (2008) put forth some skepticism regarding the hypothesis of foraging preference and reproductive potential being pleiotropically linked based on their empirical findings that there is no difference in foraging behavior or preference in both their strain of anarchistic bees and the wild type workers. These anarchistic bees contains workers that will lay eggs in a queen-right hive. The argument is that the high and low pollen hoarding strains that they perceived as the primary support for the forager RGPH had been also selected for reproductive activity solely by accident. Thus, the two traits may be only correlated by genetic drift and not pleiotropic linkage (Oldroyd and Beekman, 2008). On the other hand, Rueppell et al. (2008) found phenotypic associations between ovary size and foraging specialization even in a different species, the Japanese

honeybee (*Apis cerana*). The topic is contentious and requires further research. However, additional phenotypic studies are not sufficient to resolve the issue but direct, genetic or mechanistic studies are needed. Studies which point out specific genetic links between social behavior and reproductive characteristics lend critical support for the reproductive ground plan hypothesis.

While recently our group and collaborators were able to demonstrate direct genetic links between the *pln* QTL and worker ovary size in the selectedpollen hoarding strains (Wang et al., 2009), it is clear that the genetic linkage between social behavior and reproductive traits has to be investigated more generally. The behavioral and reproductive differences between the Africanized and the European honeybees, described above, provide an ideal model for a more general test of pleiotropy. Therefore, I studied the genetic architecture of ovary size differences between these honeybee populations and specifically investigated the *pln*, and *aff* QTL regions for pleiotropic effects on the documented ovary size differences to test the generality of genetic links between ovary size and social behavior, as predicted by the RGPH of social evolution. Furthermore, the identified QTL regions were scrutinized for plausible candidate genes that may have played a general role in social evolution.

Aims

My goal was to characterize the genetic architecture of the worker ovarian differences in crosses between Africanized and European honeybees. This was to determine any direct genetic overlap between reproductive QTLs and the previously

mapped behavioral QTLs and thus test a central prediction of the reproductive ground plan hypothesis. I predicted, based on the RGPH, that there is genetic overlap between previously identified QTLs, and QTLs identified with my data. Overlap of these QTL regions would lend support for the RGPH (Amdam et al, 2004; Amdam et al, 2006; Amdam et al, 2007): overlap with *aff* (*3*, *4*, *new*) QTLs would support the modified RGPH, and overlap with *pln* (*1-4*) QTLs would support the worker RGPH. However, this overlap was predicted to be only partial because QTLs are dependent on the segregating variation in a specific cross. The findings also contribute in general to the understanding of the genetic determination of ovary size in insects. The characterization of the genetic architecture of ovarian differences between Africanized and European worker honeybees is divided in three specific aims.

Aim 1: *My first aim was to perform quantitative trait loci mapping for ovary size in honey bees on two pre-selected backcrosses between European and Africanized honeybees (R3 and R5).*

Aim 2: *My second aim was to investigate genetic effects of the previously identified behavioral QTLs on ovary size to use as evidence to support or refute the reproductive ground plan hypothesis.*

Aim 3: *My third aim was to identify candidate genes within potentially new QTLs for ovary size.*

CHAPTER II

MATERIALS AND METHODS

Overview

My project was based on a large amount of preliminary work performed: creation of the crosses, individual honey bee dissection, ovary trait evaluation, DNA extraction, and genotyping of selected SNPs. These steps provided the mapping populations and backbone for the subsequent research carried out. My three specific aims for the project performed were achieved in four steps: 1. Establishing informative microsatellite markers to provide adequate coverage of the genome in both mapping populations. 2. Genotyping the mapping populations at the selected microsatellite loci. 3. Performing a full QTL analysis of ovary size in both mapping populations and ovary activation in R5. 4. Bioinformatic search for plausible candidate genes in the identified QTL intervals.

Cross Creation

The two crosses that were tested in this experiment were bred and housed at Arizona State University in Tempe, Arizona. Twelve colonies of Africanized honeybees (AHB) and twelve colonies of European honeybees (EHB) were evaluated for their worker ovary size (Figure 3). The most extreme colony from each group was used to breed hybrid queens: the largest ovaries in the Africanized colonies and the smallest ovaries in the European colonies. From one of those AHB/EHB hybrids, backcrosses

18

were created. Drones from the AHB and EHB parental colonies were bred with hybrid queens, to create the Africanized (AHBBC) and European (EHBBC) backcrosses, respectively. The ovary size (measured as number of ovarioles) were determined by dissecting 20 worker bees from each backcross colony. All EHBBC colonies showed a lower ovary size than their EHB and hybrid parent colonies (Figure 4). The EHBBC colonies also displayed limited variability and thus did not represent suitable populations for QTL mapping of ovary size. The AHBBC colonies, in contrast, showed over-dominance and had higher ovariole numbers than the AHB and the hybrid parents. Two AHBBC showed strongly transgressive phenotypes (Figure 4). The two most extreme colonies were "R3" and "R5". Based on these unusual preliminary observations, R3 and R5 were selected to generate mapping populations.

The mapping populations were created by inducing the queens to lay eggs, which were transferred to populous, queen-right hives of European decent. The combs with eggs were placed above a queen excluder to prevent the host colony queen from laying eggs in them. Then, the combs were bracketed by a frame of honey and a frame filled with pollen. Prior to emergence, the brood was transferred to a temperature (34°C) and humidity (70% R.H.) controlled incubator. Upon emergence, each individual was placed into a tube and stored on dry ice to preserve the specimen. They were then shipped from Arizona State University to the University of North Carolina at Greensboro, where they were stored in the -80 °C freezer for further investigation. A subset of the R5 workers was not immediately frozen but introduced into a queenless hive of EHB descent to induce

19

reproductive activation in these workers. Remaining workers were collected after three weeks for freezing and shipment to UNCG.

Dissection and DNA Extraction

Under a dissection scope, each frozen honey bee was cut to remove the head and thorax from the abdomen. Head and thorax were kept frozen for later DNA extraction. After thawing, the abdomen was dissected by peeling off the dorsal tergites or the ventral sternites and the exoskeleton folded and pinned away from the body. The gastro-intestinal organs (gut) were separated out and also pinned, in order to expose the ovaries (Figure 5). The ovaries were removed from the individual and then placed on a slide and viewed underneath a compound microscope in order to count the number of ovarioles present in the left and right ovary.

The genomic DNA was extracted from the individuals using a modified CTAB-phenol chloroform extraction (Hunt and Page 1995). The steps for the phenol chloroform DNA extractions were followed based on the standard laboratory protocol: three sets of 1.8 mL capacity microfuge tubes were labeled. Salt lysis buffer was mixed using 1% CTAB, 50 mM Tris (pH 8), 10 mM EDTA, 0.75 M NaCl, and 100 μg/ml of Proteinase K. The frozen samples were placed into tubes containing the cold lysis buffer (200 μl was used for workers) and kept on ice. The samples were then ground into the lysis buffer on ice using pestles until the solution became viscous, then they were placed into a 55-60 °C water bath for 1-5 hours. Just before the end of incubation (30 minutes at least), 50 μl of

High Salt lysis buffer (1% CTAB, 50mM, 10 mM EDTA, and 1.5 NaCl) was added to each tube.

The rest of the extraction was performed underneath the hood. The samples were extracted with equal volume of 1:1 phenol:chloroform (phenol was equilibrated to pH>7.5, using a Tris buffer with pH 7.9). The samples were inverted gently 20 times and then centrifuged at 12,000G for 10 minutes. The aqueous layer was removed and placed into a new set of tubes, and the phenol/chloroform was discarded into a waste container under the hood. Equal volumes of chloroform:isoamyl alcohol were added to the samples, inverted 20 times again, and then the mixture was centrifuged for two minutes. The upper phase was transferred to new tubes and placed on ice.

The DNA was precipitated by adding 1/10 volume of 3M sodium acetate (pH 5.2) and two volumes of cold 100% ethanol. The samples were then incubated for 15 minutes at-80 °C or two hours to overnight at -20 ºC. The samples were centrifuged at 12,000G for 20 minutes at 4 ºC. Subsequently, the ethanol was removed with a pipettor and the pellet was washed with cold 80% ethanol. The samples were then centrifuged again briefly, the last traces of the ethanol were removed and the tubes were inverted to dry the pellet. The DNA pellets were resuspended in 50-100 µL of TE per sample (10 mM Tris buffer (pH7.6) with 1 mM EDTA), heating the samples in a 65 °C water bath for 10 minutes to help redissolve the pellet. DNA concentration was quantified with a Nanodrop® spectrophotometer and all samples diluted to 100 ng/ml. These stocks were diluted to template solutions of 10 ng/ µl in low TE (0.3 mM EDTA).

Traits

Each of the Africanized backcrosses had specific ovary traits quantified. In R3, total ovary size, maximum ovary size, minimum ovary size and ovary asymmetry were scored. The R5 backcross had an additional ovary trait quantified, ovary activation, due to its reintroduction into a queenless hive and subsequent ovary activation. Every female honeybee has two ovaries, each of which is comprised of a number of ovarioles. The total ovary size of an individual bee was quantified by the total number of ovarioles present in both ovaries. The maximum ovary size was defined as the size of the larger ovary, with the most ovarioles, while the minimum ovary size was defined as the size of the smaller ovary, with the fewest ovarioles. Ovary asymmetry was calculated by dividing the difference between the maximum and minimum ovary size by the total ovary size. Therefore, an asymmetry value of 0 would signify that both ovaries have an equal number of ovarioles, and a value of 1 would indicate that one ovary contains all ovarioles, while the other contains none.

Ovary activation was measured quantitatively through a dissecting microscope using a modified version 4-point scale: 0 = undeveloped (resting ovarioles); 1 = oogenesis starting (presence of cells swelling at top of ovariole and starting to descend); 2 = slight development (eggs distinguishable from trophocytes); 3 = moderate development (egg volume exceeds that of the follicle); 4 = highly developed (eggs are fully elongated) (Pernal and Currie, 2000).

The average total ovary size for R3 and R5 were 25.16 ovarioles, and 26.14 ovarioles, respectively. The average maximum ovary size was 13.64 ovarioles in R3, and

22

14.49 ovarioles in R5. The average minimum ovary size was 11.52 ovarioles in R3, and

11.65 ovarioles in R5. The average ovary asymmetry for R3 was 0.1, and 0.12 in R5. The

average ovary activation level in R5 was 3.4.

Every trait was tested for correlation against the others using Spearman rank tests:

total ovary size, maximum ovary size, maximum ovary size, ovary asymmetry and ovary

activation (R5 only). In contrast to the mapping population R3, which had to be used

completely because few individuals could be obtained (Figure 6), R5 was selectively

genotyped, based on total ovary size (Figure 7). Selective genotyping is used in genetic

correlation studies, such as QTL mapping, because it can decrease the number of

individuals genotyped compared to the number of individuals phenotyped. It has been

shown that in continuous traits, the power of the analyses can be markedly increased

when it is based on the quantitative values of individuals in the high and low tails of the

population (Darvasi and Soller, 1992; Lander and Botstein, 1989).

Linkage Map Construction

Single nucleotide polymorphisms (SNPs) were genotyped with Sequenom's[TM]

MALDI-TOF mass spectrometry technique (Ragoussis et al., 2006). These SNPs were

genotyped in both mapping populations. These markers did not sufficiently cover the

entire honeybee genome and numerous coverage gaps existed. The first step of my thesis

was therefore to search for suitable microsatellite markers in coverage gaps and generate

genotypic data from them. This allowed an adequate QTL analysis and the evaluation of

the QTL regions for overlap with the behavioral QTL as well as the search for potential candidate genes in the QTL regions.

I determined a-priori the linkage distance between all adjacent SNP markers using a high density microsatellite genome map (Solignac et al., 2007). Any gap larger than 30 centi-Morgans (cM) between two consecutive markers was then filled using microsatellite markers to complete the coverage of the genomes in both crosses. The screening process was initiated by composing a list of primers for microsatellite loci that corresponded to the gap regions on the genome for each cross. The loci were adopted from the complete list of primers published by Solignac et al. (2007). The selected microsatellites were then amplified in eight individuals via polymerase chain reaction (PCR) to evaluate whether they amplified and were variable to indicate grandparental AHB or EHB inheritance (Figure 8). New microsatellite loci were evaluated until coverage of all significant gaps was achieved.

To amplify 8 individuals during the screening process, a PCR master mix was set up for 10 x 10 μL reactions following the laboratory protocol: 56.2 μL dH$_2$O, 10.0 μL 10x PCR buffer, 10.0 μL 2 mM dNTPs, 2.5 μL forward primer, 5.0 μL reverse primer, 5.0 μL IRD 700 or 800 – labeled M13 universal primer (explained below), and 1.3 μL 1u/μL Taq polymerase. The master mix was mixed in a 2 mL tube and kept on ice, while 1 μL of sample DNA was loaded from 8 individuals into an 8 well-strip, also on ice. In each well, 9 μL of the master mix was added to the DNA template, and then the strip was sealed and placed securely into a thermocycler set on the following program (called 50ENHANsho). The initial denaturation step was 95 °C at 4 minutes. The cycling

program started with the denaturation of the samples at 95 °C for 45 seconds, the annealing step at 68 °C for 2 minutes, and the elongation step at 70°C for 1 minute. These three steps are repeated 9 times, continually decreasing the annealing temperature by 2 °C until it reaches 50 °C. After this touchdown, the program goes through 21 more cycles with an annealing temperature of 48 °C. The last step maintains 70 °C for 5 minutes, and then stores the product for an indefinite period of time at 6 °C.

The PCR products were labeled with a Universal IRD labeled M13(-21) primer during the PCR to allow product recognition by the laser detection system of the LiCor® Automated Sequencer (Model 4300 Data Analyzer). The forward primer used in PCR has a M13(-21) tail at the 5'-end that is initially incorporated into the PCR product, and these products then become targets of the Universal IRD labeled M13(-21) primer. The Universal IRD labeled M13(-21) primer attaches to the 5'-end of the initial M13(-21) primer and takes over, incorporating itself during subsequent cycles at lower annealing temperatures (Schuelke 2000).

The microsatellite PCR products were multiplexed into sets of three (1 μL PCR product, 7 μL 1x PCR buffer for dilution, 2.5 μL dye), where in a row of 8 the first PCR was pipetted into wells 1-6, the second primer was present in wells 1-5, and 7, and the third primer was present in wells 1-5 and 8. This allowed us to genotype 6 individuals to asses whether a locus is polymorphic, with the non-multiplexed, final three wells allowing us to determine the sizes of each locus and its potential non-specific amplification products.

The screening process used the LiCor® Automated Sequencer to determine if the microsatellite primers amplified and the products were polymorphic. The polyacrylamide gel was poured and let set following the following protocol: the 25 cm Borofloat® plates were cleaned using water and 70% EtOH, then 100 µL Bind Silane and 100 µL acetic acid were mixed in a 1.0 mL tube and applied with a Kimwipe® to the inside of both Borofloat® plates where the comb was subsequently located. Two 0.25 mm spacers were then placed in between the Borofloat® plates and tighteners were secured on the right and left sides of the plates. The plate apparatus was placed in the casting stand allowing the apparatus to rest on the utmost metal posts on the casting stand.

The polyacrylamide gel was created by adding 8.4 g urea, 4 mL 5x TBE, and 2.4 mL Long Ranger™ (Lonza) into a glass beaker, and then brought to a volume of 20 mL with dH_2O. The gel mixture was mixed and poured after adding 250 µL 10% Ammonium persulfate (APS) and 15.0 µL TEMED to initiate polymerization. The 48 - well comb was then inserted and the gel allowed to polymerize for 1.5 hours.

After the gel had fully polymerized, the plates were cleaned and mounted into the LiCor® Automated Sequencer, adding 1x TBE into the buffer chambers. The wells were then flushed and the pre-run started using SAGA® software (15 minutes at 1500 V, 40 milliampere). The multiplexed PCR samples with loading dye were denatured at 95°C for 3 minutes, and then loaded into the gel, along with an IRD 700 ladder (size standard). The samples are run for 2.5 hours at 1500 V, 40 milliampere.

Markers that proved polymorphic, and appropriately covered gaps in the linkage map, were amplified, using the 100 x 10 µL master mix protocol and the PCR conditions

described above, in both mapping populations: 88 individuals in cross "R3" and 96 individuals in cross "R5". In R3, the total number of microsatellites tested was 201; of those, 79 were monomorphic/unamplified, 122 were polymorphic. In all, there were 102 polymorphic microsatellite markers and 137 SNP markers used in the analysis (Table 1). In R5, of the total number of microsatellites tested (110), 17 were monomorphic, and 93 were polymorphic. Only 68 polymorphic microsatellite and 153 SNP markers were used (Table 2). To genotype the mapping populations, the pre-screened, polymorphic microsatellite PCR products were run on the LiCor® Automated Sequencer highly multiplexed to save both time and materials. The target for multiplexing was to combine 8 PCR products with a total of 10 µL in each well (1 µL sample at 700 or 3 µL 800, 2.5 µL dye). The determined allele sizes of the markers were taken into account to ensure that none of the 8 primers overlap and interfere with each other, once multiplexed and run. The two alternative color labels were also used to avoid overlap.

Both genotyping samples were then selected to fit on one 96-well plate to be amplified simultaneously. The LiCor® Automated Sequencer was used with a polyacrylamide gel poured with a 48 well comb. Two runs were needed to genotype the entire sample population of one cross. With the "R3" population, there was room for 8 more individuals, which was filled by 8 individuals from the "R5" population. This was used to genotype 8 "R5" individuals to obtain information regarding the polymorphic status of that marker in the "R5" population and, thus, eliminate a part of the screening process for that mapping population.

Genome Linkage Map Construction: MapMaker

Once data collection was complete, the SNP and microsatellite genotypes were combined into one data set. These data were assembled into a linkage map which served as the basis for a quantitative trait loci (QTL) analysis. The initial marker order was adopted from the physical location of the markers in the honeybee genome sequence and by comparison to a previous map using 2000 microsatellite loci (Whitfield et al 2006; Solignac et al. 2007). Using Mapmaker 3.0 (Hunt and Page 1995; Lander and Botstein 1989; Lincoln et al., 1993), this marker order was then verified and linkage distances between markers estimated. The centimorgan function used was Kosambi, and the initial default linkage criteria were a LOD > 2, and a linkage distance < 50 cM. The linkage distances were compared to previous data (Solignac et al. 2007) and markers that led to a significant (>10%) extension of the linkage distance were re-evaluated by either rescoring the original gels or re-genotyping the markers in order to prevent genotyping errors. If the marker order could not be verified after rescoring or rerunning, problematic markers were removed, and not included in the final analysis.

During the map making process, the Africanized backcross R5 had poor linkage between some of the SNP and microsatellite markers that lead to inflated linkage distances between markers, and in over-estimates of genetic chromosome size. To avoid spurious QTL mapping results, any gaps ≥ 50 cM were artifically collapsed to 25 cM. Mapping analyses were performed for both the un-shortened, and shortened versions and QTLs that were only detected in the large gaps of the original map were disregarded because large gaps can lead to huge, spurious QTLs during interval mapping and also

bias permutation tests which calculate genome-wide and chromsome-wide LOD

thresholds (Rueppell, 2009).

QTL Mapping: MapQTL

The ordered and verified data sets were used for comprehensive QTL analyses to

determine correlations between marker genotypes and the investigated phenotypes and

thus localize any genetic effects on ovary size and activity. These analyses were

performed using the software package MapQTL 4.0 (Van Ooijen et al., 2002) that enable

me to identify QTLs based on a series of statistical methods of increasing complexity:

single-marker analysis, interval mapping, and multiple interval mapping (Lander and

Botstein 1989; Van Ooijen et al., 2002). The particular type of single-marker analysis that

MapQTL utilizes is the nonparametric Kruskal-Wallis test. Markers that were identified

as significant at the p≤0.01 level were considered to be statistically suggestive. Interval

mapping is similar to a single marker analysis, but the target locus is replaced by a

position in a target interval and genotype information from both flanking markers is

combined. Interval mapping (IM) scans the whole genome systematically for a putative

QTL in those target intervals. The genome was divided up into regular intervals and

maximum-likelihood test ratios of marker-trait association were calculated in each

interval. Based on the likelihood ratios, LOD scores were then calculated for the each

interval and a LOD score profile is calculated across the whole genome. In both interval

mapping and MQM, LOD ≥ 2.0 were considered suggestive QTLs and QTLs with a LOD

≥ 3.0 or a LOD $\geq$ LOD$_{GW}$ were considered to be significant (Hunt et al. 1995; Rueppell et

al. 2004). The LOD scores of these regions of interest were compared to the chromosome-wide (LOD_{ChW}) and genome-wide LOD scores (LOD_{GW}), generated by MapQTL's permutation function. However, chromosome-wide LOD significant threshold values were only given for comparison to other studies (Oxley et al., 2008).

The statistical analysis used in QTL mapping has an inherent problem since it involves multiple, independent comparisons. Thus, there needs to be an appropriate calculation of a significance threshold to detect QTLs. To determine the significance of the LOD scores, a threshold value was calculated analytically based on genome size (Hunt et al., 1995) and also determined empirically through multiple permutations of the empirical data to test the likelihood of a particular LOD score just by chance alone. The permutation approach is suitable for determining appropriate significance thresholds against which to compare the LOD scores, irrespective of marker and phenotypic distributions (Bennewitz et al., 2002; Churchill and Doerge, 1994). Therefore, when a LOD score peak exceeds the determined threshold values, a significant QTL is considered to have been found at that location.

The most complex analytical method, multiple QTL mapping (MQM) searches for multiple loci that could have simultaneous effects on phenotypic variance. Through statistical control of one locus, detection of other loci is facilitated by choosing the first locus as a cofactor: the marker with the highest calculated LOD score in an identified QTL region with the largest putative effect. This is due to elimination of genetic overshadowing and reduction of residual variance (Jansen 1993, Jansen 1994, Jansen and Stam 1994, Kao and Zeng 1997, Kao et al., 1999, Zeng et al., 1999). After the statistical

control of the cofactor effects, the QTL mapping was repeated. All chromosomes were re-investigated to identify new putative (LOD of ≥ 2.0) or significant QTLs and to assess changes in the QTLs previously identified during interval mapping. In my specific case, I selected markers at all significant and suggestive QTLs as cofactors.

For each cross, specific tests were performed: single-marker analysis and IM were run for R5, while single-marker analysis, IM, and MQM were run for R3. MQM was only run for the R3 cross because the entire population of R3 was genotyped, while the R5 cross has been subject to selective genotyping, which is in violation of the statistical assumptions of MQM (Van Ooijen et al., 2002).

In addition, Mann-Whitney U tests were performed on the markers within the *pln* and *aff* QTL regions (Hunt et al., 1995; Page et al., 2000; Rueppell et al. 2004; Rueppell 2009). Nonparametric analyses for significance at the single marker level were performed due to a non-normal distribution of the data, which could not be normalized through any transformation. As an additional measure of direct genetic links between the ovary and social behavior, I calculated the exact probability to find as much overlap between the behavioral QTLs and my significant and suggestive QTLs for each trait.

Candidate Genes

QTL regions were determined to search for functional candidate genes. The exact marker positions around each QTL were mapped onto the respective chromosome. The genomic region within 1.5 LOD units of the LOD-score peak were used to define the 95% confidence interval for each location of a QTL (Hunt et al., 2007). I used the

31

Amel4.0 genome version in the Honey Bee Genome Database by the Texas Agricultural

and Mechanical University genome browser (BEEBASE: http://genomes.tamu.edu/) and

NCBI's genome browser (http://www.ncbi.nlm.nih.gov). All genes and putative gene

entries located in the QTL confidence interval were evaluated as candidate genes on a

functional basis by searching for known functions of homologous genes in the NCBI

database, FLYBASE (http://flybase.bio.indiana.edu/), and the gene ontology database

(http://www.geneontology.org/).

CHAPTER III

RESULTS

Mapping Populations

After the loss of one individual, the R3 mapping population consisted of only 87 individuals, while R5 has a sample size of 190 individuals (Figure 6, 7). In R3, the total number of microsatellites tested was 201; of those, 79 were monomorphic/unamplified, 122 were polymorphic. Only 102 of the 122 polymorphic loci were used since some of those 20 markers were in gaps already covered by other polymorphic loci identified, or were problematic during the mapping process. Of 280 tested SNP markers, 137 SNP markers were used in the analysis (Table 1). In R5, of the total number of microsatellites tested (110), 17 were monomorphic, and 93 were polymorphic. Only 68 polymorphic markers were used in the analyses mostly due to problematic linkage with the 158 surrounding SNP markers in the mapping process (Table 2).

Trait Correlations

The Spearman rank test showed that most measured traits were highly correlated with each other. Total ovary size was positively correlated with minimum and maximum ovary size, and ovary asymmetry, while being negatively correlated to ovary activation (Table 3). Minimum ovary size was positively correlated with total and maximum ovary size, while being negatively correlated with ovary asymmetry and ovary activation.

Maximum ovary size was positively correlated with total and minimum ovary size, while being negatively correlated with ovary activation, and uncorrelated with ovary asymmetry. Ovary asymmetry was negatively correlated with total ovary, minimum and maximum ovary size, while it was not significantly correlated with ovary activation. Ovary activation was negatively correlated with total, minimum, and maximum ovary size, while not being correlated with ovary asymmetry. To the extent measured, these results were consistent in both backcrosses (Table 3).

Genetic Analysis of Total Ovary Size

For R3, single marker analysis showed markers on chromosomes 1, 2, 10 and 11 to be suggestive. Interval mapping revealed possible QTLs present on chromosomes 1, and 11. In the MQM analysis, selecting for est8460 as a single cofactor from chromosome 11.31 kept the bimodal peak (Figure 9). However, selection for est8456 as a cofactor sharpened the QTL to a singular peak. There was another suggestive QTL on chromosomes 1 (scaffold 1.06) with a LOD score of 2.1 in addition to the one significant QTL on chromosome 11 (11.30-11.31) with a LOD score of 6.45 (Table 4). The QTLs identified explained a total of 44.9% of the variation seen: chromosome 1 (1.06) with 9.9% and 11 (11.30-11.31) with 35%. The effect size of the QTL on chromosome 11 (11.30-11.31) showed that depending on a particular allele for marker est8456, there was a difference of 12 ovarioles (Figure 10).

For R5, single marker analyses indicated suggestive markers on chromosomes 1, 6, 9, 11, 13, and 15. Interval mapping confirmed possible QTLs on chromosomes 6, 11

and 13, as well as a QTL previously unidentified through single marker analysis on Chromosome 3. There were two suggestive QTLs on chromosomes 11 (11.31), and 13 (13.1), with LOD scores of 1.97, 1.89, and 2.68, respectively, and one significant on chromosome 6 (6.45-6.49) with a LOD score of 7.96 (Table 5; Figure 11). The QTLs identified explained a total of 44.6% of the variation seen: chromosome 6 (6.45-6.49) with 24%, 11 (11.30-11.31) with 13% and 13 (13.1) with 7.6%. The effect size of the major QTL on chromosome 6 (6.45-6.49) showed that depending on a particular allele for marker est4967, there was a difference of 9 ovarioles (Figure 12).

Genetic Analysis of Maximum Ovary Size

For R3, single marker analysis identified markers on chromosomes 1, 3, 10, and 11. Interval mapping showed one QTL may be present on chromosome 1, and possibly two on 11 (Table 6). The MQM mapping analyses (cofactor est8456) established one suggestive QTL on chromosome 1 with a LOD score of 2.83 (1.06), and one significant QTL on chromosome 11 (11.30-11.31) with a LOD score of 5.87 (Table 6). The QTLs identified explained a total of 43.1% of the variation seen: chromosome 1 (1.06) with 11.3% and 11 (11.30-11.31) with 31.8%.

Single marker analysis in R5 revealed suggestive markers on chromosomes 1, 6, 9, 13. Interval mapping identified one suggestive QTL on chromosome 13 (13.1) with a LOD score of 2.78 and a significant QTL on chromosome 6 with a LOD score of 7.16 (6.45-6.49) (Table 7). The QTLs identified explained a total of 46.5% of the variation

seen: chromosome 6 (6.45-6.49) with 38.1% and 13 (13.1) with 8.4%.

Genetic Analysis of Minimum Ovary Size

For R3, single marker analysis revealed markers on chromosomes 1, 10, and 11. Only Groups 1 and 11 showed possible QTLs through interval mapping with LOD scores of 2.39 and 6.71, respectively. MQM (cofactor est8456) confirmed that there were two QTLs: a suggestive on chromosome 1 (LOD = 2.39; scaffold 1.06) and a significant on chromosome 11 (LOD = 6.71; 11.30-11.31) (Table 8). The QTLs identified explained a total of 37% of the variation seen: chromosome 11 (11.30-11.31) with 29.4% and 1 (1.06) with 7.6%.

Single marker analysis in the R5 backcross showed markers showed markers on 1, 3, 6, 8, 11, 13 and 15. The markers on chromosomes 1, 3, 6, 11, and 13 were corroborated with LOD score peaks, based on chromosome-wide LOD scores. The QTL on chromosome 6 (6.45-6.49) was considered a significant QTL with a LOD score of 7.91, while the QTLs detected on chromosomes 1 (1.06), 3 (3.34), 11 (11.30-11.31), and 13 (13.1) are considered suggestive (Table 9), with LOD scores of 2.02, 2.04, 2.01, and 2.27, respectively. The QTLs identified explained a total of 64% of the variation seen: chromosome 1 (1.66) with 5.0%, 3 (3.34) with 4.8%, 6 (6.45-6.49) with 31.3%, 11 (11.30-11.31) with 16.8% and 13 (13.1) with 6.1%.

Genetic Analysis of Ovary Asymmetry Size

For R3, single marker analysis suggested markers solely on chromosome 4. Interval mapping revealed that there was a suggestive QTL on chromosome 4 (4.09-4.12). MQM analysis where est3866 (or RK0423B) was selected as a cofactor confirmed the QTL, with a LOD of 2.91 (Table 10) but suggested no other QTL. The QTL on chromosome 4 (4.09-4.12) explains 28.6% of the variation seen.

For R5, single marker analyses revealed markers on chromosomes 5, and 8. Interval mapping identified one suggestive QTLs on chromosome 8 (8.23) with a LOD score of 2.15, and one significant QTL on chromosome 5 (5.32-5.33) with a LOD score of 3.12 (Table 11; Figure 13). The QTLs identified explained a total of 14.4% of the variation seen: chromosome 5 (5.33) with 7.0% and 8 (8.23) with 7.4%.

Genetic Analysis of Ovary Activation

Only R5 was evaluated for ovary activation. The single marker analysis identified suggestive markers on chromosomes 2 and 6. However, interval mapping confirmed only the region on chromosome 6 (6.45-6.49) as a significant QTL with a LOD score of 6.86 (Table 12). The QTL on chromosome 6 (6.45-6.49) explains 30.2% of the variation seen.

Behavioral QTL and single marker analyses

The QTLs identified for pollen hoarding behavior, *pln1* (13.09-13.11), *pln2* (1.54-1.59)*, pln3* (1.24-1.29) and *pln4* (13.17), and for the age of first foraging, *aff3* (4.18-4.22)*, aff4* (5.18-5.20)*, aff new* (11.33-11.37), were assessed for their effect on ovarian traits by

investigating one genetic marker in each interval separately (with correction for multiple comparisons). Multiple effects were identified: In R3 *pln3* affected ovary asymmetry and total ovary size, and *aff4* affected maximum ovary size (Table 13). In R5, significant markers for total, minimum and maximum ovary size were found in *pln1*, for total and minimum ovary size in *pln2*, for total and maximum ovary size in *pln3*, and for total, minimum and maximum ovary size in new *aff* (Table 14; Figure 14)

Investigating directly QTL region overlap, I found one incidence of overlap for all three ovary size measurements assessed in R3 and for maximum ovary size in R5. For total ovary size and minimum ovary size, I found two incidences of QTL region overlap.

For R3, the probability of 1 or more behavioral QTLs overlapping with my QTLs is 0.0732. The probability of 1 or more behavioral QTLs overlapping for maximum ovary size in R3 is 0.0840. The probability of 2 or more behavioral QTLs overlapping for total ovary size in R5 is 0.0102. The probability of 1 or more behavioral QTLs overlapping for maximum ovary size in R5 is 0.0591. The probability of 2 or more behavioral QTLs overlapping for minimum ovary size in R5 is 0.0318. Using the conventional 5% threshold parameter ($p \leq 0.05$), only two of the five analyses are significant. However, the other three tests are also close to a 5% threshold.

Functional Assessment of Positional Candidate Genes

Candidate genes that are known to be involved in ovary development, germline development, pole-plasm development, apoptotic function and involved in the insulin-like signaling pathway were a-priori considered to be of particular interest if located in

the identified QTL regions. Due to their relative importance, only the two main QTL regions were investigated.

The QTL interval on chromosome 11 (scaffold 11.30-11.31) contained 95 genes (NCBI), several of which can be regarded as good functional candidates. I identified a promising gene: similar to fl(2) d protein (LOC552833, XM_625208). In addition, the bee homologs of *quail* (LOC410324), *cabut* (LOC410326), *delta* (LOC410351) represented good functional candidates. The QTL identified on chromosome 6, scaffolds 6.45-6.49, identified in my results, contained 55 genes (NCBI), including one good functional candidate gene, the s*even-up* homolog (LOC408872).

CHAPTER IV

DISCUSSION

Through quantitative trait loci analyses, I found two significant and several suggestive genetic influences that contribute to transgressive ovary sizes in worker honeybees. As predicted by the RGPH of social evolution, genetic linkage between reproductive traits and social behavior was identified because behavioral QTL exerted some effects on ovary size. Thus, my results have provided significant support for the RGPH, independent of the selected high and low pollen hoarding strains. However, the genetic overlap was only partial and the two main QTL for worker ovary size did not directly overlap with behavioral QTL. These two QTL presumably interact epistatically but simple genetic two-loci models fail to explain the overall crossing results. Several interesting functional candidate genes could be identified in the QTL regions that suggest hormonal signaling and processes that are also involved in ovary development in solitary species.

The Reproductive Ground Plan Hypothesis

It has been proposed that worker phenotypes were enforced initially by larval dietary restriction and subsequently genetically assimilated by allelic substitutions at genes involved in translating the nutritional signals into ovary developmental programs (Page and Amdam, 2007). Nutrition, through differential feeding, enabled the emergence

of branched developmental pathways. Consequently, it helped form the new social structure of the honey bee colony and develop foraging specialization directly from the ancestral associations of behavior within the reproductive groundwork (Page and Amdam, 2007).

The onset age of foraging, as well as the preference for pollen versus nectar, has been shown to be the culmination of effects of multiple genes across the genome. These behaviors have been theorized by the RGPH to be governed by the same genes involved in the development and maintenance of the ovary. Therefore, genes that affect ovary size should also affect these key behavioral traits. The pollen-hoarding syndrome of the honey bee consists of a common set of correlated behavioral phenotypes that include sucrose responsiveness, foraging preference and the age of first foraging (Humphries et al., 2003; Pankiw and Page, 2003). The syndrome is known to be influenced by seven significant QTLs with effects on behavior (Rüppell et al., 2004; Rueppell et al., 2006a, Rueppell 2009). Phenotypic associations between the pollen hoarding syndrome and reproductive traits prompted the modern formulation of the RGPH (Amdam et al. 2004, 2006), which has been criticized since (Oldroyd and Beekman 2008) for its over-dependency on phenotypic associations in the artificially selected high and low pollen hoarding lines. Direct genetic linkage has been demonstrated (Wang et al., 2009) but my data are the first to generalize the genetic linkage between ovary size and social behavior to wildtype populations. My results show a clear overlap between genetic effects for ovary size and *pln1*, *pln2*, *pln3*, *aff4*, and the new *aff* QTL. The overlap between the behavioral QTLs

41

and my QTLs, involved in ovary size, significantly substantiates the evidence for the RGPH.

There is a clear overlap in the QTLs identified in my study and with previously identified behavioral QTLs. It could be argued that genes involved in ovary size and genes involved in behavior (pollen hoarding, age of first foraging) are not pleiotropically linked, but, instead, are tightly linked due to genomic co-localization. Thus, the genes would segregate together and the pleiotropy would seem an artifact of linked alleles. However, *A. mellifera* is known to have the highest level of recombination across the genome of any multicellular eukaryotes (Beye et al., 2006). Linkage disequilibrium breaks down quickly under such high levels of recombination, and therefore the possibility of gene linkage is unlikely.

Oldroyd and Beekman (2008) argued that the lines selected for pollen hoarding were selected for reproductive activity solely by chance, and that the two traits are only weakly correlated. They argued that there was no direct selection of worker production, instead since the two different lines (high and low) were created from two separate colonies, the correlation is merely the result of genetic drift. The observed correlation between the tendency to forage for pollen and early onset of foraging could then simply be an artifact of selection of gene networks, which are unrelated to reproductive potential. Oldroyd and Beekman's (2008) used phenotypic data found that the selection for reproductive behavior in anarchistic bees did had an increase in the onset of foraging behavior, however it did not alter foraging preference. They claim that their lines (anarchistic bees) are more reproductively tuned because workers will lay eggs even in

42

the presence of a queen (Oldroyd and Beekman, 2008). This behavior may be due to for a mutation in odorant receptors making workers incapable of detecting queen pheromone (Oxley et al., 2008). This mechanism seems to be a completely different explanation for worker reproductive potential, unique to the anarchistic lines. In contrast, the direct genetic link between the ovary and pollen hoarding shown by Wang et al. (2009), and the overlap of my QTLs for ovary size, with the pollen hoarding QTLs (*pln1*, *pln2*, *pln3*) and age of first foraging (*aff4*, *aff new*) suggests that Oldroyd and Beekman's (2008) claim is incorrect. These results from my study show that there is genetic linkage that is unlikely to have arisen by chance alone.

My findings in unselected bees support all three aspects of the RGPH: I have identified significant QTLs that affect ovary size with functional candidate genes (see below) that could be involved in caste determination and in orchestrating the insect gonotropic cycle, which supports the original RGPH. The modified RGPH states that the worker sub – caste differentiation of nurses and foragers is akin to the gonotropic phases in solitary ancestors: this is supported because my results show overlap of reproductive traits with the *aff* QTLs, which suggests a genetic link between ovary size and age polyethism of the worker caste. Lastly, the forager RGPH states that the phases of the solitary ancestor are seen within the forager caste, as foraging preference for pollen or nectar, which is supported by the extensive overlap between *pln* QTLs and genetic effects for ovary size, which suggests direct genetic links between ovary size and foraging preference (Wang et al. 2009).

The significant QTL regions in my study, 11.30-11.31 and 6.45-6.49, contain genes that fit into the proposed signaling cascades that may underlie caste determination. They have large effects on ovariole number and may have been crucial in the initial determination of the female caste system. Therefore, the transgressive phenotypes could be seen as an atavism of earlier social evolution. Once eusociality was reached, most of the segregating additive genetic variation at these loci was probably selected against and other loci with minor effects on ovary size could have been fine-tuned to govern division of labor in the worker caste. The genes underlying the two major QTL may still have an effect on worker social behavior and adult development but would rarely be expressed. Currently, it is unknown whether segregating genetic variation in these genes causes behavioral variation but the RGPH would predict so.

Genetic Architecture of Worker Ovaries:

The significant QTLs for ovary size identified by this study were located on scaffolds 11.30-11.31 (R3/R5), and 6.45-6.49 (R5). The QTL on chromosome 11 had a significant effect in R3 but was also suggestive in R5. In R3, it had a double peak suggesting that there could be two genes responsible for this QTL. I identified also a number of additional, suggestive QTLs, located on scaffolds 1.06 (R3/R5), 4.09-4.12 (R3), 5.32-5.33 (R5), and 10.29 (R3). My analyses also showed minor effects of QTLs for pollen hoarding behavior (*pln1*, *pln2*, *pln3*) and the age of first foraging (*aff*3, *aff4*, *aff* new). The results of these direct tests suggest that many more QTLs of weak effect size

could exist but were not detected in the general QTL analyses, because of the heightened significance levels in genome-wide LOD threshold.

The ovary was quantitatively characterized as the following phenotypic traits: total, maximum, minimum, total ovary size, ovary asymmetry and ovary activation. The correlation results indicate how these traits relate to each other. I predicted that there would be a positive correlation of both maximum and minimum ovary size with total ovary size. Therefore, I expected that total ovary size would always be significant in the same QTL regions where maximum and minimum ovary size were significant.

The prediction of positive correlations between minimum, maximum, and total ovary size was supported. My results not only show that all three traits are correlated positively with each other, but they overlap readily in the same chromosomes: 1 (1.06), 10 (10.29), and 11 (11.30-11.31) in R3 and chromosome 1 (1.66), 6 (6.45-6.49), 11 (11.30-11.31) and 13 (13.1) in R5. This is not surprising because the minimum ovary size seems to establish the baseline from which the maximum ovary size can expand on, and both also add up to the total ovary size.

I predicted that the asymmetry of ovariole number between the two ovaries would be unrelated to ovary size, and, therefore, would be uncorrelated at the phenotypic and QTL level. The correlation test confirms that there is no significant association between ovary asymmetry and maximum ovary size. However, it does indicate a negative correlation with total, and minimum ovary size. Thus, the smaller the minimum ovary size, the larger the ovary asymmetry. I have already postulated that minimum ovary size creates the base ovariole number from which the maximum ovary size can add which also

affects total ovary size. This would create a situation where maximum ovary size is a less important factor for asymmetry than minimum ovary size. Thus, there would not be a correlation of maximum ovary size and asymmetry. The QTL mapping results indicate that there is no overlap in suggestive or significant QTLs identified for ovary asymmetry and ovary size measurements: The QTL identified for ovary asymmetry for R3 was on chromosome 4 (4.09-4.12), and the R5 QTLs were on chromosome 5 (5.33), and 8 (8.22-8.23). The QTLs identified on both crosses for total, maximum, and minimum ovary size did not include any QTLs on chromosome 4, 5 or 8. The phenotypic correlation between minimum ovary size and ovary asymmetry without overlap in QTL indicate that the variation explained by the identified QTLs is different from the variation that accounted for the overall correlation between ovary asymmetry and minimum ovary size. This would suggest that effects other than the QTLs identified for one trait indirectly influence the variation in the correlated trait.

I also predicted that there would be a positive correlation between ovary activation and total ovary size, minimum and maximum ovary size, but not with ovary asymmetry. A positive correlation has also been suggested by empirical data (Makert et al., 2006). My QTL analyses revealed that the significant QTL for total, maximum, and minimum ovary size in R5 was in the same region as the QTL for ovary activation. This supports the assumption that ovary activation and ovary size are correlated with each other. However, the three phenotypic traits measuring ovary size, via ovariole number, were negatively correlated with ovary activation. The data indicated that an increase in

ovariole number actually reduced the probability of any ovary activation, subsequent egg production and egg-laying.

To account for my divergent results from previous findings (Makert et al. 2006), I suggest that there is an optimal ovary size, where an increase of ovariole number is positively correlated with ovary activation, until the ovary size it too large, and then it becomes negatively correlated. Makert et al. (2006) found a positive association of ovary size and activation, where the average ovariole number per ovarioles per ovary was approximately 4, in both of his study colonies. In contrast, the average ovariole number in R5 was about 13 ovarioles per ovary. The size dependency of the correlation may be because there is a high nutritional cost for maintaining active ovaries, especially in a situation of limited resources under worker-worker competition (Velthuis, 1970).

The occurrence of the large-ovary worker phenotypes observed in my study is largely unknown in natural colonies. The results of my study imply that alleles associated with the extreme worker ovariole number segregate in honey bee populations, perhaps at low frequencies and largely shielded from selection by non-additive effects. However, a few previous studies have found corroborating evidence that alleles for large worker ovary phenotypes are segregating in Africanized honeybees (Chaud-Netto and Bueno, 1979; Thuller et al., 1996; Thuller et al., 1998). My results show that both of the Africanized backcrosses have revealed hidden genetic variation that is perhaps an atavism of ancestral bee sociality and incomplete caste differentiation (Linksvayer et al, 2009).

The two backcrosses are genetically similar, and therefore I expected to see overlap of QTLs between R3 and R5. However, this was not the case. Each backcross had its own significant QTL: R3 with 11.30-11.31 and R5 with 6.45-6.49. Although 11.30-11.31 appears in R5 as a suggestive QTL, its effect is minor and its location not identical to the QTL in R3. This suggests that one of the two significant QTLs is fixed in each backcross and therefore only the other one is detected. The overall crossing results suggest further that these two QTL need to interact epistatically to generate large worker ovaries because neither parental sources, nor the other backcrosses showed the extreme phenotypes of R3 and R5. I evaluated several simple, two-locus genetic models to account for my QTL results and these transgressive phenotypes but I was not able to formulate a satisfactory model without invoking additional modifiers or a genetic threshold model. In addition, there could be three major loci involved, as indicated by the peak on chromosome 6 (6.45-49), and the double peak on chromosome 11 (11.30-11.31).

The type of inheritance seen in my results may be due to a threshold model for multifactorial inheritance. Traits that are polygenic usually have a continuous pattern of trait expression. However, there can be a threshold, based on some physiological variable underlying the trait which determines its phenotype, creating two phenotypic classes (Falconer and Mackay, 1996). In the case of ovary size, the physiological variable could be a physiological variable such as vitellogenin or juvenile hormone that underlies the expression of two classes of ovary size (small, and large). Therefore, the combination of genes in the significant and suggestive QTLs could be involved in the sensitivity to that physiological variable and/or modulation of the two trait classes. The significant QTLs

48

identified in my results could include genes, or a combination of genes, which push the sensitivity to that variable from one sensitivity to another, thus resulting in two phenotypic classes. The suggestive QTLs could in turn be instrumental in the modulation within the sensitivity levels.

An additional possible explanation are genetic background effects. I was able to detect both significant and suggestive QTLs within the Africanized backcrosses but the variation explained by these QTLs does not account for 100% of the phenotypic variation. This suggests that the genetic background of the populations possibly contains additional modifiers which were not identified through QTL mapping. A previous study proposed the worker transgressive ovary sizes as a result of a single recessive gene with incomplete penetrance in combination with additional genes of a much smaller effect (Linksvayer et al., 2009). However, my data suggests that there are at least two loci with a major affect on ovary size. Quantitative trait studies commonly find only few QTL (Hunt et al., 2007), in contrast to the traditional quantitative genetic models, which assume many genes with small additive effects (Barton and Turelli, 1989; Fisher 1958; Orr and Coyne, 1992). In reality, most traits are presumably influenced by a mixture of major and minor QTL effects and I could show this mixture also in my study. Regardless of the strength and number of QTL involved, the genetic architecture of quantitative traits such as ovary size is further complicated by the involvement of environmental interactions and indirect genetic effects. This is particularly true for social insect colonies (Linksvayer, 2006). It is possible that the process of backcrossing has broken up co-adapted gene complexes between brood development and nurse behavior which may have been highly conserved

within the European and Africanized populations. Therefore, the gene combinations that may have been naturally selected against could have emerged and brought about the extreme worker ovariole numbers.

Genetic factors that bias development towards queen phenotypes are selected against at the colony-level, due to their selfish nature because they would result in increased production of queen-phenotypes, and decreased colony efficiency (Anderson et al., 2008; Hughes and Boommsa, 2008). Phenotypes that result in queen-like ovary sizes may be favored within the colony because it would enable workers to produce haploid drones in the event of a queen death (Markert et al, 2006). There may be an optimal worker ovariole number that is the result of the competing and variable fitness effects of inter- and intra-colonial selection (Linksvayer et al. 2009). The resulting variation in ovary size and thus reproductive physiology may be selectively reinforced through its effect on social behavior, increasing the efficiency of worker division of labor (Amdam et al., 2006).

Candidate Genes and Mechanisms

The pollen-hoarding syndrome is characterized by an individual honey bee's propensity for foraging for pollen versus nectar, as well as the age at which foraging is initiated (Rueppell et al. 2006a, b; Page and Amdam 2007). Aspects of the pollen-hoarding syndrome have been putatively linked, via QTL analysis, with genes that are involved in the insulin-like signaling (ILS) pathway (Hunt et al. 2007, Wang et al. 2009). Given the demonstrated effects of these QTLs (*pln 1*, *pln2*, *pln3*, *aff3-4*, *aff* new) on

ovary size, any of the previously identified genes can also be considered as candidate genes for ovary size (Wang et al. 2009).

In particular, Wang et al. (2009) identified two candidate genes within two behavioral QTLs: *pln2* (1.54-1.59) and *pln3* (1.24-1.29), which were experimentally linked with ovary size: *PDK1* and *HR46*. The gene *PDK1* is known to be a kinase with important functions in the ILS pathway, as a down-stream up-regulator. Neuronal ILS includes *PDK1* and affects chemotaxis behavior as well as learning, indicative of its possible effect on behavior (Tomioka et al., 2006). It was shown to vary among the two pollen-hoarding strains. *PDK1* may be somehow involved in a regulatory system where it is up-regulated in forager body fat, which is conditional on ovary size (Orgongozo et al., 2006; Wu and Brown, 2006). The other candidate gene identified was *HR46*, and is known to be a nuclear hormone receptor inducible by ecdysone, in *Drosophila*. It is especially interesting because it is essential in the regulation of a gene that affects apoptosis during fly development (Lam et al., 1997; Lam et al., 1999; Takemoto et al, 2007) and ovary size in the honey bee is also influenced by the apoptotic cascade (Schmidt-Capella and Hartfelder, 1998, 2002).

In accordance with a previous study (Linksvayer et al., 2009), I identified the region 11.30 – 11.31 as a significant QTL for ovary size with several interesting functional candidate genes. One of the most promising is the bee homolog of a gene known to affect ovary size in *Drosophila* named *quail*. It is most expressed in the testis and ovaries and generally expressed in early embryonic development in the pole cells and later germ cells (BDGP Gene Expression Report; FlyAtlas). It might also have a role in

the breakdown of the actin cytoskeleton in the germ cells leading to the apoptotic reduction of ovariole anlagen in honey bee workers, which is JH-titer-dependent (Matova et al., 1999). The other good functional candidates in this QTL interval are *delta* and *cabut* (Linksvayer et al., 2009). The protein *delta*, a ligand of *notch*, plays a role in *Drosophila* embryonic development by specifying cell fates, helping to establish oocyte polarity (Bender et al., 1993), it is also thoroughly expressed in the hindgut and ovary. It has been suggested that over expression of this gene may result in the presence of extra germline stem cells and possibly extra ovarioles (Ward et al, 2006). *Cabut* is attractive as a candidate gene because it is expressed in the pole plasm, a special cytoplasm found in the posterior portion of the egg and early embryo, is responsive to stimuli released from the endocrine system (Zhao et al., 2008), and is connected to cellular apoptosis (Mazars et al., 2001).

The final good potential candidate gene is "similar to fl(2) d protein" (LOC552833, XM_625208). In *D. melanogaster*, the gene *Female Lethal D* is involved in several biological functions related to the ovary: female germline sex determination, primary sex determination, and regulation of alternative mRNA splicing. It has 11 reported alleles in *D. melanogaster* (FLYBASE). This would make it a promising candidate in *A. mellifera*. The *Drosophila* fl(2) gene has been shown to be required for female-specific splicing of *Sxl* and *tra* (*transformer*) pre-mRNAs, which encode nuclear proteins (Ortega 2005; Penalva et al. 2000). *Sxl* is located on the X chromosome, and directs the organism to develop as a female when it is relegated to the "on" switch. Although, *A. mellifera* does not have *tra*, it has functionally similar gene called *csd*

(*complementary sex determinator*)*,* which is the primary signaling party that governs sexual development by its allelic arrangement (Beye, et al. 2003; O'Neil et al, 1992). More research has been done with *fl(2)* in *D. melanogaster* and this has yielded the information that gene *fl(2)d* in *D. melanogaster* is very critical to sex determination and has effects on several other genes via differential splicing. However, such genetic information for *similar to fl(2)d* in *A. mellifera* is non-existent which makes it hard to determine how good this functional candidate is.

The QTL interval on chromosome 6, scaffold 6.45-6.49 contains a potential candidate gene similar to the gene *seven up* or *svp* (LOC408872), in *Drosophila*. It is required for photoreceptor development in ommatidium and has been implicated in the Ras pathway, which is ubiquitous and crucial during organ development (Begemann, 1995; Kramer, 1995). The epidermal growth factor receptor (EGFR) pathway provides mitogenic tip cell signal in the Malphighian tubules that activates *svp* expression and regulates cell division. Cell ablation experiments and studies on the pattern of cell division have shown that a single large cell at the distal end of each tubule, termed the tip cell, is decisive for controlling the proliferation of its neighboring cells (Kerber, 1998). In *svp* mutants, expressions of *string* and *cyclin E* is reduced, reflecting that it is one of the receptors that transmits the mitogenic EGFR signal. Studies on ecdysone signaling pathways show that the *seven-up* protein heterodimerizes with subunits of the ecdysone receptor and regulates gene expression. Although the ecdysone pathway and ILS pathway are largely independent, ecdysteroid titers have been implicated in being directly affected by earlier JH titers (Rachinsky and Hartfelder, 1995; Tu et al., 2002). This relationship

has been suggested to have canalized JH titer differences into two distinct trajectories of the female caste and ovary size (Rachinsky and Hartfelder, 1995).

CHAPTER V

CONCLUSIONS

This project experimentally tested the extent of direct genetic overlap between previously mapped regulators of social behavior and the control of worker reproductive physiology. The overlap of behavioral and reproductive QTL suggests genetic linkage between these traits and thus supports the RGPH in general. The genetic basis of these extreme worker ovary sizes is of additional interest because the two major identified QTL may contain genes that influence the queen – worker differentiation and improve our understanding of the regulation of insect reproduction in general. The prioritized list of candidate genes can be studied in further detail at the molecular level. Two types of experiments can be done in the future on the candidate genes: expression and repression studies. Gene expression studies can evaluate when, and in what specific tissues, a gene is expressed. Repression studies, such as RNA interference (RNAi), can silence or block the expression of a specified gene, and can yield information about the effect of that gene on ovary size.

REFERENCES

Amdam G.V and Omholt S.W. (2003). The hive bee to forager transition in honeybee colonies: the double repressor hypothesis. *Journal of Theoretical Biology*. 223: 451-464.

Amdam G.V., Norberg K., Hagen A., and Omholt S.W. (2003) Social exploitation of vitellogenin. *Proceedings of the National Academy of Sciences of the United States of America*. 100: 1799-1802.

Amdam G.V., Norberg K, Fondrk K, and Page RE (2004) Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proceedings of the National Academy of Science of the United States of America* 101: 11350–11355

Amdam G.V., Csondes A., Fondrk M.K and Page RE (2006) Complex social behavior derived from maternal reproductive traits. *Nature*. 439: 76-78.

Amdam G.V., Nilsen, K.A, Norberg, K., Fondrk M.K., and Hartfelder, K. (2007) variation in endocrine signaling underlies variation in social life history. *The American Naturalist*. 170 (1): 37-46.

Ament S.E, Corona M., Pollock H.S, and Robinson G.E. (2008) Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proceedings of the National Academy of Sciences*. 11(105): 4226-4231.

Anderson, K. E., T. A. Linksvayer and C. R. Smith (2008) The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). Myrmecological News 11: 119-132.

Barton NH, and Turelli M. (1989) Evolutionary quantitative genetics: how little do we know? *Annu. Rev. Genet.* 23: 337-370.

BEEBASE. The Honeybee Genome Database. Texas Agricultural and Mechanical University. http://genomes.tamu.edu/

Bender, L. B., P. J. Kooh and M. A. Muskavitch, 1993 Complex function and expression of Delta during Drosophila oogenesis. Genetics 133: 967-978.

Begemann, G. (1995). The *Drosophila* orphan nuclear receptor *seven-up* requires the as pathway for its function in photoreceptor determination. *Development* 121:225-235

Beye M, Gattermeier I, Hasselmann M., Gempe T., Schioett M., Baines J.F, Schlipalius D., Mougel F., Emore C., Rueppell O., and Sirviö A. (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Res.* 16: 1339-1344.

Beye M, Hasselmann M, Fondrk M.K, Page R.E, Omholt S.W. (2003) The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell* 114(4): 419-429.

Butler C.G. (1957) The control of ovary development in worker honeybees (*A. mellifera*). *Experimentia*. 13: 256-257.

Chaud-netto, J., and O. C. Bueno, (1979) Number of Ovarioles in Workers of *Apis mellifera adansonii* and *Apis mellifera ligustica*: a comparative study. Journal of Apicultural Research 18: 260-263.

Corona M., Velarde R.A, Remolina S., Moran-Lauter A., Wang, Y., Hughes K.A and Robinson G.E (2007) Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proceedings of the National Academy of Sciences of the United States of America*. 104(17): 7128-7133.

Crespi B.J and Yanega D. (1995) The definition of eusociality. *Behavioral Ecology*. 6 (1): 109-115.

Darvasi A, and Soller M. (1992) Selective genotyping for determination of linkage between a marker locus and quantitative trait locus. *Theoretical and Applied Genetics*. 85 (2-3): 353-359.

Duffy J.E, Morrison C.L and Rios R. (2000) Multiple origins of eusociality among sponge-dwelling shrimps (*Synalpheus*). *Evolution*. 54 (2): 503-516.

Evans, J.D and Wheeler, D.E (2001) Gene expression and the evolution of insect polymorphenisms. *Bioessays* 23: 62-68.

Falconer D.S and Mackay T.F.C (1996). Introduction to quantitative genetics. 4th ed. Pearson Education Limited. Pp. 459.

Fisher RA (1958) *The Genetical theory of natural selection*. Dover Publications, Inc. New York.

Flybase: A Database of Drosophila Genes and Genomes <http://flybase.bio. indiana.edu/>

Foster K.R, Wenseleers T. and Ratnieks F.L.W. (2005) Kin selection is the key to altruism. *TRENDS in Ecology and Evolution*. 21(2): 57-60.

Free, J.B (1965) The allocation of duties among worker honey bees. *Symp. Zool. Soc. London* 14: 39-59.

Griffin A.S. and West S.A. (2003). Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. *Science*. 302: 634-636.

Guidugli, K.R., Nascimento A.M, Amdam G.V., Barchuk A.R., Omhold S., Simões, and Hartfelder K. (2005) Vitellogenin regulates hormonal dynamics in the worker caste of a eusocial insect. *FEBS Letters*. 4961-4965.

Hamilton W.D. (1964). The genetical evolution of social behavior. I and II. *Journal of Theoretical Biology*. 7:1-52.

Haydak MH. (1943) Larval food and development of castes in the honeybee. *J Econ Entomol*. 36: 778–792.

Hepperle C. and Hartfelder K (2001) Differentially expressed regulatory genes in honey bee caste development. *Naturwissenschaften* 88: 113-116.

Hughes, W. O. H., and J. J. Boomsma, (2008) Genetic royal cheats in leaf-cutting ant societies. Proceedings of the National Academy of Sciences of the United States of America 105: 5150-5153

Hunt G.J, and Page R.E (1995) Linage map of the honey bee *Apis mellifera*, based on RAPD markers. *Genetics* 139: 171-1382.

Hunt G.J, Page R.E, Fondrk M.K and Dullum CJ (1995) Major quantitative trait loci affecting honey bee foraging behavior. *Genetics* 141: 1537-1545.

Hunt G.J, Amdam G.V, Schlipalius D, Emore C, Sardesai N, William C.E, Rueppell O, Guzman-Novoa E, Arechavaleta-Velasco M, Chandra S, Fondrk M.K, Beye M, and Page Jr. R.E. (2007) Behavioral genomics of honeybee foraging and nest defense. *Naturwissenschaften* 94: 247-267.

Humphries MA, Muller U, Fondrk MK, Page RE (2003). PKA and PKC content in the honey bee central brain differs in genotypic strains with distinct foraging behavior. *J Comp Physiol*. 189: 155-562.

Hrassnigg N., and Crailsheim K. (2005) Differences in drone and worker physiology in honeybees (*Apis mellifera*). *Apidologie*. 36: 255-277.

Jansen, R.C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics*. 135: 205-211.

Jansen, R.C (1994). Controlling the Type I and Type II errors in mapping quantitative trait loci. *Genetics*. 138: 871-881.

Jansen, R.C and Stam, P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*. 135: 1447-1455.

Jarvis JU. (1981) Eusociality in a mammal: cooperative breeding in naked mole-rat colonies. *Science*. 212 (4494): 571-573.

Kao, C. and Z. Zeng (1997). General formulae for obtaining the MLEs and the asymptotic variance-covariance matrix in maping quantitative trait loci. *Biometrics* 53: 653-665.

Kao, C., Z. Zeng, and R. Teasdale (1999). Multiple interval mapping for quantitative trait loci. *Genetics* 152: 1203-1216.

Kelly T.J, Adams T.S, Schwarz M.B, Birnbaum M.J, Rubenstein E.C, and Imberski R.B (1987) Juvenile hormone and ovarian maturation in the diptera: a review of recent results. *Insect Biochemistry*. 17: 1089-1093.

Kerber, B., Fellert, S. and Hoch, M., (1998). Seven-up, the Drosophila homolog of the COUP-TF orphan receptors, controls cell proliferation in the insect kidney. *Genes Dev*. 12(12): 1781-1786.

Kerr, WE, Del Rio SL, and Barrionuevo MD (1982). The southern limits of the distribution of the Africanized honey bee in South America. *Am. Bee J.* 122: 196-198.

Kramer, S., West, S. R. and Hiromi, Y. (1995). Cell fate control in the Drosophila retina by the orphan receptor seven-up: its role in the decisions mediated by the ras signaling pathway. *Development* 121: 1361-1372

Johnson B.R (2003) Organization of work in the honeybee: A compromise between division of labour and behavioral flexibility. *Proceedings of the National Academy of Science of the United States of America*. 270: 147-152.

Jung-Hoffmann, L. (1966) Die determination von konigin und arbeiterin der honigbiene. *Z. Bienenforsch*. 8: 296-322.

Lam GT, Jiang C and Thummel CS (1997) A steroid-triggered transcriptional hierarchy controls salivary glad cell death during Drosophila metamorphosis. *Mol. Cell*. 5: 445-455.

Lam G, Hall BL, Bender M and Thummel CS (1999). DHR3 is required for the prepupal-pupal transition and differentiation of adult structures during Drosophila metamorphosis. *Devo. Biol.* 212:204-216.

Lander, E.S. and D. Botstein (1989). Mapping mendelian factors underlying quantitative traits using rflp linkage maps. *Genetics* 121: 185-199.

Lincoln S.E, Daly M.J, and Lander (1993) Constructing genetic linkage maps with MAPMAKER/EXP, version 3.0: A tutorial and preference manual. Cambridge, MA: A Whitehead Institute for Biological Research Technical Report.

Lindauer, M. (1953) Division of labour in the honeybee colony. *Bee World*. 34: 63-73.

Linksvayer T.A, Rueppell O., Kaftanoglu O, and Page R.E (2009) The genetic basis for transgressive ovary sizes in honey bee workers. *Genetics*. In press.

Linksvayer, T. A., (2006) Direct, maternal, and sibsocial genetic effects on individual and colony traits in an ant. *Evolution* 60**:** 2552-2561

Makert G.R., Paxton R.J., and Hartfelder K.(2006) Ovariole number – A predictor of differential reproductive success among worker subfamilies in queenless honeybee (*Apis mellifera L.*) colonies. *Behavioral Ecology and Sociobiology*. 60 (6): 815-825.

Mazars, A., F. Lallemand, C. Prunier, J. Marais, N. Ferrand (2001) Evidence for a role of the JNK cascade in Smad7-mediated apoptosis. *Journal of Biological Chemistry* 276: 36797-36803.

Michener CD (1975) The Brazilian bee problem. *Annu. Rev. Entomol.* 20: 399-416.

NCBI: National Center for Biotecnology Information. <http://www.ncbi.nlm.nih.gov/>

Nijhout, H.F (2003) The control of growth. *Development* 130: 5863-5867.

Oldroyd B.P and Beekman M (2008). Effects of selection for honey bee worker reproduction on foraging traits. *PLoS Biology*. 6(3): 463-470.

O'Neil M. T and J.M Belote. (1992) Interspecific comparison of the transformer gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics* 131(1): 113–128.

Orr, HA and Coyne JA (1992) The genetics of adaptation: a reassessment. *Am. Nat.* 140: 725-742.

Ortega A. (2005) Localization of the *Drosophila* protein FL(2)D in somatic cells and female gonads. *Cell and Tissue Research*. 320(2): 361-367.

Orgogonzo V, Broman KW, and Stern D.L (2006) High resolution quantitative trait locus mapping reveals sign of epistasis controlling ovariole number between two *Drosophila* species. *Genetics* 173: 197-205.

Oxley, P.R, Thompson G.J, and Oldroyd B.P (2008) Four quantitative trait loci that influence worker sterility in the honeybee (*Apis mellifera*). *Genetics* 179(3): 1337-1343.

Page R.E and Amdam G.V (2007) The making of a social insect: developmental architectures of social design. *Bioessays* 29(4): 334-343.

Page R.E and Fondrk M.K (1995) The effects of colony-level selection on the social organization of honey bee (*Apis mellifera* L.) colonies: colony-level components of pollen hoarding. *Behavioral Ecology and Sociobiology*. 36(2): 135-144.

Page R.E, Erber J and Fondrk M.K (1998) The effect of genotype on response thresholds to sucrose and foraging behavior of honey bees (*Apis mellifera* L.). *J. Comp. Physiol. A*. 182: 489-500.

Page R.E, Fonderk M.K, Hunt G.J, Guzaman-Novoa E., and Humphries M.A (2000) Genetic dissection of honeybee (*Apis mellifera* L.) foraging behavior. *Journal Of Heredity*. 91: 474-479.

Page R.E and Erber J. (2002) Levels of behavioral organization and the evolution of division of labor. *Naturwissenschaften*. 89: 91-106.

Page R.E and Amdam G.V (2007) The making of a social insect: developmental architectures of social design. *Bioessays* 29(4): 334-343.

Pankiw T. (2003) Directional change in a suite of foraging behaviors in tropical and temperate evolved honey bees (*Apis mellifera* L.). *Behavioral Ecology and Sociobiology*. 54(5): 458-464.

Pankiw T., and Page RE (2001) Genotype and colony environment affect honeybee (*Apis mellifera L.*) development and foraging behavior. *Behavioral Ecology and Sociobiology*. 51: 87-94.

Pankiw T and Page PE (2003). Effect of pheromones, hormones, and handling on sucrose response thresholds of honey bees (*Apis mellifera*). *J Comp Physiol.* A 189: 675-685.

Penlava LO, Ruiz MF, Ortega A, Granadino B, Vincente L, Segarra C, Valcarcel J, and Sanchez L. (2000) The *Drosophila* fl(2)d gene, required for female-specific splicing of Sxl and tra pre-mRNAs, encodes a novel nuclear protein with a HQ-rich domain. Genetics 155(1):129-39.

Queller D.C, and Strassmann J.E. (1998). Kin selection and social insects. *BioScience* 48(3): 165-175

Rachinksy, A and Hartfelder K (1995) Caste development of honeybees (*Apis mellifera*): Juvenile hormone turns on ecdysteroids. *Naturwissenschaften.* 82: 378-379.

Rachinksy A., Strambi C., Strambi A., and Hartfelder K (1990) Caste and metamorphosis: hemolymph titers of juvenile hormone and ecdysteroids in last instar honey bee larva. *General and Comparative Endocrinology.* 79L 31-38.

Ragoussis J., Elvidge GP, Kaur K., and Colella S. (2006) Matrix-assisted laser desorption/ionization, time-of-flight mass spectrometry in genomics research. *PLOS Genet.* 2: 920-929.

Ribbands, C.R. (1952) Division of labour in the honey bee community. *Proc. Roy. Soc. London.* (B) 140: 32-42.

Ribbands, C.R (1953) *The behaviour and social life of honeybees*. London, Bee Research Association. (Repr. 1964, New York, Dover).

Robinson G.E (2002) Genomics and integrative analyses of division of labor in honeybee colonies. *American Naturalist.* 160(6): s160-s172

Robinson G.E., Strambi C., Strambi A.,and Huang Z.-Y (1992) Reproduction in worker honey bees is associated with low juvenile hormone titers and rates of biosynthesis. *Gen. Comp. Endocrinol.* 87: 471-480.

Robinson G.E., and Vargo E.L. (1997) Juvenile hormone in adult eusocial hymenoptera: gonadotropin and behavioral pacemaker. *Arch. Insect Biochem. Physiolo.* 35: 559-583.

Rösch, G.A. (1925) Untersuchungen über die arbeitsteilung im bienenstaat I. Teil: Die tätigkeiten im normalen bienenstaate und ihre beziehungen zum alter der arbeisbienen. *Z. vergl. Physiol.* 2: 571-631.

Rüppell O., Pankiw T., and Page R.E. (2004) Pleiotropy, epistasis and new QTL: the genetic architecture of honey bee foraging behavior. *Journal of Heredity*. 95 (6): 481-491.

Rueppell O (2009) Characterization of quantitative trait loci for the age of first foraging in honey bee workers. *Behavior Genetics*. DOI 10.1007/s10519-009-9278-8.

Rueppell O., Pankiw T., Nielsen D.L., Fondrk M.K., Beye M., and Page R.E. (2004) The genetic architecture of the behavioral ontogeny of foraging in honeybee workers. *Genetics*. 167: 1767-1779.

Rueppell O., Chandra S.B.C, Pankiw T., Fondrk M.K., Bete M., Hunt G., and Page R.E (2006a) The genetic architecture of sucrose responsiveness in the honeybee (*Apis mellifera* L.). *Genetics*. 172: 243-251.

Rueppell O., Page R.E, and Fondrk M.K (2006b) Male behavioural maturation rate responds to selection on pollen hoarding in honeybees. *Animal Behavior*. 71(1): 227-234.

Rueppell O, Hunggims E, and Tingek S (2008) Association between larger ovarioes and pollen foraging in queenless *Apis cerana* workers supports the reproductive ground - plan hypothesis of social evolution. *J Insect Behav* 21: 317-321.

Scheiner R., Page R.E, and Erber J. (2001) Responsiveness to sucrose affects tactile and olfactory learning in preforaging honey bees of two genetic strains. *Behavioural Brain Research*. 67-73.

Schmidt-Capella IC and Hartfelder K (1998) Juvenile hormone effect on DNA synthesis and apoptosis in caste-differentiation of the larval honey bee (*Apis mellifera* L.) ovary. *J. Insect Physiol*. 44:385-391.

Schmidt-Capella IC and Hartfelder (2002). Juvenile-hormone-dependent interaction of actin and spectrin is crucial for polymorphic differentiation of the larval honey bee ovary. *Cell Tissue Res*. 301: 265-272.

Sheppard WS, Soares AEE, DeJong D, and Shimanuki H. (1991) Hybrid status of honey bee populations near the historic origin of the Africanization in Brazil. *Apidologie*. 22: 643-652.

Sherman, P.W, Lacey, E.A, Reeve, H.K, and Keller, L (1995) The eusociality continuum. *Behavioral Ecology*. 6 (1): 102-108.

Shuelke M (2000) An economic method for the fluorescent labeling of PCR fragments: A poor man's approach to genotyping for research and high-throughput diagnostics. *Nature Biotechnology*. 18: 233-234.

Snodgrass, R.E (1935). *Principles of Insect Morphology*. McGraw-Hill, New York.

Solignac, M, Mougel F, Vautrin D, Monnerot M, and Cournuet J.M. (2007) A third-genetation microsatellite-based linkage map of the honeybee, *Apis mellifera*, and it's comparison with the sequence-based physical map. *Genome Biology* 8(4): R66.

Spivak M, Fletcher DJC, and Breed MD (1991). *In: The "African" Honey Bee* (Spivak M, Fletcher DJC, Breed MD, eds) Westview, Boulder, CO, p 3.

Sullivan J.P., Fahrbach S.E., Harrison J.F., Capaldi E.A., Fewell J.H., and Robinson G.E (2003) Juvenile hormone and division of labor in honeybee colonies: effects of allatectomy on flight behavior and metabolism. *J. Exp. Biol.* 206: 2287-2296.

Taylor OR Jr. (1988) Ecology and economic impact of African and Africanized honey bees. *In: Africanized Honey Bees and Bee Mites* (Needham GR, Page RE, Delfinado-Baker M, Bowman CD, eds) Elis Horwood, Chichester, pp. 29-41.

Thuller, R. H. C., O. Malaspina, O. C. Bueno and J. Chaud-Netto (1996) Number of ovarioles in workers descendent from crossings between Africanized and Italian honeybees (Apis mellifera L.): Comparing stock, inbred and F1 colonies. Anais da Sociedade Entomologica do Brasil 25: 501-506.

Thuller, R. H. C., O. Malaspina, O. C. Bueno and J. Chaud-Netto (1998) Number of ovarioles in workers descendent from crossings between Africanized and Italian honeybees, Apis mellifera L.: Comparison among backcrosses and ancestors colonies. Anais da Sociedade Entomologica do Brasil 27: 237-243.

Tomioka M., Adachi T., Suzuki H, Kunitomo H, and Schafer WR (2006) The insulin/PI3-kinase pathway regulates salt chemotaxis in *Caenorhabditis elegans*. *Neuron*. 51: 613-625.

Trivers, R.L and Hare, H (1976). Haplodiploidy and the evolution of the social insects. *Science*. 191 (4224): 249-263.

Tu M.P, Yin C-M, and Tatar M. (2002) Impaired ovarian ecdysone synthesis of *Drosophila melanogaster* insulin receptor mutants. *Aging Cell*. 1: 158-160.

Van Ooijen J.W, Boer M.P, Jansen R.C, and Maliepaard C. (2002) MapQTL® 4.0: Software for the calculation of QTL positions on genetic maps. Wageningen, The Netherlands: Plant Research International

Velthuis H.H.W (1970) Ovarian Development in *Apis mellifera* worker bees. *Ent. Exp. and Appl.* 13:377-394.

Ward, E. J., H. R. Shcherbata, S. H. Reynolds, K. A. Fischer, S. D. Hatfield (2006) Stem cells signal to the niche through the Notch pathway in the Drosophila ovary. *Current Biology* 16: 2352-2358.

West-Eberland M.J. (1987) Flexible strategy and social evolution. *in Animal societies: theories and fact* (Y. Ito, J.L. Brown, and J. Kikkawa, eds.) Japan Scientific Societies Press, Tokyo. pp. 35-51.

West-Eberland M.L (1996) Wasp societies as microcosms for the study of development and evolution. *in Natural History and Evolution of Paper Wasps.* (S. Turillazzi and M. J. West-Eberland, eds.) Oxford University Press, New York. pp. 290-317.

Wheeler D.E. (1996) The role of nourishment in oogenesis. *Annual Review of Entomology* 41: 407-431.

Wheeler D.E, Buck N, and Evans J.D (2006) Expression of insulin pathway genes during the period of caste determination in the honeybee, *Apis mellifera. Insect Molecular Biology* 15(5): 597-602.

Whitfield C.W, Behura S, Berlocher S.H, Clark A.G, Johnston J.S, Sheppard W.S, Smith D.R, Suarez A.V, Weaver D, and Tsutsui N.D. (2006) Thrice out of Africa: ancient Expansions of the Honey Bee, *Apis mellifera. Science* 314(5799): 642-645.

Wilson E.O (1971) *The Insect Societies.* Belknap Press of Harvard Univ. Press, Cambridge, MA. 548 pp.

Winston, M.L (1987) The Biology of the Honey Bee The Harvard Univeristy Press, Cambridge, MA. 278 pp.

Wu Q. and Brown M.R (2006) Signaling and function of insulin-like peptides in insects. *Annual Review of Entomology.* 51: 1-26.

Zeng, Z., C. Kao, and C. J. Basten (1999). Estimating the genetic architecture of quantitative traits. *Genetical Research, Cambridge* 74: 279-289.

Zhao, T., T. Gu, H. C. Rice, K. L. McAdams, K. M. Roark (2008) A *Drosophila* gain-of-function screen for candidate genes involved in steroid-dependent neuroendocrine cell remodeling. *Genetics* 178: 883-901.

APPENDIX

TABLES AND FIGURES

Chromosome	Number of Markers	Mapped Length
1	31	979.8 cM
2	23	626.5 cM
3	13	393.4 cM
4	13	362 cM
5	15	478.8 cM
6	18	478.6 cM
7	15	305.8 cM
8	16	264.9 cM
9	10	323.1 cM
10	14	423.8 cM
11	19	351.2 cM
12	11	367.1 cM
13	13	227.1 cM
14	10	280.2 cM
15	12	228.1 cM
16	6	114.9 cM
Total	239	6205.3 cM

Table 1: Chromosome number, number of markers and the mapped length of each chromosome using MapMaker 3.0, in the Africanized backcross, R3.

Chromosome	Number of Markers	Mapped Length
1	30	742 cM
2	21	472.5 cM
3	10	431.4 cM
4	14	357.5 cM
5	18	379.4 cM
6	16	503.7 cM
7	11	358.2 cM
8	12	335.7 cM
9	11	458.7 cM
10	9	431.6 cM
11	18	246.1 cM
12	13	295.2 cM
13	13	271 cM
14	9	301.7 cM
15	11	205.9 cM
16	6	139.9 cM
Total	222	5930.5

Table 2: Chromosome number, number of markers and the mapped length of each chromosome using MapMaker 3.0, in the Africanized backcross, R5.

		Total	Min	Max	Asymm	Develop
Total	r		0.983	0.987	-0.263	-
	Sig.		p≤0.001	p≤0.001	0.014	-
Min	r	0.943		0.944	-0.406	-
	Sig.	p≤0.001		p≤0.001	p≤0.001	-
Max	r	0.952	0.804		-0.139	-
	Sig.	p≤0.001	p≤0.001		.198	-
Asymm	r	-0.232	-0.5	.038		-
	Sig.	p≤0.001	p≤0.001	.485		-
Activation	r	-0.255	-0.225	-0.243	.037	
	Sig.	p≤0.001	p≤0.001	p≤0.001	.491	

Table 3: Trait correlation for total, minimum, maximum ovary size, ovary asymmetry and ovary development in both of the Africanized backcrosses (R3 above and R5 below diagonal). Results of Spearman-Rank tests are shown.

Total Ovary Size (LOD_{GW}: 3.2)				
Group	Locus	K-W (SM)	IM	MQM
1.06	AT137	8.443 (p≤0.005)	2.10 (LOD_{ChW}: 2.2)	2.03 (LOD_{ChW}: 2.2)
2.31	est1833	6.758 (p≤0.01)	ns	ns
2.31	est1831	6.758 (p≤0.01)	ns	ns
10.29	ahb2105	7.455 (p≤0.01)	ns	ns
11.31	est8460	19.711 (p≤0.0001)		
11.31	K1112C	8.719 (p≤0.005)	6.45 (LOD_{ChW}: 2.0)	6.45 (LOD_{ChW}: 2.0)
11.31	AP151	8.913 (p≤0.005)		
11.31	est8456	31.915 (p≤0.0001)		

Table 4: Putative QTL regions identified for total ovary size in R3 through 3 analyses (Kruskal-Wallis single marker, interval mapping, multiple QTL mapping) with corresponding locus, group-scaffold, genome-wide LOD (LOD_{GW}) and chromosome-wide LOD scores (LOD_{ChW}). Markers that did not exceed the suggestive LOD 2.0 criterion or the chromosome-wide LOD significance threshold are marked as non-significant (n.s.).

Total Ovary Size (LOD$_{GW}$: 3.1)			
Group	Locus	K-W (SM)	IM
1.66	ahb1468	7.878 (p≤0.01)	ns
1.66	ahb1523	6.915 (p≤0.01)	ns
6.45	est4967	27.952 (p≤0.0001)	7.96 (LOD$_{ChW}$: 2.0)
6.45	SV062	23.358 (p≤0.0001)	7.96 (LOD$_{ChW}$: 2.0)
6.49	UN258	11.28 (p≤0.001)	7.96 (LOD$_{ChW}$: 2.0)
9.24	ahb10360	7.923 (p≤0.005)	ns
11.31	est8349	7.545 (p≤0.01)	1.89 (LOD$_{ChW}$: 1.8)
11.31	est8339	7.998 (p≤0.005)	1.89 (LOD$_{ChW}$: 1.8)
11.31	ahb2567	6.655 (p≤0.01)	1.89 (LOD$_{ChW}$: 1.8)
13.1	est10066	7.491 (p≤0.01)	2.68 (LOD$_{ChW}$: 1.8)
13.1	est10110	12.878 (p≤0.0005)	2.68 (LOD$_{ChW}$: 1.8)
15.36	OR15_36b	6.755 (p≤0.01)	ns

Table 5: Putative QTL regions identified for total ovary size in R5 through 2 analyses (Kruskal-Wallis single marker, interval mapping) with corresponding locus, group-scaffold, genome-wide (LOD$_{GW}$) and chromosome-wide (LOD$_{ChW}$) LOD significance thresholds. Markers that did not exceed the suggestive LOD 2.0 criterion or the chromosome-wide LOD significance threshold are marked as non-significant (n.s.).

Maximum Ovary Size (LOD_{GW}: 3.2)				
Group	Locus	K-W (SM)	IM	MQM
1.06	AT137	7.225 (p≤0.01)	2.83 (LOD_{ChW}: 2.4)	2.83 (LOD_{ChW}: 2.4)
1.82	A040	7.477 (p≤0.01)	ns	ns
3.31	AT066b	7.320 (p≤0.01)	ns	ns
10.29	ahb2105	7.789 (p≤0.01)	1.98 (LOD_{ChW}:2.0)	2.83 (LOD_{ChW}:2.0)
10.29	K1064	ns		
11.31	est8460	17.546 (p≤0.0001)	5.87 (LOD_{ChW}: 2.0)	5.87 (LOD_{ChW}: 2.0)
11.31	K1112C	7.867 (p≤0.01)		
11.31	AP151	7.922 (p≤0.005)		
11.31	est8456	29.072 (p≤0.0001)		

Table 6: Putative QTL regions identified for maximum ovary in R3 through 3 analyses (Kruskal-Wallis single marker, interval mapping, multiple QTL mapping) with corresponding locus, group-scaffold, genome-wide LOD (LOD_{GW}) and chromosome-wide LOD scores (LOD_{ChW}). Markers that did not exceed the suggestive LOD 2.0 criterion or the chromosome-wide LOD significance threshold are marked as non-significant (n.s.).

Maximum Ovary Size (LOD$_{GW}$: 3.1)			
Group	Locus	K-W (SM)	IM
1.66	Rahb1468	7.545 (p≤0.01)	-
6.45	est4967	24.909 (p≤0.0001)	
6.45	RSV062	21.336 (p≤0.0001)	7.16 (LOD$_{ChW}$: 2.0)
6.49	RUN258	13.209 (p≤0.0005)	
9.24	ahb10360	7.059 (p≤0.01)	ns
13.1	est10066	7.437 (p≤0.01)	
13.1	est10110	13.388 (p≤0.001)	2.78 (LOD$_{ChW}$: 1.8)
13.1	ahb3988	6.826 (p≤0.01)	

Table 7: Putative QTL regions identified for maximum ovary size in R5 through 2 analyses (Kruskal-Wallis single marker, interval mapping) with corresponding locus, group-scaffold, genome-wide LOD (LOD$_{GW}$) and chromosome-wide LOD scores (LOD$_{ChW}$). Markers that did not exceed the suggestive LOD 2.0 criterion or the chromosome-wide LOD significance threshold are marked as non-significant (n.s.).

Minimum Ovary Size (LOD$_{GW}$: 3.1)				
Group	Locus	K-W (SM)	IM	MQM
1.06	RAT137	9.940 (p≤0.005)	2.39 (LOD$_{ChW}$: 2.1)	ns
10.29	Rahb2105	6.979 (p≤0.01)	ns	ns
11.31	est8460	21.137 (p≤0.0001)	6.71 (LOD$_{ChW}$: 2.0)	6.71 (LOD$_{ChW}$: 2.0)
11.31	RK1112C	9.236 (p≤0.005)		
11.31	RAP151	9.488 (p≤0.005)		
11.31	Rest8456	33.790 (p≤0.0001)		

Table 8: Putative QTL regions identified for minimum ovary size in R3 through 2 analyses (Kruskal-Wallis single marker, interval mapping, multiple QTL mapping) with corresponding locus, group-scaffold, genome-wide LOD (LOD$_{GW}$) and chromosome-wide LOD scores (LOD$_{ChW}$). Markers that did not exceed the suggestive LOD 2.0 criterion or the chromosome-wide LOD significance threshold are marked as non-significant (n.s.).

Minimum Ovary Size (LOD$_{GW}$: 3.1)			
Group	Locus	K-W (SM)	IM
1.66	Rahb1468	8.176 (p≤0.005)	2.02 (LOD$_{ChW}$: 2.1)
1.66	Rahb1523	8.046 (p≤0.005)	
3.34	ahb10780	7.890 (p≤0.005)	2.04 (LOD$_{ChW}$: 1.8)
6.45	est4942	6.641 (p≤0.01)	7.91 (LOD$_{ChW}$: 1.9)
6.45	est4967	29.262 (p≤0.0001)	
6.45	RSV062	24.846 (p≤0.0001)	
6.49	RUN258	9.827 (p≤0.005)	
8.22	est6265	8.054 (p≤0.005)	-
8.22	Rahb12013	8.054 (p≤0.005)	-
8.22	Rahb12014	8.054 (p≤0.005)	-
11.31	est8349	9.553 (p≤0.005)	2.01 (LOD$_{ChW}$: 1.7)
11.31	est8339	10.48 (p≤0.005)	
11.31	ahb2567	8.839 (p≤0.005)	
13.1	est10110	10.421 (p≤0.005)	2.27 (LOD$_{ChW}$: 2.0)
15.36	OR15_36b	8.459 (p≤0.005)	-

Table 9: Putative QTL regions identified for minimum ovary size in R5 through 2 analyses (Kruskal-Wallis single marker, interval mapping) with corresponding locus, group-scaffold, genome-wide LOD (LOD$_{GW}$) and chromosome-wide LOD scores (LOD$_{ChW}$). Markers that did not exceed the suggestive LOD 2.0 criterion or the chromosome-wide LOD significance threshold are marked as non-significant (n.s.).

Ovary Asymmetry (LOD$_{GW}$: 3.3)				
Group	Locus	K-W (SM)	IM	MQM
4.09	est3866	10.913 (p≤0.001)	2.91 (LOD$_{ChW}$: 1.9)	2.91 (LOD$_{ChW}$: 1.9)
4.12	RK0423B	12.014 (p≤0.001)		

Table 10: Putative QTL regions identified for ovary asymmetry in R3 through 3 analyses (Kruskal-Wallis single marker, interval mapping, multiple QTL mapping) with corresponding locus, group-scaffold, genome-wide LOD (LOD$_{GW}$) and chromosome-wide LOD scores (LOD$_{ChW}$).

Ovary Asymmetry Size (LOD$_{GW}$: 3.0)			
Group	Locus	K-W (SM)	IM
5.33	Rest4637	12.19 (p≤0.0005)	3.12 (LOD$_{ChW}$: 2.0)
5.33	Rest4644	11.276 (p≤0.001)	
8.22	est6265	7.494 (p≤0.01)	2.15 (LOD$_{ChW}$: 1.8)
8.23	Rahb12013	7.494 (p≤0.01)	
8.23	Rahb12014	7.494 (p≤0.01)	

Table 11: Putative QTL regions identified for ovary asymmetry in R5 through 2 analyses (Kruskal-Wallis single marker, interval mapping) with corresponding locus, group-scaffold, genome-wide LOD (LOD$_{GW}$) and chromosome-wide LOD scores (LOD$_{ChW}$).

Ovary Activation (LOD_{GW}: 3.0)			
Group	Locus	K-W (SM)	IM
2.43	est1929	6.997 (p≤0.01)	ns
6.45	est4967	18.558 (p≤0.0001)	6.86 (LOD_{ChW}: 1.9)
6.49	RUN258	14.737 (p≤0.0005)	

Table 12: Putative QTL regions identified for ovary development in R5 through 2 analyses (Kruskal-Wallis single marker, interval mapping) with corresponding locus, group-scaffold, genome-wide LOD (LOD_{GW}) and chromosome-wide LOD scores (LOD_{ChW}). Markers that did not exceed the suggestive LOD 2.0 criterion or the chromosome-wide LOD significance threshold are marked as non-significant (n.s.).

QTL	pln1	pln2	pln3	pln4	aff3	aff4	aff new
Marker	est10066	K0137T	RAp047	est10185	ahb7242	RAT164*	Ahb2788
Total	ns	ns	ns	ns	ns	0.035 (p≤0.05)	ns
Min	ns	ns	ns	ns	ns	ns	ns
Max	ns	ns	ns	ns	ns	0.021 (p≤0.05)	ns

Table 13: Testing for significance of genotyped loci in Africanized backcross R3 within previously identified QTL regions for pollen-hoarding (*pln1-4*) and the age-of-first-foraging (*aff 3-4, aff new*). Mann-Whitney U Tests, ns = nonsignificant. An asterisks designates a marker that is associated with the strongest ovary effect in its chromosomal region.

QTL	pln1	pln2	pln3	pln4	aff3	aff4	aff new
Marker	est10110*	K0137T	est788*	est10185	K0416B	est4292	est8472*
Total	p≤0.001	0.048 (p≤0.05)	0.052 (p≤0.05)	ns	ns	ns	0.015 (p≤0.05)
Min	0.001 (p≤0.05)	0.032 (p≤0.05)	ns	ns	ns	ns	0.012 (p≤0.05)
Max	p≤0.001	ns	0.03 (p≤0.05)	ns	ns	ns	0.035 (p≤0.05)

Table 14: Testing for significance of genotyped loci in Africanized backcross R5 within previously identified QTL regions for pollen-hoarding (*pln1-4*) and the age-of-first-foraging (*aff 3-4, aff new*). Mann-Whitney U Tests, ns = nonsignificant. An asterisks designates a marker that is associated with the strongest ovary effect in its chromosomal region.
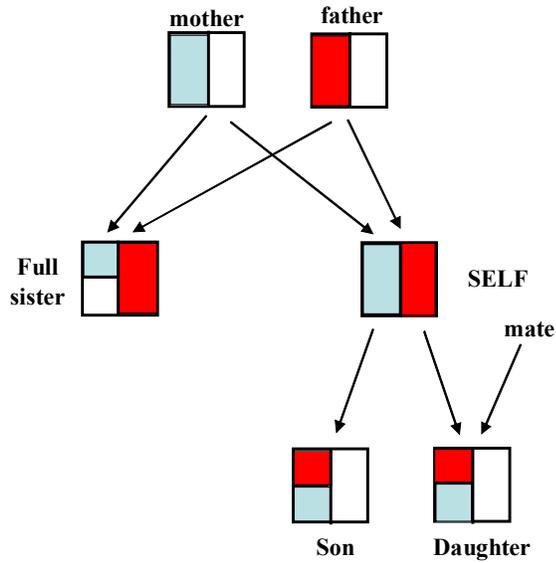
Figure 1: Haplo-diploidy entails a high relatedness among full sisters and could therefore have facilitated social evolution by kin selection in the Hymenoptera.
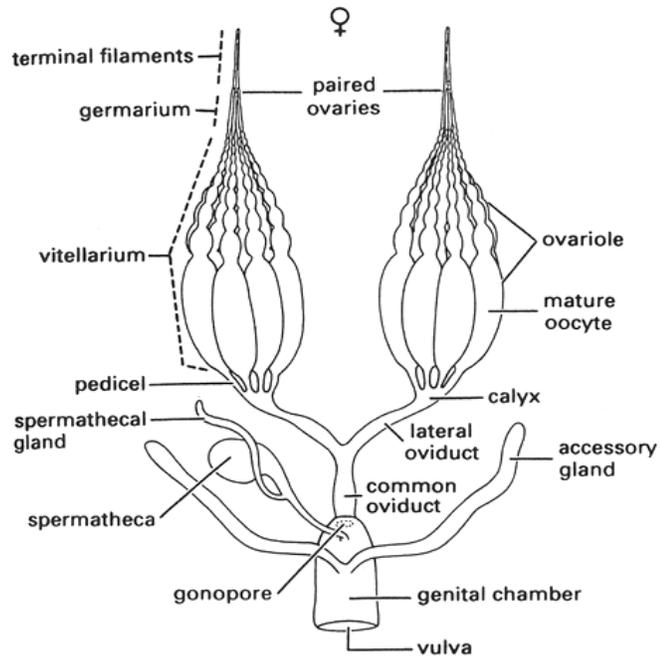
Figure 2: Generalized female reproductive system of an insect. (Snodgrass, 1935)
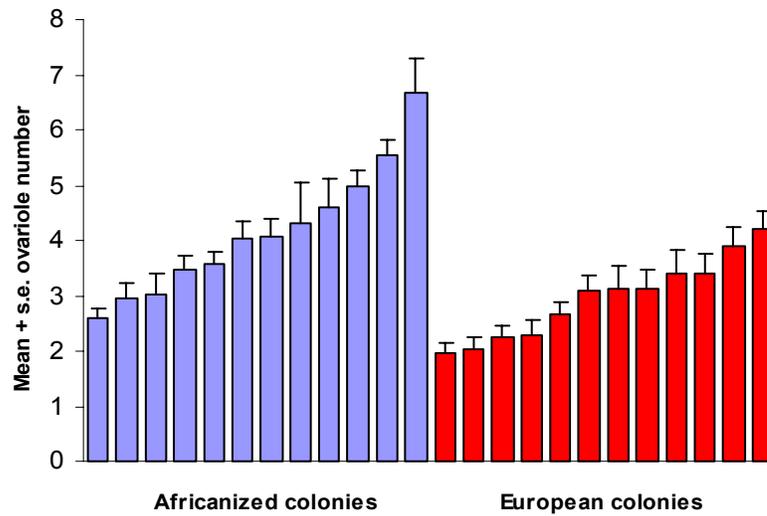
Figure 3: Mean worker ovary size (ovariole number) in the Africanized and European source colonies. Twenty workers were dissected per colony. Means and s.e. were 4.15 ± 0.34 and 2.29 ± 0.21 for Africanized and European genotypes, respectively.
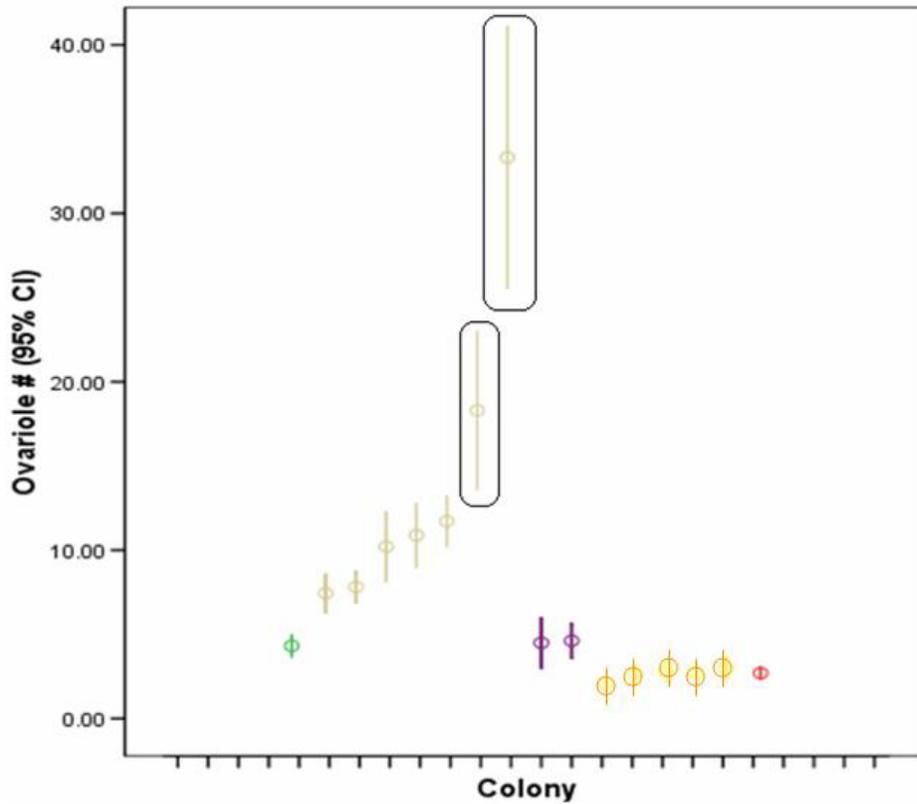
Figure 4: Total ovariole number for the parental Africanized honeybee colony (1), Africanized backcross colonies (2-8), Hybrid colonies (9-10), European backcross colonies (11-15), and parental European honeybee colony (16). The encircled colonies are the two Africanized backcross R3 and R5 used as sources for the genetic analyses.
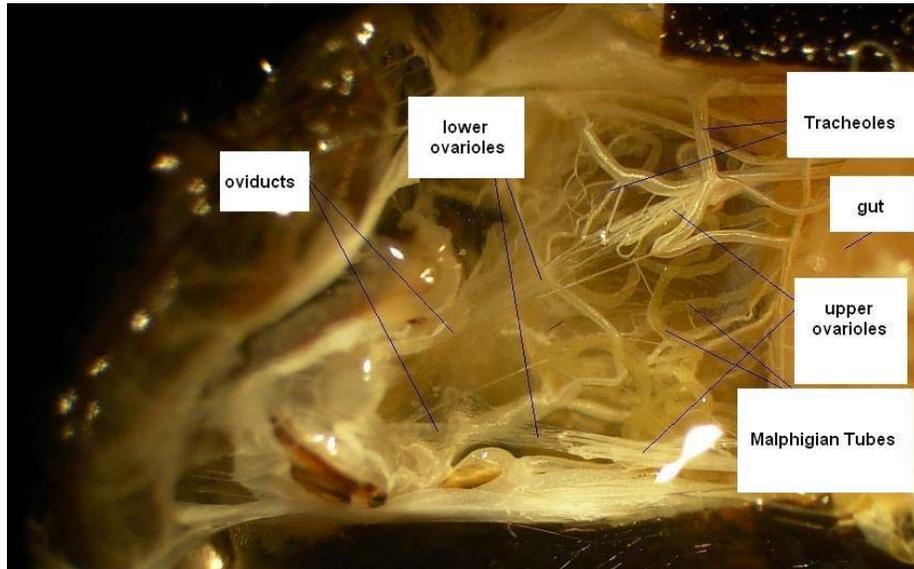
Figure 5: Labeled interior of the abdomen of a dissected honeybee, *Apis mellifera* (photo courtesy of O. Rueppell).
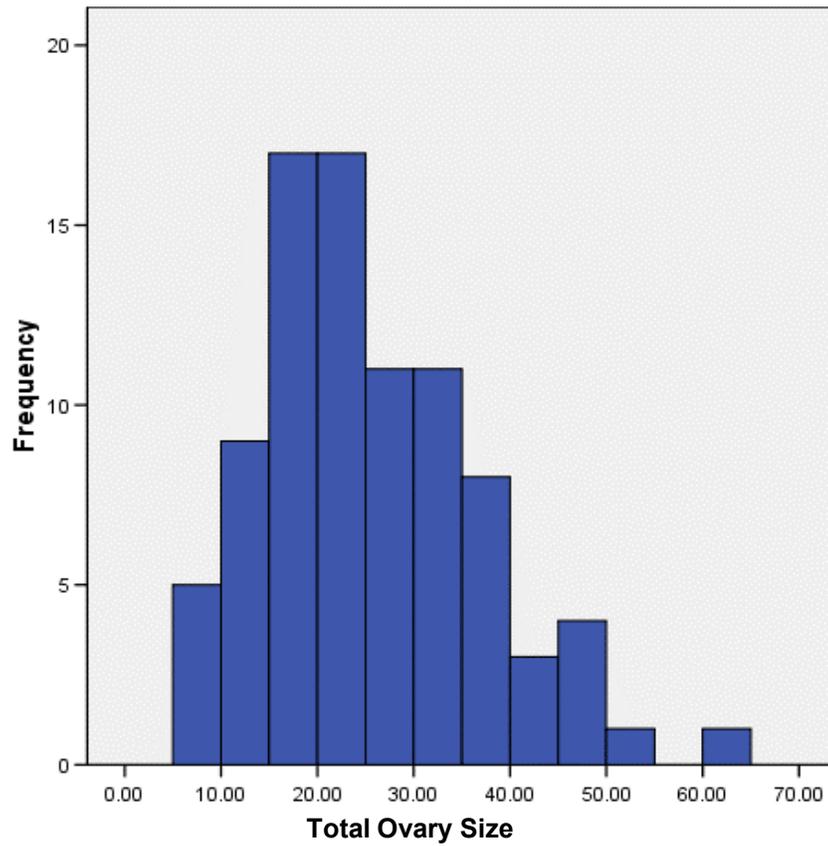
Figure 6: Total ovary size distribution within the Africanized backcross R3. Mean is 25.15, Std. Dev. Is 11.35, n = 87.
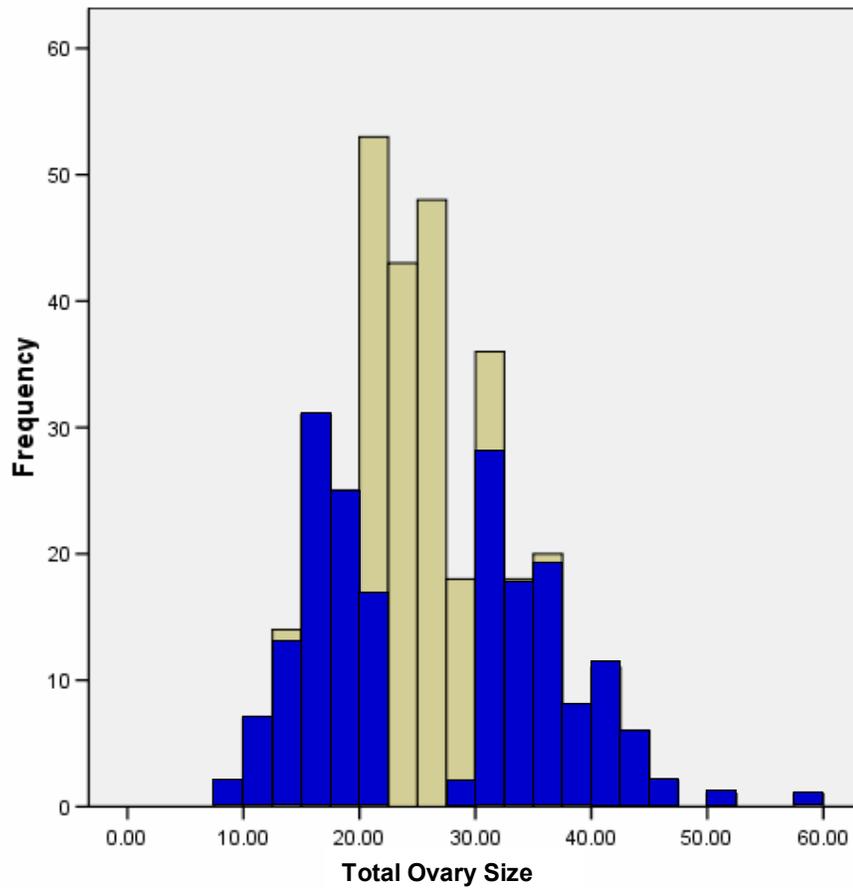
Figure 7: Total ovary size distribution within the entire Africanized backcross R5. The blue shaded bars represent the individuals selectively genotyped. Mean is 25.51, Std. Dev. Is 8.04, n = 344.
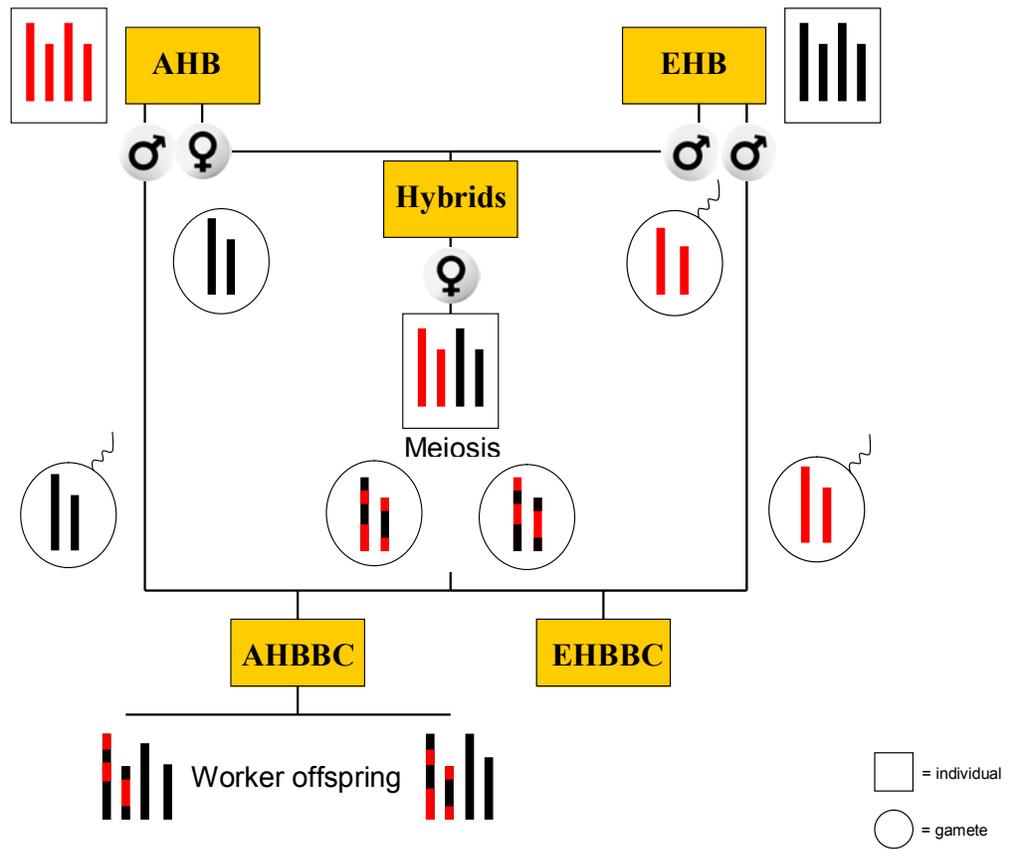
Figure 8: Pattern of inheritance of grandparental genomic material
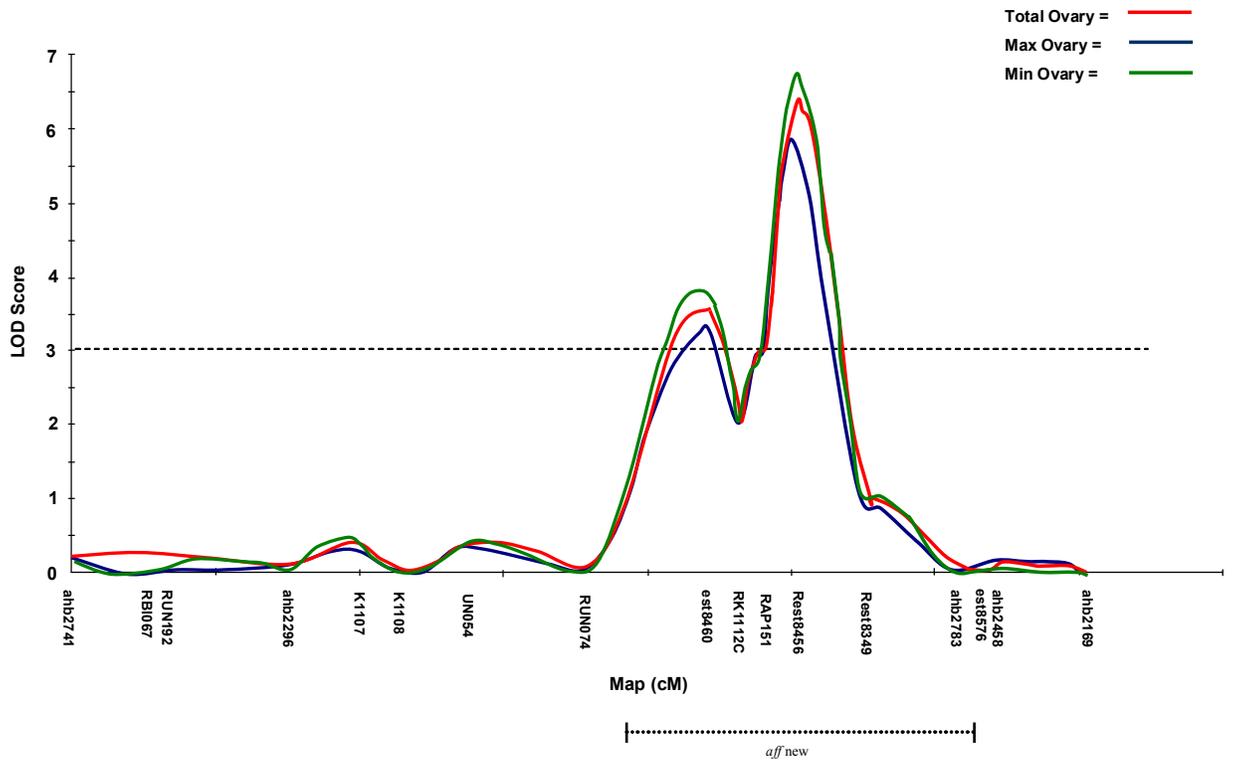
<u>Figure 9</u>: Interval Mapping for Chromosome 11, with LOD score traces for total ovary size, maximum ovary size, and minimum ovary size in the Africanized backcross R3. The genome wide threshold of LOD 3.0 is labeled as a dashed line. Overlap with a previously identified behavioral QTL (*aff* new) is labeled with a dotted line below the x-axis.
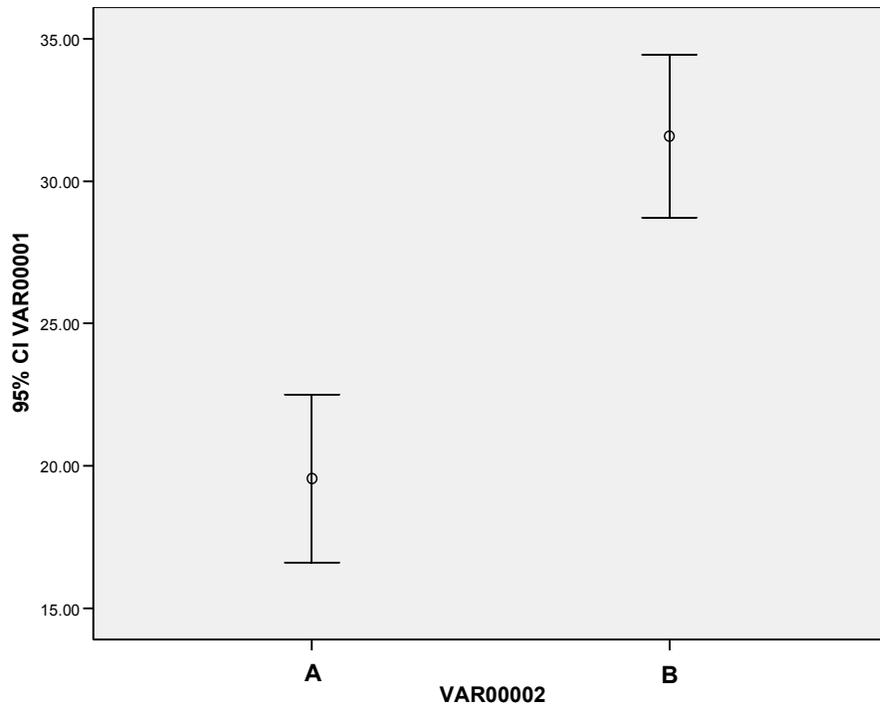
Figure 10: Effect size on total ovariole number of SNP marker est8456 on Chromosome 11, in R3
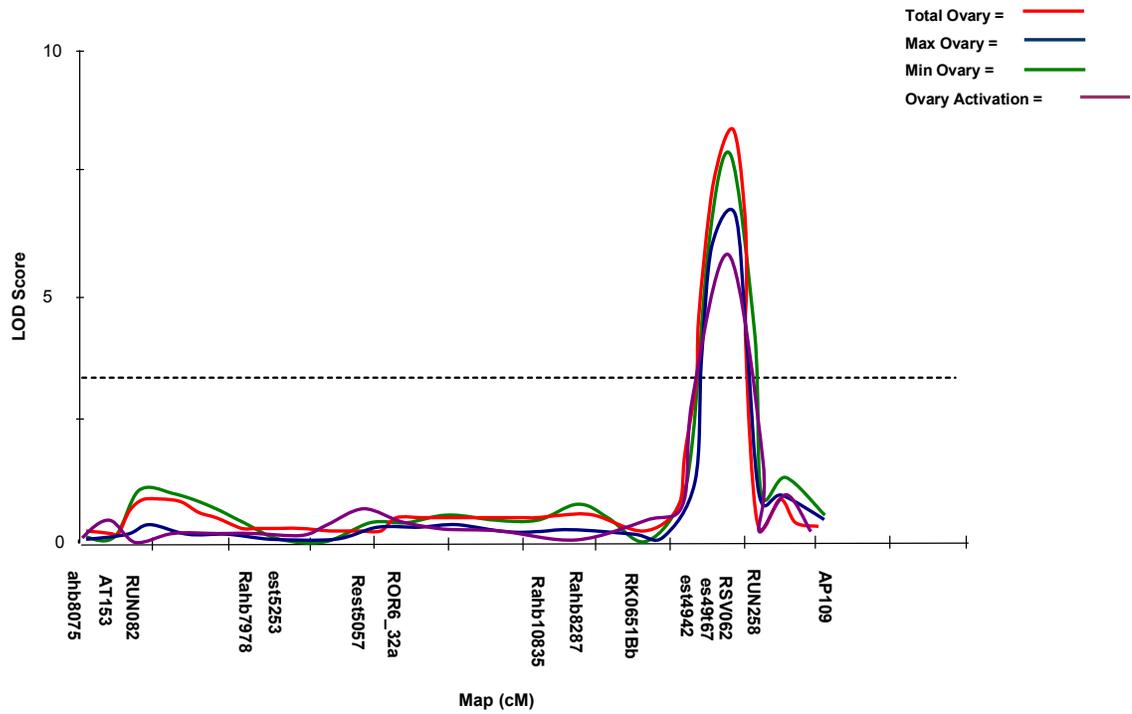
Figure 11: Interval Mapping for Chromosome 6, with LOD score traces for total ovary size, maximum ovary size, minimum ovary size and ovary activation in the Africanized backcross R5. The genome wide threshold of LOD 3.0 is labeled as a dashed line.
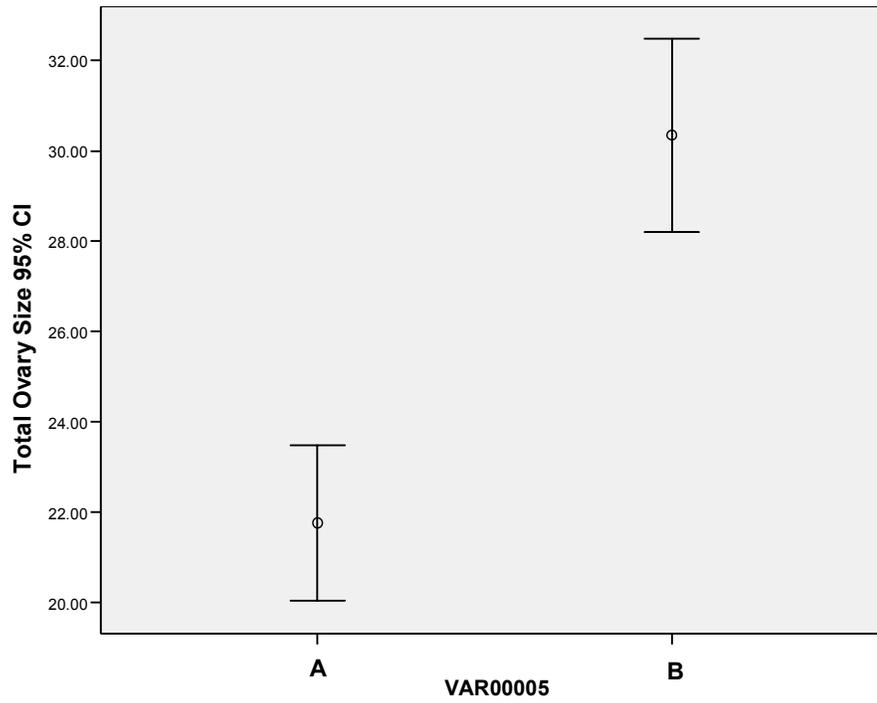
Figure 12: Effect size on total ovariole number of SNP marker est4967 on Chromosome 6, in R5.
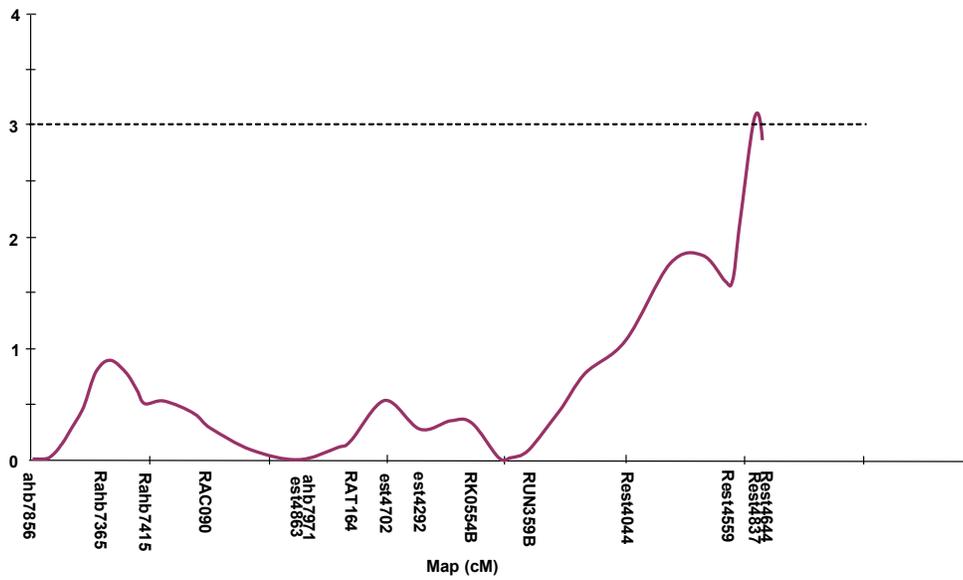
Figure 13: Interval Mapping for Chromosome 5, with LOD score for ovary asymmetry, in the Africanized backcross R5. The genome wide threshold of LOD 3.0 is labeled as a dashed line.
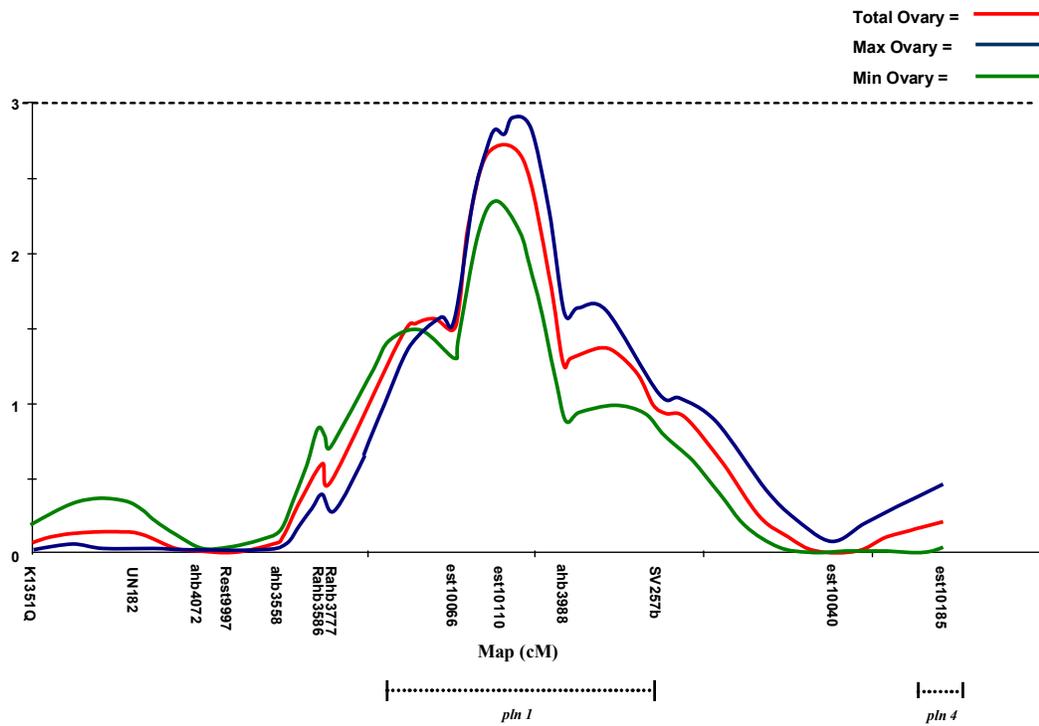
Figure 14: Interval Mapping for Chromosome 13, with total ovary size, maximum ovary size and minimum ovary size, in the Africanized backcross R5. The chromosome wide threshold of LOD 3.0 is labeled as a dashed line. Overlap with two previously identified behavioral QTLs (*pln1*, *pln4*) is labeled with a dotted line below the x-axis.