A selective sweep in a microsporidian parasite *Nosema*-tolerant honeybee population, *Apis mellifera*

Q. Huang*†, H. M. G. Lattorff*‡[1], P. Kryger§, Y. Le Conte¶ and R. F. A. Moritz*‡††**

*Institut für Biologie/Zoologie, Molekulare Ökologie, Martin-Luther-Universität Halle -Wittenberg, Hoher Weg 4, 06099, Halle (Saale), Germany. †Honeybee Research Institute, Jiangxi Agricultural University, 330045, Nanchang, China. ‡German Institute for Integrative Biodiversity Research (iDiv), Bio City, 04103 Leipzig, Germany. §Department of Agroecology, Section of Entomology and Plantpathology, Aarhus University, Flakkebjerg, 4200, Slagelse, Denmark. ¶INRA, UMR 406 Ecologie des Invertébrés, Site Agroparc, 84914, Avignon Cedex 9, France. **RoBeeTech, Universitatea de Stiinte Agricole si Medicina Vetereinaria Cluj-Napoca, 400372 Cluj-Napoca, Romania. ††Department of Zoology and Entomology, University of Pretoria, Pretoria, 0002, South Africa.

Summary

Nosema is a microsporidian parasite of the honeybee, which infects the epithelial cells of the gut. In Denmark, honeybee colonies have been selectively bred for the absence of *Nosema* over decades, resulting in a breeding line that is tolerant toward *Nosema* infections. As the tolerance toward the *Nosema* infection is a result of artificial selection, we screened chromosome 14 for a selective sweep with microsatellite markers, where a major quantitative trait locus (QTL) had been identified to be involved in the reduction in *Nosema* spores in the honeybees. By comparing the genetic variability of 10 colonies of the selected honeybee strain with a population sample from 22 unselected colonies, a selective sweep was revealed within the previously identified QTL region. The genetic variability of the swept loci was not only reduced in relation to the flanking markers on chromosome 14 within the selected strain but also significantly reduced compared with the same region in the unselected honeybees. This confirmed the results of the previous QTL mapping for reduced *Nosema* infections. The success of the selective breeding may have driven the selective sweep found in our study.

Keywords honeybee, genetic diversity, *Nosema*, resistance, selective breeding

Introduction

Nosema is a microsporidian parasite of the honeybee (*Apis mellifera*), which infects the epithelial cells of the gut (Zander 1909). *Nosema* lives as an obligate intracellular parasite. The infection starts from the ingestion of the spores, which germinates in the midgut and extrudes the polar tubes that penetrate the epithelial cells to release the sporoplasm into the cytoplasm (Higes *et al.* 2007; Fries 2010). The infected cells eventually will burst and release a new generation of spores (de Graaf *et al.* 1994; Gisder *et al.* 2011). The offspring spores either can germinate to infect new host cells or might be expelled through the feces (Gisder *et al.* 2011).

Nosema infection affects honeybees in multiple ways (Haseman 1951, 1952; Huang *et al.* 2012) including increased mortality (Mayack & Naug 2009). Two *Nosema* species are known to infect the honeybee *A. mellifera*: *N. apis* and *N. ceranae*. *N. apis* is an evolutionarily old pathogen of *A. mellifera* with a moderate virulence. The host–parasite coevolutionary relationship is well balanced, and colonies often can cure themselves under favorable environmental conditions (Zander 1909; Chen *et al.* 2009). *Nosema ceranae* originally was found in the Asian honeybee *A. cerana* (Fries *et al.* 1996) and is a newly established parasite of *A. mellifera* (Fries *et al.* 2006; Higes *et al.* 2006). Although it has been reported to have caused large colony losses (Higes *et al.* 2008, 2009), there are increasing reports of a moderate virulence of *N. ceranae* similar to that of *N. apis* (Forsgren & Fries 2010; Fries 2010; Gisder *et al.* 2010).

Irrespective of its virulence, *Nosema* adds to the pathogen load in honeybee colonies and reduces their productivity (Kralj & Fuchs 2010). This is why bee keepers in Denmark

Address for correspondence

Q. Huang, Institut für Biologie/Zoologie, Molekulare Ökologie, Martin-Luther-Universität Halle -Wittenberg, Hoher Weg 4, 06099 Halle (Saale), Germany.
E-mail: qiang.huang@zoologie.uni-halle.de

[1]Present address: Institut für Biologie, Tierphysiologie Martin-Luther-Universität Halle-Wittenberg Domplatz 4 06099 Halle (Saale) Germany.

embarked on a selection program aiming at *Nosema* absence since the 1980s (Traynor 2008). Originally, 500 colonies (*A. mellifera*) were involved in the selective breeding, reflecting an effective population size of about $n_e = 1846$ (Kerr 1975; Owen & 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 queens were replaced with queens from *Nosema* absent colonies whenever the workers were infected by *Nosema*. Hence, selection was performed at the colony level without knowing which biological mechanism drove *Nosema* absence in the sampled workers. Nevertheless, this selective breeding should have left 'footprints' of selection in the Danish breeding population whether one or a few major genes determined the free of infection phenotype. If a specific beneficial allele was selected for, this allele might eventually become fixed in the breeding population, known as a hard sweep or classic selective sweep (Palaisa *et al.* 2004; Hermisson & Pennings 2005). Neighboring neutral markers closely linked to the locus under selection also should show reduced variability because of genetic hitchhiking (Maynard Smith & Haigh 1974; Charlesworth *et al.* 1993; Chevin & Hospital 2008). Alternatively, the bee breeders might have removed a susceptibility allele (or alleles) from the population. In this case, more than one allele might be selected for, and the genetic variability of the surrounding neutral loci should also be reduced, known as soft sweep (Hermisson & Pennings 2005).

In the honeybee, selection can operate at two different levels: the individual level and the colony level. The phenotype of the colony may not just reflect the genotype of the queen but rather the combined effect of all colony members based on the genotypic composition of the entire colony (Moritz 1986). Hence, detecting *Nosema*-infected workers might be due to the individual genotype but might also be due to the genotypic composition of the entire colony. In spite of these genetic complications, Huang *et al.* (in press) detected a major quantitative trait locus (QTL) on chromosome 14, which was associated with the reduction in the *Nosema* spores in experimentally infected individual haploid male bees. This made it a promising candidate QTL region that might be responsible for the *Nosema*-tolerant phenotype at the colony level. This QTL region (6071–6409 kbp on chromosome 14, Amel 4.5) was identified by individually genotyping a mapping population of 148 drones with 221 heterozygous microsatellite markers spanning the full genome. If this QTL is responsible for the breeding success, then positive selection might have generated a selective sweep in this region given the very recent selection events. Here, we used microsatellite markers within and flanking the identified QTL region to screen for a reduction in genetic variability in the same selected Danish honeybee strain providing the mapping population for the QTL identification. Moreover, we compared the genetic variability in this region with an unselected honeybee population to analyze the potential selective sweep (Kim & Stephan 2002; Nielsen *et al.* 2005).

Materials and methods

Drones collection and DNA extraction

Ten drones per colony were collected from 10 hives from the Danish selected population in Slagelse, Denmark. Moreover, 32 drones were collected from a non-selected control population at a drone congregation area (DCA) in Halle, Germany. Sampling drones on a DCA remove the bias of sampling specific breeding lines at apiaries, because 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 a sample of the specific breeding strain. All the drones were stored in 75% ethanol at −20 °C until the DNA was extracted using 5% Chelex-100 (Bio-Rad; Walsh *et al.* 1991).

Genotyping

The drone 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. Ten fluorescence-labeled microsatellite markers within (*UN271* and *K1452*) and flanking (*UNEV2, K1453, BI116, K1418B, AT198, K1424, SV188* and *HYAL*) the QTL region were used to assess the genetic variability (Table 1). 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 (Table 1). Each multiplex PCR contained 1 μl DNA (50 ng/μl), 5 μl master mix (Promega) and 0.4 μl/primer (10 mM, 6–8 primers per reaction), adding water for a final volume of 10 μl. The PCR cycle was as follows: 95 °C for 15 s; 55 °C for 30 s; and 72 °C for 30 s for 40 cycles. The allele sizes were determined using MegaBace 1000 capillary DNA sequencer (Amersham Biosciences) and scored with MegaBACE Fragment Profiler version 1.2.

Population genetic analysis

To detect the presence of a selective sweep, we estimated the expected heterozygosity (H_e) for each locus based on the

Table 1 Locus, chromosome, physical position, expected heterozygosity and the number of effective alleles for the 16 genotyped loci.

Reference loci	Ch.	Kb	H_{eS}	A_{eS}	H_{eC}	A_{eC}	lnRθ	Mean ± SD		
AC127	1	1,109	0.72	3.52	0.69	3.22	0.12	1.09 ± 1.01		
6701	3	9,693	0.77	4.42	0.80	4.92	−0.22			
K0616	6	11,669	0.52	2.09	0.74	3.90	−1.44			
K0808	8	4,417	0.49	1.96	0.83	5.80	−2.44			
AT168	8	5,381	0.65	2.88	0.86	7.27	−1.96			
AT129	10	11,106	0.39	1.65	0.52	2.07	−0.64			
Tested loci	Ch.	Kb	H_{eS}	A_{eS}	H_{eC}	A_{eC}	lnRθ	*T* value	Original *P* value	Adjusted *P* value
UNEV2	14	4,466	0.51	2.02	0.70	3.30	−1.17	0.06	0.48	>0.05
K1453	14	4,684	0.63	2.68	0.46	1.85	0.94	−1.58	0.91	>0.05
BI116	14	4,825	0.65	2.90	0.89	8.83	−2.34	0.96	0.19	>0.05
K1418B	14	5,355	0.35	1.54	0.80	4.95	−2.83	1.34	0.11	>0.05
UN271	14	6,124	0.10	1.11	0.72	3.53	−3.89	2.17	0.04	>0.05
K1452	14	6,265	0.10	1.11	0.91	11.5	−6.33	4.06	0.0049	<0.05*
AT198	14	6,953	0.10	1.11	0.88	8.12	−5.62	3.51	0.0085	>0.05
K1424	14	7,174	0.53	2.11	0.53	2.13	−0.02	−0.83	0.78	>0.05
SV188	14	7,432	0.64	2.75	0.52	2.08	0.68	−1.38	0.89	>0.05
HYAL	14	8,319	0.52	2.09	0.37	1.59	0.78	−1.45	0.90	>0.05

Kb, physical position on the chromosome in Kbp (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; and lnRθ, natural logarithm transformation of the ratio between the two populations.
*$P < 0.05$, Bonferroni adjusted for 10 comparisons, one tailed *t*-test.

reconstructed queen genotypes and corrected for the sample size (Alam *et al.* 2011) as follows:

$$H_e = \frac{n}{n-1}(1 - \sum p_i^2),$$

where n = the number of sampled chromosomes and $p_i = i^{th}$ allele frequency at a locus.

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

$$A_e = 1/(1 - H_e),$$

where H_e = the expected heterozygosity of a locus.

To account for the initial population differences 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 & Kimura 1973; Wiehe *et al.* 2007) as follows:

$$\text{lnR}\theta = \ln \frac{\left(\frac{1}{1-H_{\text{selected}}}\right)^2 - 1}{\left(\frac{1}{1-H_{\text{control}}}\right)^2 - 1},$$

where H_{selected} = expected heterozygosity of the selected population and H_{control} = expected heterozygosity of the unselected control population.

As lnRθ of unlinked loci have been shown to follow a normal distribution (Kauer *et al.* 2003; Wiehe *et al.* 2007), we compared the lnRθ of the reference loci with each of the 10 target loci on chromosome 14 using a *t*-test. The *t* value of 10 tested loci on chromosome 14 was calculated to indicate the existence of the selective sweep as follows:

$$t = \frac{\text{Mean}_1 - \text{Mean}_2}{\sqrt{\left(\frac{N_1 * S^2 + N_2 * S^2}{N_1 + N_2 - 2}\right) * \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}},$$

where $t = t$ value of each locus in chromosome 14, Mean_1 = mean lnRθ value of reference loci, Mean_2 = lnRθ value of each locus in chromosome 14, N_1 = number of lnRθ value of reference loci, N_2 = number of lnRθ value of the tested loci and S = standard deviation of reference loci.

Results

Expected heterozygosity

In the selected population, the genotypes of 10 queens were successfully reconstructed. Because 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 10 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 (Table S1). In all cases, both maternal alleles were detected in each set of 10 drones. By calculating the relatedness among drones of the control population, 29 sets of

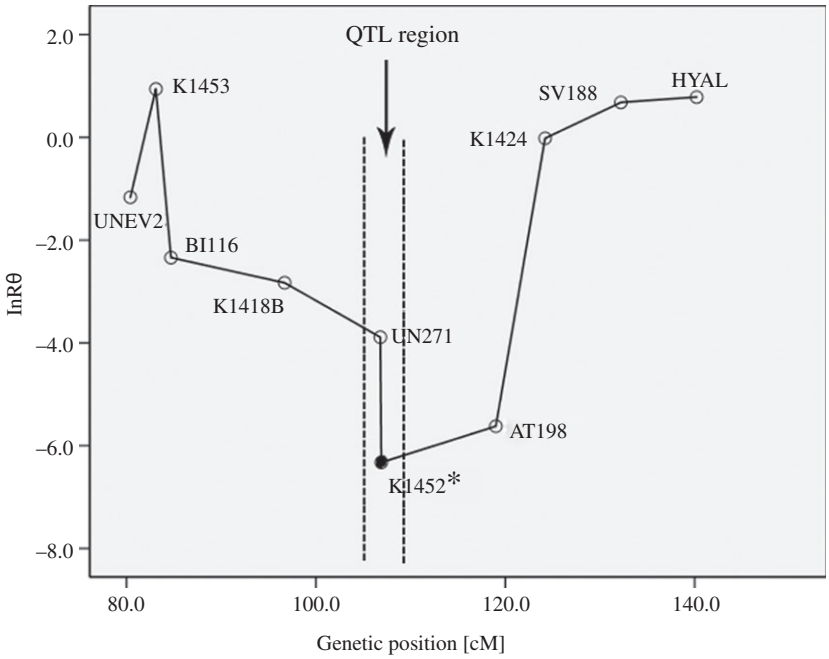

Figure 1 The lnRθ between the selected and unselected control populations on chromosome 14. The locus *K1452* was swept due to the drastically reduced genetic variability in the selected population in relation to both randomly chosen marker loci and the same loci in the unselected population. The previously mapped quantitative trait locus region for reduced number of *Nosema* spores showed a selective sweep. *$P < 0.05$, Bonferroni adjusted for 10 comparisons.

chromosomes were constructed from 32 drones that were classified into 22 ± 1.1 colonies (Table S1). The allele frequencies were calculated based on the 29 sets of chromosomes (Table S1) to avoid an estimation bias due to sampling drones with the same chromosomal set that originated from the same mother queen. The loci *UN271, K1452* and *AT198* had the same lowest genetic variability in the selected population. The expected heterozygosity was $H_e = 0.1$ for these three loci, which was much lower 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 did 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 1).

The number of effective alleles

To visualize the potential allele fixation, we compared the number of effective alleles (A_e) of the two populations. 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 was 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 ($A_e = 4.5 \pm 0.76$) in the selected population (Table 1).

Selective sweep

To calculate the t value to detect the existence of the selective sweep, we compared the six reference loci (-1.09 ± 1.01, mean $\pm$ SD) with each of 10 target locus with a two-sample t-test by assuming equal variance. The locus *K1452* showed a significantly lower genetic diversity in the selected population than did the control population compared with the reference loci (df = 5, $t = 4.06$, one-tailed t-test, $P < 0.05$; Bonferroni adjusted for 10 comparisons), which gave evidence for the existence of the selective sweep to reduce *Nosema* infection (Fig. 1 and Table 1). To obtain confidence limits testing the power of the six reference loci, we used a jack-knifing resampling approach over loci. Thus, the mean and standard deviation of five reference loci was calculated six times, leaving out one locus each. Testing this jack-knifed overall mean (-1.0873 ± 1.0108) vs. the swept locus *K1542* confirmed the selective sweep of the locus *K1452* (df = 4, $t = 3.88$, $P < 0.01$, one-tailed t-test). Furthermore, if the sweep was

false positive due to a monomorphic locus in the initial breeding population, the diversity of the neighboring loci should not have been affected. A regression of the diversity ($\ln R\theta$) of seven adjacent loci on the genetic distance (cM) to the target locus yielded $R^2 = 0.7$ ($P = 0.015$), which gave additional support for the swept region.

Discussion

The low expected heterozygosity and the low number of effective alleles clearly revealed that the genetic diversity in the QTL region for *Nosema* tolerance was strongly reduced in the selected honeybee population. The comparison with an unselected control population revealed a selective sweep in the same genomic region that was associated with the reduction in *Nosema* spores in the selected honeybee population. The statistic method we used to detect the selective sweep was based on Kauer *et al.* (2003). They used a large marker set spanning two chromosomes including the loci under selection as reference loci, because in their case, the candidate region under selection was unknown. We used fewer reference loci because we already had information on a candidate sweep region from a QTL mapping study. The genetic variance was most strongly reduced spanning 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 in both genetic diversity and effective alleles (more than an order of magnitude).

If locus *K1424* reflects the true sweep, the reduced diversities of loci *UN271* and *AT198* could be a result of hitchhiking with the actual locus under selection (Fay & Wu 2000). Surprisingly, the locus *K1424* was heterozygous in all queens of the selected strain. But the effective number of alleles and expected heterozygosity were similar between the selected and unselected population. Based on our current knowledge on pedigree information, we have no other plausible explanation for the consistent heterozygosity other than chance. It seems very unlikely that we found a genetic system similar to the genetic load at the sex locus with lethal selection against homozygotes (Mackensen 1951, 1955). Given that the selected allele detected in the previous QTL study is close to fixation, it appears that the classic selective sweep has driven the success of breeding programs at the colony level. As we do not have the genetic information about the initial breeding population, we cannot unambiguously conclude whether this selected allele is from the standing variation or a new mutation.

The low levels of heterozygosity obviously resulted from the low number of alleles in the selected population. The selective breeding dramatically decreased the number of effective alleles at locus *K1425* from 11.5 in the unselected population to 1.1 in the selected population. This extreme reduction clearly is more likely a result of a hard sweep rather than a soft sweep (Chevin & Hospital 2008; Stephan 2010). Even though a soft sweep can decrease the genetic diversity of the selected and flanking loci, it is unlikely to lead to fixation (Charlesworth *et al.* 1993; Charlesworth 1996).

After over 20 years of selective breeding, the number of alleles of reference loci in the breeding population is smaller but not significantly different from the unselected population (*t*-test, $P > 0.05$). Because 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 (19 cM/Mb) in the honeybee genome (The Honeybee Genome Sequencing Consortium 2006), the selective sweep in the selective Danish honeybee population must have occurred extremely fast. Otherwise, the recombination should have eroded the trace footprints of selection before the markers linked to the selected locus became monomorphic (Moritz & Evans 2008). 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 & Norris 2011).

In our study, the selection was conducted at the colony level, but it showed clear footprints of positive selection at the individual level. So the simple procedure of replacing queens in susceptible colonies with queens from colonies which lacked *Nosema* left this selective footprint in the population structure. The colony level selection has selected a locus that confers tolerance against *Nosema* to individual bees. The effect of a major QTL on the individual bee was identified through the reduction in the *Nosema* spores in guts of individual bees. Experimentally infected selected bees also showed a significantly higher tolerance toward *Nosema* infection than did the unselected control bees (Huang *et al.* 2012). The results presented in this study revealed a selective sweep as a result of colony level selection (Palaisa *et al.* 2004; Chevin & Hospital 2008), with the swept loci and the QTL in the same genomic region. Hence, two studies (individual level QTL mapping and the colony level selective sweep analysis) used different approaches but obtained the same result, which provides strong support to the notion that the identified genetic region is indeed associated with the tolerance toward the *Nosema* infection.

Funding

This work was supported by the European Union's Seventh Framework Program BeeDoc, Deutsche Forschungsgemeinschaft project DFG SPP 1399 (MO 373/26-1), Romanian Centre for Bee Biotechnology (RoBeeTech) and China Scholarship Council.

References

Alam M.T., de Souza D.K., Vinayak S., *et al.* (2011) Selective sweeps and genetic lineages of Plasmodium falciparum drug-resistant alleles in Ghana. *Journal of Infectious Diseases* **203**, 220–7.

Charlesworth B. (1996) Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genetical Research* **68**, 131–49.

Charlesworth B., Morgan M.T. & Charlesworth D. (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–303.

Chen Y.P., Evans J.D., Zhou L., Boncristiani H., Kimura K., Xiao T.G. *et al.* (2009) Asymmetrical coexistence of *Nosema ceranae* and *Nosema apis* in honey bees. *Journal of Invertebrate Pathology* **101**, 204–9.

Chevin L.M. & Hospital F. (2008) Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* **180**, 1645–60.

Fay J.C. & Wu C.I. (2000) Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–13.

Forsgren E. & Fries I. (2010) Comparative virulence of *Nosema ceranae* and *Nosema apis* in individual European honey bees. *Veterinary Parasitology* **170**, 212–7.

Fries I. (2010) *Nosema ceranae* in European honey bees (*Apis mellifera*). *Journal of Invertebrate Pathology* **103**, 73–9.

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). *European Journal of Protistology* **32**, 356–65.

Fries I., Martin R., Meana A., García-Palencia P. & Higes M. (2006) Natural infections of *Nosema ceranae* in European honey bees. *Journal of Apicultural Research* **45**, 230–3.

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*? *Applied and Environmental Microbiology* **9**, 3032–8.

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. *Environmental Microbiology* **13**, 404–13.

de 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*). *Journal of Invertebrate Pathology* **63**, 74–81.

Haseman M.H. (1951) The influence of *Nosema apis* on the larval honeybee. *Annals of Applied Biology* **38**, 844–6.

Haseman M.H. (1952) The effects of infection with *Nosema apis* on the pharyngeal salivary glands of the worker honey-bee. *Proceedings of the Royal Entomological Society of London. Series A, General Entomology* **27**, 22–7.

Hermisson J. & Pennings P. (2005) Soft sweep: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–52.

Higes M., Martín R. & Meana A. (2006) *Nosema ceranae*, a new microsporidian parasite in honeybees in Europe. *Journal of Invertebrate Pathology* **92**, 93–5.

Higes M., García-Palencia P., Martín-Hernández R. & Meana A. (2007) Experimental infection of *Apis mellifera* honeybees with *Nosema ceranae* (Microsporidia). *Journal of Invertebrate Pathology* **94**, 211–7.

Higes M., Martin-Hernandez R., Botias C., *et al.* (2008) How natural infection by *Nosema ceranae* causes honeybee colony collapse. *Environmental Microbiology* **10**, 2659–69.

Higes M., Martin-Hernandez R., Garrido-Bailón E., González-Porto A.V., García-Palencia P., Meana A., del Nozal M.J., Mayo R. & Bernal J.L. (2009) Honeybee colony collapse due to *Nosema ceranae* in professional apiaries. *Environmental Microbiology Report* **1**, 110–3.

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.

Huang Q., Kryger P., Le Conte Y., Lattorff H.M.G., Kraus F.B. & Moritz R.F.A. (2013) Four quantitative trait loci associated with low *Nosema ceranae* (Microsporidia) spore load in the honeybee *Apis mellifera*. *Apidologie* **44**, 729–37.

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**, 1137–48.

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**, 765–77.

Kralj J. & Fuchs S. (2010) *Nosema* sp. Influences flight behavior of infected honey bee (*Apis mellifera*) foragers. *Apidologie* **41**, 21–8.

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**, 223–9.

Lynd A., Weetman D., Barbosa S., Yawson A.E., Mitchell S., Pinto J., Hastings I. & Donnelly M.J. (2010) Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in *Anopheles gambiae s.s.* *Molecular Biology and Evolution* **27**, 1117–25.

Mackensen O. (1951) Viability and sex determination in the honey bee. *Genetics* **36**, 500–9.

Mackensen O. (1955) Further studies on a lethal series in the honey bee. *Journal of Heredity* **46**, 72–4.

Mayack C. & Naug D. (2009) Energetic stress in the honeybee *Apis mellifera* from *Nosema ceranae* infection. *Journal of Invertebrate Pathology* **100**, 185–8.

Maynard Smith J.M. & Haigh J. (1974) Hitch-hiking effect of a favorable gene. *Genetical Research* **23**, 23–5.

Moritz R.F.A. (1986) Estimating the genetic variance of group characters: social behavior of the honeybee (*Apis mellifera*). *Theoretical and Applied Genetics* **72**, 513–7.

Moritz R.F.A. & Evans J.D. (2008) Honeybee genomics and breeding for resistance to virus infections. In: *Virology and the Honey Bee* (Ed. by M. Aubert, B. Ball, I. Fries, R. Moritz, N. Milani, I. Bernardinelli), pp. 347–70. Europ Comm Brussels, Luxembourg.

Nagylaki T. (1985) Homozygosity, effective number of alleles, and interdeme differentiation in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 8611–3.

Nielsen R., Williamson S., Kim Y., Hubisz M.J., Clark A.G. & Bustamante C. (2005) Genomic scans for selective sweeps using SNP data. *Genome Research* **15**, 1566–75.

Norris L.C. & Norris D.E. (2011) Insecticide resistance in *Culex quinquefasciatus* mosquitoes after the introduction of insecticide-treated bed nets in Macha, Zambia. *Journal of Vector Ecology* **36**, 411–20.

Ohta T. & Kimura M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**, 201–4.

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

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. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9885–90.

Schlötterer C. (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**, 753–63.

Schlüns H., Moritz R.F.A., Neumann P., Kryger P. & Königer G. (2005) Multiple nuptial flight, sperm transfer and the evolution of extreme polyandry in honeybee queens. *Animal Behavior* **70**, 125–31.

Stephan W. (2010) Genetic hitchhiking versus background selection: the controversy and its implications. *Philosophical Transactions of the Royal Society B Biological Science* **365**, 1245–53.

The Honeybee Genome Sequencing Consortium. (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**, 931–49.

Traynor K. (2008) Bee breeding around the world. *American Bee Journal* **148**, 135–9.

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**, 506–13.

Wang J.L. (2004) Sibship reconstruction from genetic data with typing errors. *Genetics* **166**, 1963–79.

Wiehe T., Nolte V., Zivkovic D. & Schlötterer C. (2007) Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. *Genetics* **175**, 207–18.

Zander E (1909) Tierische parasiten als krankenheitserreger bei der biene. *Münchener Bienenzeitung* **31**, 196–204.

Supporting Information

Table S1 queen genotypes of the selected and the control honeybee populations

	Honey bee populations of Denmark and 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
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
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