Genetic analysis of *Nosema* tolerance in the honey bee, *Apis mellifera*

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Naturwissenschaftlichen Fakultät I-

Biowissenschaften

der Matrin-Luther-Universität Halle-Wittenberg

von

Herrn Qiang Huang

geb.am 21.01.1983 in Jiujiang

Gutachter / in:

1. Prof. Robin F.A. Moritz

2. Prof. Ingemar Fries

3. Prof. Dirk de Graaf

Promotionsgesuch eingereicht am:

Tag der öffentlichen Verteidigung: 30-07-2013

Table of Content

1 INTRODUCTION 1

1.1 Honey bee pathology 1

1.1.1 The role of the honey bee's life history for its pathology 1

1.1.2 The role of beekeeping for honey bee pathology 2

1.2 Honey bee immunity 3

1.2.1 Honey bee social immunity 3

1.2.2 Honey bee innate immunity 4

1.3 *Nosema*, a gut parasite of honey bees 6

1.3.1 Biology of *Nosema* 6

1.3.2 Two *Nosema* species in the honey bee 7

1.4 Identification of Quantitative Trait Locus (QTL) 8

1.4.1 QTL mapping procedure 8

1.4.2 Advantage of using drones for the QTL mapping 9

1.5 Selective breeding for the *Nosema* resistant honey bees 10

1.5.1 Selective sweep and background selection 10

1.5.2 *Nosema* resistant honey bees 10

1.6 Aim of this work 11

2 MATERIAL AND METHOD 14

2.1 Honey bee drones 14

2.1.1 Drones used to quantify the innate immune response and survival towards *N. ceranae* infection 14

2.1.2 Drones used to map the QTLs associated with the *Nosema* tolerance 14

2.1.3 Drones used to trace the selective 'foot print' 15

2.2 *N. ceranae* infection 16

2.2.1 *N. ceranae* extraction 16
2.2.2 *N. ceranae* infection to determine the innate immune response and survival 16
2.2.3 *N. ceranae* infection to map the QTLs 17

2.3 DNA extraction and genotyping 18
2.3.1 DNA extraction 18
2.3.2 Genotyping 18

2.4 Gene expression analysis 18
2.4.1 RNA extraction, cDNA synthesis and qPCR procedure 18
2.4.2 Innate immune genes and candidate gene for the qPCR 19

2.5 QTL mapping procedure 23
2.5.1 Marker selection 23
2.5.2 Phase determination 23
2.5.3 Major QTL analysis 23
2.5.4 Additive and epistatic QTLs analysis 24

2.6 Statistics 24
2.6.1 Spore load dynamics and the survival towards the *N. ceranae* infection 24
2.6.2 Relative gene expression quantification 25
2.6.3 Population genetic diversity statistics 25

3 RESULTS 29

3.1 Survival and innate immune responses towards *N. ceranae* infection 29
3.1.1 Spore load dynamic and survival 29
3.1.2 Reference genes 31
3.1.3 Expression levels of innate immune genes of Toll pathway 31
3.1.4 Expression levels of innate immune genes of IMA and JAK/STAT pathway 32

3.2 QTLs associated with the *N. ceranae* spore load 34
3.2.1 Interval mapping 34
3.2.2 Fine mapping 34
3.2.3 Additive and epistatic QTLs analysis 40
3.2.3 Candidate genes 43

3.3 Selective sweep within the QTL_m region 49
3.3.1 Expected heterozygosity (H_e) 49
3.3.2 Effective number of alleles (A_e) 52
3.3.3 Selective sweep 52

4 DISCUSSION 54

4.1 Responses to the *N. ceranae* infection in drones of selected strain comparing to the unselected strain 54
4.1.1 Responses of innate immune genes 54
4.1.2 Spore load dynamic and survival 55

4.2 QTLs associated with the tolerance towards *N. ceranae* 57
4.2.1 Major QTL in Chromosome 14 57
4.2.2 Epistatic effects 57
4.2.3 Candidate gene 58

4.3 'Foot print' of the selective breeding 60
4.3.1 Selective sweep 60
4.3.2 Combination of colony level and individual level analysis of the tolerance towards *Nosema* 61

5 SUMMARY AND OUTLOOK 63

6 REFERENCE 65

7 APPENDIX 78

7.1 Acknowledgements 78

7.2 Curriculum vitae 79

7.3 Publication list 80

7.4 Oral presentations 81

7.5 Erklärung 82

1 Introduction

1.1 Honey bee pathology

1.1.1 The role of the honey bee's life history for its pathology

The honey bee *A. mellifera*, is an important economical insect not only for honey production but also for pollination services for crops in natural ecosystems (Ish-Am and Lahav, 2011; Breeze et al., 2011). The honey bee is a eusocial insect, which lives in large colonies composed of a single reproductive queen, hundreds of drones in the mating season (spring) and tens of thousands of workers (Seeley 1995). The queen and the workers develop from fertilized eggs whereas the males are hemizygous and develop from haploid unfertilized eggs. The drones are usually only produced during the reproductive swarming and mating season (spring in temperate Europe). Although the drones do not engage in any specific behavior other than mating, their haploid genome provides a powerful tool for genetic analyses. Genes expressed in drones can be easily analyzed because the haploid genome is free of interactions between two homologous chromosomes in diploid females (Moritz and Evans, 2008).

Within the honey bee colony, thousands of individuals live in a homeostatic nest environment. Sociality increases fitness of individuals, including cooperative brood care, defense of the enemy and the ability to modify the micro-environment (Moritz and Southwick, 1992; Wilson-Rich et al., 2009). But the high density also increases the risk of disease out-breaks. Any pathogens can easily spread within the colonies due to the intensive contacts among individual honey bees. Moreover the high colony temperature and humidity provides ideal conditions for many bacterial and fungal pathogens. (Schmid-Hempel 2005, 2009; Koch and Schmid-Hemepl 2011; Li et al. 2012)

1.1.2 The role of beekeeping for honey bee pathology

Beekeeping (apiculture) is the management of honey bees in hives. A beekeeper keeps honey bees in order to collect honey (and other products such as beeswax, propolis, pollen, royal jelly) and pollinate plants. The honey bee colonies can be either transported to different locations following the flowering time (migratory beekeeping) or fixed at one location (stationary beekeeping). Irrespective the two different ways to manage the honey bees, the colonies were usually located at a very high density (apiary), which facilitates the transmission of the pathogens and parasites among colonies.

With recently large honey bee colonies deaths (colony collapse disorder, CCD), unprecedented efforts have been invested to improve the honey bee health status from multiple perspectives (Cox-Foster et al., 2007; Ribiere et al., 2008; Anderson and East, 2008; Johnson et al., 2009; Dainat et al., 2012). Multiple factors were believed to cause the honey bee death, among which pathogens and parasite were believed to be the main impetus to drive the colony loss in the managed honey bees (Cox-Foster et al., 2007).

The “Short Sighted Selection Hypothesis” suggests a trade-off between virulence and transmission of pathogens (Frank, 2012). Pathogens with high virulence are expected to have a reduced transmission and vice versa. This trade-off is important to develop the balancing host-pathogen co-evolutionary relationship, which is important for the long term survival of both host and pathogen. However, apiculture breaks this trade-off and will select for high virulence since colonies are located at very high densities and the beekeeper may even facilitate pathogen spread by switching combs among colonies. So transmission will not be selected for, which gives high virulence pathogens a selective advantage. The highly virulent pathogen could be easily picked

up by the wild honey bees by foraging the polluted followers and the robbing between bees. The infected wild honey bees will bring the pathogens back to the hive to infect more individuals. Hence, the pathogens could easily spill over to wild honey bee populations.

1.2 Honey bee immunity

1.2.1 Honey bee social immunity

The honey bees are not defenseless against pathogens. In order to prevent disease outbreaks, they evolved two levels of immune systems: the social immune system (colony level) and the individual immune system (individual level). At the colony level, workers can identify and remove the infected larvae (even adults) to prevent the disease spread known as hygienic behavior. Rothenbuhler (1964) was the first to interpret the heredity of the complex social behavior - hygienic behavior. A resistance strain and a susceptible strain towards *Bacillus larvae* which causes American foulbrood disease were used. He crossed the two trains to produce the F1 hybrid queens to test the ability of removing *Bacillus larvae* infected larvae. Then he further backcrossed the F1 drones with two parental strains respectively. Based on the observed proposition of removed larvae, a two loci model of hygienic behavior was proposed. One locus stands for the uncapping ability and the other stands for the removing ability. Recessive alleles r for removing ability and u for uncapping ability were postulated that the homozygous workers of both loci (r/r, u/u) performing compete hygienic behavior. Homozygous at one of the two loci, the workers should express the corresponding phenotype of the homozygous locus. Heterozygous at both loci, the workers do not perform hygienic behavior. Rothenbuhler's work has become the classic study for the complex social behavior study. Moritz (1988) re-analyzed the

data from Rothenbuhler (1964) and conceded that the hereditary pattern of the hygienic behavior better matched a three-locus model. By re-analyzing the variance among the tested colonies in more detail, the hygienic behavior is in generally unlikely to follow simple Mendelian segregations. The more recently QTL (quantitative trait locus) studies of hygienic behavior confirmed a multifactorial inherited way and at least six quantitative trait loci were association with the hygienic behavior (Oxley et al. 2010).

1.2.2 Honey bee innate immunity

1.2.2.1 Cellular Immunity

Cellular defense refers to blood cells (hemocytes) mediated immune responses against pathogens. Hemocytes recognize the invading pathogens either by direct contact with the pathogen with surface receptors or indirectly recognizing the humoral receptor binding to the surface of the pathogen. Plasmocytes and granulocytes are the major hemocytes. They are capable of adhering to the pathogen surface to react by phagocytosis, nodulation and encapsulation (Strand and Pech, 1995). Phagocytosis refers to the engulfment of biotic target (e.g. bacteria, yeast) and small abiotic particles by a single hemocyte cell. Nodulation refers to multiple hemoctye cells binding to aggregation of bacterial and encapsulation refers to binding of hemocytes cell to large targets. The other hemocytes are non-adhesive spherule cells, oenocytoids and prohemocytes. Spherule cells have been suggested to transport cuticular components (Sass et al., 1994), while oenocytoids contain cytoplasmic phenoloxidase precursors that likely play a role in melanization of hemolymph (Jiang et al., 1997). Prohemocytes are stem cells which differentiate into one or more of above hemocyte cell types. Hemocytes first developed during embryogenesis from

mesoderm. The honey bee continuously produces hemocytes during larval stage via cell division in mesodermally derived organs (Snodgrass 1956).

1.2.2.2 Humoral Immunity

Humoral defenses include the synthesis and secretion of anti-microbial peptides (AMP), reactive intermediates of oxygen or nitrogen and complex enzymatic cascades that regulated coagulation or melanization (Muta and Iwanaga, 1996; Gillespie et al., 1997; Bogdan et al., 2000; Vass and Nappi, 2001). AMPs could be synthesized in fat body and released into hemolymph. AMPs could also be synthesized by epithelial cells of the gut to protect the gut from the pathogens (Tsakas and Marmaras 2010). AMPs are synthesized upon activation of one or both of Toll pathway and IMD pathway (Evans 2006). The IMD and Toll pathways recognize the peptidoglycan which is a constituent of the bacterial and fungi cell wall by peptidoglycan recognition protein. The signal would be transmitted inside the cell to increase the expression of the *Relish* and *Dorsal* which regulate and production of the AMP and melanizing agent phenoloxidase (PO). The AMPs are small peptides which bind bacterial and fungal membranes leading to the destruction of the pathogen. For example Defensin could destroy the bacterial and fungi by forming channels in the plasma membrane which leads to the cell lysis. Prophenoloxidase is an important immune protein which is produced by the hemocytes. The microbial infection activates the conversion of Prophenoloxidase to phenoloxidase (PO) which induces the production of melanin (melanization) around the microorganisms to first trap them off from the circulation and then kill them (Shao et al. 2012).

1.3 *Nosema*, a gut parasite of honey bees

1.3.1 Biology of *Nosema*

The Microsporidia constitute a phylum of spore forming unicellular organisms with an obligate parasitic life history (Keeling and Fast 2002; Williams 2009). Mirosporidia are restricted to animal hosts and mostly infect insects. In honey bees, *Nosema* (Figure 1) infects the epithelial cells of the gut (Zander, 1909) where is lives as an obligate intracellular parasite. The infection starts from the ingestion of the spores, which germinate in the mid-gut lumen and extrude the polar tubes that penetrate the epithelial cells to release the sporoplasm into the cytoplasm (Zander, 1909; Higes et al., 2007; Gisder et al., 2011). The intracellular life cycle has two phases, the proliferative phase (merogony) and the sporogonic phase (sporogony), and ends with the formation of spores (Figure 2). The infected cells eventually burst and release a new generation of spores (Graaf et al., 1994). The offspring spores can either germinate to infect new host cells or can be expelled through the feces. The mechanisms which control the germination of the spores are unknown.

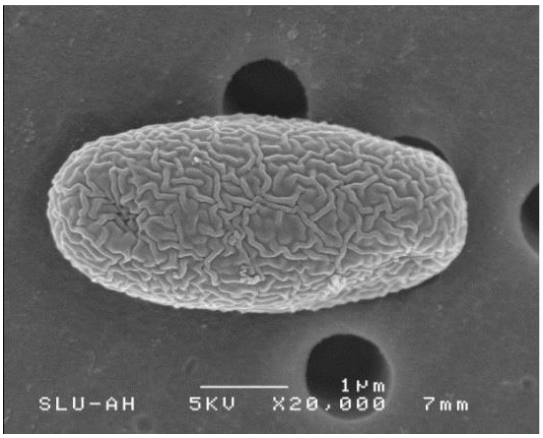


Figure 1 *Nosema* spore under the electric microscope

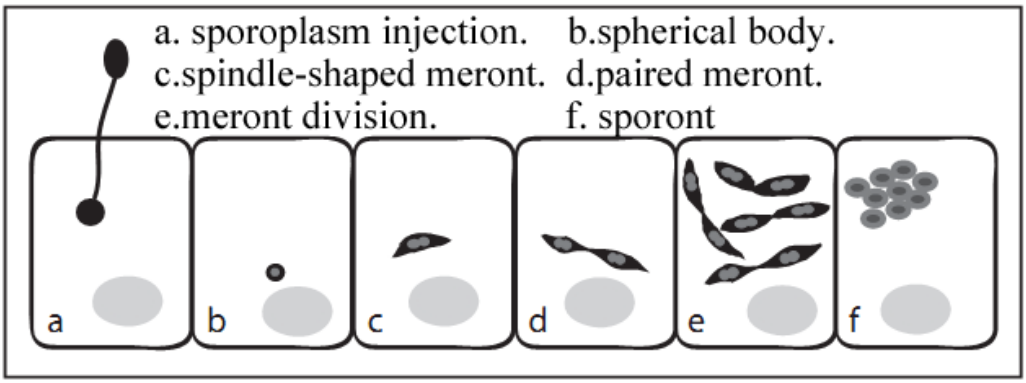


Figure 2 *Nosema* life cycle (Gisder et al. 2011)

1.3.2 Two *Nosema* species in the honey bee

There are two *Nosema* species known to infect the honey bee *A. mellifera*: *Nosema apis* (*N. apis*) and *Nosema ceranae* (*N. ceranae*). *N. apis* is an evolutionarily old pathogen of the honey bee *A. mellifera* with a moderate virulence. The host-parasite co-evolutionary relationship is well balanced and colonies can often cure themselves under favorable environmental conditions (Zander, 1909; Klee et al. 2007; Chen et al., 2009). *N. ceranae* was first found in the Asian honey bee *A. cerana* (Fries et al., 1996). It is a newly established parasite of *A. mellifera* (Higes et al., 2006; Fries et al., 2006). The parasite *N. ceranae* has been reported to be highly virulent and may have caused large colony losses in Spain (Higes et al., 2008; Higes et al., 2009). However, there are increasing reports suggesting a moderate virulence of *N. ceranae* similar as that of *N. apis* (Gisder et al., 2010; Forsgren and Fries, 2010; Fries, 2010). The specific *N. ceranae* strain and honey bee strain used for the virulence experiment might be the reason for the discrepancy, as the same *N. ceranae* showed moderate virulence in French bees, but killed Spanish bees (Dussaubat et al. 2012, 2013).

It has been reported that the *Nosema* infection impeded the honey bee foraging behavior, nursing ability and reducing its life expectancy (Hassanein 1951, 1952; Mayack and Naug, 2009). Furthermore, the infection of *N. ceranae* suppressed the innate immune response in adult honey bees (Antúnez et al. 2009; Huang et al., 2012). The destroyed peritrophic membrane due to *N. ceranae* infection makes the infected honey bees more vulnerable to other pathogens (Dussanbat et al. 2012; Higes et al. 2013).

1.4 Identification of Quantitative Trait Locus (QTL)

Classically, a quantitative trait locus (QTL) is a region of DNA which is associated with a quantitative trait. The trait could follow Mendelian inheritance or multifactorial inheritance (Cook 1908; Mackay and Fry 1996; Han and Denlinger 2009). QTL mapping is based on the linkage of genetic markers with a specific trait, which forms the initial step to locate the functional genes on the genome (Buzzatitraverso 1959; Barton et al. 2009). In the honey bees, QTL mapping has been widely used to address both quantitative traits and complex social behavior (Rothenbuhler 1964; Hunt et al. 1989; Rüppelll et al. 2004; Ammons et al. 2008; Rüppelll 2009; Oxley et al. 2010).

1.4.1 QTL mapping procedure

In general, there are two steps to identify the QTLs. The first step is to identify the major QTL without considering any interaction effects. The second is to identify the interaction effects of QTLs, such as epistatic effects and additive effects.

The simplest method for the analysis of major QTL mapping is marker regression which considers each marker individually (Weller et al. 1987). The advantage of the marker regression is its simplicity where an ANOVA is performed for each marker

genotype and covariates can be easily incorporated. But the individual with missing marker genotypes was omitted. The interval mapping improves the marker regression by taking account the missing genotype information (Oxley et al. 2010). The interval mapping is now the most commonly used method for the QTL mapping, in which one posits the presence of a single QTL and considers each location at a time as the putative QTL (Bromam et al., 2003). With the interval mapping, the major QTL could be identified without considering any interaction effects. The second step is searching epistatic and additive QTLs. The principle of this method is to pairwise analyze all the markers and estimate interaction effects (Behrens et al. 2012).

1.4.2 Advantage of using drones for the QTL mapping

In the case of *Nosema* tolerance, we can take advantage of haploid drones for QTL mapping because the trait is also expressed in the males. The male haploid genetic system is particularly suited for QTL mapping studies as interpretational problems resulting from dominance interactions between alleles on different homologous chromosomes cannot occur (Behrens et al., 2011). Epistatic effects are generally difficult to reveal and typically a very large sample size is required because the effect of one locus depends on the interaction with another one (Noor and Chang 2009; Gertz et al. 2010). In diploid organisms already the interactions between of two homologous alleles at the same locus are in itself complex, making the quantification of epistatic effects exceptionally sample size intensive. The haploid drones fortunately provide us with an extremely simple and highly efficiently model system to dissect genetic interaction among loci.

1.5 Selective breeding for the *Nosema* resistant honey bees

1.5.1 Selective sweep and background selection

Selective breeding is widely used to achieve a desired trait such as parasite resistance. During the selection program, beneficial alleles (either from the standing variation or as novel mutations) rise in frequency in the breeding strain. The beneficial allele can eventually be fixed in the breeding population. The fixation will reduce the variability of neighboring neutral loci linked to the selected locus, a process interpreted as genetic hitchhiking which results from selective sweep (Maynard and Haigh 1974; Berryman, 1992; Kim and Stephan 2002; Palaisa et al., 2004; Chevin and Hospital, 2008). Alternative, a deleterious allele could be selected against in the breeding population. The genetic variability of the surrounding neutral loci should also be reduced, known as background selection (Charlesworth et al. 1993; Charlesworth 1996). The background selection is a form of genetic hitchhiking as both selective sweep and background selection resulting changes in the allele's frequency due to selection operating upon linked locus. The difference of the two is that the selective sweep could result in an extremely low number of alleles of the selected locus, such as single effective allele could be observed (fixation). For the background selection, the decreased genetic diversity also could be observed, but fixation is less likely (Stephan 2010).

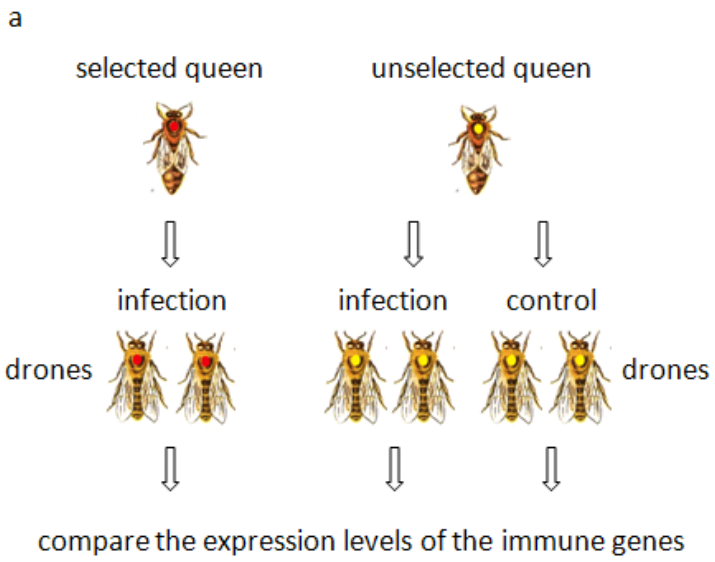
1.5.2 *Nosema* resistant honey bees

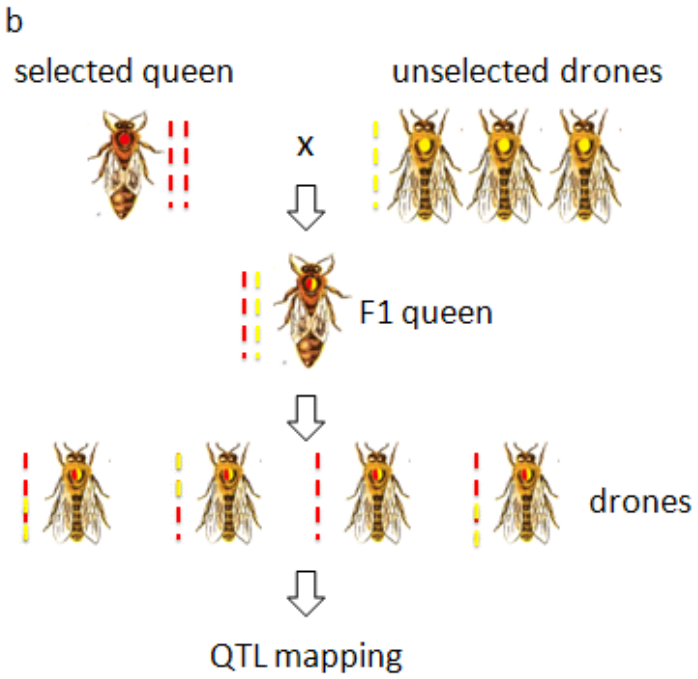
Nosema adds to the pathogen load in honey bee colonies and reduces their productivity (Haseman 1951, 1952; Kralj and Fuchs 2010). This is why beekeepers in Denmark embarked on a selection program to obtain *Nosema* resistant honey bees since 1980s (Traynor, 2008). Originally, 500 colonies (*A. mellifera*) were involved in

the selective breeding program which reflects an effective population size of about N_e = 1800 (Kerr 1975; Owen and Owen 1989) given each queen is mated with an average of about 12 males (Schlüns et al. 2005). Queens were naturally mated on an island and the colonies were checked every year for *Nosema* infections. The worker honey bees were used to determine the *Nosema* infection levels. The queens were replaced whenever the workers were infected by *Nosema* with queens from *Nosema* free colonies. Hence, the selection was performed at the colony level without knowing which biological mechanism drove the lack of *Nosema* infection in the sampled workers. Beekeepers could have selected a resistant allele (selective sweep) or removed the susceptible allele (background selection) from the breeding population. Nevertheless, this selective breeding should have left "foot prints" if one or few major genes determine the resistance phenotype.

1.6 Aim of this work

Due to selective breeding, the *Nosema* is now rarely found in honey bee colonies from the breeding strain, which provides us a unique opportunity to study the genetic basis of *Nosema* tolerance, I here (1) Screen how the innate immune system of the selected honey bees negotiate with this novel pathogen (Huang et al. 2012) (Figure 3 a). (2) Identify the QTLs associated with the *Nosema* tolerance and search the candidate genes (Huang et al. in revision) (Figure 3 b). (3) Track the selective "foot print" by comparing the genetic variability within and flanking the identified QTL region of the selected honey bees comparing to an unselected honey bee population (Huang et al. in revision) (Figure 3 c).


a
selected queen
unselected queen
infection
infection
control
drones
drones
compare the expression levels of the immune genes


b
selected queen
unselected drones
x
F1 queen
drones
QTL mapping

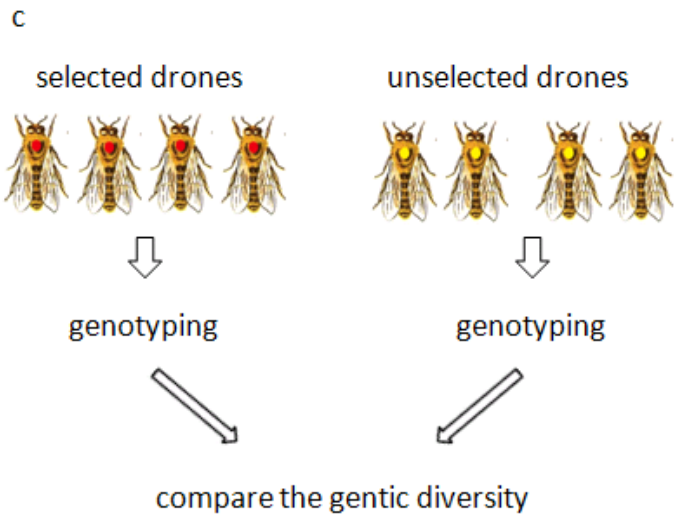


Figure 3 The experimental setup of this work. a: the drones of the selected and unselected strains were infected with *N. ceranae* spores to compare the differential expression levels of the innate immune genes (Huang et al. 2012). b: the selected queen was inseminated with the unselected drones to rear the hybrid F1 queen. The drone offspring of the F1 queen were individually infected with the *N. ceranae* spores to identify the QTLs associated with low spore load towards the *N. ceranae* infection. We then searched the candidate genes within the identified QTLs regions. The two red vertical lines represent diploid genome of the queen and the one yellow vertical line represents the haploid genome of the drone (Huang et al. in revision). c: drones from the selected and unselected population were sampled respectively. These drones were genotyped within and flanking the identified QTL region to compare the genetic diversity to trace the selective foot print (Huang et al. in revision).

2 Material and Method

2.1 Honey bee drones

2.1.1 Drones used to quantify the innate immune response and survival towards *N. ceranae* infection

The same age two sister virgin queens (*A. mellifera*, Danish strain) selected for the *Nosema* resistance were provided by the Department of Integrated Pest Management Research Centre Flakkebjerg, Denmark. The queens (*A. mellifera*, German strain) that had not been unselected for *Nosema* resistance were kept at the apiary of the Martin-Luther-University Halle-Wittenberg served as controls. All queens were introduced into small colonies composed of ~ 2000 newly emerged workers. The queens were treated with CO_2 to initiate ovary activation (Mackensen, 1947) without mating so that they exclusively produced haploid drone offspring (Mackensen, 1951). We decided to screen the immune response of haploid drones rather than that of diploid workers because of their simple hemizygous genetic design which is free of allelic dominance interactions that might complicate any gene expression patterns of diploids. The drones were reared in full sized drone frames. Frames of sealed drone brood and the worker brood were kept in an incubator (34±1 ℃). Freshly emerged drones of the selected queens and unselected queens were collected daily from the brood frames to provide *Nosema* free, age standardized individuals (0-24h) for the artificial infection experiments. Freshly emerged workers of unselected queens in the apiary were served as nurse bees.

2.1.2 Drones used to map the QTLs associated with the *Nosema* tolerance

Ten virgin queens of the selected Danish strain were artificially inseminated with the same mixed sperm (Moritz 1984) of thirty drones of an unselected strain (French

strain) kept at the Laboratoire de Biologie et Protection de l'abeille, INRA Avignon, France. The inseminated queens were introduced into small colonies composed of ~4000 freshly emerged workers at the apiary of the Martin-Luther-University Halle-Wittenberg, Germany. We reared F1 hybrid queens from the inseminated queens and treated the F1 queens with CO_2 to initiate ovary activation without mating so that the F1 queens exclusively produced unfertilized eggs developing into drones. The drones were reared in full sized drone frames. We decided to use the haploid drones for QTL mapping rather than diploid workers, because the male haploid genetic system is particularly suited in QTL mapping studies as problems resulting from interactions between gene effects of different homologous chromosomes cannot occur. A single F1 queen who first started to lay eggs was chose to reproduce the QTL mapping population.

2.1.3 Drones used to trace the selective 'foot print'

Ten drones per colony were collected from ten hives from the selected Danish honey bee strain. Moreover, 32 drones were collected from the unselected control population at a drone congregation area (DCA), Germany. Sampling drones on a DCA, removes the bias of sampling specific breeding lines at apiaries since the drones originate from many different apiaries surrounding the DCA in a range of 2 km (Kraus et al. 2005). Hence the population sample is less apiary biased and better reflects an unselected population that might even include feral colonies. The Danish sample could not be sampled in the same way because we explicitly needed to have the sample of the specific breeding strain. The drone genotypes were used to infer the mother queen genotypes. These queen alleles were used to determine the allele frequencies of the population.

2.2 *N. ceranae* infection

2.2.1 *N. ceranae* extraction

Fresh *N. ceranae* infected workers were provided by Laboratoire de Biologie et Protection de l'abeille, INRA Avignon, France. The abdomens of infected workers were homogenized in distilled water, filtered through filtering paper and centrifuged at 3220G for 10 min. The pellet was re-dissolved in distilled water and centrifuged at 8700G for 5 min to purify the *N. ceranae* spores. Spores were counted using a Fuchs-Rosenthal haemocytometer and the *Nosema* species was verified by a standard PCR protocol (Hamiduzzaman et al., 2010).

2.2.2 *N. ceranae* infection to determine the innate immune response and survival

Drone offspring (drones described in **2.2.1**) originated from the selected queens and unselected queens were individually fed with 2 µl sucrose solution containing ~10^5 *N. ceranae* spores. Drones that did not consume the entire volume were discarded. For each experimental group, 45-50 drones and 20 uninfected nurse workers were housed in a wooden cage (depth13cm x width10cm x height11.5cm) at 34±1 ℃, 60% rel. hum. Three cage replicates were conducted for selected strain, unselected strain and control group respectively. Drones of the unselected strain received 2µl sucrose solutions without any *Nosema* spores served as the control group. Drones and workers were fed with 50% sucrose solution *ad libitum* without pollen during the remaining time of the experiment.

Three drones were sampled from each cage every 24 hours. Dead bees were removed and recorded until day eight post infection. The mid-guts of all sampled drones were

removed and homogenized in 500µl distilled H_2O individually. 10µl of this solution was used to count *N. ceranae* spores in a Fuchs-Rosenthal haemocytometer.

2.2.3 *N. ceranae* infection to map the QTLs

Freshly emerged drones (drones described in **2.1.2**) originated from the single F1 queen were collected daily from the brood frames to provide age standardized individuals (0-24h). Freshly emerged workers from brood frames kept in the incubator served as nurse bees. Freshly emerged drones of F1 queen were individually fed with 2µl sucrose solution containing ~10^5 *N. ceranae* spores. Drones that did not consume the entire solution were discarded. Infected drones and uninfected nurse workers were housed in a wooden cage (depth 13.0cm x width 10.0cm x height 11.5cm) at 34±1℃ and 60% rel. humidity. Drones receiving 2µl sucrose solution without any *Nosema* spores served as controls for the analysis of the candidate gene expression level. Drones and workers were fed with 50% sucrose solution *ad libitum* without pollen during the remaining time of the experiment.

In total, 319 drones were infected with *N. ceranae*. All the drones were sampled on day six post infection to allow *Nosema* to complete the first reproduction cycle. Dead drones were removed and recorded daily. The guts of all sampled drones were individually removed and homogenized in 500µl distilled H_2O. 10µl homogenized solution was again diluted by 90µl distilled water to count the *N. ceranae* spores in a Fuchs-Rosenthal haemocytometer. By day six post infection, 64 out of the 319 infected drones had died. It is unclear the 64 drones were killed by the infection or the behavior of the workers in the same cage. As the *N. ceranae* requires ~4 days to complete the infection cycle. 55 out of 64 dead drones were dead before day four post infection, the dead drones were discarded from analysis. The mid-guts of the

remaining 255 drones were individually removed to determine the spore load (median & range, $5.6*10^5$ & $2.4*10^6$). No *N. ceranae* spores were found in the control groups. 76 drones with < $3.6*10^5$ spores (low; median: $8*10^4$, range: $3.4*10^5$) and 72 drones with >8.2 $*10^5$ spores (high; median: $1.1*10^6$, range: $1.7*10^6$) were selected for the QTL mapping (Figure 7). The number of spores were significantly different between these two groups (Mann-Whitney U test, P <0.001).

2.3 DNA extraction and genotyping

2.3.1 DNA extraction

The drones were preserved in 75% ethanol and stored at -20 ℃. The DNA was extracted from the leg with 5% Chelex 100 (Walsh et al. 1991) for the genotyping.

2.3.2 Genotyping

Fluorescence labeled microsatellite markers were used for the multiplex PCR. Each multiplex PCR reaction contained 1 µl DNA (50ng/µl), 5 µl master mix (Promega), 0.4 µl/primer (10mM, 6-8primers per reaction), adding water for a final volume of 10 µl. The PCR cycle was as follows: 95 ℃ for 15 sec; 55 ℃ for 30 sec; 72 ℃ for 30 sec for 40 cycles. The allele sizes were determined using MegaBace 1000 capillary DNA sequencer (Amersham Biosciences, Germany) and scored with MegaBace Fragment Profiler Version 1.2.

2.4 Gene expression analysis

2.4.1 RNA extraction, cDNA synthesis and qPCR procedure

The abdomens of drones were preserved in RNA later (Sigma-Aldrich) and stored at -80 ℃ until RNA was extracted with RNeasy Mini kit (Qiagen). RNA concentration

and quality (absorption ratio 260nm/280nm) were spectrophotometrically measured (Nanodrop 1000, Pequlab).

cDNA was synthesized by mixing 80ng of the RNA with 0.8 µl (0.5 µg/ µl) oligo-dT18 primer (Fermentas, St. Leon-Rot, Germany) in 10 µl RNAse free water and incubated 5min at 70 ℃. Next 3.4 µl 5xRT buffer (Promega, Mannheim, Germany), 0.8 µl dNTPs(10mM), 0.4 µl M-MLV Revertase (200u/ µl, Promega, Mannheim, Germany), 0.2 µl Ribolock RNAse inhibitor (40 u/ µl, Fermentas, St. Leon-Rot, Germany), 1.6DTT(0.1mM) were added and incubated at 42 ℃ for 2h and extending 15min at 70 ℃. Finally, the synthesized cDNA was diluted 1:10 with DEPC-water (DNase & RNase free water) for subsequent qPCR reactions.

1 µl of this diluted cDNA was added to 5 µl SensiMix of the SYBR & Fluorescein kit (SYBR-Green, Bioline, Luckenwalde, Germany), 0.2 µl of the specific gene primer (10mM) (table 1) and 3.6 µl DEPC water for qPCR assay. After an initial phase of 95 ℃ for 3min, the temperature cycle was as follows: 95 ℃ for 15sec; 60 ℃ for 30sec; 72 ℃ for 30sec for 40 cycles including plate read in every cycle. Finally melting curves were recorded by increasing the temperature from 50 ℃ to 90 ℃. Target genes and reference genes were run in the same 96 well plate to avoid plate variance. Each gene per sample was run in duplicate.

2.4.2 Innate immune genes and candidate gene for the qPCR

In order to quantify the responses of innate immune system towards the *N. ceranae* infection, 33 immune genes (Evans, 2006) covering all four innate immune pathways were amplified together with five reference genes (Yang and Cox-Foster, 2005; Van Hiel et al., 2009; Jarosch and Moritz, 2011) (Table 1) for the infected drones from day one to six post infection (drones described in **2.2.1**). The equal amount of RNA of

every three drones was pooled and three replicates were conducted for each day. In total, three pools of drones of selected strain, three pools of drones of unselected strain and three pools for the controls were constructed for each post infection day. Each pool contains three drones. The uninfected control group was used as standard to infer to which extend the infection regulate the expression level of the immune genes.

The expression level of the candidate gene *Aub* from the statistically significant major QTL region (NCBI, Map viewer, Amel 4.5) were quantified (drones described in **2.2.2**). Out of 148 drones used for the QTL mapping, RNA of 50 drones was individually extracted from the abdomen excluding the gut. Within the 50 drones, 25 drones (from 72 drones with high spore load) had high spore load and 25 drones (from 76 drones with low spore load) had low spore load. Additionally, 10 uninfected drones of the same F1 queen were used as controls. The RNA of each five drones was pooled to synthesize the cDNA with the same method described above. In total, five pools with high spore load drones, five pools with low spore load drones, two pools for the controls were constructed respectively. The uninfected control group was used as standard to infer to which extend the infection regulate the expression level of the candidate gene.

Table 1 Tested immune genes, associated pathways and primers. Categories: P = pathogen or bacteria receptor; T = transmembrane signal receptor; E = end product; K = kinase; TR = transcription regulator; R = reference gene.

Gene name	pathway	Category	F. primer	B. primer
Abaecin	Toll	E	CAGCATTCGCATACGTACCA	GACCAGGAAACGTTGGAAAC
AmPPO	Toll	E	AGATGGCATGCATTTGTTGA	CCACGCTCGTCTTCTTTAGG
Apidaec	Toll	E	TAGTCGCGGTATTTGGGAAT	TTTCACGTGCTTCATATTCTTCA
Apisimin	Toll	E	TGAGCAAAATCGTTGCTGTC	AACGACATCCACGTTCGATT
Cactus-1	Toll	K	CACAAGATCTGGAGCAACGA	GCATTCTTGAAGGAGGAACG
Cactus-2	Toll	K	TTAGCAGGACAAACGGCTCT	CAGAAAGTGGTTCCGGTGTT
Defensin-2	Toll	E	GCAACTACCGCCTTTACGTC	GGGTAACGTGCGACGTTTTA
Defensin-1	Toll	E	TGCGCTGCTAACTGTCTCAG	AATGGCACTTAACCGAAACG
Dorsal-2	Toll	TR	TCACCATCAACGCCTAACAA	AACTAACACCACGCGCTTCT
Hymenopt	Toll	K	CTCTTCTGTGCCGTTGCATA	GCGTCTCCTGTCATTCCATT
Lys-1	Toll	E	GAACACACGGTTGGTCACTG	ATTTCCAACCATCGTTTTCG
Lys-2	Toll	E	CCAAATTAACAGCGCCAAGT	GCAATTCTTCACCCAACCAT
Lys3l	Toll	E	ATCTGTTTGCGGACCATTTC	TCGATGAATGCGAAGAAAATC
Myd88	Toll	K	TCACATCCAGATCCAACTGC	CAGCTGACGTTTGAGATTTTTG
PGRP9710	Toll	P	TTTGAAAATTTCCTATGAAAGCA	TTTTTAATTGGTGGAGATGGAAA
PGRPSC2505	Toll	P	TAATTCATCATTCGGCGACA	TGTTTGTCCCATCCTCTTCC
PGRPSC4300	Toll	P	GAGGCTGGTACGACATTGGT	TTATAACCAGGTGCGTGTGC
PPPOact	Toll	E	GTTTGGTCGACGGAAGAAAA	CCGTCGACTCGAAATCGTAT

Spaetzle	Toll	K	TGCACAAATTGTTTTTCCTGA	GTCGTCCATGAAATCGATCC
Toll	Toll	T	TAGAGTGGCGCATTGTCAAG	ATCGCAATTTGTCCCAAAAC
Dredd	Imd	K	GCGTCATAAAGAAAAAGGATCA	TTTCGGGTAATTGAGCAACG
Imd	Imd	T	TGTTAACGACCGATGCAAAA	CATCGCTCTTTTCGGATGTT
Kenny	Imd	K	GCTGAACCAGAAAGCCACTT	TGCAAGTGATGATTGTTGGA
Relish	Imd	TR	GCAGTGTTGAAGGAGCTGAA	CCAATTCTGAAAAGCGTCCA
Tak-1	Imd	K	ATGGATATGCTGCCAATGGT	TCGGATCGCATTCAACATAA
Demeless	JAK/STAT	P	TTGTGCTCCTGAAAATGCTG	AACCTCCAAATCGCTCTGTG
Hopscotch	JAK/STAT	K	ATTCATGGCATCGTGAACAA	CTGTGGTGGAGTTGTTGGTG
TepA	JAK/STAT	E	CAAGAAGAAACGTGCGTGAA	ATCGGGCAGTAAGGACATTG
Basket	JNK	K	AGGAGAACGTGGACATTTGG	AATCCGATGGAAACAGAACG
Dscam	JNK	E	TTCAGTTCACAGCCGAGATG	ATCAGTGTCCCGCTAACCTG
EGFlikeA	JNK	E	CATTTGCCAACCTGTTTGT	ATCCATTGGTGCAATTTGG
Hemipterous	JNK	K	CACCTGTTCAGGGTGGATCT	CCTTCGTGCAAAAGAAGGAG
ß-actin		R	ATGCCAACACTGTCCTTTCTGG	GACCCACCAATCCATACGGA
RP49		R	CGCTACAAGAAGCTTAAGAGGTCAT	CCTACGGCGCACTCTGTTG
Gpdh56		R	GGATCAGGAAATTGGGGTTC	CGGAAGCTTATGTCCTGGAA
Gpdh-1		R	GCTGGTTTCATCGATGGTTT	ACGATTTCGACCACCGTAAC
EF		R	GATGCTCCAGGCCACAGAGA	TGCACAGTCGGCCTGTGAT

2.5 QTL mapping procedure

2.5.1 Marker selection

The single F1 queen that was used to reproduce the mapping population was genotyped with 732 fluorescence labeled microsatellite markers by multiplex PCR to select heterozygous markers for the individual genotyping. Out of 732 microsatellite markers, 216 microsatellite markers were heterozygous. 148 drones originate from the single F1 queen were individually genotyped with all 216 microsatellite markers.

2.5.2 Phase determination

As the mother queen of the F1 queen was accidently killed and removed by the worker honey bees, we could not use her directly to determine the phase (Danish alleles and French alleles) of the used markers. We sampled ten drones each from ten colonies of the selected Danish honey bee strain, and genotyped them to determine the phase. We identified all alleles for all 216 heterozygous loci of the selected drones. As all the drones used for the QTL mapping were from a single F1 queen with a Danish and a French allele at every locus, Danish alleles could be unambiguously identified, whenever they were unique and different from the French strain. We used these unique markers as anchor loci to determine the phase of all other loci in the mapping population based on linkage analysis.

2.5.3 Major QTL analysis

We used the number of spores in 1 µl of the diluted gut homogenates as the mapping phenotype. Interval mapping was performed to identify the significant QTL associated with the number of *N. ceranae* spores without considering any interactive effects using Window QTL Cartographer 2.5 (Wang et al. 2011). The threshold for

statistical significance of the putative QTL was calculated with a 5000 times permutation test walking along the genome with the genotyping error probability of 0.01. By doing interval mapping, we could identify the QTL with relatively large effects without considering the interaction effects. If a locus showed a significant association ($p < 0.05$) with the spore load variance, the nearest markers flanking (up-stream and down-stream) this significant locus were used as the criteria to define a QTL candidate region.

2.5.4 Additive and epistatic QTLs analysis

We searched for potential epistatic and additive QTLs using the R/qtl package (Broman et al. 2003; R development core team 2008) with haploid diploid genome type, a genotyping error probability of 0.01 and 1cM walking steps along the genome. All the loci using for the QTL mapping were two loci paired analyzed for the potential additive and epistatic interaction. The significance of the interactive effect was calculated by ANOVA (Bonferroni adjusted for the multiply comparison). If the interaction effects of the paired loci showed a significant association ($p < 0.05$) with the spore load variance, the nearest markers flanking each of the two loci involved were used to define the two interacting QTL regions.

2.6 Statistics

2.6.1 Spore load dynamics and the survival towards the *N. ceranae* infection

The temporal analysis of spore loads in the guts of drones (drones described in **2.2.1**) were analyzed by Univariate Analysis of Variance, with strain and day post infection as fixed factor and replicates as a random factor using the SPSS 16.0 package. Survival of the selected and unselected honey bee strains and the control group were analyzed with Kaplan-Meier procedure (SPSS 16.0).

2.6.2 Relative gene expression quantification

The Opticon Monitor 3 (Bio-Rad, Munich, Germany) software was used to compute the C_t values. The PCR efficiency was calculated for individual reaction with the qpcR package (Spiess and Ritz, 2010), R (Hornik, 2011). The mean amplification efficiency of each gene over all samples was used for subsequent computations. The relative gene expression levels were analyzed with an ANOVA (SPSS 16.0).

The relative gene expression was computed as follows:

$$r = \frac{\sqrt[n]{\prod_{i=1}^{n} E_i^{Ct(i)}}}{E^{Ct(\mathrm{target})}}$$

Where:

E = PCR amplification efficiency

C_t = cycle threshold

i = i^{th} reference gene

n = the number of reference gene

r = relative gene expression level

2.6.3 Population genetic diversity statistics

The drone (described in **2.1.3**) genotypes were used to infer the mother queen genotypes. These queen alleles were used to determine the allele frequencies of the population. The colonies of the unselected control population contributing drones of the DCA were reconstructed using the maximum likelihood algorithm of COLONY 1.3 (Wang 2004) to avoid an estimation bias due to sampling drones with the same chromosomal set that originated from the same mother queen. To detect the presence

of a selective sweep within the significant QTL region, ten fluorescence labeled microsatellite markers within (UN271 and K1452) and flanking (UNEV2, K1453, BI116, K1418B, AT198, K1424, SV188, HYAL) the QTL region were used to assess the genetic variability. Additionally, six randomly chosen and unlinked fluorescence labeled microsatellite markers on chromosomes 1, 3, 6, 8 and 10, served as reference loci to estimate the background genetic diversity of the two populations. The calculation of expected heterozygosity (H_e) was based on the reconstructed queen genotypes and corrected for the sample size (Alam et al. 2011) as:

$$H_e = \frac{n}{n-1}\left(1-\sum P_i^2\right)$$

Where:

n = the number of sampled chromosomes.

p_i = i^{th} allele frequency at a locus.

In addition we determined the number of effective alleles (A_e) for the given locus (Nagylaki 1985) as:

$A_e = 1/(1-H_e)$

Where:

H_e = the expected heterozygosity of a locus.

In order to account for the initial genetic diversity difference and locus specific effects on the pattern of genetic diversity (low mutation rate, selective sweep or background selection), we used a control population that had not been exposed to the artificial selection as a comparison. The population variation estimator ($\theta=4\mu N_e$) of each locus

was calculated according to Schlötterer (2002) and Kauer et al. (2003) assuming that the marker loci in both populations had the same mutation rate. We used the natural logarithm transformation of the ratio (lnRθ) between the two populations to indicate the genetic diversity difference for each locus (Ohta and Kimura 1973; Wiehe et al. 2007) as:

$$\ln R\theta = \ln \frac{\left(\frac{1}{1-.H_{selected}}\right)^2 - 1}{\left(\frac{1}{1-H_{control}}\right)^2 - 1}$$

Where:

$H_{selected}$ = expected heterozygosity of the selected population.

$H_{control}$ = expected heterozygosity of the unselected control population.

As lnRθ of the unlinked reference loci have been proved to follow the normal distribution (Kauer et al. 2003; Wiehe et al. 2007), we compared the lnRθ of the six reference loci with that of the ten target loci within and flanking QTL region. As the number of the reference loci is not high, we used t distribution to estimate the mean and standard deviation of the six reference loci. The t value of ten target loci was calculated as an estimator indicating the existence of the selective sweep for each locus as:

$$t_i = \frac{y_i - m}{s/\sqrt{n}}$$

where:

m = mean of lnRθ of six reference loci.

s = standard deviation of lnRθ of six reference loci of t distribution.

y_i = lnRθ value of locus i.

t_i = t value for the locus i.

3 Results

3.1 Survival and innate immune responses towards *N. ceranae* infection

3.1.1 Spore load dynamic and survival

Drones of selected and unselected strain (drones described in **2.2.1**) were initially fed with *N. ceranae* spores. Drones fed with sugar solution were served as controls. Spores were not found in the uninfected control group. During the first four days after the infection, the *N. ceranae* spore load in the infected drones' guts showed no significant differences between the two strains. However, from day five onward the number of spores in the drones of the selected strain was significantly higher than in the unselected strain ($p < 0.001$). This difference was largest on day six with more than an order of magnitude more spores in drones of the selected strain ($8.2*10^5$ vs $7*10^4$, selcted vs unselected, $p < 0.001$) (Figure 4). At day eight post infection, the spore titers became more similar again. However, in spite of this much higher spore load, the drones of the selected strain had a significantly lower mortality than those of the unselected strain (Log Rank, $p < 0.001$). Moreover the mortality of the selected drones was not significantly different from the uninfected control group (Log Rank, $p = 0.25$) (Figure 5). The mortality in the unselected strain was significantly higher than the control group ($p < 0.001$). More than 90% of the drones of the selected strain were still alive even on day eight post infection whereas 40% the drones of unselected strain were dead. The highest mortality occurred on day two post infection in drones of the unselected strain (Figure 5).

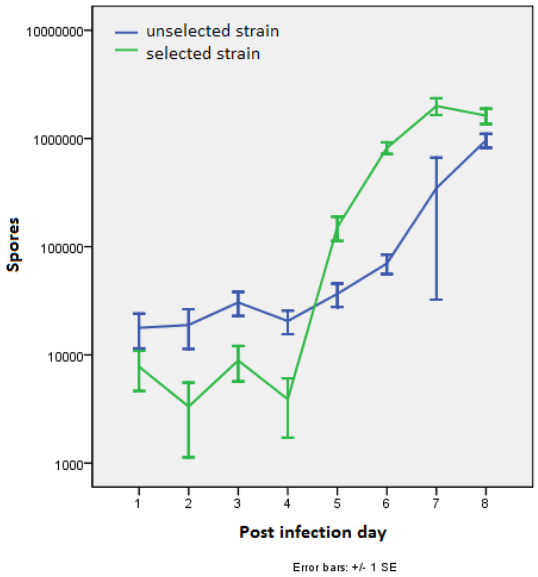


Figure 4 Dynamic of *N. ceranae* spore load in the gut of artificially infected drones of the selected and unselected strain. Both strains were fed with the same number of *N. ceranae* spores. From 5 days post infection, the spore load in selected strain was much higher than in the unselected strain.

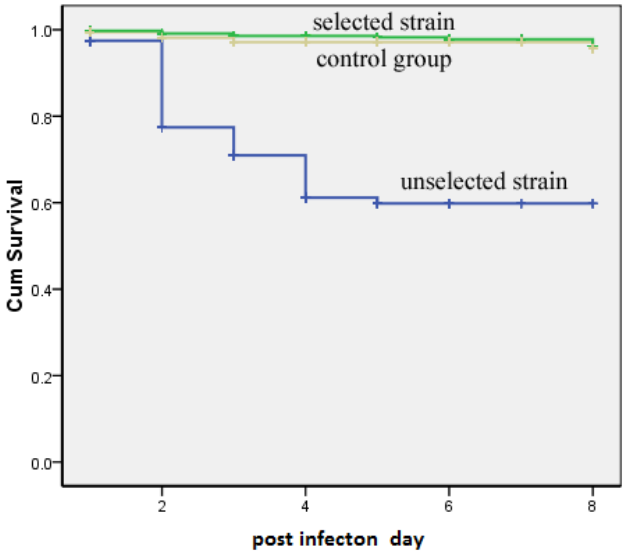


Figure 5 Time series mortality. Three drones that were sampled every day from each cage. These three drones were treated as sensored in the survival analysis. Drones in the unselected strain had significantly higher mortality (Kaplan-Meier procedure) whereas there was no significant difference in survival between the uninfected control group and the drones from the selected lineage.

3.1.2 Reference genes

Three genes from the five chosen reference genes (*ß-Actin*, *RP49* and *GAPDH56*) were significantly differentially expressed between drones of selected and unselected population ($p < 0.05$) and hence unsuited and excluded as the reference in the further analyses. The two remaining reference genes *GAPDH-1* and *EF* showed no co-regulation with the treatment and were used for the normalization of the immune gene expression levels.

3.1.3 Expression levels of innate immune genes of Toll pathway

All 33 innate immune genes tested, showed no significant differences in expression levels between the selected strain and unselected strain during the first five days after the infection. However, on day six post infection, six immune genes from three pathways showed significantly differential regulation between the two strains (Bonferroni adjusted for multiple comparisons) (Figure 6).

Immune genes of Toll pathway were up-regulated after the infection in both the selected and the unselected strain. The peptidoglycan recognition protein-SA (*PGRPSC4300*, GB15371) was significantly up regulated on day one post infection compared to the control group ($p < 0.001$). This up-regulation decreased steadily over the subsequent days in both strains until this gene was close to the expression level of the uninfected control group in the selected strain on day 6 post infection, and even significantly down regulated in the unselected strain ($p < 0.05$). *Toll* (GB18520), *lysozyme-2* (*lys-2*, GB15106) and *dorsal-2* (GB18032) showed a similar but less pronounced expression dynamics and only on day six was there a significant difference in the expression levels between the selected and the unselected strain ($p < 0.05$) (Figure 6).

3.1.4 Expression levels of innate immune genes of IMA and JAK/STAT pathway

Genes of IMD pathway and JAK/STAT pathway were up-regulated in the infected drones of both strains during the first five days after the infection. Again significant differences only occurred on day 6 post infection when *Dredd* (GB17683, IMD pathway) up regulated in the selected strain ($p < 0.01$) but down regulated in the unselected strain. The gene *domeless* (GB16422) from the JAK/STAT pathway was similarly up-regulated after the infection in both strains, but again significantly down-regulated in the unselected strain on day six post infection ($p < 0.01$) (Figure 6).

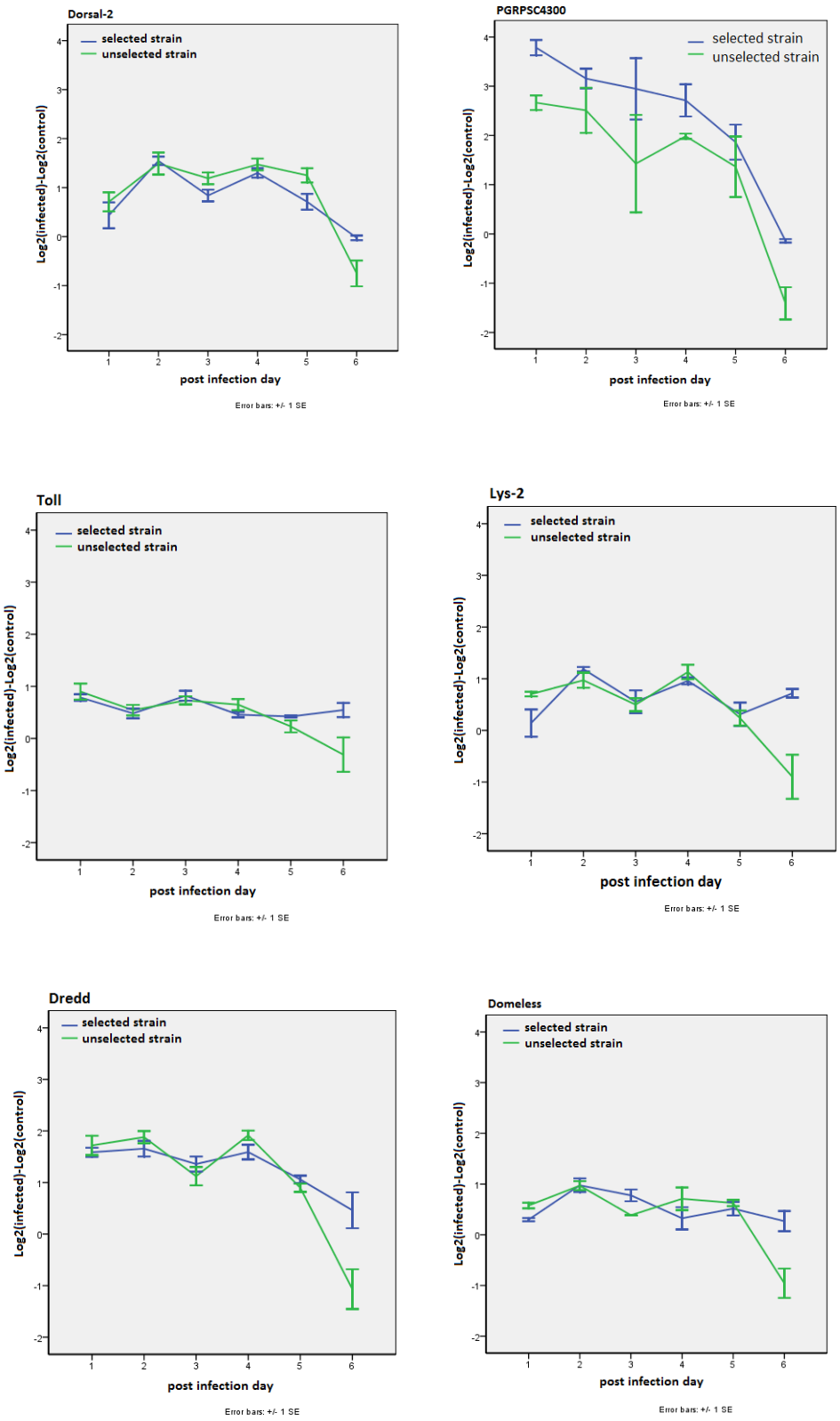


Figure 6 Immune gene expression levels (log transformed) in relation to the uninfected controls and normalized for two reference genes. In the Y axis, the value 0 represents the express level of uninfected control group.

3.2 QTLs associated with the *N. ceranae* spore load

3.2.1 Interval mapping

Of the 732 screened microsatellite markers, 216 were identified as polymorphic. 60 anchor loci were identified where we could unambiguously phase the marker alleles. The average distance of 216 polymorphic markers was 20 cM. In total, 90% of the genome was less than 20 cM and 99.5% of the genome was less than 50 cM from a marker. So the chance of omitting a significant QTL was low due to the linkage gap. A total of 148 drones (Figure 7) were individually genotyped with all 216 heterozygous markers (Figure 8). The linkage to the spore load was quantified for each marker by interval mapping using the LOD scores. The locus UN271 on chromosome 14 showed a significant association with the spore load (LOD = 2.5, $p < 0.05$). The significant QTL region spanned 1598kbp (between the locus K1418 at 5355kbp and AT198 at 6953kbp).

3.2.2 Fine mapping

To further narrow down the detected QTL region, we designed and genotyped the 148 drones with five additional heterozygous microsatellite markers, which are flanking the locus UN271 within the statistically significant QTL region (between the locus K1418 and the locus AT198). The association between the markers and the spore load was re-calculated by interval mapping. The locus UN271 again showed a significant association with the *Nosema* spore load (LOD = 2.6, $p < 0.05$), confirmed by the permutation test along the genome (Figure 9). By adding additional five markers, the statistically significant QTL (QTL_m) region was narrowed down to 338kbp, spanning from locus HQ1414 at 6071kbp to BI103 at 6409kbp explaining 7.7% of total variance. Looking at the actual segregation of the alleles at this locus, 46 out of 76 low spore load drones had the selected Danish allele and 49 out of 72 high spore

drones had the unselected French allele. This is a significant biased allelic distribution towards the predicted phenotype ($p < 0.001$, χ^2 test). The locus UN271 showed significant segregation, but its neighboring loci HQ1411 and HQ1414 did not, suggesting the target region is between these two loci. The frequency of the selected Danish allele decreased with the increasing number of spores (Figure 10).

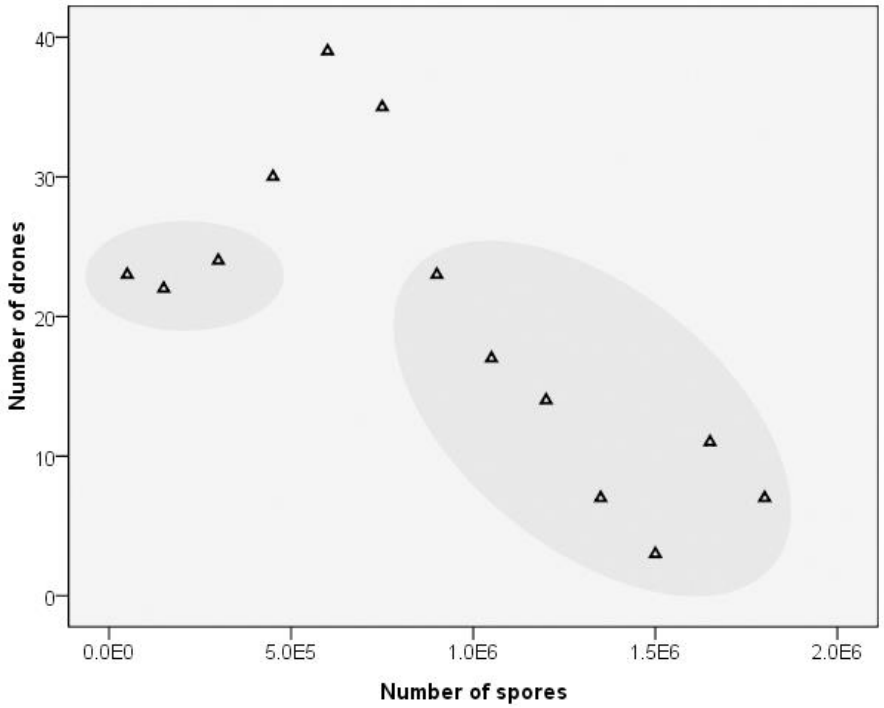


Figure 7 Spore loads of the infected drones. The Y axis represents the number of the infected drones. The guts of all drones were removed on day six post infection and the spore load was counted using Fuchs-Rosenthal haemocytometer. The guts of 255 drones were individually removed to determine the spore load (median & range, $5.6*10^5$ & $2.4*10^6$) (Figure 7). No *N. ceranae* spores were found in the control groups. 76 drones with the relatively lowest (ellipse on the left side, < $3.6*10^5$; median: $8*10^4$, range: $3.4*10^5$) and 72 drones with the relatively highest spore loads (ellipse on the right side, >8.2 $*10^5$ spores, median $1.1*10^6$, range: $1.7*10^6$) were selected for the QTL mapping. The number of spores were significantly different between these two groups (Mann-Whitney U test, $p < 0.001$).

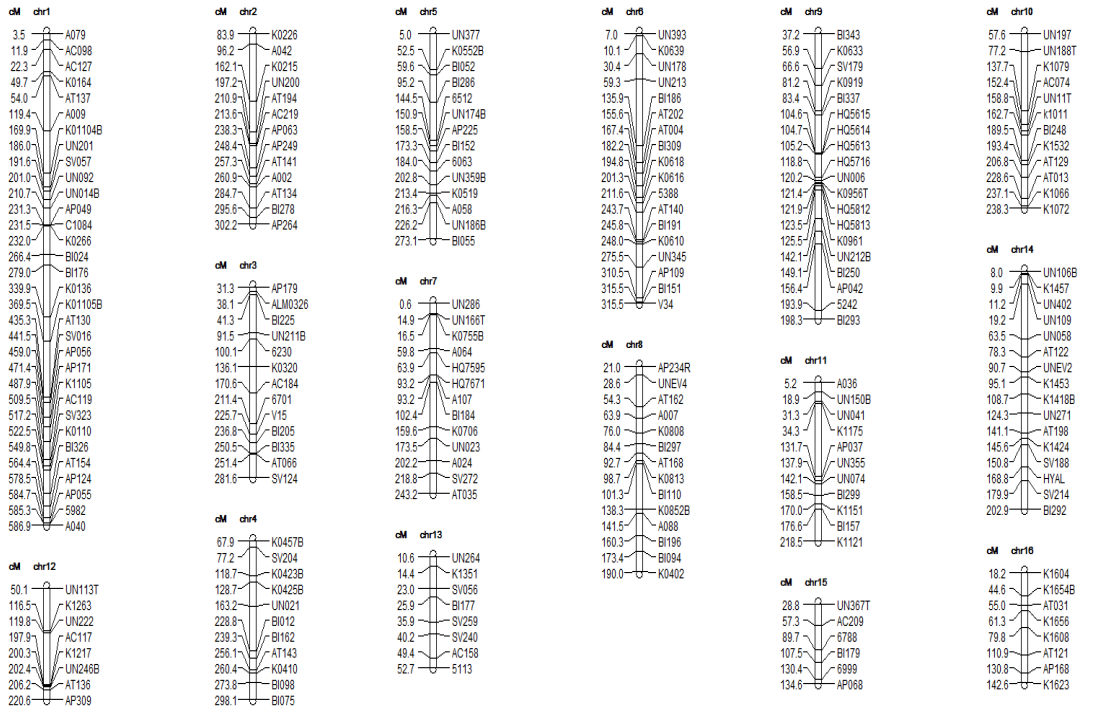


Figure 8 216 heterozygous genetic marker trees along the genome. We screened 732 fluorescence labeled microsatellite markers with the F1 queen to search heterozygous markers by multiplex PCR. The microsatellite markers' names are on the right side of each chromosome and the genetic distance (cM) is one the left side.

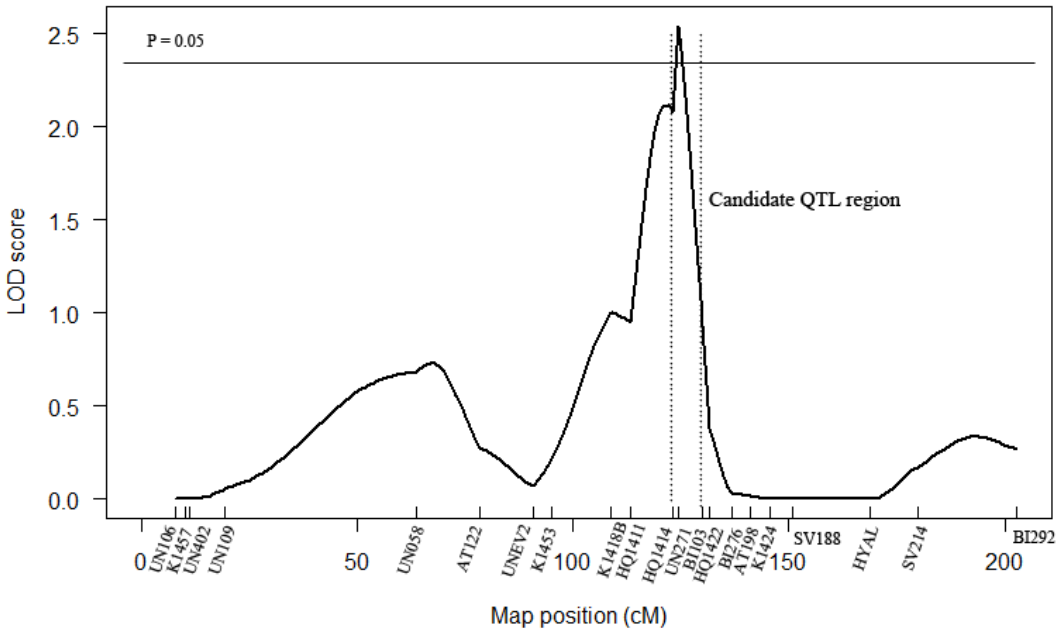


Figure 9 Significant QTL on Chromosome 14. The threshold for the significant QTL was calculated by 5000 time permutation test along the genome with the genotyping error probability of 1%. The QTL is statistically significant at P = 0.05 level with LOD score of 2.4. By interval mapping, the marker UN271 (LOD=2.6) showed a significant association with the *Nosema* spore load. The significant QTL region was located between the marker HQ1414 and BI103 (spanning 338kbp) explaining 7.7% of the total variance. The locus UN271 showed significant segregation, but its neighboring loci HQ1411 and HQ1414 did not, suggesting the target region is between these two loci. We used the interval between these two neighboring markers to search the candidate genes (represented by the vertical dashed line).

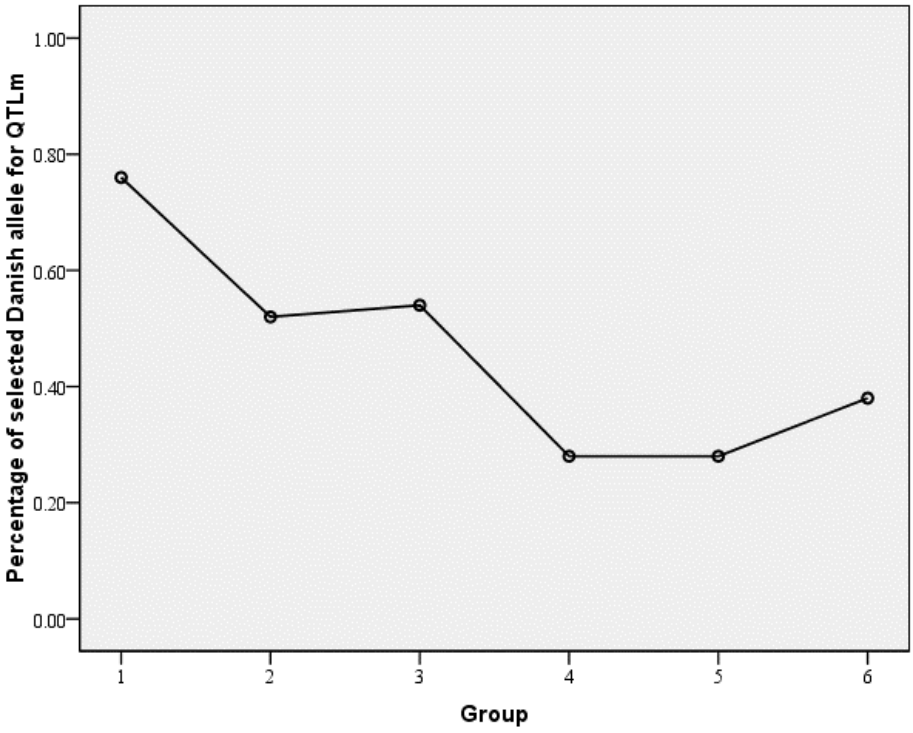


Figure 10 Percentage of the selected Danish alleles at the QTL_m in chromosome 14. The mapping population of 148 drones was divided into 6 groups according to the spore load from low to high. Group 1 (n=25), group 2 (n=25) and group 3 (n=22) represent three groups of drones with low spore load. Group 4 (n=25), group 5 (n=25) and group 6 (n=26) represent three groups of drones with high spore loads. Group 1 had the lowest spore load and group 6 had the highest spore load. The allele was significantly biased distributed within the drones with low and high spore load.

3.2.3 Additive and epistatic QTLs analysis

Locus AC184 on chromosome 3 (QTL_{ep3}) significantly interacted with the locus AT129 (QTL_{ep10}) on chromosome 10 in a two dimensional two QTLs scan, which explained 6.3% of the total variance (LOD = 3.6, $p < 0.01$). Drones with either the Danish or the French alleles at both loci (QTL_{ep3} & QTL_{ep10}), had a higher spore load than those with a combination of two alleles at either locus (Figure 11, Figure 12). This suggests that the interactions between the Danish allele and French allele were important for a low spore load. Additionally, we identified an additive QTL on chromosome 6 (K0616, QTL_{ad}). The additive effect between the Danish allele of QTL_{ad} and the Danish allele of the major QTL_m was also associated with a low spore load explaining 6.4% of the total variance (LOD = 2.5, $p < 0.05$) (Figure 11).

From the actual allelic segregation, 86% of the drones with a low spore load carried at least one Danish allele at either locus of QTL_{ad} or QTL_m. 79% of the drones that carried the Danish alleles at both loci showed a low spore load, which was a strong and significant bias compared to the random combination of two alleles towards the predicted phenotype ($p < 0.01$, χ^2 test).

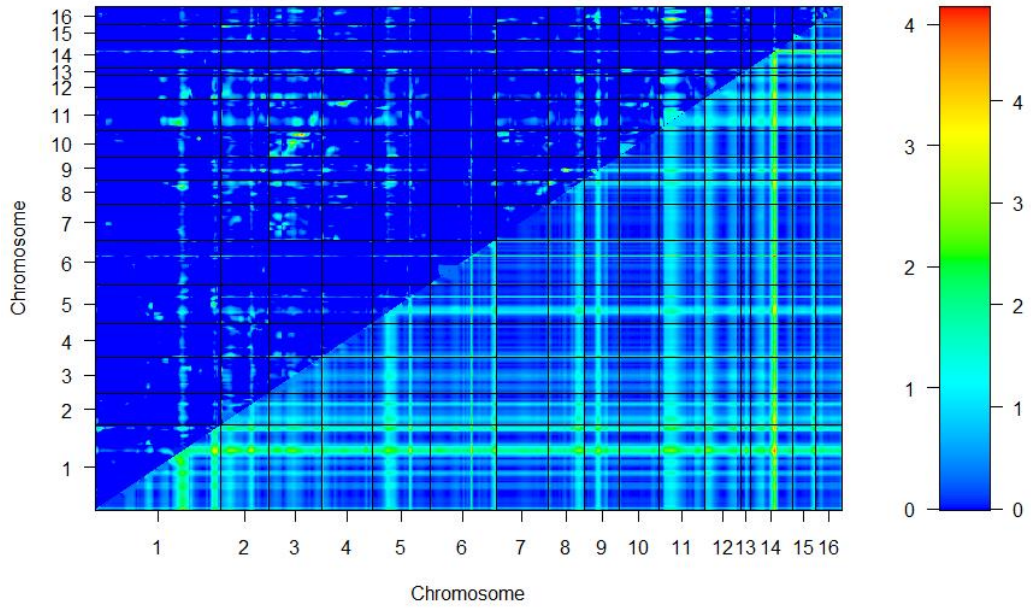


Figure 11 LOD scores along the genome for the evidence of the epistatic and additive QTLs by paired two loci scan. Epistatic effect displayed in the upper left triangle. Additive effect is displayed in the lower right triangle. The numbers of left and right in the color scale on the right correspond to LOD score for epistatic effect and additive effect respectively. Chromosome 3 and chromosome 10 showed significant epistatic effects (LOD = 3.6, $p < 0.01$) and chromosome 6 and chromosome 14 showed significant additive effects (LOD = 2.5, $p < 0.05$).

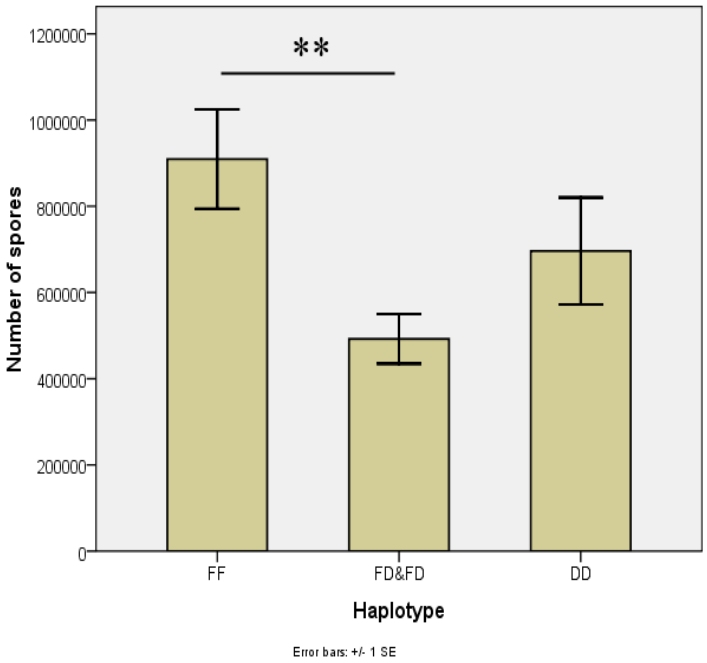


Figure 12 Observed spore load for two-locus genotype groups at the location of two epistatically interactive QTLs. AC184 which is in chromosome 3 significantly interacted with AT129 on chromosome 10 by explaining 6.3% of total variance. D and F represent the allele originated from Danish and French strains respectively. The genotype 'DD' represents the Danish allele at both AC184 in chromosome 3 and AT129 in chromosome 10. The genotype 'DF' represents the Danish allele at AC184 and French allele at AT129. Drones with the genotype of "DF" and "FD" had significantly lower spore load than the drones with "FF" allele at both loci which suggests the direction of the epistatic effect is to reduce the spore load. ** represents the significant level at $p < 0.01$, ANOVA, two tailed test, Bonferroni adjusted for the multiply comparison.

3.2.3 Candidate genes

We searched for candidate genes within QTL_m (between the loci HQ1414 and BI103) and QTL_{ad} (between the loci K0618 and 5388). 31 and 11 annotated genes were identified respectively (Table 2, Table 3). The candidate gene *Aubergine* (*Aub*) within QTL_m was of particular interest, as *Aub* is a member of the Agronaut family containing the typical active piwi domain, which functions as a nucleic acid binding domain and is involved in RNA interference (Kawaoka et al. 2008; Liao et al. 2010). Subsequently we quantified the expression levels of *Aub* in 60 drones of the F1 queen towards *N. ceranae* infection. Comparing with the controls, the expression level of *Aub* was up-regulated towards *N. ceranae* infection in drones with both high and low spore load. Hence *Nosema* infection enhanced the *Aub* expression. This effect was however stronger in the low than high spore load group. *Aub* was significantly more over expressed in drones with a low spore load than in those with a high spore load (two tailed t test, $p < 0.05$) (Figure 13).

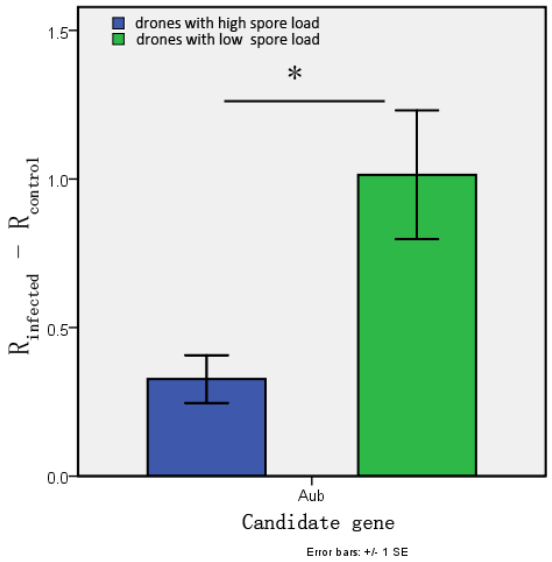


Figure 13 Candidate gene expression profiles between drones with high spore load and low spore load. $R_{infected}$ and $R_{control}$ represent the relative gene expression value of the infected drones and control drones respectively. Comparing with the controls, the expression level of *Aub* was up-regulated towards *N. ceranae* infection in drones with both high and low spore load. *Aub* was significantly over expressed in drones with low spore load than drones with high spore load. * significant level, $p < 0.05$, two tailed t test.

Table 2 candidate genes, fly orthologs and human orthologs identified within the statistically significant QTL_m by fine mapping. The QTL region was between the loci HQ1411 and HQ1414 which flanked the significant locus UN271 in chromosome14.

Gene name	**Description / function in the honey bee**	**Orthology in the fly**	**Orthology in human**
LOC100577208	hypothetical protein LOC100577208	unknown	unknown
LOC 724365	UPF0539 protein CG14977 - like	unknown	unknown
LOC551191	coiled-coil domain-containing protein 130 homolog	Protein structure motif; gene expression regulation	Regulation of germ cells differentiation during spermatogenesis
LOC724506	heat repeat containing protein 3-like	Binding	Involved in ribosome biosynthesis
Aub aubergine	RNA interference	DNA/RNA binding; expression interference	RNA-mediated gene silencing; Translation regulation
LOC 4099300	hypothetical LOC 4099300	Microtubule binding; microtubule cytoskeleton organization	Assembly of kinetochore proteins, mitotic progression and chromosome segregation.
LOC100577050	Hypothetical protein LOC100577050	Zinc ion binding	Regulation of spermatogenesis
LOC724688	Pallidin-like	unknown	Intracellular vesicle trafficking
O-fut2	GDP fucose protein O-fucosyltransferase 2	GDP-fucose protein O-fucosyltransferase	Carbohydrate metabolism; Fucose metabolism
LOC409902	Liprin-alpha-2-like	Sterile alpha motif domain;	Cell-matrix adhesion; regulating the disassembly of focal adhesions
LOC100576983	Hypothetical protein LOC100576983	unknown	Extracellular matrix structural constituent; transferase activity, transferring acyl groups
LOC724846	Hypothetical protein LOC724846	unknown	unknown

LOC100577029	Hypothetical protein LOC100577029	Ionotropic glutamate receptor activity; extracellular-glutamate-gated ion channel activity	Glycine catabolic process; aminomethyltransferase activity; aminomethyltransferase activity
Shrb	Transport and catabolism; Endocytosis	Protein transport	Protein transport
LOC552492	Dual specificity protein phosphatase 3-like	Protein tyrosine / serine / threonine phosphatase activity; protein tyrosine phosphatase activity; protein dephosphorylation	Tyrosine-protein phosphate and serine-protein phosphate
Kinesin-6A	Kinesin motor domain; ATPase sctivity and involved in mitosis.	ATP binding; microtubule motor activity; microtubule-based movement; Kinesin, motor domain	Cell division; ATP-binding; microtubule-based movement
RpL13	Ribosomal protein; involved in translation	Structural constituent of ribosome; involved in translation	SRP-dependent cotranslational protein targeting to membrane; Ribonucleoprotein; structural constituent of ribosome
LOC552532	probable leucyl-tRNA synthetase, mitochondrial-like	Leucine-tRNA ligase activity; ATP binding; leucyl-tRNA aminoacylation	Leucyl-tRNA aminoacylation; tRNA aminoacylation for protein translation
LOC725055	gamma-glutamylcyclotransferase-like	Gamma-glutamylcyclotransferase activity; Butirosin biosynthesis, BtrG-like; Gamma-glutamylcyclotransferase	Glutathione homeostasis process; gamma-glutamylcyclotransferase activity
LOC552569	Butirosin biosynthesis, BtrG-like; Gamma-glutamylcyclotransferase	Transferase activity, transferring phosphorus-containing groups; ABC-1; Protein kinase-like domain	Transferase activity, transferring phosphorus-containing groups
LOC100576113	Hypothetical protein LOC100576113	Kinetochore assembly; mitotic spindle organization; neurogenesis	Histone acetyltransferase complex;

LOC100576082	Hypothetical protein LOC100576082	unknown	ATP binding; protein serine/threonine kinase activity; signal transduction
LOC725200	IQ and AAA domain-containing protein 1-like	ATP binding; nucleoside-triphosphatase activity	ATP-binding
LOC551378	slit homolog 1 protein-like	Cysteine-rich flanking region; Leucine-rich repeat; Leucine-rich repeat, typical subtype	Protein stabilization; enzyme regulator activity
LOC725292	homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 2 protein-like	Ubiquitin; Ubiquitin supergroup	Response to unfolded protein; integral to membrane
Aprt	Aprt adenine phosphoribosyltransferase	Adenine phosphoribosyltransferase activity; adenine salvage; nucleoside metabolic process	Catalyzes a salvage reaction resulting in the formation of AMP; Purine salvage
LOC724206	WD repeat-containing protein 67-like	Rab GTPase activator activity; regulation of Rab GTPase activity	Positive regulation of Rab GTPase activity
LOC409984	DET1- and DDB1-associated protein 1-like	unknown	Transcription regulation; proteasomal degradation; interaction with CUL4A and WD repeat proteins
LOC724158	Hypothetical protein LOC724158	unknown	DNA damage response, signal transduction resulting in transcription; double-strand break repair
LOC100576385	Hypothetical protein LOC100576385	unknown	Regulation of Rho protein signal transduction

Table 3 candidate genes, fly orthologs and human orthologs, identified within the statistically significant additive QTL region in chromosome 6. The QTL region is defined between the neighboring markers (K0618 and 5388) of the statistically significant additive locus K0616.

Gene name	**Description / function in the honey bee**	**Orthology in the fly**	**Orthology in human**
LOC100576573	Hypothetical protein LOC100576573	Acetylation, Phohphorprotein	unknown
LOC100576598	Hypothetical protein LOC100576598	Calcium ion binding, zinc ion binding	unknown
LOC551803	Spalt-like	System development; sensory organ development; neuron differentiation	Inductive cell-cell signaling; olfactory bulb development; positive regulation of neuron differentiation
LOC100576253	Hypothetical protein LOC100576253	Retinitis pigmentosa GTPase regulator, nucleosomal DNA binding	unknown
LOC100576206	Hypothetical protein LOC100576206	Hemophilic cell adhesion, calcium ion binding	Importing protein into nucleu
LOC552589	ENSANGP00000002131-like	Establishment or maintenance of actin cytoskeleton polarity	Interacting between neurites derived from specific subsets of neurons during development
LOC409263	glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1-like	Protein modification, protein glycosylation	Lycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase activity
LOC410159	mediator of RNA polymerase II transcription subunit 13-like	A coactivator involved in the regulated transcription of nearly all RNA	Regulation of transcription from RNA polymerase II promoter

		polymerase II-dependent genes	
LOC724516	Hypothetical protein LOC724516	Cytokine activity	unknown
LOC100577622	Hypothetical protein LOC100577622	ATP binding, nucleotide binding	unknown
mam	Neurogenic protein mastermind	Methyltransferase activity, zinc ion binding	Transition metal ion binding; oxidoreductase activity

3.3 Selective sweep within the QTL_m region

3.3.1 Expected heterozygosity (H_e)

In the selected honey bee population, the genotypes of ten queens were successfully reconstructed (Table 4). Since drones develop from the unfertilized eggs of the queen, they carry only one maternal chromosomal copy comprising the QTL and the entire linkage group under study. In a set of ten drones per colony, the allelic composition at the series of the 10 linked loci at both of the queens' chromosomes can be reconstructed with a probability of 99.8% for the two entire linkage groups. In all cases, both maternal alleles were detected in each set of ten drones. By calculating the relatedness among drones of the control population, 29 sets of chromosomes were constructed from 32 drones that were classified into 22 ± 1.1 colonies (Table 4). The loci UN271, K1452 and AT198 had the lowest genetic variability in the selected population. The expected heterozygosity was $H_e = 0.1$ for these three loci, which was much smaller than the seven neighboring flanking loci ($H_e = 0.55 \pm 0.04$) and the six reference loci ($H_e = 0.58 \pm 0.05$). In the control population, these three loci had a higher average expected heterozygosity ($H_e = 0.84 \pm 0.06$) than the neighboring ($H_e = 0.61 \pm 0.07$) and reference loci ($H_e = 0.74 \pm 0.05$). By comparing the two populations, the expected heterozygosity was most strongly reduced at the loci UN271, K1452 and AT198, whereas the neighboring and reference loci were similar (Table 5).

Table 4 Re-constructed queen genotypes of the selected and unselected honey bee populations

	Honey bee populations of Denmark (data deposit)																
marker		AC127	6701	K0616	K0808	AT168	AT129	UNEV2	K1453	BI116	K1418B	UN271	K1452	AT198	K1424	SV188	HYAL
chromosome		1	3	6	8	8	10	14	14	14	14	14	14	14	14	14	14
Physical position (kp)		1109	9693	11669	4417	5381	11106	4466	4684	4825	5355	6124	6265	6953	7174	7432	8319
Denmark colonies	alleles																
queen N.4	allele 1	170	103	140	148	184	192	205	146	182	181	171	331	288	216	187	379
	allele 2	175	109	142	154	193	192	205	152	000	187	171	331	288	218	189	379
queen N.16	allele 1	170	103	142	148	182	192	205	146	182	181	171	331	288	216	187	369
	allele 2	175	113	142	154	184	192	205	148	182	187	171	331	282	218	189	379
queen N.17	allele 1	170	103	142	148	184	192	205	146	182	181	171	331	288	216	187	369
	allele 2	172	113	142	154	184	192	205	148	220	187	171	331	288	218	189	379
queen N.25	allele 1	172	103	140	154	184	192	205	148	220	181	171	331	288	216	187	369
	allele 2	175	113	142	154	184	192	205	148	220	199	173	313	288	218	187	369
queen N.72	allele 1	175	97	140	146	182	192	207	148	176	181	171	331	288	216	187	369
	allele 2	180	103	140	154	182	192	207	148	176	181	171	331	288	218	189	379
queen N.123	allele 1	172	109	140	154	182	179	207	146	176	181	171	331	288	216	181	379
	allele 2	175	113	142	154	193	192	207	152	182	181	171	331	288	218	189	369
queen N.159	allele 1	172	109	140	154	182	179	205	148	176	181	171	331	288	216	187	369
	allele 2	175	113	142	154	184	192	207	152	182	181	171	331	288	218	189	379
queen N.162	allele 1	173	109	140	146	184	179	205	148	176	181	171	331	288	216	181	369
	allele 2	175	113	140	154	193	192	207	148	182	181	171	331	288	218	187	379
queen N.184	allele 1	172	97	140	154	182	179	207	146	176	181	171	331	288	216	181	369
	allele 2	175	109	142	154	193	192	207	152	182	181	171	331	288	218	187	379
queen N.196	allele 1	172	97	140	146	184	179	205	148	176	181	171	331	288	216	187	379
	allele 2	175	109	140	154	184	192	205	148	176	181	171	331	288	218	189	379

	Honey bee populations of Germany (data deposit)																
marker		AC127	6701	K0616	K0808	AT168	AT129	UNEV2	K1453	BI116	K1418B	UN271	K1452	AT198	K1424	SV188	HYAL
chromosome		1	3	6	8	8	10	14	14	14	14	14	14	14	14	14	14
Physical position (kp)		1109	9693	11669	4417	5381	11106	4466	4684	4825	5355	6124	6265	6953	7174	7432	8319
Germany colonies	alleles																
queen N.1	allele 1	179	97	132	153	185	192	201	137	170	194	170	331	171	218	197	370
	allele 2	172	97	140	153	185	192	201	147	172	194	173	331	167	212	197	387
queen N.2	allele 1	172	107	136	153	189	192	206	147	180	175	177	000	169	218	197	370
	allele 2	172	107	138	155	187	192	206	147	180	175	177	000	169	218	197	370
queen N.3	allele 1	172	97	132	149	183	192	201	137	167	196	177	357	182	218	197	385
	allele 2	172	97	132	153	189	192	206	147	167	196	173	357	180	212	188	387
queen N.4	allele 1	172	120	141	143	193	179	208	147	182	196	170	329	184	218	182	370
	allele 2																
queen N.5	allele 1	174	97	141	145	180	192	206	147	174	181	170	347	196	218	197	370
	allele 2																
queen N.6	allele 1	174	109	132	147	183	192	201	147	172	181	177	331	182	218	197	370
	allele 2																
queen N.7	allele 1	174	000	140	145	187	192	206	147	178	181	175	351	184	212	188	370
	allele 2	179	97	136	147	184	192	206	152	178	181	170	351	180	216	188	380
queen N.8	allele 1	172	97	136	147	189	181	210	147	172	192	177	353	180	216	188	370
	allele 2																
queen N.9	allele 1	168	000	140	147	183	192	206	150	172	194	177	333	186	218	188	370
	allele 2																
queen N.10	allele 1	174	120	136	155	189	181	206	147	174	192	173	000	182	218	188	370
	allele 2	172	117	136	145	183	192	206	147	174	192	170	313	169	218	188	370
queen N.11	allele 1	172	109	140	147	183	181	201	137	180	181	170	000	180	218	188	370
	allele 2																
queen N.12	allele 1	174	109	144	155	000	194	224	147	170	181	000	313	186	218	188	370
	allele 2																
queen N.13	allele 1	176	97	140	145	185	192	208	137	176	181	175	345	167	216	188	391
	allele 2																
queen N.14	allele 1	172	97	136	145	185	192	208	147	188	185	177	331	182	218	188	370
	allele 2																
queen N.15	allele 1	172	113	140	153	183	188	206	147	172	192	170	000	180	218	188	370
	allele 2																
queen N.16	allele 1	174	107	140	147	183	192	206	147	176	192	177	313	169	218	188	385
	allele 2	176	97	140	158	189	192	210	147	176	196	177	313	186	216	182	370
queen N.17	allele 1	172	109	136	147	191	194	224	137	176	175	173	327	184	218	182	370
	allele 2																
queen N.18	allele 1	174	113	132	145	189	192	206	147	190	181	170	329	180	218	188	370
	allele 2	174	117	140	155	189	192	206	147	180	181	177	329	180	212	188	370
queen N.19	allele 1	172	117	136	155	180	192	206	143	172	181	170	337	196	212	188	370
	allele 2																
queen N.20	allele 1	172	113	140	145	195	192	208	147	180	185	173	329	196	218	188	370
	allele 2																
queen N.21	allele 1	176	000	136	143	180	194	218	147	174	181	173	339	184	216	188	370
	allele 2																
queen N.22	allele 1	176	000	140	145	190	197	206	147	172	196	177	329	184	218	188	370
	allele 2																
000 represents ungenotyped allele																	

Table 5 Locus, chromosome, physical position, expected heterozygosity, the number of effective alleles and for the 16 genotyped loci. bp = the physical position on the chromosome in bp (based on Amel 4.5); $H_{(eS)}$ = expected heterozygosity in the selected population; $A_{(eS)}$ = number of effective alleles in the selected population; $H_{(eC)}$ = expected heterozygosity in the unselected control population; $A_{(eC)}$ = number of effective alleles in the unselected control population. lnRθ = natural logarithm transformation of the ratio between the two populations. ** represent P < 0.01; *** represent P < 0.001, Bonferroni adjusted for multiple comparisons, two tails test.

Loci	Chrom.	bp	$H_{(eS)}$	$A_{(eS)}$	$H_{(eC)}$	$A_{(eC)}$	lnRθ
AC127	1	1,109,005	0.72	3.52	0.69	3.22	0.12
6701	3	9,693,772	0.77	4.42	0.80	4.92	-0.22
K0616	6	11,669,265	0.52	2.09	0.74	3.90	-1.44
K0808	8	4,417,956	0.49	1.96	0.83	5.80	-2.44
AT168	8	5,381,724	0.65	2.88	0.86	7.27	-1.96
AT129	10	11,106,906	0.39	1.65	0.52	2.07	-0.64
UNEV2	14	4,466,449	0.51	2.02	0.70	3.30	-1.17
K1453	14	4,684,327	0.63	2.68	0.46	1.85	0.94
BI116	14	4,825,535	0.65	2.90	0.89	8.83	-2.34
K1418B	14	5,355,209	0.35	1.54	0.80	4.95	-2.83
UN271	14	6,124,407	0.10	1.11	0.72	3.53	-3.89**
K1452	14	6,265,487	0.10	1.11	0.91	11.50	-6.33***
AT198	14	6,953,129	0.10	1.11	0.88	8.12	-5.62***
K1424	14	7,174,765	0.53	2.11	0.53	2.13	-0.02
SV188	14	7,432,688	0.64	2.75	0.52	2.08	0.68
HYAL	14	8,319,697	0.52	2.09	0.37	1.59	0.78

3.3.2 Effective number of alleles (A_e)

In order to visualize the potential allele fixation, we compared the number of effective alleles (A_e) of the two populations. Again, the loci UN271, K1452 and AT198 showed the lowest number of effective alleles of $A_e = 1.11$ in the selected population that was almost fixed. The average number of effective alleles of the seven flanking loci were more than twice as high ($A_e = 2.3 \pm 0.18$), but slightly lower than the six reference loci ($A_e = 2.8 \pm 0.43$). In the unselected population, the locus K1452 had the highest number of effective alleles with $A_e = 11.5$. The loci UN271, K1452 and AT198 had a high average number of effective alleles ($A_e = 7.72 \pm 0.23$), which was even higher than the neighboring ($A_e = 3.5 \pm 0.99$) and the reference loci ($A_e = 4.5 \pm 0.76$) (Table 5).

3.3.3 Selective sweep

In order to detect the existence of the selective sweep, we estimated the mean and standard deviation of $\ln R\theta$ from the six reference loci of -1.087 ± 1.03. By comparing with ten loci on chromosome 14, the loci UN271 ($p < 0.01$), K1452 ($p < 0.001$) and AT198 ($p < 0.001$) crossed the threshold, which gave evidence for the existence of the selective sweep (Figure 14, Table 5). The threshold was Bonferroni adjusted for ten comparisons on chromosome 14.

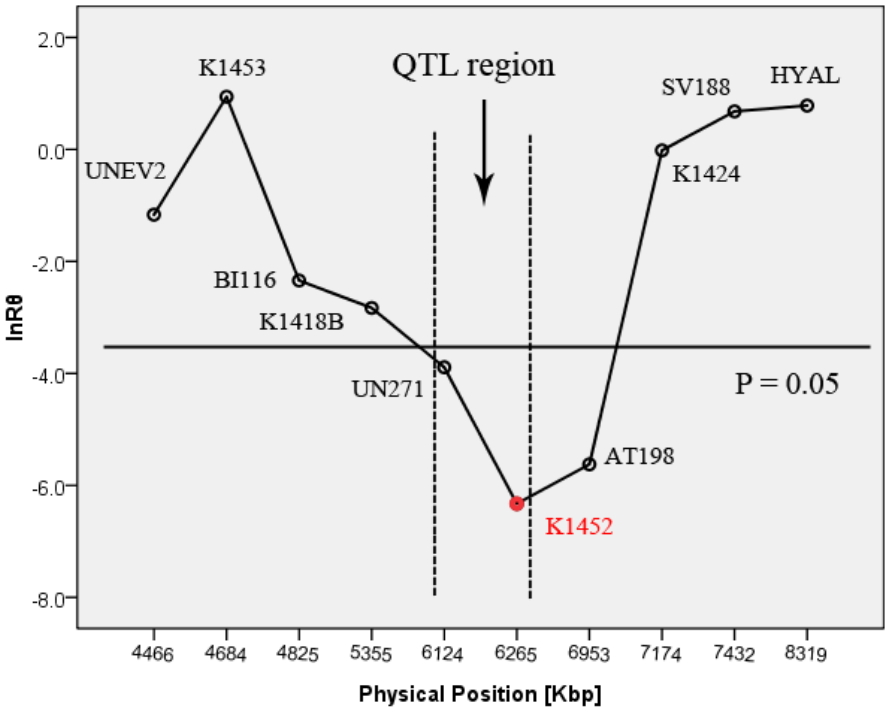


Figure 14 The lnRθ between the selected and unselected control populations on chromosome 14. The horizontal line represents the significance threshold (p = 0.05) to detect the selective sweep (Bonferroni adjusted for multiple comparisons). Three loci (UN271, K1452, AT198) are below the threshold and are suggestive for a selective sweep due to the drastically reduced genetic variability in the selected population in relation to both, randomly chosen reference loci and the same loci in the unselected population. The previously mapped major QTL_m region (6071kbp - 6409kbp between vertical dash lines) showed a selective sweep.

4 Discussion

4.1 Responses to the *N. ceranae* infection in drones of selected strain comparing to the unselected strain

4.1.1 Responses of innate immune genes

As shown before for honey bee workers (Antúnez et al., 2009), also in our experiments with honey bee drones immune genes were involved in the response towards *N. ceranae* infection, resulting in a general increase in immune activity in response to the infection. The innate immune gene expression levels from day one to five post infection in both the selected and the unselected strain were higher than in the uninfected controls. Hence the infection increased the expression levels of immune genes. However, the expression level between the selected strain and unselected strain was not significantly different. Only on day six post infection could we observe significantly reduced expression levels of innate immune genes in the unselected strain, similar to the depression of the immune system in workers (Antúnez et al. 2009).

Although *Nosema* was classified into the fungi (Liu et al., 2006) and is not a bacterial pathogen, the strongest response towards the *N. ceranae* infection was that of the bacterial recognition receptor *PGRPSC4300* of the Toll pathway. The membrane-spanning Toll receptor known to initiate the cell immune cascade (Imler et al., 2004) was also up-regulated as well as *Dorsal* which is involved in the production of melanization and antimicrobial effectors. After day four post infection, the overall immune activity decreased to reach expression levels close to the uninfected control group (for the selected strain) or even significantly less (for the unselected strain) on day six post infection. In spite of the receptor activation, the infection did apparently not further enhance the subsequent immune gene cascade of the drones at this stage.

In addition the trans-membrane receptor *Domeless* in JAK/STAT pathway was also up-regulated immediately at the beginning of the infection. But we did not find any expression differences neither on *hopscotch* nor *thioester* containing protein between the selected and unselected strain. *Dredd* was the only gene from the IMD pathway that showed a different expression on day six post infection between the selected and unselected strain. There is no clear cascade pattern whether the IMD and /or the JAK/STAT pathway initiated the response towards the *N. ceranae* infection, or had been secondarily activated by other triggers. Nevertheless, the *Toll* pathway showed a consistent up-regulating pattern compared to the unselected group, which suggest that the *Toll* pathway might be the major pathway responding the *N. ceranae* infection.

4.1.2 Spore load dynamic and survival

Higes et al. (2007) showed that the number of infected epithelial gut cells markedly increase on day six post infection and start degeneration after day seven post infection. Also in our study, the number of *Nosema* spores rapidly increased after four days post infection in both strains. The *N. ceranae* spore load was even higher in drones of the selected strain in conjunction with lower mortality than those of the unselected strain. The high mortality in the first few days after infection is not due to the replication of *Nosema* which occurs at a later stage. But adding *N. ceranae* spores to the inoculation diet did significantly increase the mortality, because the control drones survived. This may be due to the stress of the gut damage of the initial infection, but it may also be due to the behavior of the host workers towards the infected drones. It is important to note that these highly sensitive drones died and could not establish high spore loads after *Nosema* replication, which is why we found more spores in the selected strain. The variance in survival might result from the variance in the initial feeding of spore solutions. Although we tried to control the initial spore dose, variance among the

actual numbers of spore administered to the drones and germinated is inevitable. If only those susceptible drones which had received a low spore number survived, this would also in a low spore load in drones that survived until the end of the experiment. Hence spore number is a more complex trait than initially thought because it is confounded with survival. We checked the spore load in dead drones, which was not significantly different from the alive drones neither in the selected nor in the unselected strains. We do not know the actual mechanisms how drones of selected strain fight against *N. ceranae* and how an initial *N. ceranae* infection can kill the drones of unselected strain. Based on our preliminary histological dissections, the number of lysing epithelia cell was significantly higher in the unselected drones, which might generate a very high stress and eventually killed the infected drones.

The results from the experimental infections suggested, that the selected honey bees are tolerant to *N. apis* not only at the colony level but also at the individual level. The drones of the selected strain showed a very low mortality and a high tolerance to *N. ceranae* spore load. If not kept in the cages, under natural conditions at the apiary, the drones would simply expel the high spore load outside the colony during defecation flights.

4.2 QTLs associated with the tolerance towards *N. ceranae*

4.2.1 Major QTL in Chromosome 14

The selected honey bees showed a strong tolerance towards the *Nosema* infections (Huang et al. 2012). The genetic basis of this trait could also be reflected by the lower coefficient of variance for spore load within the selected (c_v=0.37) and the unselected strain (c_v= 0.43), compared to the coefficient of spore load variance in the mapping population (c_v=0.88) which originated from the F1 hybrid queen. Four QTLs were significantly associated with the *N. ceranae* spore load variance. With the average marker distance of ~20 cM, the power to identify a QTL explaining 7% of the total variance exceeds 95% (Rebai et al. 1995). Although QTLs with small effects might remain undetected, the chance is low that we have missed QTLs with larger effect than the identified significant QTLs. The QTL on chromosome 14 might play a major role (major QTL = QTL_m) for the low number of *Nosema* spores in the selected Danish honey bees, since it explained 7.7 % of the total variance. Even though the Danish allele of QTL_{ad} on chromosome 6 showed a significant association with the low spore load, this effect only occurred together with the Danish allele of the major QTL_m. Nevertheless, the combination of both loci might help to implement marker assisted breeding in the Danish population. In our study, as the *Nosema* tolerance was affected by this major QTL_m, a specific allele might have been positively selected during the breeding process. As a result, we expect a reduced genetic diversity in this major QTL_m region in the breeding population.

4.2.2 Epistatic effects

In addition to the additive gene effects also the epistatic interactions between loci on chromosomes 3 and 10 interfered with the spore load. In this case, it was the combination of Danish and French alleles that caused a low spore load. Epistatic

interactions had also been involved in *Varroa destructor* resistance, where the two interactive alleles were both from the same selected strain and hence suitable for the selective breeding (Behrens et al. 2011). In our case, the epistatic interactions cannot be used for selective breeding because it is exceptionally difficult to maintain the specific allele combination at the two loci to conserve the same epistatic effects in a selected strain. Nevertheless, this example showed again the power of using drones for QTL studies as it is very likely we would have missed any epistatic effects if we had used diploid workers instead.

4.2.3 Candidate gene

Even though the *Aub* was the prominent candidate gene, we can of course not exclude that also other candidate genes might be involved in the tolerance towards *N. ceranae*. From a functional perspective, *Aub* is of particular interest, as it is involved in regulating via RNA interference machinery known to interfere with foreign RNA (Liao et al. 2010; Teo et al. 2011). In *Drosophila*, *Aub* has been shown to be involved in the silencing of retrotransposons via RNA interference (Kawaoka et al. 2008). Moreover, it was reported to be involved in the resistance towards the gut bacterium *Serratia marcescens* infection in the *Drosophila* (Cronin et al. 2009). A reduction in *Aub* transcript levels in *Drosophila* resulted in an increase of mortality due to bacterial infection. This fits to our findings of increased transcript levels in low spore load drones. Unfortunately we do not know the exact mechanisms in which *Aub* is involved in the context of tolerance towards *N. ceranae* infections. *Aub* may trigger an apoptosis of the infected cells in order to prevent the production of spores, potentially with the help of immune genes (Higes et al 2013). Indeed six innate immune genes from three different immune pathways had significantly higher

expression levels in drones of the selected strain that in those of unselected strain (Huang et al. 2012). We also found the expression level of *Aub* to be higher in the selected strain, albeit not statistically significant which might be due to a low sample size. Nevertheless, this might reflect a general enhanced immune response of honey bees of the selected strain against *Nosema* infections which may have considerably contributed to the selection success of the Danish bee breeders.

4.3 'Foot print' of the selective breeding

4.3.1 Selective sweep

As the major QTL_m has been identified associated with the *Nosema* spore load (Huang et al. in revision), we expect a reduced genetic diversity within the major QTL_m region as a result of selective breeding. The low expected heterozygosity and the number of effective alleles clearly revealed that the genetic diversity was strongly reduced in the selected honey bee population. The comparison with an unselected control population revealed a selective sweep in the major QTL_m. The statistic method we used to detect the selective sweep was based on Kauer et al. (2003) who used all the markers spanning two chromosomes including the loci under selection as reference loci, because the candidate region under the selection is unknown. We used less reference loci because we had already information on a candidate QTL_m region which might be under selection. We further tested the influence of the number of reference loci on the ability to detect the selective sweep. One reference locus was dropped at a time for six times. So each reference locus was dropped once. We took the average of the mean and standard deviation of the rest five reference loci (-1.0873±1.0108) to re-detect the selective sweep and the same selective sweep was detected.

The genetic variance was most strongly reduced spanning across the three loci (UN271, K1452 and AT198) in the selected population. When we compared the genetic variance with the same loci in the control population, the locus K1452 showed the highest reduction of both genetic diversity and effective alleles (more than an order of magnitude). If this reflects the true locus under selection, the reduced diversities of loci UN271 and AT198 could be a result of hitchhiking with the actual locus under selection (Fay and Wu 2000). Given the selected allele detected in the

QTL mapping study is close to fixation, it appears that the classic selective sweep has driven the success of breeding program at the colony level. As we do not have the genetic information about the initial breeding population, we cannot conclude whether this selected allele is from the standing variation or a new mutation.

After over 20 years selective breeding, the number of alleles of reference loci in the breeding population is smaller but not significantly different from the unselected population (two tailed t test, $P > 0.05$). Since we included both neighboring loci and a set of unlinked reference loci to detect the selective sweep, we provided two independent sets of information to show that selection rather than random drift caused the reduction in the number of alleles in the sweep region. Given the high recombination rate (19cM/Mb) in the honey bee genome (The honeybee Genome Sequencing Consortium, 2006), the selective sweep in the selective Danish honey bee population must have occurred extremely fast and very recent. Otherwise the high recombination rate should have eroded the trace foot prints of the selection before the markers linked to the resistance locus became monomorphic. A similar case of extreme positive selection has been reported for various species including vector mosquitoes for malaria disease. The mosquitoes quickly became resistant to insecticide treatment and the genomic region controlling drug resistance could be identified by a selective sweep (Lynd et al. 2010; Norris and Norris, 2011).

4.3.2 Combination of colony level and individual level analysis of the tolerance towards *Nosema*

In our study, the selection was conducted at the colony level and it showed clear foot prints of positive selection. The effect of the major QTL_m on the individual bee was identified through the number of the *Nosema* spores in guts of individual bees (Huang et al. in revision). Experimentally infected selected bees also showed a significantly

higher tolerance towards *Nosema* infection than the unselected honey bees (Huang et al. 2012). Furthermore, a selective sweep was revealed as a result of colony level selection, with the swept loci and the QTL_m in the same genomic region (Huang et al. in revision). Hence we studied the tolerance towards the *Nosema* step by step, from the individual level artificial infection and QTL mapping to the colony level selective sweep analysis. The individual level and colony level analyses used different approaches, but the consistent results leading to the conclusion that the identified genetic region (QTL_m) is indeed associated with the tolerance towards the *Nosema*.

5 Summary and Outlook

- In Denmark, beekeepers have been selecting honey bee colonies that have had no *Nosema* infection for decades and today *Nosema* is rarely found in these breeding colonies. By the artificial infection of *N. ceranae* spores, the drones of the selected strain showed a significantly enhanced tolerance compared to the drones of the unselected strain.

- After artificial infection with *N. ceranae* spores, the differences in the expression levels of the innate immune genes between the selected and unselected drones were strongest at day six post infection. Six innate immune genes (*peptidoglycan recognition protein-SA, Toll, lysozyme-2, dorsal-2, Dredd, Domeless*) were significantly over-expressed in the selected drones. In particular, genes of the *Toll* pathway were consistently up-regulated in the selected honey bees, which is probably the main immune pathway involved in *N. ceranae* infection response.

- In order to further locate the genetic region involved in *N. ceranae* tolerance, we fed *N. ceranae* spores to drones of a hybrid queen, resulting from a cross of a selected queen and unselected drones. The spore loads of the infected drones were used as the phenotype to identify the QTL associated with the low number of *N. ceranae* spores. Four QTLs were identified significantly associated with the low *N. ceranae* spore load and explaining 20.4% of total spore load variance including a major QTL in chromosome 14, an additive QTL in chromosome 6 and two epistatic QTLs in chromosome 3 and 10 respectively.

- Within the identified major QTL region, the candidate gene *Aubergine* (*Aub*) was particular interest and significantly over expressed in drones with lower spore loads than those of higher spore load.

- In order to trace the "foot print" of the selective breeding, markers loci within and flanking the major QTL region were used to analyze the reduced genetic variability in selected Danish honey bees comparing with an unselected honey bee population. A selective sweep was identified within the major QTL region and the swept loci were close to fixation, which confirmed the importance of the major QTL for the *Nosema* tolerance.

- In this PhD thesis, individual level artificial infection and QTL mapping and the colony level selective sweep analysis used different approaches but obtain the same result which supports the notion that the identified genetic region is indeed associated with the tolerance towards the *N. ceranae*. Our results indicated the existence of a genetic basis of *N. ceranae* infection tolerance in honey bee *A. mellifera* and provided insights into the general genetic mechanisms underlying insect immunity.

- The work depicted in this PhD thesis represents the beginning of the study of the tolerance towards *Nosema*. Future work would be focused on the following aspects: (i) Silencing the candidate gene *Aub* with dsRNA to confirm its function of *Nosema* tolerance. (ii) Searching the tolerant allele of *Aub* and studying the impact of alternative splicing on the *Nosema* tolerance. (iii) Studying the expression pattern of the *Nosema* genome to uncover the impact of *Aub* on the *Nosema* reproduction.

6 Reference

Alam M.T., de Souza D.K., Vinayak S., Griffing S.M. *et al*. (2011) Selective sweeps and genetic lineages of Plasmodium falciparum drug -resistant alleles in Ghana. J. Infect. Dise. 203: 220-227.

Ammons A.D., Hunt G.J. (2008) Identification of quantitative trait loci and candidate genes influencing ethanol sensitivity in honey bees. Behav. Genet. 38: 531-553.

Anderson D., East I.J. (2008) The latest buzz about colony collapse disorder. Science 319: 724-725.

Antúnez K., Martín-Hernández R., Prieto L., Meana A., Zunino P., Higes M. (2009) Immune suppression in the honey bee (*Apis mellifera*) following infection by *Nosema ceranae* (Microsporidia). Environ. Microbiol. 11(9): 2284-2290.

Barton N.H., de Vladar H.P. (2009) Statistic mechanics and the evolution of polygenic quantitative trait. Genetics 181: 997-1011.

Behrens D., Huang Q., Geßner C., Rosenkranz P., et al. (2011). Three QTL in the honey bee *Apis mellifera* L. suppress reproduction of the parasitic mite *varroa destructor*. Ecol. Evol .doi: 10.1002/ece3.17.

Berryman A.A. (1992) The origins and evolution of predator-prey theory. Ecology 73: 1530-1535.

Bogdan C., Rollinghoff M., Diefenbach A. (2000) Reactive oxygen and reactive nitrogen intermediates in innate and specific immunity. Curr. Opin. Immunol. 12: 64-76.

Breeze T.D., Bailey A.P., Balcombe K.G. et al. (2011) Pollination services in the UK: How important are honeybees? Agric. Ecosyst. Environ. 142: 137-143.

Broman, K.W., Wu, H., Sen, S., Churchill, G.A. (2003) QTL mapping in experimental crosses. Bioinfromatics 19, 889-890.

Buzzatitraverso A.A. (1959) Quantitative triats and polygenic systems in evolution. Cold Spring harbor Symp. Quant. Biol. 24: 41-46.

Charlesworth B. (1996) Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. 68, 131-149.

Charlesworth B., Morgan M.T., Charlesworth D. (1993) The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303.

Chen Y.P., Evans J.D., Zhou L., Boncristiani H., Kimura K., Xiao T.G., Litkowski A.M., Pettis J.S. (2009) Asymmetrical coexistence of *Nosema ceranae* and *Nosema apis* in honey bees. J. Invertebr. Pathol. 101: 204-209.

Chevin L.M., Hospital F. (2008) Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180: 1645-1660.

Cook O.F. (1908) The Mendelian inheritance of mutation. Science 28: 86-88.

Cox-Foster D.L., Conlan S., Holmes E.C., Palacios G. et al. (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. Science 318: 283-287.

Cronin S.J.F., Nehme N.T., Limmer S., Liegeois S., Pospislik J.A., Schramek D. *et al*. (2009). Genome-wide RNAi screen identifies genes involved in intestinal pathogenic bacterial infection. Science 325: 340-343.

Dainat B., van Engelsdorp D., Neumann P. (2012) Colony collapse disorder in Europe. Env. Microbiol. Rep. 4: 123-125.

Dussaubat C., Brunet J.L., Higes M., Colbourne J.K., Lopez J. et al. (2012) Gut Pathology and Responses to the Microsporidium *Nosema ceranae* in the Honey Bee *Apis mellifera*. PLoS ONE 7(5): e37017. doi:10.1371/journal.pone.0037017

Dussaubat C., Sagastume S., Gomez-Moracho T., Botias C., Garcia-Palencia P. et al. (2013) Comparative study of *Nosema ceranae* (Microsporidia) isolates from two different geographic origins. Veter. Microbiol. 162: 670-678.

Evans J.D. (2006) Beepath: An ordered quantitative-PCR array for exploring honey bee immunity and disease. J. Invertebr. Pathol. 93:135-139.

Evans J.D., Aronstein K., Chen Y.P., Hetru C., Imler J.L., Jiang H. et al. (2006) Immune pathways and defence mechanisms in honey bees *Apis mellifera*. Insect. Mol. Biol. 15: 645-656.

Fay J.C., Wu C.I. (2000) Hitchhiking under positive Darwinian selection. Genetics 155: 1405-1413.

Forsgren E., Fries I. (2010) Comparative virulence of *Nosema ceranae* and *Nosema apis* in individual European honey bees. Vet. Pathol. 170: 212-217.

Frank S.A. (2012) Natural selection. III. Selection versus transmission and the levels of selection. J. Evol. Biol. 25:227-243

Fries I., Feng F., Silva A., Slemenda S.B., Pieniazek N.J., (1996) *Nosema ceranae* n sp (Microspora, Nosematidae), morphological and molecular characterization of a microsporidian parasite of the Asian honey bee *Apis cerana* (Hymenoptera, Apidae). Europ. J. Protistol. 32: 356-365.

Fries I. (2010) *Nosema ceranae* in European honey bees (*Apis mellifera*). J. Invertebr. Pathol. 103: 73-79.

Fries I., Martin R., Meana A., Garc á-Palencia P., Higes M. (2006) Natural infections of *Nosema ceranae* in European honey bees. J. Apic. Res. 45: 230-233.

Gertz J., Gerke J.P., Cohen B.A. (2010) Epistasis in a quantitative trait captured by a molecular model of transcription factor interactions. Theor. Popul. Biol. 77: 1-5.

Gillespie J.P., Kanost M.R., Trenczek T. (1997) Biological mediators of insect immunity. Annu. Rev. Entomol. 42: 611–643.

Gisder S., Hedtke K., Möckel N., Frielitz M.C., Linde A., Genersch E. (2010) Five-Year Cohort Study of *Nosema* spp. in Germany: Does Climate Shape Virulence and Assertiveness of *Nosema ceranae*? Appl. Environ. Microbiol. 9: 3032-3038.

Gisder S., Möckel N., Linde A., Genersch E. (2011) A cell culture model for *Nosema ceranae* and *Nosema apis* allows new insights into the life cycle of these important honey bee-pathogenic microsporidia. Environ. Microbiol. 13: 404-413.

Graaf D.C., Sabbe H.R., De Rycke P.H., Jacobs F.J. (1994) Early development of *Nosema apis* (Microspora: Nosematidae) in the Midgut Epithelium of the Honeybee (*Apis mellifera*). J. Invertebr. Pathol. 63: 74-81.

Hamiduzzaman, M.M., Guzman-Novoa, E., Goodwin, P.H. (2010) A multiplex PCR assay to diagnose and quantify *Nosema* infections in honey bees (*Apis mellifera*). J. Invertebr. Pathol. 105(2): 151-155.

Han B., Denlinger D.L. (2009) Mendelian inheritance of pupal diapauses in the flesh fly, Sarcophage bullata. J. Hered. 100: 251-255.

Haseman M.H. (1951) The influence of *Nosema apis* on the larval honeybee. Ann. Appl. Biol. 38: 844-846.

Haseman M.H. (1952) the effects of infection with *Nosema apis* on the pharyngeal salivary glands of the worker honey-bee. Proc. Roy. Entomol. Soc. London Ser. A 27: 22-27.

Higes M., Martín R., Meana A. (2006) *Nosema ceranae*, a new microsporidian parasite in honeybees in Europe. J. Invertebr. Pathol. 92: 93–95.

Higes M., García-Palencia P., Martín-hernández R., Meana A. (2007) Experimental infection of *Apis mellifera* honeybees with *Nosema ceranae* (Microsporidia). J. Invertebr. Pathol. 94: 211-217.

Higes M., Martin-Hernandez R., Botias C., Bailon E.G., Gonzalez-Porto A.V., Barrios L. et al. (2008) How natural infection by *Nosema ceranae* causes honeybee colony collapse. Environ. Microbiol. 10: 2659-2669.

Higes M., Martín-Hernández R., Garrido-Bailón E., González-Porto A.V., García-Palencia P., Meana A. et al. (2009) Honeybee colony collapse due to *Nosema ceranae* in professional apiaries. Environ. Microbiol. 1: 110-113.

Higes M., Juarran A., Dias-Almeida J., Lucena S., Bitias C. et al. (2013) Apoptosis in the pathogenesis of *Nosema ceranae* (*Microsporidia: Nosematidae*) in honey bees (*Apis mellifera*). Environ. Microl. Rep. DOI: 10.1111/1758-2229.12059

Hornik K., 2011. The R FAQ. *ISBN 3-900051-08-9* 2011, http://CRAN.R-project.org/doc/FAQ/R-FAQ.html.

Huang Q., Kryger P., Le Conte Y., Moritz R.F.A. (2012) Survival and immune response of drones of a *Nosemosis* tolerant honey bee strain towards *N. ceranae* infections). J. Invertebr. Pathol. 109: 397-302.

Huang Q., Kryger P., Le Conte Y., Lattorff H.M.G., Kraus F.B., Moritz R.F.A. (in revision) Four QTLs associated with low *Nosema ceranae* (Microsporidia) spore load in the honey bee *Apis mellifera*.

Huang Q., Lattorff H.M.G., Kryger P., Le Conte Y., Moritz R.F.A. (in revision) A selective sweep in a Microspordian parasite *Nosema* tolerant honeybee population, *Apis mellifera*.

Hunt G.J., Collins A.M., Rivera R., Page R.E., Guzman-Novoa E. (1999) Quantitative trait loci influencing honeybee alarm pheromone levels. J. Hered. 90: 585-589.

Imler J.L., Ferrandon D., Royet J., Reichhart J.M., Hetru C., Hoffmann J.A. (2004) Toll-dependent and Toll-independent immune responses in *Drosophila*. J. Endotoxin. Res. 4(10): DOI 10.1179/096805104225005887.

Ish-Am G., Lahav E. (2011) Evidence for a major role of honeybee (*Apis mellifera*) rather than wind during avocado (*Persea americana* Mill.) pollination. J. Hortic. Sci. Biotech. 86: 589-594.

Jarosch A., Moritz R.F.A. (2011) Systemic RNA-interference in the honeybee *Apis mellifera*: Tissue dependent uptake of fluorescent siRNA after intra-abdominal application observed by laser-scaning microscopy. J. Insect. Physiol. 57(7): 851-857.

Jiang H., Wang Y., Ma C., Kanost M.R. (1997) Subunit composition of pro-phenoloxidase from Manduca sexta: molecular cloning of subunit proPO–P1. Insect Biochem. Mol. Biol. 27: 835-850.

Johnson R.M., Evans J.D., Robinson G.E., Berenbaum M.R. (2009) Changes in transcript abundance relating to colony collapse disorder in honey bees (*Apis mellifera*). PNAS 106: 14790-14795.

Kawaoka S., Minami K., Katsuma S., Mita K., Shimada T. (2008) Developmentally synchronized expression of two *Bombyx mori* Piwi subfamily genes, SIWI and BmAGO3 in germ-line cells. Biochem. Biophys. Res. Commun. 367: 755–760.

Kauer M.O., Dieringer D., Schlötterer C. (2003) A micro-satellite variability screen for positive selection associated with the ‘‘out of Africa’’ habitat expansion of *Drosophila melanogaster*. Genetics 165: 1137–1148.

Keeling P.J., Fast N.M. (2002) Microsporidia: Biology and evolution of highly reduced intracellular parasites. Annu. Rev. Microbiol. 56: 93-116.

Kerr W.E. (1975) Evolution of population structure in bees. Genetics 79: 73-84.

Kim Y., Stephan W. (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160: 765–777.

Klee J., Besana A.M., Genersch E., Gisder S., Nanetti A., Tam D.Q. et al. (2007) Widespread dispersal of the microsporidian *Nosema ceranae*, an emergent pathogen of the western honey bee, *Apis mellifera*. J. Invertbr. Pathol. 9: 1-10.

Koch H., Schmid-Hempel P. (2011) Bacterial communitites in central European Bumblebees: Low diversity and high specificity. Microb. Ecol. 62: 121-133.

Kralj J., Fuchs S. (2010) *Nosema* sp. Influences flight behavior of infected honey bee (*Apis mellifera*) foragers. Apidologie 41: 21-28

Kraus F.B., Koeniger N., Tingek S., Moritz R.F.A. (2005) Using drones for estimating colony number by microsatellite DNA analyses of haploid males in *Apis*. Apidologie 36: 223-229.

Liao Z., Jia Q.D., Li F., Han Z.J. (2010). Identificaion of two Piwi genes and their expression profile in honeybee, *Apis mellifer*a. Arch. Insect. Biochem. Physiol. 74: 91-102.

Liu Y.J., Hodson M.C., Hall B.D. (2006) Loss of the flagellum happened only once in the fungal lineage: phylogenetic structure of kingdom Fungi inferred from RNA polymerase II subunit genes. BMC. Evol. Biol. 6: 74.

Lynd A., Weetman D., Barbosa S., Yawson A.E., Mitchell S. *et al*. (2010) Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in *Anopheles gambiae s.s*. Mol. Biol.Evol. 27: 1117-1125.

Mackensen O. (1947) Effect of carbon dioxide on initial oviposition of artificially inseminated and virgin queen bees. J. Econ. Ent. 40: 344-349.

Mackensen O. (1951) Viability and sex determination in the honey bee (*Apis mellifera* L.). Genetics 36: 500-569.

Mackay T.F.C., Fry J.D. (1996) Polygenic mutation in *Drosophila melanogaster*: Genetic interactions between selection lines and candidate quantitative trait loci. Genetics 144: 671-688.

Mayack C., Naug D. (2009) Energetic stress in the honeybee *Apis mellifera* from *Nosema ceranae* infection. J. Invertebr. Pathol. 100: 185-188.

Maynard S.J., Haigh J. (1974) Hitch-hiking effect of a favorable gene. Genet. Res. 23: 23-25.

Moritz R.F.A. (1984) The effect of different diluents on insemination success in the honeybee using mixed semen. J. Apicul. Res. 23, 164-167

Moritz R.F.A. (1986) Estimating the genetic variance of group characters: social behavior of the honeybee (*Apis mellifera*). Theor. App. Genet. 72: 513-517.

Moritz R.F.A. (1988) A reevaluation of the two-locus model for hygienic behavior in honeybees (*Apis mellifera* L.). J. Hered. 79: 257-262.

Moritz R.F.A., Southwick E.E. (1992) Bees as superorganisms. Springer-Verlag, Heidelberg.

Moritz R.F.A. and Evans J.D. (2008) Honeybee genomics and breeding for resistance to virus infections. In: Virology and the Honey Bee (eds Aubert M, et al) Europ. Comm. Brussels. pp: 347-370.

Muta T., Iwanaga S. (1996) The role of hemolympyh coagulation in innate immunity. Curr. Opin. Immunol. 8: 41–47.

Nagylaki T. (1985) Homozygosity, effective number of allleles, and interdeme differentiation in subdivided populations. PNAS 82: 8611-8613.

Noor M.A.F., Chang A.S. (2009) Epistasis modifies the dominance of loci causing hybrid male sterility in the *Drosophila pseudoobscura* species group. Evolution 64: 253-260.

Norris L.C., Norris D.E. (2011) Insecticide resistance in Culex quinquefasciatus mosquitoes after the introduction of insecticide-treated bed nets in Macha, Zambia. J. Vect. Ecol. 36: 411-420.

Ohta T., Kimura M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. 22: 201–204.

Owen R.E., Owen A.R.G. (1989) Effective population size in social Hymenoptera with worker-produced males. Heredity 63: 59-65.

Oxley P.R., Spivak M., Oldroyd B.P. (2010) Six quantitative trait loci influence task threshold for hygienic behavior in honeybees (*Apis mellifera*). Mol. Ecol. 19: 1452-1461.

Palaisa K., Morgante M., Tingey S., Rafalski A. (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. PNAS 101: 9885-9890.

R Devolopment Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rana T.M. (2007) Illuminating the silence: understanding the structure and function of small RNAs. Mol. Cell Biol. 8: 23-36

Rebai A., Goffinet B., Mangin B. (1995) Comparing power of different methods for QTL detection. Biomethics 51: 87-99.

Ribiere M., Olivier V., Blanchard P., Schurr F. (2008) The collapse of bee colonies: the CCD case ("Colony collapse disorder") and the IAPV virus (Israeli acute paralysis virus). Virology 12: 319-322.

Rothenbuhler W.C. (1964) Resistance to American foulbrood in honey bees: I. Differential survival of larval of different genetic lines. Am. Zool. 4: 111-123.

Rüppelll O., Pankiw T., Page R.E. (2004) Pleiotropy, epistasis and new QTL: The genetic architecture of honey bee foraging behavior. J. Hered. 95: 481-491.

Rüppell O. (2009) Characterization of quantitative trait loci for the age of first foraging in honey bee workers. Behav. Genet. 39: 541-553.

Sass M., Kiss A., Locke M. (1994) Integument and hemocyte peptides. J. Insect Physiol. 40: 407-421.

Schlötterer C. (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics 160: 753–763.

Schlüns H., Moritz R.F.A., Neumann P., Kryger P., Koeniger G (2005) Multiple nuptial flight, sperm transfer and the evolution of extreme polyandry in honeybee queens. Anim. Behav. 70: 125-131.

Schmid-Hempel P. (2005) Evolutionary ecology of insect immune defenses. Annu. Rev. Entomol. 50: 529-551.

Schmid-Hempel P. (2009) Immune defence, parasite evasion strategies and their relevance for 'macroscopic phenomena' such as virulence. Phil. Trans. R. Soc. B 364: 85-98.

Seeley T.D. (1995) The wisdom of the hive, Harvard University Press, Cambridge, Massachusetts.

Shao Q.M., Yang B., Xu Q.Y. et al. (2012) Hindgut Innate Immunity and Regulation of Fecal Microbiota through Melanization in Insects. J. Biol. Chem. 287: 14270-14279.

Snodgrass R.E. (1956) Anatomy of the honey bee. Cornell University Press, UAS.

Spiess A.N., Ritz C. (2010) qpcR: Modelling and analysis of real-time PCR data. R package version 1.3-4.

Stephan W. (2010) Genetic hitchhiking versus background selection: the controversy and its implications. *Phil. Trans. R Soc. B* 365: 1245-1253.

Strand M.R., Pech L.L. (1995) Immunological basis for compatibility in parasitoid–host relationships. Annu. Rev. Entomol. 40: 31-56.

Teo C.H., Pui H.P., Othman R.Y., Harikrishna J.A. (2011). Comparative analysis of Argonaute gene sequences in bananas (Musa sp.) shows conserved species-specific Ago-7 PIWI domains. Genet. Resour. Crop Evol. 58: 713-725.

The Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443, 931-949.

Traynor K. (2008) Bee breeding around the world. Am. Bee. J. 148: 135-139.

Tsakas S., Marmaras V.J. (2010) Insect immunity and its signalling: an overview. Invertebr. Survival 7: 228-238.

Van Hiel M.B.V., Van Wielendaele P., Temmerman L., Van Soest S., Vuerinckx K. et al. (2009) Identification and validation of housekeeping genes in brains of the desert locust *Schistocerca gregaria* under different developmental conditions. BMC Mol. Bilol. doi: 10.1186/1471-2199-10-56

Vass E., Nappi A.J. (2001) Fruit fly immunity. BioEssays 51: 529-535.

Walsh P.S., Metzger D.A., Higuchi R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10: 506-513.

Wang S., Basten C.J., Zeng Z.B. (2011) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Wang J.L. (2004) Sibship reconstruction from genetic data with typing errors. Genetics 166: 1963-1979.

Weller J.I., Soller M., Brody T. (1987) Linkage analysis of quantitative traits in an interspecific cross of Tomato (*Lycopersicon esculentum* x *Lycopersicon pimpinellifolium*) by means of genetic markers. Genetics 118: 329-339.

Wiehe T., Nolte V., Zivkovic D., Schlötterer C. (2007) Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. Genetics 175: 207-218.

Williams B.A. (2009) Unique physiology of host–parasite interactions in microsporidia infections. Cell Microbiol. 11: 1551-1560.

Wilson-Rich N., Spivak M., Fefferman N.H., Starks P.T. (2009) Genetic, individual, and group facilitation of disease resistance in insect societies. Annu. Rev. Entomol. 54: 405-423.

Yang X., Cox-Foster D.L. (2005) Impact of an ectoparasite on the immunity and pathology of an invertebrate: evidence for host immunosuppression and viral amplification. PNAS 21: 7470–7475.

Zander E. (1909) Tierische Parasiten als Krankenheitserreger bei der Biene. Münchener Bienenzeitung 31: 196–204.

Zander E. (1909) Tierische Parasiten als Krankenheitserreger bei der Biene. Münchener Bienenzeitung 31: 196–204.

7 Appendix

7.1 Acknowledgements

First of all, I give my deepest gratitude to my supervisor Prof. Robin F.A. Moritz for providing me the opportunity to work in his lab. I made mistakes. I addressed the comments of the reviewers poorly. But Robin always guides me to the right direction with extreme patience. So I here especially appreciate his extreme patience on the data discussion and the revision of the manuscripts. His enthusiasm and conscientious attitude on science deeply inspire me with of great benefit and encouragement for my future career.

I would like to give my special thanks to Dr. H. Michael G. Lattorff. He can never spend over his energy and enthusiasm on the scientific research and his intelligent suggestions are essential to the success of this thesis.

I am grateful to Dr. Per Kryger providing critical honey bees and Prof. Yves Le Conte for organizing the cross breeding. Without their help, this work cannot be done.

I give my special thanks to Molecular Ecology group. The relaxing and amusing atmosphere makes the work enjoyable. They helped me a lot both in the work and the living in Halle. My special regards to Dieter Behrens, Silvio Erler, Bernhard Kraus, Stephan Wolf, Petra Weber, Petra Leibe, Bertrand Fouks, Mattias Müller, Susann Parsche, Gaetano Solazzo, Anett Huth Schwarz, Eckart Stolle, Hans-Hinrich Kaatz, Holger Scharpenberg.

Last but not least, I thank all my friends, my family, especially my super wonderful wife and super active son. Without their support, this work cannot be done.

The work was conducted in the Institute of Biology within the Molecular Biology research group of Prof. Robin F.A. Moritz and finically supported by European Union's Seventh Framework Programme BeeDoc; Deutsche Forschungsgemeinschaft (DFG) and China Scholarship Council.

7.2 Curriculum vitae

Personal Information

Name: Qiang Huang

Nationality: China

Born: 21 Jan. 1983 in Jiujiang

Spouse: Jia Lan, Zheng

Son: Le Chen, Huang

Education

Since 2009 PhD student

Martin-Luther Universität Halle-Wittenberg

Institute of Biology/Zoology, Molecular Ecology

Supervisor: Prof. Robin F.A. Moritz

Thesis Title: Genetic analysis of *Nosema* tolerance in the honey bee, *Apis mellifera*

2006 - 2009 Master of Science

Jiangxi Agricultural University, College of Animal Science and Technology

Supervisor: Prof. Zhi Jiang Zeng

Thesis title: Study on the natural swarming behavior of *Apis cerana cerana*

2001 - 2005 Bachelor

Jiangxi Agricultural University, College of Animal Science and Technology

7.3 Publication list

Huang Q., Kryger P., Le Conte Y., Moritz R.F.A. (2012) Survival and immune response of drones of a *Nosemosis* tolerant honey bee strain towards *N. ceranae* infections. Journal of Invertebrate Pathology, 109: 297-302.

Behrens D., **Huang Q**., Geßner C., Rosenkranz P., Frey E., Locke B., Moritz R.F.A., Kraus F.B. (2011) Three epistatic QTL in the honeybee *Apis mellifera* L. suppress reproduction of the parasitic mite Varroa destructor. Ecology and Evolution, doi: 10.1002/ece3.17.

Huang Q., Kryger P., Le Conte Y., Lattorff H.M.G., Kraus F.B., Moritz R.F.A. (in revision) Four QTLs associated with low *Nosema ceranae* (Microsporidia) spore load in the honey bee *Apis mellifera*.

Huang Q., Lattorff H.M.G., Kryger P., Le Conte Y., Moritz R.F.A. (in revision) A selective sweep in a Microspordian parasite *Nosema* tolerant honeybee population, *Apis mellifera*.

7.4 Oral presentations

Bioinformatics Workshop, Münster, Germany (12/2010): QTL screening towards *Varroa destructor* tolerance in honeybee *Apis mellifera*.

2011 International Social Insect Genomics Research Conference, Shenzhen, China (07/2011): An insight into host-parasite co-evolution from *Nosema ceranae* tolerant honey bees.

5th European conference of Apidology, Halle, Germany (09/2012): Quantitative trait locus and selective sweep analysis for *Nosema ceranae* tolerane in the honeybee *Apis mellifera*.

7.5 Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst sowie keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen wurden als solche von mir kenntlich gemacht. Ich erkläre, dass ich mich bisher noch nicht um den Doktorgrad beworden habe.

Halle (Saale), 8 May 2013

Qiang Huang