



Original Article

Genome-Wide Association Study of Insect Bite Hypersensitivity in Swedish-Born Icelandic Horses

**Merina Shrestha, Susanne Eriksson, Anouk Schurink,
Lisa S. Andersson, Marie Sundquist, Rebecka Frey, Hans Broström,
Tomas Bergström, Bart Ducro, and Gabriella Lindgren**

From the Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden (Shrestha, Eriksson, Bergström, and Lindgren); Animal Breeding and Genomic Centre, Wageningen University, 6700 AH Wageningen, The Netherlands (Shrestha, Schurink, and Ducro); Capilet Genetics AB, SE-725 93 Västerås, Sweden (Andersson); Östra Greda Research Group, SE-387 91 Borgholm, Sweden (Sundquist); Norsholms Animal Hospital, SE-602 37 Norrköping, Sweden (Frey); and Department of Clinical Sciences, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden (Broström).

Address correspondence to Gabriella Lindgren at the address above, or e-mail: gabriella.lindgren@slu.se.

Received October 22, 2014; First decision January 11, 2015; Accepted May 8, 2015.

Corresponding editor: Susan J. Lamont

Abstract

Insect bite hypersensitivity (IBH) is the most common allergic skin disease in horses and is caused by biting midges, mainly of the genus *Culicoides*. The disease predominantly comprises a type I hypersensitivity reaction, causing severe itching and discomfort that reduce the welfare and commercial value of the horse. It is a multifactorial disorder influenced by both genetic and environmental factors, with heritability ranging from 0.16 to 0.27 in various horse breeds. The worldwide prevalence in different horse breeds ranges from 3% to 60%; it is more than 50% in Icelandic horses exported to the European continent and approximately 8% in Swedish-born Icelandic horses. To minimize the influence of environmental effects, we analyzed Swedish-born Icelandic horses to identify genomic regions that regulate susceptibility to IBH. We performed a genome-wide association (GWA) study on 104 affected and 105 unaffected Icelandic horses genotyped using Illumina® EquineSNP50 Genotyping BeadChip. Quality control and population stratification analyses were performed with the GenABEL package in R ($\lambda = 0.81$). The association analysis was performed using the Bayesian variable selection method, Bayes C, implemented in GenSel software. The highest percentage of genetic variance was explained by the windows on X chromosomes (0.51% and 0.36% by 73 and 74 mb), 17 (0.34% by 77 mb), and 18 (0.34% by 26 mb). Overlapping regions with previous GWA studies were observed on chromosomes 7, 9, and 17. The windows identified in our study on chromosomes 7, 10, and 17 harbored immune system genes and are priorities for further investigation.

Subject areas: Genomics and gene mapping

Key words: Bayes C, Bayesian variable selection, *Culicoides* spp., single nucleotide polymorphism, summer eczema

Insect bite hypersensitivity (IBH) is an allergic seasonal dermatitis that reduces the welfare and commercial value of affected horses. IBH is known by different names in different parts of the world: “Queensland itch” in Australia, “Kasen disease” in Japan, “Sweet itch” in Great Britain, “Sommareksem” in Nordic countries, and the similar “Sommerekzem” in Germany (Broström et al. 1987; Schaffartzik et al. 2012). IBH is an immediate (type I) hypersensitivity reaction (Wagner et al. 2006; Wagner et al. 2009) against the allergen(s) in the saliva of female biting midges of the genus *Culicoides* (Hellberg et al. 2006) and potentially against female black flies of the genus *Simulium* (Hellberg et al. 2009; Schaffartzik et al. 2010). Allergic dermatitis due to *Culicoides* has also been reported in other species, such as sheep and cattle (Yeruham et al. 1993; Corrêa et al. 2007).

The allergic reaction is pruritic and extensive scratching leads to hair loss, self-excoriations, and in severe cases results in skin thickening, crusting, and scaling. These clinical signs are mainly observed on the mane and tail, which are the preferred feeding sites for the insects (Schaffartzik et al. 2012). The average age of onset has been found to be in the range of 2–4 years (Schurink et al. 2013). The clinical signs tend to become worse with increasing number of grazing seasons (Broström et al. 1987).

IBH affects different horse breeds worldwide. A review by Schaffartzik et al. (2012) suggested that the worldwide prevalence of IBH for different breeds ranges from 3% in Great Britain to 60% in Australia, and the prevalence is more than 50% in horses exported from Iceland to the European continent (Schaffartzik et al. 2012). *Culicoides* do not habituate in Iceland; thus, IBH is not observed in Iceland. A study by Broström et al. (1987) in 441 Swedish-born Icelandic horses suggested a prevalence of 6.7%, and a more recent study by Eriksson et al. (2008) with 1250 horses suggested a prevalence of 8%.

IBH is multifactorial and influenced by genetic, environmental, and possibly epigenetic factors (Marti et al. 2008). Heritability on the liability scale was estimated to be 0.16 in Friesian horses (Schurink et al. 2011), 0.24 in Dutch Shetland ponies (Schurink et al. 2009), and 0.27 (IBH as a binary trait) in Swedish-born Icelandic horses and 0.30 when severity was taken into account (Eriksson et al. 2008).

A candidate gene study by Andersson et al. (2012) showed an association of the equine leukocyte antigen (ELA) class II region with IBH. Vychodilova et al. (2013) also observed associations between non-ELA genes that influence immunity and allergy and IBH. A genome-wide association (GWA) study for IBH in Shetland ponies and Icelandic horses in the Netherlands revealed breed-specific and common genomic regions associated with IBH (Schurink et al. 2012).

This study aimed to identify genomic regions associated with IBH in Swedish-born Icelandic horses. The long-term goal of our research is to understand the genetic mechanisms involved in IBH development. Identifying genes and mutations would enable efficient selection of horses to reduce IBH prevalence in different horse breeds.

Materials and Methods

Horse Samples

This study included 209 Icelandic horses. Out of these, 104 were affected by IBH and 105 were unaffected. The gender distribution was even, with 104 females (51 affected and 53 unaffected) and 105 males (53 affected and 52 unaffected). Ages ranged from 4 to 27 years with similar average ages for the affected (mean = 11.6 years, standard deviation [SD] = 3.7 years, median = 11 years) and unaffected (mean = 11.4 years, SD = 4.8 years, median = 11 years) horses.

Most horses (97%) were born in Sweden, and the remaining horses were imported from Iceland, the Netherlands, and Norway. In Sweden, samples were collected from 6 different regions: South, South-west, South-east, Mid-west, Mid-east, and North (see [Supplementary Table S1](#) online). In some parts of Sweden, such as the North region, the climate is less beneficial for *Culicoides*. To ensure that horses had been exposed to *Culicoides*, avoiding false-negatives, unaffected horses were not selected from the North region. Thus, the selection of horses influenced the proportion of affected and unaffected horses in different regions.

The presence of subgroups within a population can lead to false-positive associations. To reduce the formation of subgroups, the horses should be matched geographically and genealogically (Karlsson et al. 2008). Thus, the majority of data in our study consisted of paternal half-sib groups. Exceptions were 2 horses that were full-sibs and 2 other horses that were maternal half-sibs. One sire had only 1 affected offspring, and 3 sires had 1 unaffected offspring each in the data (without matching affected horses). The horses were the offspring of 42 stallions and 207 mares. The number of offspring per sire ranged from 1 to 24 (Figure 1).

Phenotype Assessment

Data collection started in 2005 by disseminating information regarding the study (questionnaire) through websites, horse magazines, and the Icelandic horse association. Information about the horse (name, registration number, birth year, sex, country of origin, current location), IBH (status, season and age of onset of IBH, severity of IBH, affected areas on body), other allergies, and pedigree (IBH status of dam) were collected by a questionnaire.

Horse owners classified their horses as unaffected or affected based on whether they had observed signs of IBH. The horse owners observed the horses' condition over a longer period of time and had knowledge about their management; therefore, we can rely on their answers regarding the affected or unaffected status for their horse (Eriksson et al. 2008). The distinct seasonal character of this disease facilitates diagnosis. Seasonal signs of IBH for 2 preceding grazing seasons were criteria for selecting affected horses for the study. Horses classified as unaffected had no history of pruritic skin problems and were at least 4 years old, providing sufficient time for signs of IBH to develop.

Based on the questionnaire, horses from paternal half-sib groups with an equal number of affected and unaffected horses were included in the study. Veterinarians collected blood samples from each horse.

DNA Isolation, SNP Array Genotyping, and Quality Control

To prepare DNA from the blood samples, 200 μ L of blood was isolated using the QIAamp DNA blood mini kit. Genotyping was performed using the Illumina® EquineSNP50 Genotyping BeadChip containing 54602 single nucleotide polymorphisms (SNPs) evenly distributed across the genome. The SNP chip was derived from the EquCab2 Assembly, and the average distance between SNPs on the chip was 43.2 kb. Quality control was performed using the GenABEL package (version 1.7–6; Aulchenko et al. 2007) in R statistical software (version 3.0.1). All of the horses passed the quality control of $\geq 95\%$ genotyping call-rate. Out of 54602 SNPs, 2111 SNPs (3.86%) had a call-rate of less than 95% and 14602 SNPs (26.74%) had a minor allele frequency of less than 5%; these SNPs

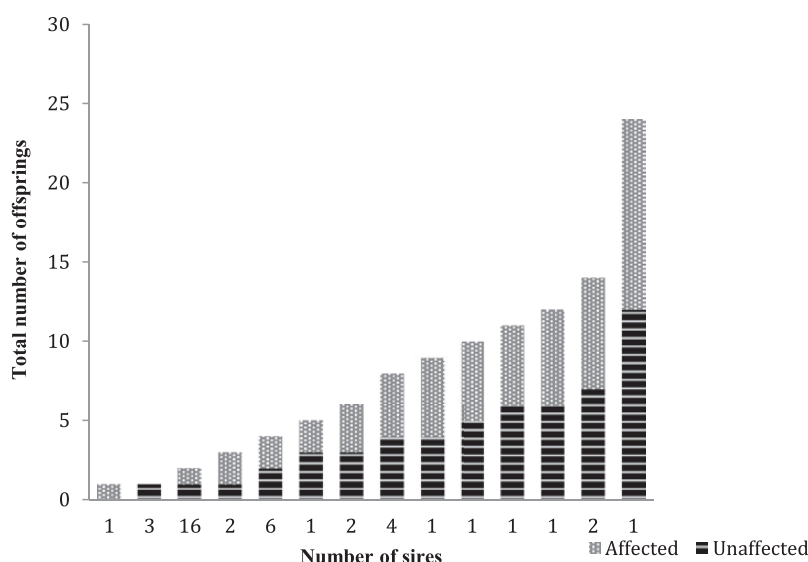


Figure 1. The number of sires on the x axis referring to the number of affected and unaffected offspring on the y axis. For example, third bar from the left on the x axis refers to 16 out of 42 sires who have 2 offspring in total (1 affected and 1 unaffected horse). The right-most bar represents 1 sire with 24 offspring (12 affected and 12 unaffected horses).

were removed from the subsequent analysis. A genome-wide case-control association study was performed with the remaining 38 486 SNPs and 209 horses.

None of the SNPs were removed based on deviation from Hardy-Weinberg equilibrium (HWE) criteria during quality control. Significant deviation ($P < 0.05$) from HWE may occur due to genotyping errors, population stratification, or nonrandom selection of samples and may lead to spurious associations (Kimura 2014). The significance values (P -values) for deviation from HWE were calculated for the top SNPs associated with IBH. Two SNPs (Table 1) deviated significantly from HWE (SNP on chromosome 5, P -value = 0.01 and X chromosome, P -value = 0.046). However, the cluster plots (data not shown) for these SNPs suggested no genotyping errors. Therefore, all the SNPs were further analyzed and the associations with IBH were not considered spurious associations.

Data Archiving

In fulfillment of data archiving guidelines (Baker 2013), we have deposited the primary data for 209 Icelandic horses on Dryad as mentioned below:

- SNPs genotypes (.ped and .map format) on Illumina 50K SNP chip,
- IBH status (affected/unaffected), and
- pedigree, age, gender, and region.

Population Stratification Analysis

Population stratification is the most important cause of spurious associations (Balding 2006). A multidimensional scaling (MDS) plot was used to visualize the relationships among the horses (Purcell et al. 2007). A genomic kinship matrix (209 × 209) containing pairwise identity by state (IBS) values (f_{ij}) was constructed based on autosomal markers with the GenABEL package. IBS was computed with the *ibs* () function as:

$$f_{ij} = \sum_{k=1}^N \frac{(x_{i,k} - p_k)(x_{j,k} - p_k)}{(p_k(1 - p_k))},$$

where f_{ij} is IBS between horses i and j ; k changes from 1 to N , where N is the number of SNPs remaining after excluding SNPs located on the X chromosome (36 758 autosomal SNPs were used in the stratification analysis out of 38 486 SNPs); $x_{i,k}$ is the genotype coded as 0, 1/2, or 1 in the k th SNP marker of the i th horse and $x_{j,k}$ is the corresponding genotype of the j th horse; and p_k is the frequency of the top strand allele.

The genomic kinship matrix was converted to a distance matrix using the *as.dist* () function in R. To obtain the principal components (PCs) from the distance matrix, the MDS function *cmdscale* () was used. This function converts the set of dissimilarities into a set of points. The PCs represent approximate distances among individuals. The 2 components (PC1 and PC2) that explained the largest variance among individuals based on marker information were used to create a MDS plot to detect any subgroups within the population and observe the relationships based on SNPs between affected and unaffected horses.

The result from a preliminary mixed model approach with single SNPs using the GenABEL package in R was compared with the expected distribution of P -values to obtain the genomic inflation factor (λ).

Genome-Wide Association Study

For the association study, the Bayesian variable selection method Bayes C was used (Kizilkaya et al. 2010; Habier et al. 2011), which is implemented in version 4.0 of GenSel software (<http://big.ansci.iastate.edu/>). A study comparing the single marker test, mixed model approach, haplotype-based test, and Bayesian estimation concluded that Bayesian estimation with variable selection performed better in terms of power, precision on locating quantitative trait loci (QTLs), and type I error rates (Sahana et al. 2010).

In the analysis, the observed binary (affected/unaffected) phenotype was linked to the linear predictor “ η ” through a probit link function. A multilocus Bayesian method was used in which all SNPs were fitted simultaneously as random effects in the model with only a few SNPs assigned to having an effect and the remaining SNPs

Table 1. Twenty-nine, 1-mb pair nonoverlapping windows that explained $\geq 0.14\%$ of genetic variance, along with risk allele frequency of SNPs associated with IBH within a window

1-Mb windows explaining highest genetic variance (%)					Associated SNP within 1-mb window			
Chr ^a	Mb ^b	No. SNP in window	Gen. var (%) ^c	P > average ^d	SNP name	Model frequency ^e	Position of SNP (bp)	Risk allele frequency
								Unaffected Affected
1	24	18	0.20	0.021	BIEC2-11435	0.0022	24 217 339	0.37 0.63
1	175	19	0.14	0.021	BIEC2-84371	0.0073	175 563 573	0.40 0.60
1	181	23	0.15	0.020	BIEC2-88925	0.0016	181,696,461	0.47 0.53
3	87	16	0.19	0.030	BIEC2-799492	0.0049	87,610,119	0.45 0.55
5	71	17	0.22	0.031	BIEC2-917852	0.0046	71,366,629	0.45 0.55
6	33	15	0.16	0.023	BIEC2-946836	0.0028	33,126,397	0.38 0.62
6	48	20	0.16	0.027	BIEC2-953643	0.0021	48,848,753	0.31 0.69
7	57	21	0.18	0.036	BIEC2-1002972	0.0025	57,294,948	0.38 0.62
9	50	19	0.14	0.021	BIEC2-1093056	0.0034	50,013,570	0.44 0.56
9	65	21	0.14	0.027	BIEC2-1101171	0.0017	65,054,810	0.45 0.55
9	66	20	0.15	0.025	BIEC2-1102462	0.0014	66,606,644	0.46 0.54
10	19	13	0.15	0.026	BIEC2-107346	0.0036	19,507,420	0.42 0.58
10	30	19	0.15	0.021	BIEC2-112128	0.0024	30,107,713	0.46 0.54
10	35	10	0.24	0.024	BIEC2-115349	0.0069	35,942,977	0.42 0.58
10	41	17	0.33	0.036	BIEC2-119028	0.0086	41 699 776	0.35 0.65
10	49	14	0.23	0.022	BIEC2-121644	0.0042	49 192 501	0.47 0.53
11	16	19	0.15	0.026	BIEC2-141319	0.0026	16 480 388	0.46 0.54
11	18	16	0.15	0.021	BIEC2-142253	0.0040	18 386 393	0.39 0.61
17	77 ¹	17	0.34	0.025	BIEC2-385377	0.0163	77 343 575	0.43 0.57
18	26 ¹	15	0.34	0.025	BIEC2-409581	0.0113	26 766 516	0.43 0.57
26	3	18	0.14	0.027	BIEC2-678688	0.0050	3 370 525	0.44 0.56
X	4	16	0.14	0.018	BIEC2-1108625	0.0022	4 294 248	0.36 0.64
X	67	19	0.14	0.025	BIEC2-1129129	0.0032	67 182 568	0.40 0.60
X	73 ²	20	0.51	0.040	BIEC2-1132313	0.0090	73 129 947	0.34 0.66
X	74 ³	16	0.36	0.038	BIEC2-1133113	0.0041	74 332 689	0.45 0.55
X	76	17	0.27	0.024	BIEC2-1133893	0.0087	76 174 026	0.40 0.60
X	78	15	0.23	0.030	BIEC2-1134793	0.0055	78 794 348	0.42 0.58
X	82	12	0.22	0.021	BIEC2-1138724	0.0098	82 842 940	0.43 0.57
X	100	20	0.14	0.020	BIEC2-1147674	0.0029	100 809 575	0.34 0.66

^aChr: Chromosome number.^bMb: Mega base pair, 24 mb specifies a region from 24 to 25 (1 mb) on chromosome 1.^cGen. var (%): Percentage of genetic variance explained by the window.^dP > average: Proportion of cycles where the window explained more than the expected genetic variance 0.042%.^eModel frequency: proportion of post burn-in cycles, where the SNP included in the model had an effect.^{1,2,3}Top 3 windows that explained the highest, second highest, and third highest percentage of the genetic variance.

assigned to having no effect. The results were also summarized by 1-mb nonoverlapping windows; the genome was divided into 2373 of these windows following the approach described by Wolc et al. (2012). The model was as follows:

$$\eta = \mu + \sum_{j=1}^K Z_j u_j \delta_j,$$

where μ is mean, K is the number of SNPs in the data set ($K = 38\,486$), Z_j is a vector representing the genotypes of SNP covariates at locus j fitted in the model (scores for genotypes are described below), u_j is the random allele substitution effect for locus j conditional on the variance σ_u^2 (common to all loci that were assigned to have an effect) with normal distribution $N(0, \sigma_u^2)$, and δ_j is a 0/1 variable that indicates the absence (with probability π) or presence (probability $1 - \pi$) of the SNP at locus j to be included in the model. When δ_j is 0, then random allele substitution effect for SNP at locus j (u_j) will also be 0. SNPs were scored in Z_j as -10 for AA, 0 for AB, and +10 for BB, where B is the minor allele. Missing genotypes

were replaced by an average for that locus. The average value was calculated by multiplying each score by its frequency and adding the values for all 3 scores.

The probability “ π ,” indicating the percentage of SNPs in each iteration assigned to having no effect on the trait, was set to 0.999. On average, 37.35 SNPs (SD = 6.03) were included in the model across the post burn-in iterations. The marker variance σ_u^2 was assumed to follow an inversed scale chi-square distribution with $v_u = 4$ degrees of freedom, and scale parameter S_u^2 (Kizilkaya et al. 2010), where the scale parameter is $S_u^2 = \frac{\sigma_u^2 (v_u - 2)}{K(1 - \pi)2\bar{p}\bar{q}v_u}$, and the expected marker variance is $E(\sigma_u^2) = \frac{\sigma_u^2}{K(1 - \pi)2\bar{p}\bar{q}}$.

The prior marker variance was derived from the additive genetic variance ($\sigma_a^2 = 0.369$) based on the estimated heritability of 0.27 for IBH as a binary trait (Eriksson et al. 2008). The marker effects were estimated as described by Kizilkaya et al. (2010).

A total of 200 000 Markov Chain Monte Carlo (MCMC) cycles were performed in which the first 50% were discarded as a burn-in

period. The horse genome build (EquCab 2.0) was divided into 2373 nonoverlapping 1-mb windows. When a SNP was included in the model, the window related to that SNP was also considered to be included in the model.

In every 100th MCMC cycle, SNP effects were used to estimate breeding values. The SNP effects were multiplied by the SNP genotypes and summed over all SNPs within a window to obtain a posterior distribution of the breeding values for the individuals in that window. The variance in breeding values across all horses was used to construct the posterior distribution for the genetic variance captured by that window. The posterior mean was divided by the total genetic variance (i.e., the variance explained by all SNPs across the genome in that particular cycle) to obtain the proportion of genetic variance captured by that window. The total genetic variance divided by the total number of windows ($n = 2373$) in our study gave the expected percentage of genetic variance explained by each window ($100\%/2373 \text{ windows} = 0.042\%$). The windows explaining more than the expected percentage (0.042%) of genetic variance were considered nonrandom associations. However, among these nonrandomly associated windows, only those that explained the highest percentage of genetic variance ($\geq 0.14\%$) were considered for further investigation. The marker with the highest model frequency within a window was taken as the candidate SNP associated with IBH. Model frequency refers to the proportion of post burn-in iterations for which a SNP was included in the model. In version 4.0 of the GenSel software, the samples were obtained directly from the posterior distribution of parameters (e.g., variance explained by the window). Therefore, the significance level of the variance explained by the window was not determined (Wolc et al. 2012).

Results

Population Stratification Analysis

Two groups were observed in the MDS plot (Figure 2): a small group on the left ($PC1 \leq -0.1$ and $PC2 < 0.1$) with 38 horses and a large group on the right with 171 horses. The number of affected and unaffected horses was distributed evenly within each group, with 19

of each in the small group and 85 affected and 86 unaffected in the large group. PC1 accounted for 6.18% of the total variance and PC2 accounted for 3.19%.

The genomic inflation factor was less than 1 ($\lambda = 0.81$), suggesting no population stratification.

Genome-Wide Association Study

A Manhattan plot of the percentage of genetic variance captured by different windows across the genome using Bayesian variable selection method Bayes C (Figure 3) revealed 29 windows that explained $\geq 0.14\%$ of the genetic variance (Table 1). These windows were located on 12 chromosomes and considered for further analyses.

The largest percentage of genetic variance, 0.51%, was explained by a window located at 73 mb on X chromosome. The SNP with highest model frequency (0.009) within this window was BIEC2-1132313. The frequency of the risk allele of this SNP was 0.34 in unaffected horses and 0.66 in affected horses (Table 1). The window explaining the second largest percentage of genetic variance (0.36%) was located at 74 mb on X chromosome. The model frequency of the SNP associated with IBH within this window, BIEC2-1133113, was 0.0041. The windows on chromosomes 17 (77 mb) and 18 (26 mb) explained 0.34% of the genetic variance, which is the third largest percentage of genetic variance explained by a window. The model frequencies for SNPs BIEC2-385377 and BIEC2-409581 associated within these windows were 0.0163 and 0.0113, respectively, and represented the SNPs with the highest and second highest model frequencies across the genome. The model frequencies for SNPs within 4 windows explaining the greatest proportion of genetic variance are given in Figure 4. The heritability based on the posterior means of the proportion of genetic variance was 0.114, which is lower than the heritability of 0.27 on the underlying scale estimated by Eriksson et al. (2008).

The number of SNPs within 1-mb windows ranged from 1 to 25 (average 16.1). Approximately 611 windows explained 50% of the genetic variance. The total genetic variance exceeded 100% (105.95%), possibly due to covariance between window variances (Wolc et al. 2012).

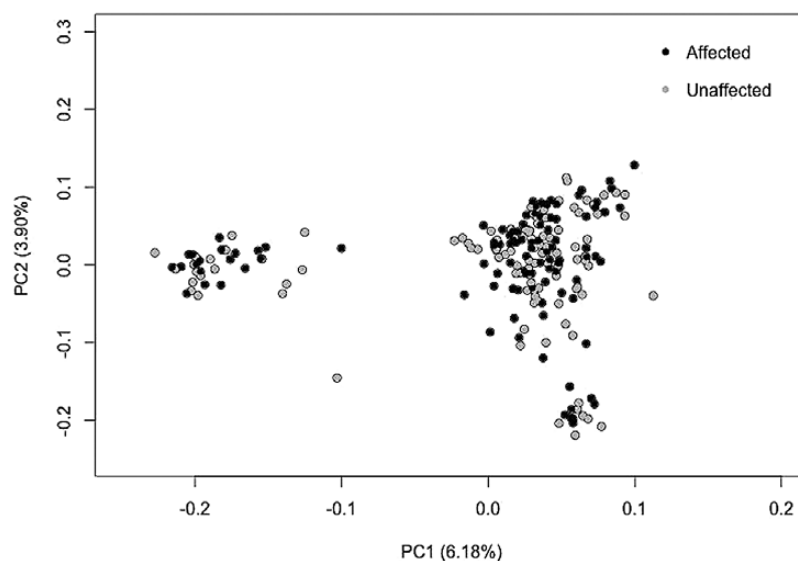


Figure 2. MDS plot for 209 horses. PC1 explained 6.18% of the variance and PC2 explained 3.90%. Each dot represents an individual.

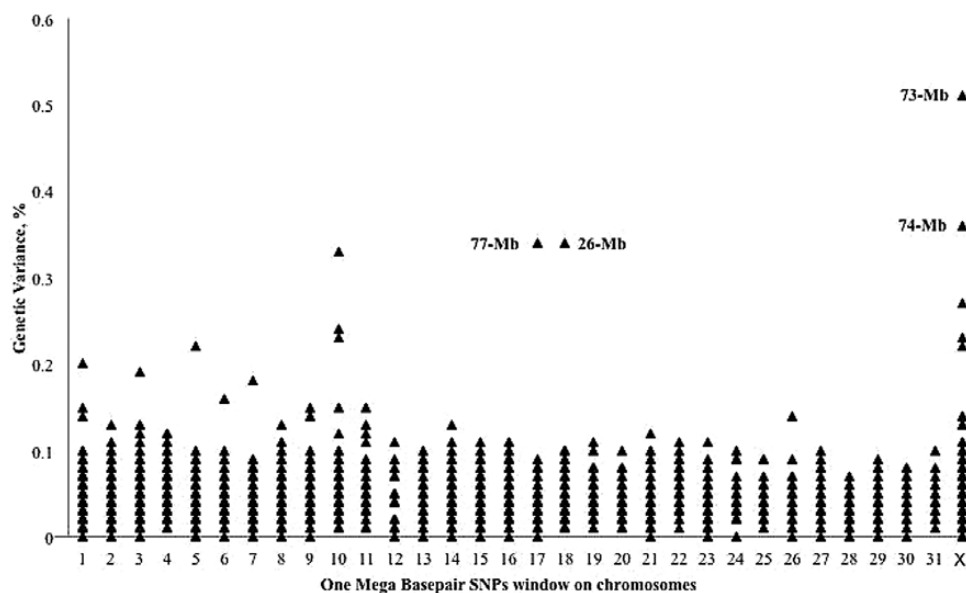


Figure 3. The percentage of genetic variance explained by 1-mb windows against the chromosomes. Each triangle represents the percentage of genetic variance (y axis) explained by 1-mb windows present along the chromosomes on the x axis.

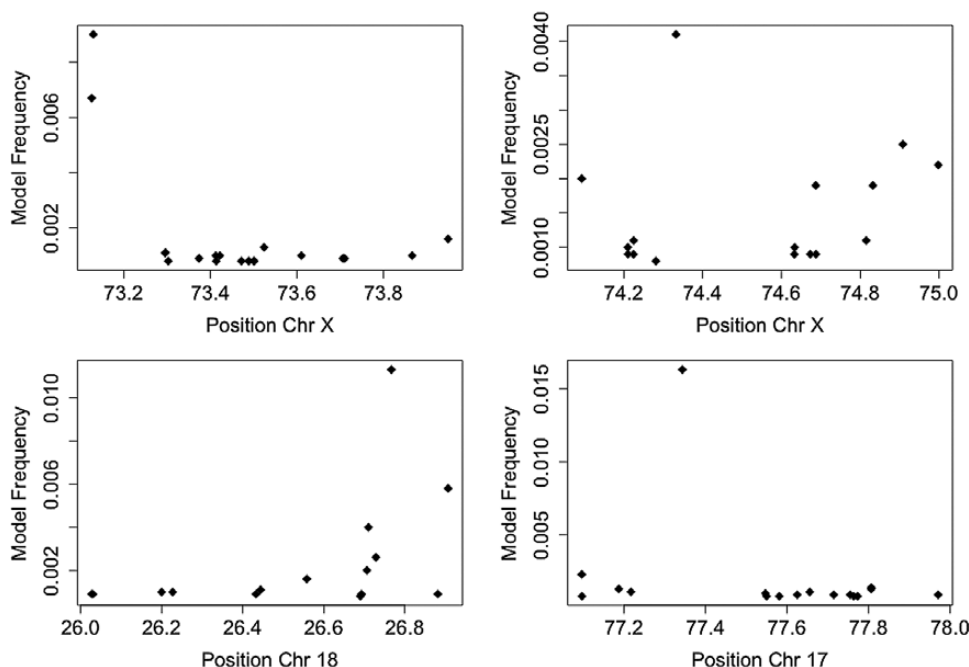


Figure 4. The model frequencies of SNPs are plotted on the y axis against the genomic location for the SNPs within the windows (in mb) explaining the largest percentage of genetic variance.

Discussion

We performed a GWA study to identify and quantify genomic regions associated with IBH using the Bayesian variable selection method Bayes C on 209 Swedish-born Icelandic horses. The study identified 29 windows on 12 chromosomes that explained $\geq 0.14\%$ of the genetic variance (Table 1). The results from this study were compared with previous studies on IBH in different horse populations. Overlapping regions across different horse populations were observed on chromosomes 7, 9, and 17. Immunity-related genes were also observed in the genomic regions on chromosomes 7, 10, and 17.

Investigation of the Windows Associated with IBH

The horse genome browser EquCab2.0 was used to investigate the genomic regions of the windows associated with IBH in our study. Some associations were previously reported between equine IBH and genes located in the ELA region and genes related to allergy and immunity that are located in other regions (Andersson et al. 2012; Klumplerova et al. 2013; Vychodilova et al. 2013). The allergic skin disorders known as extrinsic atopic eczema or atopic dermatitis (AD) in humans and IBH in horses have overlapping features (Schaffartzik et al. 2012). Most of the genes associated with AD in

humans are involved in skin barrier function and innate and adaptive immune responses (Boguniewicz and Leung 2011). Tengvall et al. (2013) suggested an association between AD in dogs and a genomic region containing *PKP2*, the gene encoding plakophilin 2, a protein that takes part in pathways that drive interactions required for normal desmosome assembly. Changes in the corneodesmosome (modified desmosomes in the epidermis) degradation process influence the thickness of the stratum corneum and surface of the skin. Abnormal corneodesmosome degradation has been found in common skin diseases, including AD (Ishida-Yamamoto et al. 2011).

The associated windows containing immunity-related genes, allergy-related genes, and genes directly or indirectly involved in skin function were considered interesting genomic regions. More detailed information on the windows harboring interesting genes is provided below.

In this study, a window located at 77 mb on chromosome 17 explained 0.34%, the third highest percentage, of the genetic variance. The associated SNP within this window was in the intergenic region at 77 343 575 bp. Based on the horse genome browser EquCab2.0, this 1-mb window harbors 8 protein-coding genes. One gene, insulin receptor substrate 2 (*IRS2*), is located 105 kb upstream of the associated SNP. Acevedo et al. (2009) observed that *IRS2* is associated with the total level of Immunoglobulin E (IgE) in asthmatic patients and plays a role in the modulation of antibody levels. According to the human gene database GeneCards (www.genecards.org), *IRS2* interacts with the immune response interleukin 9 (*IL-9*) signaling pathway and one of the related super-pathways (Rebhan et al. 1998). A study in mice reported increased production of IgE by *IL-9* (Petit-Frere et al. 1993) and showed that IgE mediates type I hypersensitivity reactions.

A window located at 35 mb on chromosome 10 explained 0.24% of the genetic variance. This window harbors the gene inhibitor of bruton agammaglobulinemia tyrosine kinase (*IBTK*) 159 kb downstream of the SNP associated with IBH. *IBTK* regulates the activity of Bruton's tyrosine kinase (*BTK*), which is required for B-cell development in humans and mice (Spatuzza et al. 2008). Progenitor B cells transform into mature B cells and lead to the production of different immunoglobulins: IgG, IgM, IgA, IgD, and IgE. B cells play a role in the adaptive immune system, and defects in B-cell development, selection, and function can lead to the development of immunodeficiency, allergies, and autoimmunity (Pieper et al. 2013).

A window at 57 mb on chromosome 7 explained 0.18% of the genetic variance in this study. The gene cathepsin C (*CTSC*) located 600 kb downstream of the associated SNP within this window has been shown to play a role in major histocompatibility complex (MHC) class II antigen presentation in humans (Rebhan et al. 1998).

No known gene related to immunity, allergy, or skin function is located in the other windows found in this study. However, causal variants may be distantly located regulatory regions governing other nonimmunity genes related to IBH. Therefore, careful investigation of these regions is necessary. Also, improved annotation of the horse genome build will facilitate further studies aiming to identify genes that regulate IBH.

Comparison with Previous Genome-Wide Association Studies

Few GWA studies have been performed for equine IBH. Schurink et al. (2013) analyzed Illumina® EquineSNP50 Genotyping BeadChip data from 188 Dutch Shetland pony mares with known IBH status using a logistic regression model that fitted 1 SNP at a time. Their study suggested 24 significantly associated SNPs on

12 chromosomes. In another study, Schurink et al. (2012) used the equine HD chip (Illumina® Inc., San Diego, CA) containing 65 157 SNPs to genotype 200 Shetland pony mares and 146 Icelandic horses, which were analyzed separately using the Bayesian variable selection method. The study identified associations with IBH on chromosomes 3, 7, 11, 20, and 23 that were common to both breeds.

For most horse breeds, linkage disequilibrium (LD) declines rapidly over the first 50–100 kb and then reaches a background level within 1–2 mb (McCue et al. 2012). Therefore, when comparing the results of the present in Icelandic horses to previous GWA studies, we only considered the region as overlapping if it was present within 2 mb of the genomic regions in this study. Schurink et al. (2013) reported a significantly associated SNP at 66 153 825 bp on chromosome 9 in Dutch Shetland ponies. In this study, a window located at 66 mb on chromosome 9 explained 0.15% of the genetic variance. The SNP associated with IBH within this window was 452 819 bp downstream of the SNP identified in the previous study by Schurink et al. (2013). None of the genes within this window have previously been reported to be involved in allergy, immunity, or skin function.

In the study by Schurink et al. (2012), a window on chromosome 7 located at 55 mb was associated with IBH in Icelandic horses and explained 0.179% of the genetic variance. In this study, a window at 57 mb on chromosome 7 was associated with IBH.

Schurink et al. (2012) also suggested regions associated with IBH on chromosome 17 in Shetland pony mares, with the 75 and 76 mb windows explaining 0.30% and 0.16% of the genetic variance, respectively. In our study, a window located at 77 mb on chromosome 17 explained 0.34% of the genetic variance and was one of the top 3 windows associated with IBH. The identification of these neighboring windows on chromosome 17 in 2 separate breeds further strengthens the potential role of *IRS2* in IBH.

In our study, the 2 windows explaining the highest (0.51%) and second highest (0.36%) percentage of genetic variance for IBH were located at 73 and 74 mb on X chromosome. The SNPs within these windows (BIEC2-1132313 at 73 mb and BIEC2-1133113 at 74 mb) were intergenic variants. As these windows are located next to each other, analysis was performed with these windows merged as 1 window. The percentage of genetic variance explained by the 2-mb window was 0.56% and the $P > \text{average}$ (i.e., proportion of cycles in which the window explained more than the expected genetic variance of 0.042%) was 0.059. Schurink et al. (2012) merged 2 consecutive windows located on chromosome 17 in a GWA study of Shetland ponies and estimated almost the same percentage of genetic variance explained by the 2-mb window as the sum of the percentages explained by 2 consecutive 1-mb windows separately. Thus, the 2 consecutive windows were assumed to explain a proportion of the total QTL variance. In our study, the percentage of genetic variance explained by the window increased from 0.51% to 0.56% when the consecutive windows were merged. However, this is lower than the expected sum of the percentage ($0.51 + 0.36 = 0.87$) explained by the 2 consecutive 1-mb windows (73 and 74 mb) separately. Thus, the QTL effect is spread across the two 1-mb windows. Therefore, to estimate the QTL effect, these windows should be merged. The optimal window size depends on the LD between SNPs, which varies among different chromosomes and among regions within a chromosome (Schurink et al. 2012).

Schurink et al. (2012) reported that the window associated with IBH that explained the highest percentage of genetic variance (0.658%) was located at 59 mb on X chromosome in Icelandic horses. Schurink et al. (2013) also observed an association between a SNP located on X chromosome and IBH in Dutch Shetland ponies.

In a study by Eriksson et al. (2008), male Swedish-born Icelandic horses seemed to be at higher risk of developing IBH than females, but Broström et al. (1987) did not find a gender effect. In humans, the X chromosome plays a role in autoimmune diseases (Selmi 2008; Zhang et al. 2015). However, in horses, the studies reporting an association between genes located on the X chromosome and allergic diseases are limited.

In summary, the comparison with previous studies identified common genomic regions associated with IBH on chromosomes 7, 9, and 17. Differences in methodology could be one plausible reason why we did not find additional overlapping regions with the previous study by Schurink et al. (2013), or the differences may reflect the existence of breed-specific loci. Schurink et al. (2013) fitted single SNPs in a logistic regression model, whereas we used the Bayesian variable selection method with a 1-mb window approach that fitted multiple SNPs simultaneously. The Bayesian variable selection method was developed for genomic selection, but it has been successful in identifying associated genomic regions using both simulated and real data sets (Dekkers 2012).

Schurink et al. (2012) used the equine HD SNP chip, whereas this study is based on data from the EquineSNP50 Genotyping BeadChip. The 2 chips had 45 986 overlapping SNPs. As many as 8616 SNPs on the EquineSNP50 Genotyping BeadChip and 19 171 SNPs on the equine HD SNP chip did not overlap. Therefore, SNPs associated with IBH in one study may not be present in the other study.

The phenotype assessment may be another reason for differences between the results of different studies. However, the phenotypic data used in this study were also used by Andersson et al. (2012), who associated IBH with a microsatellite marker located in the ELA region in this population.

Comparison with Candidate Gene Studies

The equine MHC (i.e., ELA) recognizes foreign molecules and evokes immune response that can also cause autoimmune disease and allergies (Bailey et al. 2000). Several serological studies investigated the role of ELA in IBH and found segregation of ELA class II serotypes with IBH (Halldórsdóttir et al. 1991; Marti et al. 1992; Lazary et al. 1994). The ELA class II region is located on chromosome 20 and spans 32–33 mb. Andersson et al. (2012) performed an association study using microsatellite markers and the sequence of the highly polymorphic region of exon 2 in the *DRA* and *DRB3* genes within the ELA class II region in Icelandic horses and Exmoor ponies. A microsatellite marker was associated with susceptibility for IBH in both breeds, but causal variants of ELA class II remain to be identified. Furthermore, the GWA study by Schurink et al. (2012) observed an association of windows on chromosome 20 located within 2 mb of the ELA class II region in both Shetland pony mares and Icelandic horses. No association between the ELA class II region and IBH was observed in this study, likely due to the poor coverage of the ELA region by the EquineSNP50 Genotyping BeadChip. In a candidate gene study, Vychodilova et al. (2013) reported IBH associations for SNPs in candidate non-ELA genes influencing allergy and immunity located on chromosomes 5, 13, 14, 23, and 24. In our study, associations were not observed with these genomic regions.

In our study of 209 Swedish-born Icelandic horses, we observed an association of IBH with genomic regions on different chromosomes. The genomic regions on chromosomes 7, 9, and 17 overlapped between our study and previous GWA studies (Schurink et al. 2012; Schurink et al. 2013). These genomic regions are very interesting and will be further investigated in additional horses and

different breeds for validation. Genomic regions on chromosomes 7, 10, and 17 also harbored genes that directly or indirectly participated in immunity.

Ideally, more animals per breed, more markers, and additional horse breeds are needed for increased power in the analysis. Our ongoing work deals with genotyping additional horse breeds with a SNP chip that is approximately 10 times more dense in order to compare and validate the identified genomic regions in this study. Based on the results from the combined data sets, we will select genomic regions for fine mapping and sequence cases and controls to identify causative mutations.

Supplementary Material

Supplementary material can be found at <http://www.jhered.oxfordjournals.org/>.

Funding

European Union's Seventh Framework Programme managed by the Research Executive Agency (REA; <http://ec.europa.eu/research/rea> [FP7/2007–2013]) under grant agreement no. 606142, The Swedish Research Council (FORMAS), and European Graduate School of Animal Breeding and Genetics (EGS-ABG).

Acknowledgments

The authors would like to acknowledge Bert Dibbits for his contribution in editing the data and providing the cluster plots.

References

- Acevedo N, Mercado D, Vergara C, Sánchez J, Kennedy MW, Jiménez S, Fernández AM, Gutiérrez M, Puerta L, Caraballo L. 2009. Association between total immunoglobulin E and antibody responses to naturally acquired *Ascaris lumbricoides* infection and polymorphisms of immune system-related LIG4, TNFSF13B and IRS2 genes. *Clin Exp Immunol*. 157:282–290.
- Andersson LS, Swinburne JE, Meadows JR, Broström H, Eriksson S, Fikse WF, Frey R, Sundquist M, Tseng CT, Mikko S, et al. 2012. The same ELA class II risk factors confer equine insect bite hypersensitivity in two distinct populations. *Immunogenetics*. 64:201–208.
- Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 23:1294–1296.
- Bailey E, Marti E, Fraser DG, Antczak DF, Lazary S. 2000. Immunogenetics of the Horse. In: Bowling AT, Ruvinisky A, editors. *The genetics of the Horse*. New York: CABI Publishing. p. 123–154.
- Baker CS. 2013. Journal of heredity adopts joint data archiving policy. *J Hered*. 104:1.
- Balding DJ. 2006. A tutorial on statistical methods for population association studies. *Nat Rev Genet*. 7:781–791.
- Boguniewicz M, Leung DY. 2011. Atopic dermatitis: a disease of altered skin barrier and immune dysregulation. *Immunol Rev*. 242:233–246.
- Broström H, Larsson A, Troedsson M. 1987. Allergic dermatitis (sweet itch) of Icelandic horses in Sweden: an epidemiological study. *Equine Vet J*. 19:229–236.
- Corrêa TG, Ferreira JM, Riet-Correa G, Ruas JL, Schild AL, Riet-Correa F, Guimaraes A, Felipe-Bauer ML. 2007. Seasonal allergic dermatitis in sheep in southern Brazil caused by *Culicoides insignis* (Diptera: Ceratopogonidae). *Vet Parasitol*. 145:181–185.
- Dekkers JC. 2012. Application of genomics tools to animal breeding. *Curr Genomics*. 13:207–212.
- Eriksson S, Grandinson K, Fikse WF, Lindberg L, Mikko S, Broström H, Frey R, Sundquist M, Lindgren G. 2008. Genetic analysis of insect bite hypersensitivity (summer eczema) in Icelandic horses. *Animal*. 2:360–365.

- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*. 186:12.
- Halldórsdóttir S, Larsen HJ. 1991. An epidemiological study of summer eczema in Icelandic horses in Norway. *Equine Vet J*. 23:296–299.
- Hellberg W, Mellor PS, Torsteinsdóttir S, Marti E. 2009. Insect bite hypersensitivity in the horse: comparison of IgE-binding proteins in salivary gland extracts from *Simulium vittatum* and *Culicoides nubeculosus*. *Vet Immunol Immunopathol*. 132:62–67.
- Hellberg W, Wilson AD, Mellor P, Doherr MG, Torsteinsdóttir S, Zurbriggen A, Jungi T, Marti E. 2006. Equine insect bite hypersensitivity: immunoblot analysis of IgE and IgG subclass responses to *Culicoides nubeculosus* salivary gland extract. *Vet Immunol Immunopathol*. 113:99–112.
- Ishida-Yamamoto A, Igawa S, Kishibe M. 2011. Order and disorder in corneocyte adhesion. *J Dermatol*. 38:645–654.
- Karlsson EK, Lindblad-Toh K. 2008. Leader of the pack: gene mapping in dogs and other model organisms. *Nat Rev Genet*. 9:713–725.
- Kimura A. 2014. Departure from the Hardy-Weinberg equilibrium. *Gene*. 537:357.
- Kizilkaya K, Fernando RL, Garrick DJ. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J Anim Sci*. 88:544–551.
- Klumplerova M, Vychodilova L, Bobrova O, Cvanova M, Futas J, Janova E, Vyskocil M, Vrtkova I, Putnova L, Dusek L, et al. 2013. Major histocompatibility complex and other allergy-related candidate genes associated with insect bite hypersensitivity in Icelandic horses. *Mol Biol Rep*. 40:3333–3340.
- Lazary S, Marti E, Szalai G, Gaillard C, Gerber H. 1994. Studies on the frequency and associations of equine leucocyte antigens in sarcoid and summer dermatitis. *Anim Genet*. 25:75–80.
- Marti E, Gerber H, Lazary S. 1992. On the genetic basis of equine allergic diseases: II. Insect bite dermal hypersensitivity. *Equine Vet J*. 24:113–117.
- Marti E, Gerber V, Wilson AD, Lavoie JP, Horohov D, Cramer R, Lunn DP, Antczak D, Björnsdóttir S, Björnsdóttir TS, et al. 2008. Report of the 3rd Havemeyer workshop on allergic diseases of the Horse, Hólar, Iceland, June 2007. *Vet Immunol Immunopathol*. 126:351–361.
- McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, Distl O, Guérin G, Hasegawa T, Hill EW, et al. 2012. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet*. 8:e1002451.
- Petit-Frère C, Dugas B, Braquet P, Mencia-Huerta JM. 1993. Interleukin-9 potentiates the interleukin-4-induced IgE and IgG1 release from murine B lymphocytes. *Immunology*. 79:146–151.
- Pieper K, Grimbacher B, Eibel H. 2013. B-cell biology and development. *J Allergy Clin Immunol*. 131:959–971.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81:559–575.
- Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. 1998. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics*. 14:656–664.
- Sahana G, Guldbbrandtsen B, Janss L, Lund MS. 2010. Comparison of association mapping methods in a complex pedigreed population. *Genet Epidemiol*. 34:455–462.
- Schaffartzik A, Hamza E, Janda J, Cramer R, Marti E, Rhyner C. 2012. Equine insect bite hypersensitivity: what do we know? *Vet Immunol Immunopathol*. 147:113–126.
- Schaffartzik A, Marti E, Cramer R, Rhyner C. 2010. Cloning, production and characterization of antigen 5 like proteins from *Simulium vittatum* and *Culicoides nubeculosus*, the first cross-reactive allergen associated with equine insect bite hypersensitivity. *Vet Immunol Immunopathol*. 137:76–83.
- Schurink A, Ducro BJ, Bastiaansen JW, Frankena K, van Arendonk JA. 2013. Genome-wide association study of insect bite hypersensitivity in Dutch Shetland pony mares. *Anim Genet*. 44:44–52.
- Schurink A, Ducro BJ, Heuven HC, van Arendonk JA. 2011. Genetic parameters of insect bite hypersensitivity in Dutch Friesian broodmares. *J Anim Sci*. 89:1286–1293.
- Schurink A, van Grevenhof EM, Ducro BJ, van Arendonk JA. 2009. Heritability and repeatability of insect bite hypersensitivity in Dutch Shetland breeding mares. *J Anim Sci*. 87:484–490.
- Schurink A, Wolc A, Ducro BJ, Frankena K, Garrick DJ, Dekkers JC, van Arendonk JA. 2012. Genome-wide association study of insect bite hypersensitivity in two horse populations in the Netherlands. *Genet Sel Evol*. 44:31.
- Selmi C. 2008. The X in sex: how autoimmune diseases revolve around sex chromosomes. *Best Pract Res Clin Rheumatol*. 22:913–922.
- Spatuzza C, Schiavone M, Di Salle E, Janda E, Sardiello M, Fiume G, Fierro O, Simonetta M, Argiriou N, Faraonio R, et al. 2008. Physical and functional characterization of the genetic locus of IBtk, an inhibitor of Bruton's tyrosine kinase: evidence for three protein isoforms of IBtk. *Nucleic Acids Res*. 36:4402–4416.
- Tengvall K, Kierczak M, Bergvall K, Olsson M, Frankowiack M, Farias FH, Pielberg G, Carlborg Ö, Leeb T, Andersson G, et al. 2013. Genome-wide analysis in German shepherd dogs reveals association of a locus on CFA 27 with atopic dermatitis. *PLoS Genet*. 9:e1003475.
- Vychodilova L, Matiasovic J, Bobrova O, Futas J, Klumplerova M, Stejskalova K, Cvanova M, Janova E, Osickova J, Vyskocil M, et al. 2013. Immunogenomic analysis of insect bite hypersensitivity in a model horse population. *Vet Immunol Immunopathol*. 152:260–268.
- Wagner B, Miller WH Jr, Erb HN, Lunn DP, Antczak DF. 2009. Sensitization of skin mast cells with IgE antibodies to *Culicoides* allergens occurs frequently in clinically healthy horses. *Vet Immunol Immunopathol*. 132:53–61.
- Wagner B, Miller WH, Morgan EE, Hillegas JM, Erb HN, Leibold W, Antczak DF. 2006. IgE and IgG antibodies in skin allergy of the horse. *Vet Res*. 37:813–825.
- Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, Habier D, Fernando R, Garrick DJ, Hill WG, et al. 2012. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. *Anim Genet*. 43:87–96.
- Yeruham I, Braverman Y, Orgad U. 1993. Field observations in Israel on hypersensitivity in cattle, sheep and donkeys caused by *Culicoides*. *Aust Vet J*. 70:348–352.
- Zhang Y, Zhang J, Yang J, Wang Y, Zhang L, Zuo X, Sun L, Tam PK-H, Sham PC, Yang S, et al. 2015. Meta-analysis of GWAS on two Chinese populations followed by replication identifies novel genetic variants on the X chromosome associated with systemic lupus erythematosus. *Hum Mol Genet*. 24:274–278.

Supplementary Table

Table S1. The number of affected and unaffected horses, Mean and Standard Deviation (S.D.) for different regions in Sweden

Region	Affected	Unaffected	Total	Mean of IBH ¹	S.D. ²
South	20	26	46	0.43	0.50
South west	20	6	26	0.77	0.43
South east	19	37	56	0.34	0.48
Mid east	32	34	66	0.48	0.50
Mid west	8	2	10	0.80	0.42
North	5	0	5	1.00	0.00

¹Mean of IBH: Mean of insect bite hypersensitivity