
Article:
Höglund, J., Wang, B., Saether, S.A. et al. (7 more authors) (2017) Blood transcriptomes and de novo identification of candidate loci for mating success in lekking great snipe (Gallinago media). Molecular Ecology, 26 (13). pp. 3458-3471. ISSN 0962-1083

https://doi.org/10.1111/mec.14118

Blood transcriptomes and de novo identification of candidate loci for mating success in lekking great snipe (Gallinago media)

Jacob Höglund[1], Biao Wang[1], Stein Are Sæther[3], Mozes Pil Kyu Blom[123], Peder Fiske[3], Peter Halvarsson[1], Gavin J. Horsburgh[4], Terry Burke[4], John Atle Kålås[3], Robert Ekblom[1,4]

1) Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-75236 Uppsala, Sweden

2) Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia

3) Norwegian Institute for Nature Research (NINA), PO Box 5685 Sluppen, 7485 Trondheim, Norway

4) Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

Address for correspondence:

J. Höglund, Dept of Ecology and Genetics, Uppsala University, Norbyv 18D, SE-75236, Uppsala Sweden

jacob.hoglund@ebc.uu.se

+46-18-4712671, +46-70-4250126

2016-11-25

Abstract

We assembled the great snipe blood transcriptome using data from fourteen lekking males, in order to de novo identify candidate genes related to sexual selection, and determined the expression profiles in relation to mating success. The three most highly transcribed genes were encoding different haemoglobin subunits. All tended to be overexpressed in males with high mating success. We also called Single Nucleotide Polymorphisms (SNPs) from the transcriptome data and found considerable genetic variation for many genes expressed during lekking. Among these we identified 14 polymorphic candidate SNPs that had a significant genotypic association with mating success (number of females mated with) and/or mating status (mated or not). Four of the candidate SNPs were found in HBAA (encoding the haemoglobin α-chain). Heterozygotes for one of these and one SNP in the gene PABPC1 appeared to enjoy higher mating success compared to males homozygous for either of the alleles. In a larger dataset of individuals we genotyped 38 of the identified SNPs but found low support for consistent selection since only one of the zygosities of previously identified candidate SNPs and none of their genotypes were associated with mating status. However, candidate SNPs generally showed lower levels of spatial genetic structure compared to non-candidate markers. We also scored the prevalence of avian malaria in a sub-sample of birds. Males infected with avian malaria parasites had lower mating success in the year of sampling than non-infected males. Parasite infection and its interaction with specific genes may thus affect performance on the lek.

Introduction

How genetic variation for female preferences of male traits is maintained has been a long-standing issue in evolutionary biology (Hamilton and Zuk 1982; Andersson 1994) and is related to the more general problem of how genetic variation is upheld in natural populations (reviewed by Hedrick and Kim 1975; Radwan 2008). Leks are mating systems with a high potential for intense sexual selection which have played a pivotal role in sexual selection studies (e.g. Höglund and Alatalo 1995; Shorey et al. 2000; Ryder et al. 2010; Alonso et al. 2010; DuVal 2013). In most lek species males defend small aggregated non-resource based territories that females visit for the sole purpose of mating, after which the females leave and raise the offspring on their own without the aid of the male partner (Höglund and Alatalo 1995). Male mating success is typically highly skewed and thus only a fraction of the males reproduce during any given mating season (Höglund and Alatalo 1995). The lek mating behaviour of males is commonly physically and energetically costly, signified by intense fighting and display behaviour (Vehrencamp et al. 1989; Höglund et al. 1992). Intrasexual selection is generated by male-male competition for access to territories and fertilisable females, while intersexual selection is generated by female choice for certain kinds of males (Darwin 1871; Andersson 1994). Both of these forms of selection are prevalent and intense in lek-mating species, since male territory defence and combat is readily observed and females are relatively free to choose among the males (Höglund and Alatalo 1995). Furthermore, any benefit of female mate choice is likely indirect and mediated by genetic effects, since lekking males, by definition, provide no resources but the sperm necessary for successful fertilisation of the eggs (Höglund and Alatalo 1995). Thus, with the exception of avoiding sexually transmitted disease, female preferences cannot be based on immediate material benefits nor parental care considerations. Given that just a few males get most of the matings, this sets the scene for the so-called lek paradox. In this paper we attempt at identifying candidate genes for sexual selection in lekking great snipe (*Gallinago media*) using transcriptome sequencing to study genetic varaiation at such loci.

The lek paradox is founded on the expectation that sexual selection favours certain kinds of males (strong directional selection on male phenotypes) and that part of this selection is due to female choice. Given that strong directional selection should lead to fixation of the genetic variation coding for the favoured traits, how is it that females are still selective? Thus the paradox refers to situations where directional selection on male traits via female choice should

deplete the genetic variation and therefore, in turn, relax or eliminate the original basis of the female preference (Borgia 1979; Taylor and Williams 1982; Kirkpatrick and Ryan 1991). Since the formulation of the paradox, there have been numerous suggestions on how it might be solved (Rowe and Houle 1996; Tomkins et al. 2004; Miller and Moore 2007, Fromhage et al. 2009; Greenfield et al. 2012). Spatio-temporal variation in selection has for example, been suggested as one of the mechanisms which may uphold genetic variation for sexually selected traits (reviewed by Radwan 2008). However, to date, empirical studies addressing these issues remain scarce (Kotiaho et al. 2001; 2008, Cornwallis and Uller 2012), partly because it is often unclear which traits, if any, are favoured by females in lekking species. It is therefore of pivotal importance to identify candidate genes subjected to sexual selection in lekking species in the wild.

Transcriptome sequencing (also known as RNA-seq) of phenotyped individuals is one method which would allow for identification of candidate loci for sexual selection. First, this will highlight genes which are expressed at high levels in males when they are competing and females are selective (i.e. the lekking period). Second, it will be possible to identify genes differentially expressed among males with varying mating success. Third, it will also reveal sequence data from which functional genotype information on the studied individuals can be extracted. This in turn will allow de novo identification of candidate genes that have sequence variation co-segregating with mating success (Pardo-Diaz et al. 2015). If genetic polymorphism is found for such candidate genes related to mating success, this calls for further in-depth studies examining how this variation is maintained in the population (i.e. solving the lek-paradox). Genotyping of a different set of individuals could, for example, be used to examine temporal and spatial variation in selection regimes and/or possible covariation with levels of parasitism.

In lekking great snipe, we have previously shown that males with more centrally located territories which are more active in display are more often involved in mating (Höglund and Lundberg 1987). Females are often polyandrous and may mate repeatedly with the same male (Fiske and Kålås 1995; Sæther et al. 2001). Studies of individually marked females suggest that the relationship between the centrality on the lek of a male's territory and mating success is more complex than a direct female preference for central males (Höglund and Robertson 1990; Sæther et al. 2005). Instead, female preference seems to favour healthy and vigorous

males able to sustain the energetically demanding lekking behaviour (Kålås et al. 1995; Sæther et al. 2000). This interpretation has been supported by direct measurements of the energetic costs of male display using doubly labelled water to estimate energy consumption which show that male lekking behaviour is indeed energetically costly (Höglund et al. 1992). Furthermore, studies of male immune response (Ekblom et al. 2005) and Major Histocompatibility Complex (MHC) genotyping (Ekblom et al. 2004; 2009) suggest that there might be sexual selection for immunocompetence (Folstad and Karter 1992), parasite resistance (Hamilton and Zuk 1982) and particular MHC genotypes that may be linked to health status (Ekblom el al. 2004, 2009). Great snipes also perform extraordinary annual migrations from their African winter grounds to the breeding areas. Males have been shown to fly more or less in one go at high speed and altitude (Lindström et al. 2016). Being able to perform energetically demanding behaviour may thus be important in this species during other parts of the year and not only in the mating period.

In order to determine expression profiles of genes transcribed during episodes of mate choice, we sequenced the transcriptomes (RNA-Seq) of peripheral blood cells from male great snipe caught during lekking. The main aim was to identify candidate genes for sexual selection, and in particular to find sequence variation related to the probability of mating. We thus compared gene expression between males with different mating success. We also called SNPs from the transcriptome data, and tested for associations between mating success (number of females mated with) or mating status (mated or not) and SNP genotype in this same sample of birds. Finally, we wanted to investigate the selective mechanisms maintaining genetic variation for such genes in the population. To address this, we tested if a subset of candidate and non-candidate SNPs were correlated with mating status in a larger data set consisting of males studied during several years, and investigated spatial and temporal genetic variation across different loci. Avian malaria is known to have detrimental effects in other birds, and that it is likely that snipe also have it. In this larger data set, we also studied associations between parasite prevalence and SNP genotypes by scoring the prevalence of avian malaria.

Methods

Sampling and sequencing

Fieldwork for the RNA-seq study was conducted during the spring of 2010 in Gåvålia (62º 17’N, 9º 36’ E), Central Norway (Løfaldli et al. 1992; Sæther et al. 2005). The two largest

great snipe leks in the study area were followed during the main mating period in May and June. At the onset of the breeding season, displaying birds attending the leks were captured using mist nets and individually marked using coloured plastic rings (see Løfaldli et al. 1992). Birds were sexed and blood samples were taken from the brachial vein and stored in RNAprotect Animal Blood Tubes (Qiagen). Behavioural observations were made each night during the mating season by two to four trained observers per lek, and the mating success of individual males (minimum number of females mated, or soliciting mating, with the male during the entire field season) was estimated using our previously described approach (Fiske and Kålås 1995; Sæther et al. 2000). Only males observed in detail for at least 5 days during the mating season were assigned mating success.

RNA was extracted from blood samples of 14 individual males with known mating success using RNeasy Protect Animal Blood Kit (Qiagen) and cDNA was synthesised using the MINT kit (Evrogen). cDNA was purified through QIAquick PCR purification columns (Qiagen) and diluted to a concentration of 100 ng/µL (measured on a Nanodrop, Thermo Scientific). Tagged sequencing libraries of each individual were prepared and sequenced on one full plate of GS FLX Titanium (Roche 454) at the Uppsala SNP&SEQ Technology Platform (www.sequencing.se). The 454 plate was divided into two sections, and 7 individuals with tags were pooled in each section, respectively. The raw sequencing reads have been placed in the NCBI Sequence Read Archive (SRA) under accession number SRA060814.

A de-novo assembly of the great snipe transcriptome, utilising data from all individuals simultaneously, was conducted with the default parameters (minimum read length = 20 bp, minimum overlap length = 40 bp, minimum overlap identity = 90%) of the cDNA/transcriptome algorithm implemented in gsAssembler (Newbler, version 2.6, 454 Life Sciences). Prior to assembly, the multiplex identifiers used to discern between individual libraries, as well as sequencing adaptors, were removed and reads were quality-filtered using default settings in Newbler. To trim away the cDNA synthesis primers, a separate file containing the primer sequences was entered into the trimming database. We used the 'cDNA' mode of gsAssembler, which generated two types of results: one was contigs, i.e. the assembled sequences split by the inferred alternative splicing sites, equivalent to the exons; the other was isotigs, i.e. the splice variants consisting of the affiliated contigs, equivalent to

the transcript isoforms. The program used isogroups to classify the contigs and isotigs that were from the same inferred genes. All computations were performed using the computational resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (Uppmax).

Assembled contigs, isotigs and the singletons larger than 100 bp were functionally annotated. Inferences on their functional properties were first made by conducting BLAST searches (BLAST 2.2.24+; Altschul et al. 1997) of the isotigs and size-filtered singletons. The functional annotation of the contigs was then obtained by their correspondent isotigs through the isogroup identifiers. The BLASTX program with an e-value cut-off of 10e-10 was used to ensure the quality of the analysis. The BLAST search was performed against both the chicken (Gallus gallus) genes (WASHUC2) International Chicken Genome Sequencing Consortium and the zebra finch (Taeniopygia guttata) genes (taeGut3.2.4)(Warren et al. 2010). The BLAST databases were downloaded from the Ensembl server through BioMart (Hubbard et al. 2002, Smedley et al. 2009). A list of annotated genes and their corresponding expression levels can be found in Supplementary Information Table S1.

Gene expression analyses

To investigate gene expression levels, each individual sequencing library was mapped back to the isotigs generated from the assembly. Since some singletons might be from the same genes as the isotigs, but were not assembled because of low sequencing coverage, all the singletons with annotation information were also included in the mapping reference file. The mapping was conducted by gsMapper (Newbler, version 2.6, 454 Life Sciences) with 'cDNA' mode and its default parameter settings (minimum read length = 20 bp, minimum overlap length = 40 bp, minimum overlap identity = 90%). Reads that mapped to alternative splice variants of the same gene were only counted once and reads that mapped equally well to two or more genes were removed.

Statistical analyses and handling of large datasets were performed using R 2.15 and later versions (R-core team 2014). Information on the individual males included in the transcriptome analyses can be found in Table S2. Read counts per each gene and individual were extracted from the mapping results. Genes that were only expressed in single individuals, or with a total read count number of less than 10, were filtered out from the read

count libraries. Differential gene expression analyses were conducted using the R/Bioconductor (Gentleman 2004) and edgeR (Robinson and Oshlack 2010) packages. Normalization of the read counts is critical when comparing gene expression between treatments or groups, since it adjusts for overall RNA population differences between the samples (Robinson et al. 2010). For the edgeR analysis, normalization across all samples was performed using the Trimmed Mean of M values (TMM) protocol. Dispersion estimation was calculated based on all the genes jointly. False discovery rate (FDR) correction (Benjamini & Hochberg 1995) was applied to control for multiple testing. We tested for an association between mating success (number of females mated) and normalised gene expression levels using Spearman rank correlations.

Identification of genotypes associated with mating success

To identify SNPs in the transcript sequences, reads from all individuals were mapped to the contigs using gsMapper with the 'cDNA' mode and its default settings. High-confidence SNPs were extracted applying the stringency criteria of: coverage ≥ 10, minor allele frequency (MAF) $\geq 5\%$, and coverage of minor allele ≥ 3, to guarantee the quality of the SNP calling. The sequencing data for each individual were then mapped back onto the assembled contigs to estimate the SNP genotypes for each individual using gsMapper. For the SNP genotypes of each individual, only those from the known SNP panel were kept, and further criteria of coverage ≥ 5, minor allele frequency (MAF) $\geq 5\%$, and coverage of minor allele ≥ 2 were applied. Since the homologous SNP sites were not reported in the SNP variants analysis, we examined the mapping quality and coverage of the known SNP sites for each individual and accepted the ones with a mapping quality ≥ 10 and a coverage*MAF ≥ 5 as homologous SNPs.

The effect on mating success of the genotype for each SNP were tested with Generalized Linear Models, fitted with a Poisson error distribution and a log link function (PGLM). In order to investigate the congruence between candidate genes from the gene expression analysis and the genotype associations, the effect sizes from this model (deviance explained, $D^2_{Adjusted}$, adjusted for the number of fitted regression parameters and the number of observations, Weisberg 1980, Guisan and Zimmermann 2000) for each locus were plotted against the coefficient from the correlation between mating success and individual gene expression levels. Significance thresholds for tests involving all genes from the transcriptome

dataset were set to 0.01 in order to limit the number of false positives. We also investigated genetic differentiation between mated and unmated males (mating status) using the 'population differentiation' option in GenePop (http://genepop.curtin.edu.au/) (Raymond and Rousset 1995), testing for both genic differentiation (differences in allele frequencies) and genotypic differentiation (differences in allele combinations) for each locus separately using the G-tests. Candidate SNPs were defined as loci with a significant association in at least one of these comparisons (Table 1) all other SNPs were treated as non-candidates. The genotype data is available at Dryad (http://datadryad.org/review?doi=doi:10.5061/dryad.p42r8).

Testing the effects of the candidate SNPs in a larger data set

In order to validate the candidate genes for mating success identified from the RNA-Seq data and to investigate patterns of spatial and temporal genetic structure, we submitted 48 SNP markers (including both candidates and non-candidates, as inferred from the analyses on genotype and mating success and/or mating status) for multiplex genotyping using the Illumina Golden Gate assay available at the SNP & SEQ Technology Platform at Uppsala University (www.genotyping.se). 38 of these yielded reliable genotype scores and were polymorphic in the genotyped individuals. Markers were chosen from polymorphic positions in the transcriptome data based on Illumina design scores. Primers for multiplex genotyping were designed for SNPs flanked by regions of, at least, 100 bp on both sides. Information on SNP markers including flanking sequences can be found in Supplementary Information Table S3. To avoid pseudo replication and year effects when testing for effects of genotype on mating success we only entered individual males once in any analysis, and transformed yearly mating success to lifetime mating status ("1" = >0 matings and "0" = no observed matings). We obtained genotypes for seven of the candidate loci listed in Table 1 and for 31 non-candidate loci, and analysed if the genotype or zygosity of SNPs were associated with mating status (mated or not) using likelihood tests (G-tests) of the contingency tables, and using GenePop tests for genetic differentiation (as described above). This data set consisted of 130 males caught and studied at the Gåvålia site during the mating seasons of 1994-2002 (see Fiske and Kålås 1995; Sæther et al. 2000, Sæther et al. 2005). The genotype data for this and the below analyses is available at Dryad (http://datadryad.org/review?doi=doi:10.5061/dryad.p42r8).

Spatio-temporal population genetic structure of candidate and control SNPs

If genetic variation for sexually selected traits are upheld by spatially or temporally varying selection pressures we should be able to observe this as increased genetic population structure or temporal variation for selected loci compared to neutral loci. If on the other hand balancing sexual selection is acting across time and space we would expect decreased structure in the selected SNPs. We thus examined spatio-temporal population genetic structure of both candidate and non-candidate markers (as defined from the analysis of the RNA-Seq genotypes; Table 1), by genotyping 304 birds caught at 10 sites in Norway, Sweden (7 sites) and Poland/Estonia (3 sites) during 1994–2002, (see Ekblom et al. 2007, Sæther et al. 2007 for details) for genetic variation at the same 38 loci as above. Except for the males caught at the main site (Gåvålia) in Norway, we had no data on mating status of these birds.

In order to identify marker specific signatures of selection (Storz 2005, Beaumont 2005) we ran outlier analyses using the software LOSITAN (Antao et al. 2008), and tested for outlier loci with respect to spatial population structure (deviant F_{ST} among regions, East (Poland, Estonia East and Estonia West) and West (the Scandinavian sites), and temporal structure by analysing two time periods (early: 1994–98 and late: 1999–2002). A few males whose lifespan encompassed both time periods were ascribed to the first period.

To further explore spatial and temporal structure we used part of the code of the R package ADEGENET (Jombart 2008) to create PCA plots to illustrate the multidimensional relationships between each individual genotype in two dimensional plots (PC1 vs. PC2). These analyses were done for spatial structure by examining regional (East vs. West) and temporal differences (Early vs. Late) for both candidate and non-candidate loci, respectively. We also examined the corresponding differentiation by calculating F_{ST} (with 95% confidence limits determined by 1000 bootstrap replicates over individuals) among these "populations" in the R package Hierfstat (Goudet and Jombart 2015). We furthermore ran corresponding STRUCTURE analyses with 100,000 burn-in followed by 1,000,000 chains for K = 1–10, with 10 repetitions for each K using the admixture model with loc prior option. We ran separate analyses using candidate and non-candidate SNPs respectively in STRUCTURE 2.3.4. (Pritchard et al. 2000) and used the R package POPHELPER v1.2.0 (Francis 2016) to extract the relevant data from all STRUCTURE output files and plotted the posterior estimate of the likelihoods and ΔK for each K (Evanno et al. 2005). The assignment probability (Q) for

each K was summarised using the software CLUMPP (Jakobsson and Rosenberg 2007) and visualised using POPHELPER.

Parasite prevalence

In the same larger data set, we examined associations between parasites and SNP genotypes by scoring the prevalence of avian malaria using a nested-PCR protocol. More specifically we scored the prevalence of Plasmodium/Haemoproteus and Leucocytozoon infections for all birds caught in 1994–2002. The external primer pair HaemNFI/ HaemR2L (Bensch et al. 2000) was used to amplify a part of the haemosporidian mtDNA cyt b gene. The PCR included 0.125 mM of each dNTP, 1.5 mM MgCl2, 10x NH4 buffer, 0.6 µM of each primer, 0.1 units of Biotaq DNA polymerase (Bioline, Luckenwalde, Germany) and 1 µl gDNA in a 10-µl reaction. Detecting avian malaria of the genera Plasmodium and Haemoproteus was done with the primer pair HaemF/HaemR2 (Hellgren et al. 2004) and Leucocytozoon with primer pair Haem FL/HaemR2L (Bensch et al. 2000). The external PCR product was used as a template for the internal PCRs. The internal PCRs had the same composition as the external PCR, but for Leucocytozoon 0.125 units of Biotaq was used. PCR conditions were as described in Bensch et al. (2000). The final PCR product was visualized on an ethidium bromide-stained agarose gel (1.5%) to determine the initial presence/absence of parasite DNA.

We used the birds from the Gåvålia site in contingency table tests (G-tests) for associations between malaria infection and mating status, using mating data from the year a given male was screened for parasites. We used birds from all sites for investigating relationships between genotype and malaria. Finally, we analysed the interaction effect between malaria and genotype/zygosity of candidate SNPs on mating status (in the screening year) using logistic regression.

Results

Expression data

The three most highly-expressed transcripts in peripheral blood cells from male great snipe corresponded to genes encoding haemoglobin subunits in the chicken and zebra finch

genomes (HBAA, HBG1 and HBAD)(Fig. 1). Individual expression levels in all of these tended to be positively correlated to mating success (but non-significantly so; Fig. 2).

We found 34 genes with a significant interaction between male mating success and gene expression levels (23 with positive correlations and 11 with negative correlations). However only one of these remained significant after correcting for false discovery rates (Fig.1). This gene codes for "EF-hand calcium-binding domain-containing protein 14" (EFCAB14) and was expressed at a relatively low level in most samples, but had no expression at all in some of the un-mated males (Fig. 2).

SNP analyses of expressed genes

We identified a total of 2,874 gene-linked SNPs, covering 618 contigs, in our transcriptome data. For downstream analyses we used 288 SNPs, in which we could make reliable genotype calls for at least 10 individuals (mean Minor Allele Frequency=0.29). Genotypes of individuals and analysis of relationship with mating success is presented for all SNPs in Table S4. Among these markers we found 14 SNPs with a significant association of mating success or mating status with genotype (genic and genotypic). These 14 candidate SNPs were defined by showing a significant association in at least one of the comparisons provided in Table 1. Of these, three were non-synonymous substitutions, the rest were synonymous SNPs (Table 1). Four of the SNPs were found in HBAA (one non-synonymous and three synonymous). The two remaining non-synonymous SNPs were associated with acidic ribosomal phosphoprotein (RPLB0) and polyubiquitin-B (UBB) (Table 1, Fig. 3). Some of the top candidate genes, like HBAA and polyadenylate-binding protein PABPC1, showed signs of heterozygote advantage (Fig. 3, inserts).Markers situated in genes with a positive correlation between male mating success and gene expression level, showed an increased relationship between mating success and SNP genotype ($r = 0.29$, $df = 52$, $p = 0.033$, Fig. 3).

Analyses of candidate and non-candidate SNPs in a larger dataset

To check the generality of the above results, we tested for genotypic differences and heterozygote advantage for a subset of SNPs, genotyped in 130 males sampled during 1994-2002 in the core study area. This subset included seven of the previously identified candidate SNPs from Table 1. We did not find any evidence that individuals heterozygous for any of the these seven previously identified candidate SNPs had different mating status than

homozygous individuals nor any consistent effects of genotype on mating status (Table 2; Supplementary Information, Table S5) – with one exception: there was a significant effect of zygosity for a SNP in Ribosomal protein S14, but with higher proportion of mated homozygous males rather than heterozygous males. When testing for genic and genotypic differentiation (in the same way as done to identify candidates using GenePop), we found that the two most highly differentiated markers were in candidate genes (RHD and RPS3). But these were not statistically significant (Table 1, Table S5).

Outlier tests for differential spatial selection (F_{ST} among regions vs. heterozygosity), revealed two loci, unknown protein 1487_1016 and Rh blood group D antigen (RHD), which showed signs of being differentially selected among regions (Fig. 4a). Among these, the RHD gene was among the previously identified candidate loci. DAPC and STRUCTURE analyses suggested the presence of weak population structure (F_{ST} = 0.034; 95% confidence intervals: 0.024 - 0.047; P < 0.0001) among regions when using the 31 non-candidate loci (Fig. 4b, Supplementary information, Fig. S6). But we found no differentiation (F_{ST} = 0.002; 95% confidence intervals: -0.006 - 0.007; NS) when limiting the analysis to only the 7 candidate loci (Fig. 4c). To obtain a similar number of individuals from each geographic region, and avoid ascertainment bias due to overrepresentation of Gåvålia birds, we did these analyses using all 51 individuals from the Eastern region (Poland/Estonia) and randomly selected 51 birds from Norway/Sweden. Using all the genotyped birds did not alter our conclusions.

We then divided the data into two time periods (first observation up to and including 1998 and after) and tested for temporal outliers. We found the same SNP as above with no annotation information available (1487_1016) showing evidence of being differentially selected over time (Fig. 5). As we observed only one P-value <0.05, this may be coincidental given the number of tests and the chosen significance level. PCA and STRUCTURE analyses suggested absence of genetic differentiation for both non-candidate (F_{ST} =-0.001; 95% confidence intervals: -0.006 - -0.002; NS) and candidate SNPs (F_{ST} =-0.004; 95% confidence intervals: -0.006 - 0.007; NS) among the two time periods (Fig. 5 and Supplementary information, Fig. S7).

Avian malaria

Males infected with Plasmodium/Haemoproteus had lower mating status (in the year of screening) than uninfected males: only nine of 32 infected birds (28%) mated while 77 out of 162 uninfected birds (48%) did so, indicating that parasite infection may affect performance on the lek (G-test, $\chi^2 = 4.2$, P= 0.04).

Genotype appeared to have low or no effect on infection status of Plasmodium/Haemoproteus. (Table 3a). Leucocytozoon prevalence was generally too low (<5%) to allow meaningful analyses. The strongest associations are given in Table 3b and the full prevalence data is presented in Table S8.

Interaction between genotype and malaria infection

Since malaria had a negative effect on obtaining matings, it could be that males with certain genotypes have an advantage if infected and that this can obscure relationships between genotype and mating success. We therefore analysed the joint effect of malaria and genotype/zygosity of candidate SNPs on mating status. This reduced the sample size of available males somewhat, and malaria infection did not any longer show a significant effect on its own. Two of the seven candidate genes showed a significant interaction effect of malaria and genotype and/or zygosity on mating status, both of these were located in HBAA (Table 4, Figure S9). One was a synonymous (6311_93) and one was non-synonymous substitution (5447_210). The SNP 5192_296 located in the RPS14-gene showed a direct effect of zygosity on mating success with homozygotes experiencing higher mating status compared to heterozygotes regardless of infection status (Figure S9).

Discussion

By comparing the gene-expression profiles and nucleotide variation in relation to mating success and mating status we attempted at a de novo identification of candidate genes that are related to sexual selection in male great snipe. We were able to assemble the great snipe transcriptome using data from fourteen sequenced males simultaneously. The assembly included almost six thousand previously annotated bird genes and another seventeen hundred unknown great snipe genes. This is comparable to other de-novo transcriptome

433 characterisation studies with a similar sequencing effort (Künstner et al. 2010).
434
435 Expression analyses and downstream SNP variation analyses of expressed genes have
436 previously been utilised in ecological RNA-Seq studies to detect potential loci under
437 selection. For example, several genes were detected to be associated with morphological
438 differentiation of lake trout (Goetz et al. 2010) and some SNP loci were identified to be
439 associated with growth rate in rainbow trout (Salem et al. 2012). In this study, we used a
440 similar approach to perform a comprehensive identification of candidate genes correlated with
441 mating performance in the great snipe. We found one differentially expressed gene and 14
442 candidate SNP loci with genetic differentiation in respect to male mating success.
443
444 The most highly expressed gene across all individuals was, not surprisingly since we studied
445 transcription levels in peripheral blood cells, coding for alpha 1 globin (HBAA), a protein
446 involved in oxygen transport (Ashburner et al. 2000). This and other haemoglobin genes also
447 tended to show a positive correlation between expression levels and male mating success (but
448 not significantly after multiple test correction). We did find significant positive relationships
449 between mating success and the expression of the EFCAB14 gene, but the functional link
450 remains obscure here.
451
452 We also found a correlation between the differential expression level of genes and the effect
453 of SNP genotype on mating success, suggesting that markers situated in genes with higher
454 expression in mated males showed a stronger relationship between genotype and mating
455 success. This indicates that there is a functional link between genotype and expression level,
456 also known as allele specific expression. A possible explanation is that mated males were able
457 to “turn on” and upregulate certain genes and thus improved their physical condition allowing
458 them access to females and matings, while males that become unmated may be forced to
459 upregulate alternative genes. Alternatively, mating directly triggers the expression of certain
460 genes, but this is a less likely explanation since the male blood was sampled in the beginning
461 of the mating season.
462
463 Among the SNPs found to correlate with mating success/status in our limited data set based
464 on the SNP-calling of the RNA-transcripts, three out of 14 were non-synonymous, suggesting
465 functional changes in the downstream proteins. The 11 synonymous SNPs associated with

mating could possibly be explained by hitchhiking with linked, functional variants. However, other studies of fitness correlations or outlier tests have also found a signal of selection for synonymous substitutions, and thus prompting alternative functional explanations such as selection on alternative splicing and miRNA stability (Milano et al. 2014).

When analysing SNP genotypes among the expressed genes we found one non-synonymous SNP in HBAA. HBAA polymorphisms will be discussed in detail below. The two remaining non-synonymous SNPs were associated with the genes encoding acidic ribosomal phosphoprotein (RPLB0) and polyubiquitin-B (UBB). GO-terms for RPLP0 includes "host-virus interaction" and it has been associated with the innate immune response in human corectal cancer patients (Benvenuto et al. 2015). UBB is coding for one of the most conserved proteins known among eukaryotes and is involved in protein degradation and possibly regulation of gene expression. Mutations in this gene has been associated with a number of human disorders (see Conoway et al. 2002).

Comparative and functional studies of haemoglobin, the oxygen-transporting molecule of nearly all vertebrates, have been performed for more than half a century (Perutz 1972; 1983; Barra et al. 1981). Haemoglobin carries oxygen from the respiratory organs to the rest of the tissues where it releases the oxygen to burn nutrients to provide energy to power the functions of the organism during metabolism. Genetic polymorphisms for the genes encoding haemoglobin have been detected among a large number of species and many of these polymorphisms have been shown to correlate with, and explain, local adaptations to hypoxia (Jessen et al. 1991; Storz et al. 2007; Weber 2007), physical work ability (Gardner et al. 1977), anaemia (Ingram et al. 1957) and malaria resistance (Allison 1954). Thus, a large body of evidence suggests that haemoglobin molecules are subjected to selection and that different variants may be favoured in different situations.

Both HBAA and a synonymous SNP for polyadenylate-binding protein PABPC1, showed signs of heterozygote advantage in our RNA-Seq data. Genetic polymorphism in the human homologues of HBAA, HBA1 and HBA2 causes α-thalassemia. The worldwide distribution of this and other disorders corresponds to areas of human malaria exposure, and heterozygotes have been shown to be protected against severe malaria (Flint et al. 1986, 2001, Kwiatkowski 2005).

Our above analyses suggest a number SNPs associated with mating success in the limited dataset of 14 males for which transcriptome data were accessible. However, this result could not be replicated in the larger dataset, although we had to rely only on the cruder mating score (as mated or not) in the latter analyses. This could suggest that the signals found in the transcriptome data may have been a result of falsely inferred candidate genes in this first dataset (which is sensitive to random noise due to a small sample size, and a high degree of multiple testing due to the large number of markers used). Finding the genetic basis for a complex, fitness related trait such as "mating success" may need a better control of environmental confounding effects or experimental manipulation – which is not possible in our studied system. Studies of sexual selection have been trying to find the traits associated with mating success for decades (Andersson 1994, Jones and Ratterman 2009) and consensus is that such traits most likely are polygenic (Andersson 1994) and also that sexual selection involves many different traits and signalling modes (Schaefer and Ruxton 2015). What makes a successful great snipe is likely condition dependent (Höglund et al. 1992, Ekblom et al. 2004, 2009), and if driven by disease resistance the underlying genetics is most certainly very complex.

As in previous studies (Ekblom et al. 2007) we found a weak but significant genetic structure between the eastern and the western population of great snipes when analysing non-candidate SNP markers. In contrast there was no geographic structure in the pre-identified candidate markers. This pattern is consistent with directional or balancing sexual selection acting on these genes, but does not support the hypothesis that spatially varying selection patterns are preserving variation at the sexually selected loci. It should however be noted that many processes other than sexual selection could potentially generate within genome variance in spatial structure (Wolf and Ellegren 2016). As we analysed very few candidate loci compared to non-candidates this could also be an effect of lower statistical power. The only one of the candidate genes for which we found increased spatial structure compared to neutral expectations was the Rh blood group D antigen (RHD). This gene is associated with the Rh blood group system in humans and non-coding polymorphisms have been shown to be inflammatory markers (Flegel and Wagner 2000).

Our results, even if not congruent among data sets, do hint at relationships with disease

resistance genes. In this study we showed that malaria infected males were less likely to belong to the 'mated' class of birds indicating mated birds were less often carrying avian malaria. The two SNPs with a significant interaction between genotype and malaria prevalence with mating status are both found in the gene encoding for HBAA, one being a synonymous substitution and the other non-synonymous. For the synonymous substitution there were also signs of a significant interaction between zygosity and malaria with mating status. However, the relationships are weak and we cannot completely rule out that they may be coincidental. For the SNP in the gene RPS14 we found a direct effect of heterozygosity but here birds with matings appeared to more often be homozygous. RPS14 is reported to be involved in human erythrocyte differentiation (Schneider et al. 2016).

More than thirty years ago it was suggested that parasites play a key role in generating sexual selection (Hamilton and Zuk 1982). It was suggested that sexual reproduction and recombination is a defence against antagonistically co-evolving parasites. Under this hypothesis, selecting males that carry parasite-resistant genes plays a key function in mate choice and species are predicted to evolve complex behaviours that allow the choosy sex to reveal such genes. Lekking behaviours could be indicators of such resistance, since it is difficult to perform courtship and territory defence while ill. How genetic variation is maintained for such traits has been a longstanding question in evolutionary biology (Hamilton and Zuk 1982, Andersson 1994). Various explanations have been put forward, including fluctuating selection in time and space, antagonistic selection between the sexes (sexual conflict), antagonistic pleiotropic effects and balancing selection favouring heterozygote genotypes (reviewed by Hedrick and Kim 1975, Radwan 2008). In this study we find no conclusive evidence for any of these mechanisms but we do show that there is genetic variation in genes expressed during lekking (which is when mate choice takes place). We document genetic variation for the most highly transcribed gene in peripheral blood cells, HBAA and several other disease- and condition related genes in a wild lekking bird species. Furthermore, males infected with avian malaria parasites had lower mating success than non-infected males. Our results thus suggest that physically exhausting lek displays could be used by females as indicators of 'good genes' while selecting their mates and that genetic variation for such traits is present in the population. This does not exclude the possibility that the same genetic variants that allow good health status during lekking may also be advantageous during other parts of the great snipe life cycle (for example during migration).

Acknowledgements

We thank three anonymous reviewers for very constructive criticism on former versions of this paper. Sten L. Svartaas, Magnus Johansson, Urs Kormann, Robert Mannelqvist, Claire Morandin, Lenka Vysinova and many other field assistants provided excellent assistance during field work. Library preparation, sequencing and SNP genotyping were completed at the Uppsala University SNP&SEQ Technology Platform. The high-performance computing facility at Uppsala University (SNIC-UPPMAX) was utilised for bioinformatic analyses. Ethical permits for capture and blood sampling of birds were obtained from the Stavanger Museum and the Norwegian Animal Research Authority. Sample preparation in Sheffield was supported by the UK Natural Environment Research Council. Funding for this study was granted by the Carl Trygger Foundation (to RE) and the Swedish Research Council (VR) to JH.

References

Allison AC (1954) Protection afforded by sickle-cell trait against subtertian malarial infection. British Medical Journal 6, 290–294.

Alonso JC, Magana M, Martin CA, et al. (2010) Sexual traits as quality indicators in lekking male great bustards. Ethology 116, 1084-1098.

Altschul SF, Madden TL, Schäffer AA et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389-3402.

Andersson M (1994) Sexual selection. Princeton Univ Press, Princeton.

Antao T, Lopes A, Lopes RJ et al. (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{st}-outlier method. BMC Bioinformatics 9, 323.

Ashburner M, Ball CA, Blake JA, et al. (2000) Gene ontology: tool for the unification of biology. The gene onthology consortium. Nature Genetics 25, 25-29.

Barra D, Bossa F, Brunori M (1981) Structure for binding sites for heterotropic effectors in fish haemoglobins. Nature 293, 587-588.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. B 57, 289-300.

598 Bensch S, Stjernman M, Hasselquist D et al. (2000) Host specificity in avian blood parasites:
599 a study of Plasmodium and Haemoproteus mitochondrial DNA amplified from birds.
600 Proceedings of the Royal Society B 267, 1583–1589.
601 Benvenuto M, Sileri P, Rossi P (2015) Natural humoral immune response to ribosomal P0
602 protein in corectal cancerpatients. Journal of Translational Medicine 13, 101.
603 Beaumont MA (2005). Adaptation and speciation: what can F-st tell us? Trends in Ecology
604 and Evolution 20, 435–440.
605 Borgia G (1979) Sexual selection and the evolution of mating systems In: Sexual
606 selection and reproductive competition (eds Blum M, Blum A) pp 19-80
607 Academic Press, New York
608 Conaway RC, Brower CS, Conaway JW (2002) Emerging roles of ubiquitin in transcription
609 regulation. Science 296, 1254–1258.
610 Cornwallis CK, Uller T (2012) Towards an evolutionary ecology of sexual traits. Trends in
611 Ecology and Evolution 25, 145-152
612 Darwin CR (1871) The descent of man, and selection in relation to sex. John Murray, London.
613 DuVal EH (2013) Female mate fidelity in a lek mating system and its implications for the
614 evolution of cooperative lekking behavior. American Naturalist 181, 213-222
615 Ekblom R, Sæther SA, Grahn M, et al. (2004) Major histocompatibility complex variation
616 and mate choice in a lekking bird, the great snipe (Gallinago media). Molecular Ecology
617 13, 3821-3828.
618 Ekblom R, Sæther SA, Hasselquist D, et al. (2005) Female choice and male humoral immune
619 response in the lekking great snipe (Gallinago media). Behavioral Ecology 16, 346-351.
620 Ekblom R, Sæther SA, Jacobsson P, et al. (2007) Spatial pattern of Mhc Class II variation in
621 the great snipe (Gallinago media). Molecular Ecology 16: 1439-1451.
622 Ekblom R, Sæther SA, Fiske P, et al. (2009) Balancing selection, sexual selection and
623 geographic structure in MHC genes of Great Snipe. Genetica 138, 453–461.
624 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using
625 the software STRUCTURE: a simulation study. Molecular Ecology 14, 2611–20.
626 Fiske P, Kålås JA (1995) Mate sampling and copulation behaviour of great snipe females
627 Animal Behaviour 49, 209 – 219
628 Flegel WA, Wagner FF (2000) Molecular genetics of RH. Vox Sanguinis 78, 109–15.
629 Flint J, Hill AV, Bowden DK, et al. (1986). High frequencies of α-thalassemia are the result
630 of natural selection by malaria. Nature 321 (6072), 744–50.

631 Flint J, Tufarelli C, Peden J, et al. (2001) Comparative genome analysisdelimits a
632 chromosomal domain and identifies key regulatory elements in the α globin
633 cluster. Human Molecular Genetics 10, 371-382.
634 Folstad I, Karter AJ (1992) Parasites, bright males, and the immunocompetence
635 handicap. American Naturalist 139, 603–622.
636 Francis RM (2016) POPHELPER: an R package and web app to analyse and visualize
637 population structure. Molecular Ecology Resources, doi:10.1111/1755-0998.12509
638 Fromhage L, Kokko H, Reid JM (2009) Evolution of mate choice for genome-wide
639 heterozygosity. Evolution 63: 684-694.
640 Gardner GG, Edgerton VR, Senewiratne B, et al. (1977) Physical work capacity and stress in
641 subjects with iron deficiency anemia. American Journal of Clinical Nutrition 30, 910-
642 917.
643 Gentleman R, Carey V, Bates D, et al. (2004) Bioconductor: open software development for
644 computational biology and bioinformatics. Genome Biology 5, R80
645 Greenfield MD, Danka RG, Gleason JM, et al. (2012) Genotype × environment interaction,
646 environmental heterogeneity and the lek paradox. Journal of Evolutionary Biology 25,
647 601-613.
648 Goetz F, Rosauer D, Sitar S, et al. (2010) A genetic basis for the phenotypic differentiation
649 between siscowet and lean lake trout (Salvelinus namaycush). Molecular Ecology 19,
650 176-196.
651 Goudet J, Jombart T (2015) Package hierfstat. http://github.com/jgx65/hierfstat.
652 Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology.
653 Ecological Modelling 135, 147-186.
654 Hamilton WD, Zuk M (1982) Heritable true fitness and bright birds: a role for parasites.
655 Science 218, 384-387.
656 Hedrick PW, Kim TJ (1975) Genetics of complex polymorphisms: parasites and the
657 maintenance of the major histocompatibility complex variation. In: Evolutionary
658 Genetics: From molecules to morphology (eds. Singh RS, Krimbas CB). pp 713.
659 Cambridge Univ Press, Cambridge.
660 Hellgren O, Waldenström J, Bensch S (2004) A new PCR assay for simultaneous studies of
661 Leucocytozoon, Plasmodium, and Haemoproteus from avian blood. Parasitology 90,
662 797–802.

663 Höglund J, Lundberg A (1987) Sexual selection in a monomorphic lek-breeding bird:
664 correlates of male mating success in the great snipe Gallinago media. Behavioural
665 Ecology and Sociobiology 21, 211-216.
666 Höglund J, Robertson JGM (1990) Female preferences, male decision rules and the evolution
667 of leks in the Great snipe. Animal Behaviour 40, 15-22.
668 Höglund J, Kålås JA, Fiske P (1992) The costs of secondary sexual characters in the lekking
669 great snipe (Gallinago media). Behavioural Ecology and Sociobiology 30, 309-315
670 Höglund J, Alatalo RV (1995) Leks. Princeton Univ Press, Princeton.
671 Hubbard T, Barker D, Birney E, et al. (2002) The Ensembl genome database project. Nucleic
672 Acids Research 30, 38-41.
673 Ingram VR (1957) Gene mutations in human haemoglobin: the chemical difference between
674 normal and sickle cell haemoglobin. Nature 458: 326-328.
675 Jakobsson M, Rosenberg N (2007) CLUMPP: a cluster matching and permutation program
676 for dealing with label switching and multimodality in analysis of population structure.
677 Bioinformatics 23, 1801–1806.
678 Jessen TH, Weber RE, Fermi G, et al. (1991) Adaptation of bird hemoglobins to high-
679 altitudes – demonstration of molecular mechanism by protein engineering. Proceedings
680 of the National Academy of Sciences USA 88, 6519–6522.
681 Jombart T (2008) ADEGENET: a R package for the multivariate analysis of genetic markers.
682 Bioinformatics 24,1403–1405.
683 Jones AG, Ratterman NL (2009) Mate choice and sexual selection: what have learned since
684 Darwin? Proceedings of the National Academy of Sciences 106, 10001-10009.
685 Kålås JA, Fiske P, Sæther SA (1995) The effect of mating probability on risk taking: an
686 experimental study in lekking great snipe. American Naturalist 146, 59-71.
687 Künstner A, Wolf JBW, Backström N et al. (2010) Comparative genomics based on massive
688 paralell transcriptome sequencingreveals patterns of substitution and selection across 10
689 bird species. Molecular Ecology 19, 266-276.
690 Kirkpatrick M, Ryan M (1991) The evolution of mating preferences and the paradox of the
691 lek Nature 350, 33–39.
692 Kotiaho J. Simmons LW. Tomkins JL (2001) Towards a resolution of the lek paradox Nature
693 410, 684-686.
694 Kotiaho J, LeBas NR, Puurtinen M, et al. (2008) On the resolution of the lek paradox. Trends
695 in Ecology and Evolution 23, 1–3.

Kwiatkowski DP (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. American Journal of Human Genetics 77, 171-192.

Lindström Å, Alerstam T, Bahlenberg P, et al**.** (2016) The migration of the great snipe Gallinago media: intriguing variations on a grand theme. Journal of Avian Biology 47, 321-334.

Løfaldli L, Kålås JA, Fiske P (1992) Habitat selection and diet of Great Snipe Gallinago media during breeding. Ibis 134, 35-43.

Milano I, Babbucci M, Cariani A, et al. (2014) Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (Merluccius merluccius). Molecular Ecology 23, 118-135.

Miller CV, Moore AJ (2007) A potential resolution to the lek paradox through indirect genetic effects Proceedings of the Royal Society B 274, 1279-1286.

Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. Methods in Ecology and Evolution 6, 445-464.

Perutz MF (1972) Nature of Haem-Haem interaction Nature 237, 495-499.

Perutz MF (1983) Species adaptation in a protein molecule. Molecular Biology and Evolution 1, 1–28.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Radwan J (2008) Maintenance of genetic variation in sexual ornaments: a review of the mechanisms. Genetica 134, 113-127.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and eumenicism. Journal of Heredity 86, 248–249.

Robinson M, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology 11, R25.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140. e13

Rowe L, Houle D (1996) The lek paradox and the capture of genetic variance by condition dependent traits. Proceedings of the Royal Society B 263, 1415–1421.

728 Ryder TB, Tory WP, Blake JG, et al. (2010) Mate choice for genetic quality: a test of the
729 heterozygosity and compatibility hypotheses in a lek-breeding bird. Behavioral Ecology
730 21, 203-210.
731 Salem M, Vallejo RL, Leeds TD et al. (2012) RNA-Seq identifies SNP markers for growth
732 traits in rainbow trout. PLoS ONE 7, e36264
733 Sæther SA, Fiske P, Kålås JA, et al. (2000) Females of the lekking great snipe do not prefer
734 males with whiter tails. Animal Behaviour 59, 273–280.
735 Sæther SA, Fiske P, Kålås JA (2001) Male mate choice, sexual conflict and strategic
736 allocation of copulations in a lekking bird. Proceedings of the Royal Society B 268,
737 2097-2102.
738 Sæther SA, Baglo R, Fiske P, et al. (2005) Direct and indirect mate choice on leks. American
739 Naturalist 166, 145-157.
740 Sæther SA. Fiske P. Kålås JA. et al. (2007) Inferring local adaptation from Qst-Fst
741 comparisons: neutral genetic and quantitative trait variation in European populations
742 of great snipe. Journal of Evolutionary Biology 20, 1563-1576.
743 Schaefer HM, Ruxton GD (2015) Signal diversity, sexual selection and speciation. Annual
744 Review of Ecology, Evolution and Systematics 46, 573-592.
745 Schneider RK, Schenone M, Ventura Ferreira M, et al. (2016) Rps 14 haploinsufficiency
746 causes a block in erythroid differentiation mediated by S100A8 and S100A9. Nature
747 Medicine 22, 288-297.
748 Shorey L. Piertney S. Stone J. et al. (2000) Fine-scale genetic structuring on Manacus
749 manacus leks. Nature 408, 352-353.
750 Smedley DS, Haider B, Ballester R et al. (2009) BioMart - biological queries made easy.
751 BMC Genomics 10, e22.
752 Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population
753 divergence. Molecular Ecology 14, 671–688.
754 Storz JF, Sabatino SJ, Hoffmann FG et al. (2007) The molecular basis of high-altitude
755 adaptation in deer mice. PLoS Genetics 3, 448–459.
756 Taylor PD, Williams GC (1982) The lek paradox is not resolved. Theoretical
757 Population Biology 22, 392-409
758 Tomkins J, Radwan J, Kotiaho J, et al. (2004) Genic capture and resolving the lek paradox
759 Trends in Ecology and Evolution 19, 323–328.

760 Vehrencamp SL, Bradbury JW, Gibson RM (1989) The energetic costs of display in male
761 sage grouse. Animal Behaviour 38, 885-896.
762 Warren WC, Clayton DF, Ellegren H, et al. (2010) The genome of a songbird. Nature 464,
763 757-762.
764 Weber RE (2007) High-altitude adaptations in vertebrate hemoglobins. Respiratory
765 Physiology and Neurobiology 158, 132–142.
766 Weisberg S (1980) Applied Linear Regression. New York, Wiley
767 Wolf JBW, Ellegren H (2016) Making sense of genomic islands of differentiation in light of
768 speciation. Nature Review Genetics, doi:10.1038/nrg.2016.133
769
770 Author contribution
771 The study was designed by R.E and J.H. J.H, R.E and B.W wrote the manuscript with the
772 assistance of all other authors. S.A.S, J.A.K and P.F led the field work and maintained the behavioural
773 data base. M.K.P.B, B.W and G.H produced and analysed the genetic data under the supervision of
774 R.E, J.H and T.B. S.A.S and P.H produced and analysed the parasite data. S.A.S, M.K.P.B,
775 and B.W. conducted the statistical analyses with the assistance of J.H. and R.E.
776
777
778 Supporting information
779 Table S1. A list of annotated genes and their corresponding expression levels.
780
781 Table S2. Information on the individual males in the transcriptome part of the study.
782
783 Table S3. Information on SNP markers including flanking sequences.
784
785 Table S4. Genotypes of individual males (coded as 0. 1,5 and 1) and results of GLMs
786 (Poisson regression) for the effect of genotype on mating success (number of mated females).
787 P-values from GenePop analysis of alleles (P_Genic) and genotypes (P_Genotypic) are also
788 included. Analyses based on the 288 SNPs in 14 males called from the expression data.
789
790 Table S5. Results from tests of the association between male mating status (mated/unmated)
791 and SNP genotype and zygosity, as inferred from all SNP markers from the larger data set (G-
792 tests and p-values on differentiation from GenePop).

Figure S6. Results of STRUCTURE runs for genetic differentiation among regions. Analyses were run on 31 non-candidates and 7 candidate SNP-loci, respectively, with the extended data set from all sites.

Figure S7. Results of STRUCTURE runs for genetic differentiation among two time periods (Early vs. Late). Analyses were run on 31 non-candidates and 7 candidate SNP-loci, respectively, with the extended data set from the Gåvålia site.

Table S8. Data on avian malaria prevalence and genotype data for 38 SNP in individual birds.

Figure S9. Probability of mating in relation to malaria infection and genotype of seven candidate genes. Estimates and confidence intervals modelled using logistic regression. Mating status measured in the year of screening for malaria (i.e. the year the blood was sampled). “Rugs” show individual males.

Table 1. Top candidate SNPs for association between male mating success/status and genotype, as inferred from great snipe transcriptome sequencing data from 14 males in 2010. SNPs with a significant association in at least one of the tests performed are listed. MAF = Minor Allele Frequency, D^2_{Adj} = Adjusted Deviance Explained from the GLM analyses (Poisson regression, quantified mating success), P_{PGLM} = significance value from the GLM analyses, H_o = observed heterozygosity, H_e = expected heterozygosity, P_{Genic} = significance value for G test of genic differentiation between mated and un-mated males, $P_{Genotypic}$ = significance value for G test of genotypic differentiation between mated and un-mated males. For full data on all SNPs see Supplementary Table S4.

SNP_ID	MAF	D^2_{Adj}	P_{PGLM}	H_o	H_e	P_{Genic}	$P_{Genotypic}$	ENSEMBL number	Type	Gene	Description
5447_197	0.29	0.43	0.0024	0.43	0.41	0.0145	0.011	ENSGALG00000007468	Syn	HBAA	Alpha 1 globin
5447_210	0.25	0.47	0.0006	0.5	0.38	0.0316	0.0109	ENSGALG00000007468	nonSyn	HBAA	Alpha 1 globin
6308_135	0.41	0.56	0.0137	0.27	0.48	0.0159	0.0538	ENSGALG00000007468	Syn	HBAA	Alpha 1 globin
4291_136	0.41	0.42	0.0061	0.27	0.48	0.0015	0.01	ENSGALG00000017330	Syn	RPS3	Ribosomal protein S3
4035_467	0.41	0.2	0.0334	0.27	0.48	0.0262	0.0675	ENSGALG00000004818	Syn	RPL6	Ribosomal protein L6
6311_93	0.29	0.11	0.0686	0.42	0.41	0.0331	0.029	ENSGALG00000007463	Syn	HBAA	Alpha 1 globin
3776_761	0.29	0.29	0.0042	0.58	0.41	0.1719	0.0719	ENSGALG00000023294	nonSyn	RPLP0	Acidic ribosomal phosphoprotein
2233_202	0.43	0.59	0.0005	0.43	0.49	0.1132	0.1448	ENSGALG00000014450	Syn	PABPC1	Polyadenylate-binding protein 1
1276_857	0.23	0.37	0.0088	0.27	0.35	0.1132	0.1748	ENSGALG00000001233	Syn	RHD	Rh blood group D antigen
5394_129	0.46	0.38	0.004	0.79	0.5	0.2568	0.084	ENSGALG00000004509	nonSyn	UBB	Polyubiquitin-B
4662_109	0.31	0.45	0.002	0.46	0.43	0.1873	0.186	ENSGALG00000017299	Syn	RPS11	40S ribosomal protein S11
5192_296	0.36	-0.11	0.494	0.57	0.46	0.0484	0.0211	ENSGALG00000004588	Syn	RPS14	Ribosomal protein S14
2530_298	0.15	0.41	0.0037	0.15	0.26	1	1	ENSGALG00000014023	Syn	H2AFZ	Histone H2A.Z
3776_914	0.23	0.42	0.006	0.27	0.35	1	1	ENSGALG00000023294	Syn	RPLP0	Acidic ribosomal phosphoprotein

Table 2. Association between male mating status (mated/unmated) and SNP genotype (df = 2) and zygosity (df = 1), as inferred from SNP typing data from the larger data set for seven of the candidate loci listed in Table 1 (G-tests and p-values on differentiation from GenePop). Results from all markers are presented in Table S5.

SNP_ID	n	Genotype $\chi 2$	P	Zygosity $\chi 2$	P	GenePop p-value		Type	Gene	Description
						Genic	Genotypic			
5447_210	125	3.57	0.16	0.01	0.94	0.89	0.89	nonSyn	HBAA	Alpha 1 globin
6308_135	129	2.26	0.32	1.57	0.21	0.79	0.78	Syn	HBAA	Alpha 1 globin
4291_136	128	0.54	0.76	0.02	0.88	0.12	0.12	Syn	RPS3	Ribosomal protein S3
4035_467	128	0.62	0.73	0.58	0.45	0.76	0.74	Syn	RPL6	Ribosomal protein L6
6311_93	130	4.14	0.12	0.18	0.67	0.89	0.89	Syn	HBAA	Alpha 1 globin
1276_857	130	0.09	0.95	0.01	0.96	0.077	0.079	Syn	RHD	Rh blood group D antigen
5192_296	123	4.67	0.10	4.65	0.03	0.33	0.32	Syn	RPS14	Ribosomal protein S14

Table 3. SNP genotypes showing the highest association with avian malaria infection status (SNP id followed by Gene abbreviation). For each genotype the number of uninfected individuals is given followed by the number of infected individuals after the slash. a) SNPs associated with Plasmodium/Haemoproteus infection and b) with Leucocytozoon infection.

SNP ID	Gene	Genotype			χ^2	P
a)		AA	AB	BB		
3159_891	Anon SNP	23/8	86/20	72/35	5.33	0.07
3236_232	EIF4H	68/19	63/32	45/12	4.37	0.13
b)						
3366_1091	UNK Prot	223/5	14/3	-	11.96	0.001
3766_604	VEZF1	72/0	115/7	49/1	5.02	0.06
4045_521	SRSF5A	17/1	81/0	136/7	4.18	0.09

Table 4. Effects of candidate SNP genotype, malaria infection status and their interaction on mating status. Sequential analysis of deviance of logistic regression models. Mating status modelled as a function of infected or not with malaria (Plasmodium/Haemoproteus), genotype or zygosity of the SNP, and interaction between malaria and genotype or zygosity. For each SNP, a model is tested against the one above it showing the reductions in the residual deviance as each term of the formula is added in turn. Significant models highlighted in yellow.

	Genotype						Zygosity					
Candidate SNP	Model	Df	Deviance	Residual Df	Residual Deviance	Pr(>Chi)	Model	Df	Deviance	Residual Df	Residual Deviance	Pr(>Chi)
1276_857	NULL			89.000	118.288		NULL			89	118.288	
RHD	Malaria	1	0.834	88.000	117.454	0.361	Malaria	1	0.834	88	117.454	0.361
	Genotype	2	2.250	86.000	115.205	0.325	Zygosity	1	1.514	87	115.940	0.218
	Malaria:Genotype	1	1.280	85.000	113.925	0.258	Malaria:Zygosity	1	1.327	86	114.613	0.249
4035_467	NULL			88.000	117.368		NULL			88	117.368	
RPL6	Malaria	1	0.782	87.000	116.586	0.376	Malaria	1	0.782	87	116.586	0.376
	Genotype	2	1.338	85.000	115.248	0.512	Zygosity	1	1.146	86	115.440	0.284
	Malaria:Genotype	1	0.957	84.000	114.291	0.328	Malaria:Zygosity	1	1.017	85	114.424	0.313
4291_136	NULL			88.000	116.262		NULL			88	116.262	
RPS3	Malaria	1	0.931	87.000	115.331	0.335	Malaria	1	0.931	87	115.331	0.335
	Genotype	2	2.334	85.000	112.998	0.311	Zygosity	1	1.784	86	113.547	0.182
	Malaria:Genotype	2	1.594	83.000	111.404	0.451	Malaria:Zygosity	1	0.246	85	113.301	0.620
5192_296	NULL			86.000	114.454		NULL			86	114.454	
RPS14	Malaria	1	0.823	85.000	113.631	0.364	Malaria	1	0.823	85	113.631	0.364
	Genotype	2	4.465	83.000	109.166	0.107	Zygosity	1	4.377	84	109.255	0.036
	Malaria:Genotype	2	2.034	81.000	107.132	0.362	Malaria:Zygosity	1	0.387	83	108.868	0.534
5447_210	NULL			86.000	114.454		NULL			86	114.454	
HBAA	Malaria	1	1.464	85.000	112.990	0.226	Malaria	1	1.464	85	112.990	0.226

	Genotype	2	2.677	83.000	110.313	0.262	Zygosity	1	0.001	84	112.990	0.976
	Malaria:Genotype	2	6.341	81.000	103.972	0.042	Malaria:Zygosity	1	3.164	83	109.826	0.075
6308_135	NULL			88.000	116.262		NULL			88	116.262	
HBAA	Malaria	1	0.931	87.000	115.331	0.335	Malaria	1	0.931	87	115.331	0.335
	Genotype	1	1.912	86.000	113.419	0.167	Zygosity	1	1.912	86	113.419	0.167
	Malaria:Genotype	1	0.328	85.000	113.090	0.567	Malaria:Zygosity	1	0.328	85	113.090	0.567
6311_93	NULL			89.000	118.288		NULL			89	118.288	
HBAA	Malaria	1	0.834	88.000	117.454	0.361	Malaria	1	0.834	88	117.454	0.361
	Genotype	2	3.772	86.000	113.682	0.152	Zygosity	1	0.124	87	117.330	0.725
	Malaria:Genotype	2	8.220	84.000	105.462	0.016	Malaria:Zygosity	1	4.903	86	112.427	0.027

Figure legends

Fig 1. Correlation between mating success (number of females mated with) and gene expression levels, for all genes studied in peripheral blood from 14 great snipe males, in relation to total gene expression levels. Points above the grey dotted line represent transcripts with a positive correlation. Points in red are ESTs with a significant correlation. The gene highlighted in red (EFCAB14) is the only one that remained significant after the false discovery rate correction. The three most highly expressed genes, all coding for haemoglobin subunits are highlighted in black.

Fig 2. Correlation between mating success and individual gene expression levels plotted for the transcript with a robust significant relation (EFCAB14: r_s=0.89, n=14, P<0.0001), as well as for the three highly expressed haemoglobin genes (HBAA: r_s=0.38, n=14, P=0.185; HBG1: r_s=0.07, n=14, P=0.810; HBAD: r_s=0.53, n=14, P=0.049). Larger points denotes overlapping observations. Statistically significant relations are marked with a dashed regression line.

Fig 3. Effect of SNP Genotype on mating success plotted against differential gene expression in relation to mating success of males sampled in 2010. Red dots indicate the top candidate markers listed in Table 1. Boxplots show mating success for males with the different Genotypes of four of the candidate genes (gene symbols as in Table 1) in dark red inserts. Diamonds in the insert graphs shows predicted mating success from the GLM. Solid lines are connecting each insert plot with the SNP in question, and dashed lines with another SNP from the same gene.

Fig 4. a) Outlier analyses showing two loci (the non-candidate Anonymous SNP 1487_1016 (left) and the candidate 1276_857- in the RHD gene (right)) being differentially selected among sites. Candidate SNPs in yellow and non-candidates in black. The sites are: Ånnsjön, Bruksvallarna, Gåvålia, Hemavan, Nord-Trondelag, Røros, and Valdres in Norway/Sweden and Eastern Estonia, Western Estonia and Poland. b) PCA-plot (PC1 vs. PC2) showing SNP differentiation among regions (East vs. West) for non-candidate SNPs. c) PCA-plot plot (PC1 vs. PC2) showing absence of SNP differentiation among regions for the seven candidate SNPs.

Fig 5. a) Outlier analyses showing one locus (the Anonymous SNP 1487_1016) being differentially selected among two time periods. Candidate SNPs in yellow and non-candidates in black. For these analyses 1994–2002 was divided into two periods (early and late). Dotted lines refer to 95% confidence limits for neutral loci (upper and lower bound), the line in the middle is the median expectation. b) PCA-plot (PC1 vs. PC2) showing absence of SNP differentiation among the two time periods for non-candidate and c) candidate SNPs.

Figure 1

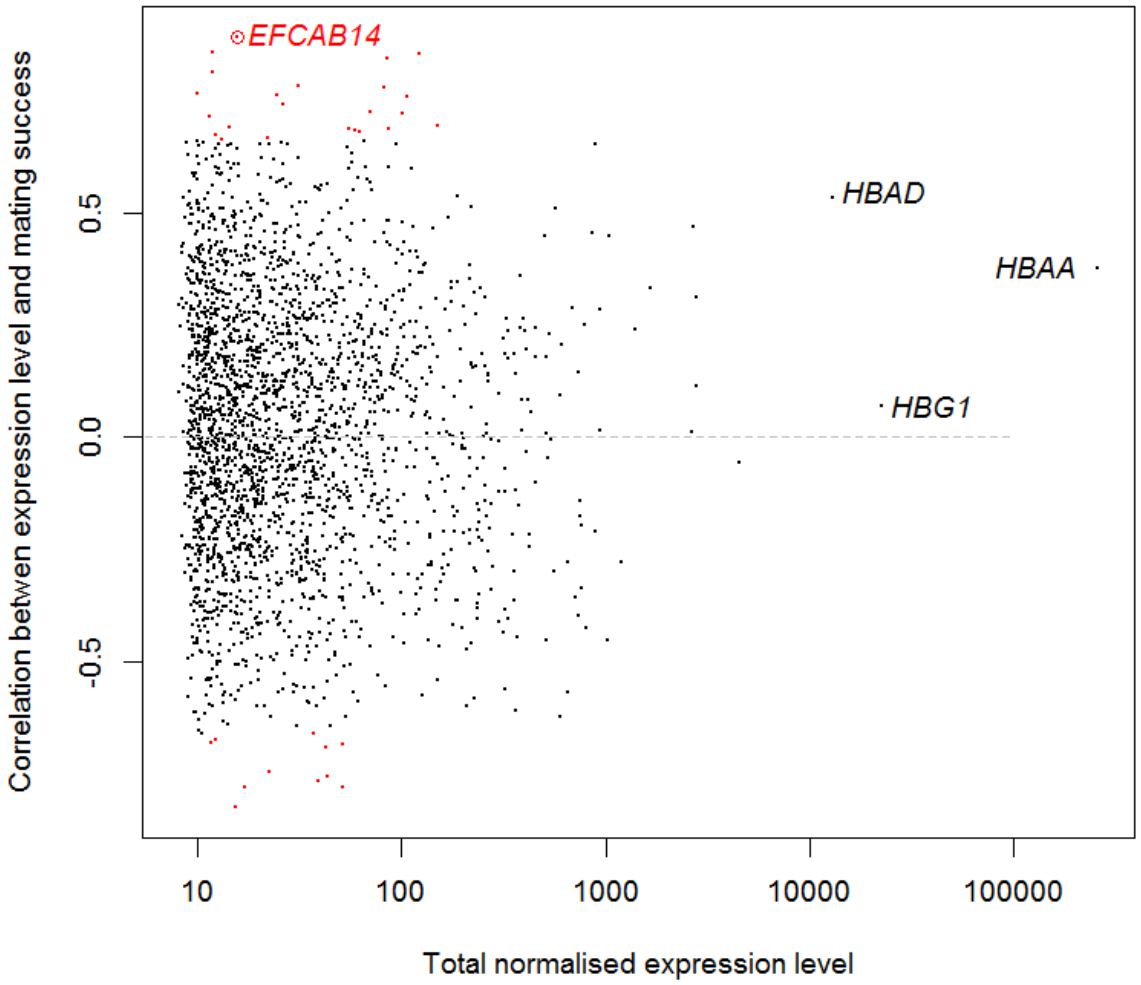

EFCAB14
HBAD
HBAA
HBG1
Correlation between expression level and mating success
0.5
0.0
-0.5
10
100
1000
10000
100000
Total normalised expression level

Figure 2

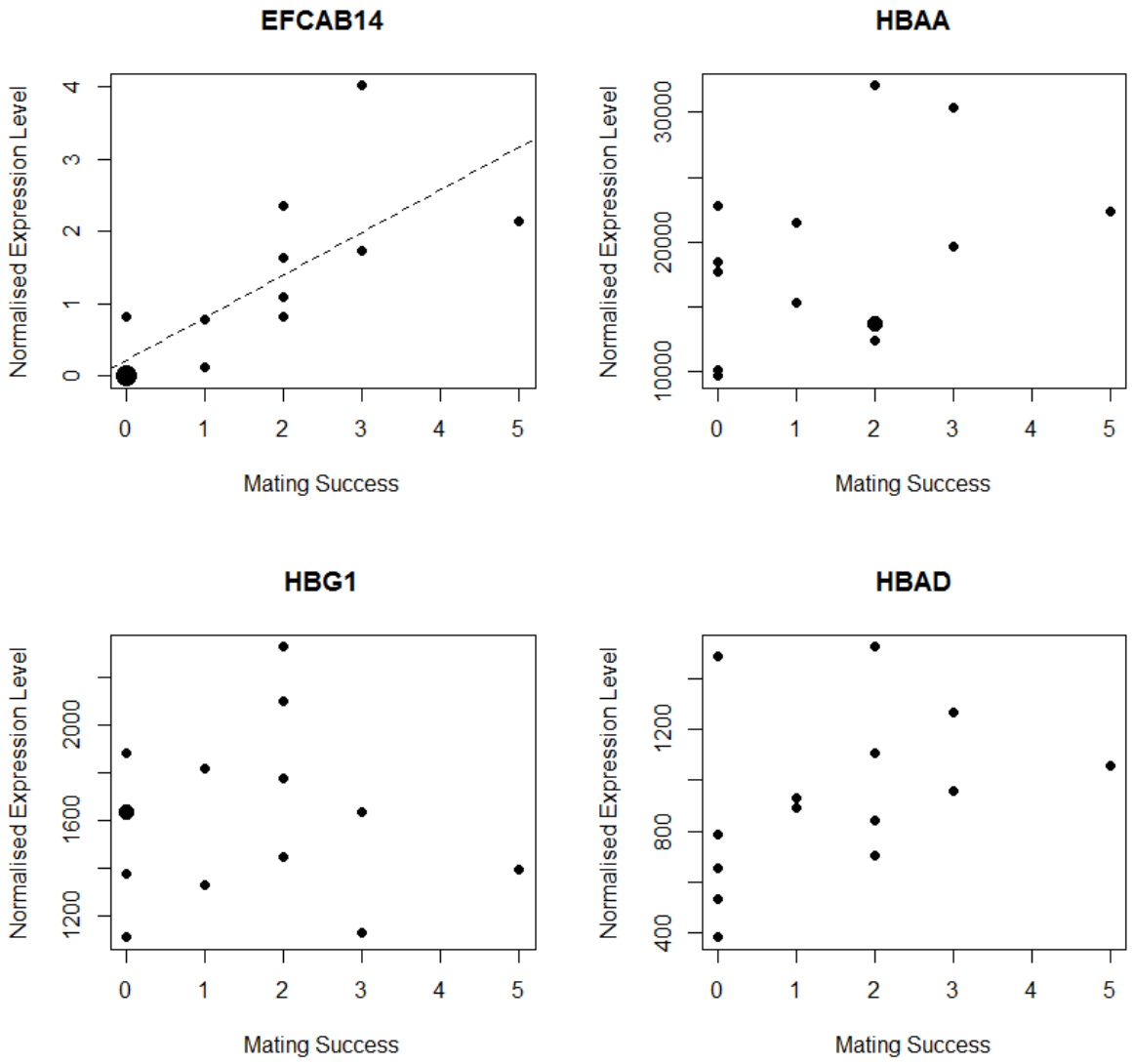

EFCAB14
Normalised Expression Level
Mating Success
HBAA
Normalised Expression Level
Mating Success
HBG1
Normalised Expression Level
Mating Success
HBAD
Normalised Expression Level
Mating Success

Figure 3

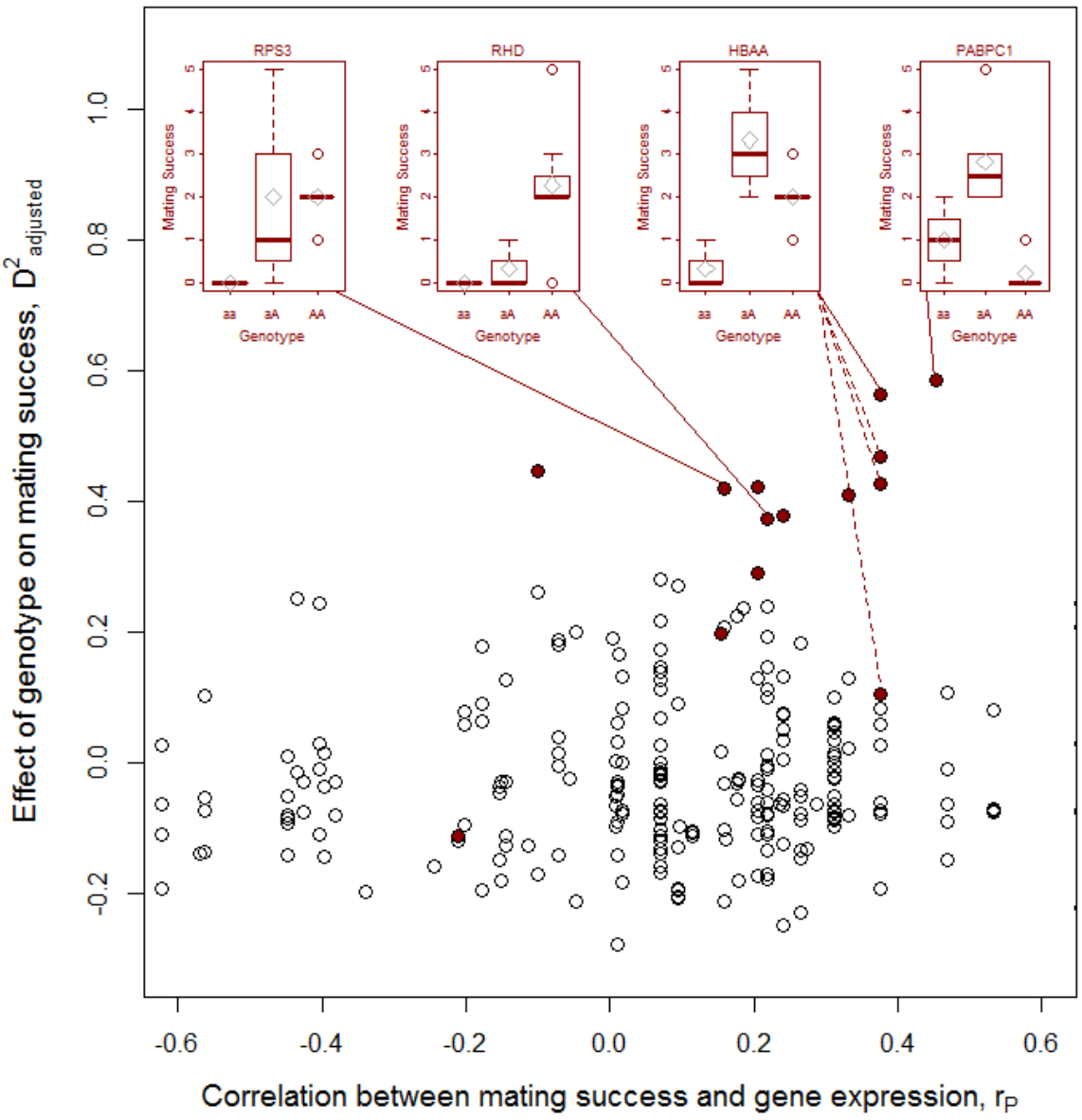
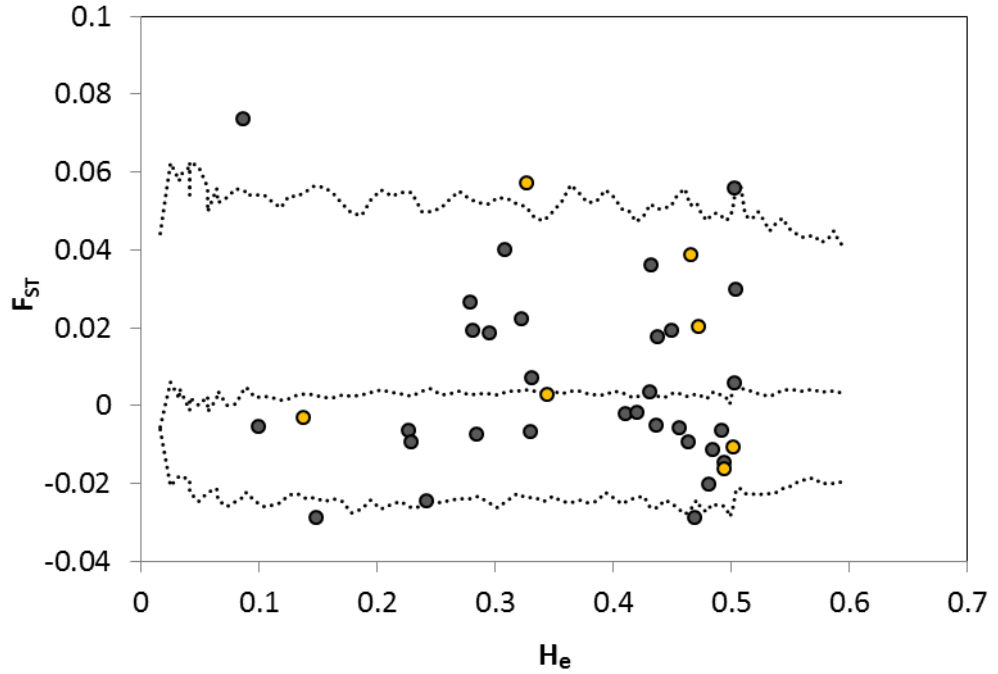
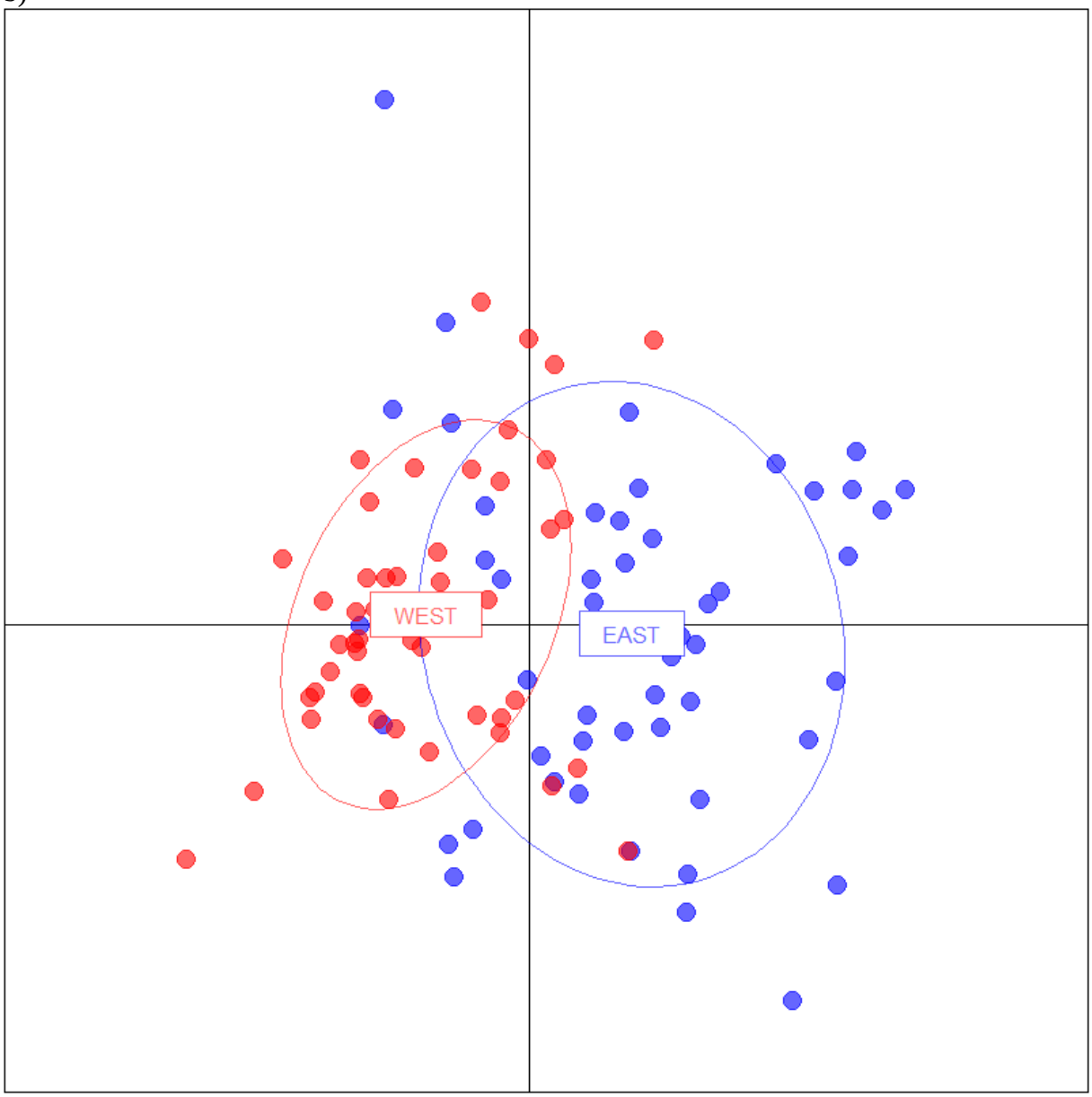

RPS3
RHD
HBAA
PABPC1
Mating Success
Genotype
aa
aA
AA
Effect of genotype on mating success, $D^2_{adjusted}$
Correlation between mating success and gene expression, r_P
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6

Figure 4
a)



a

b)



c)

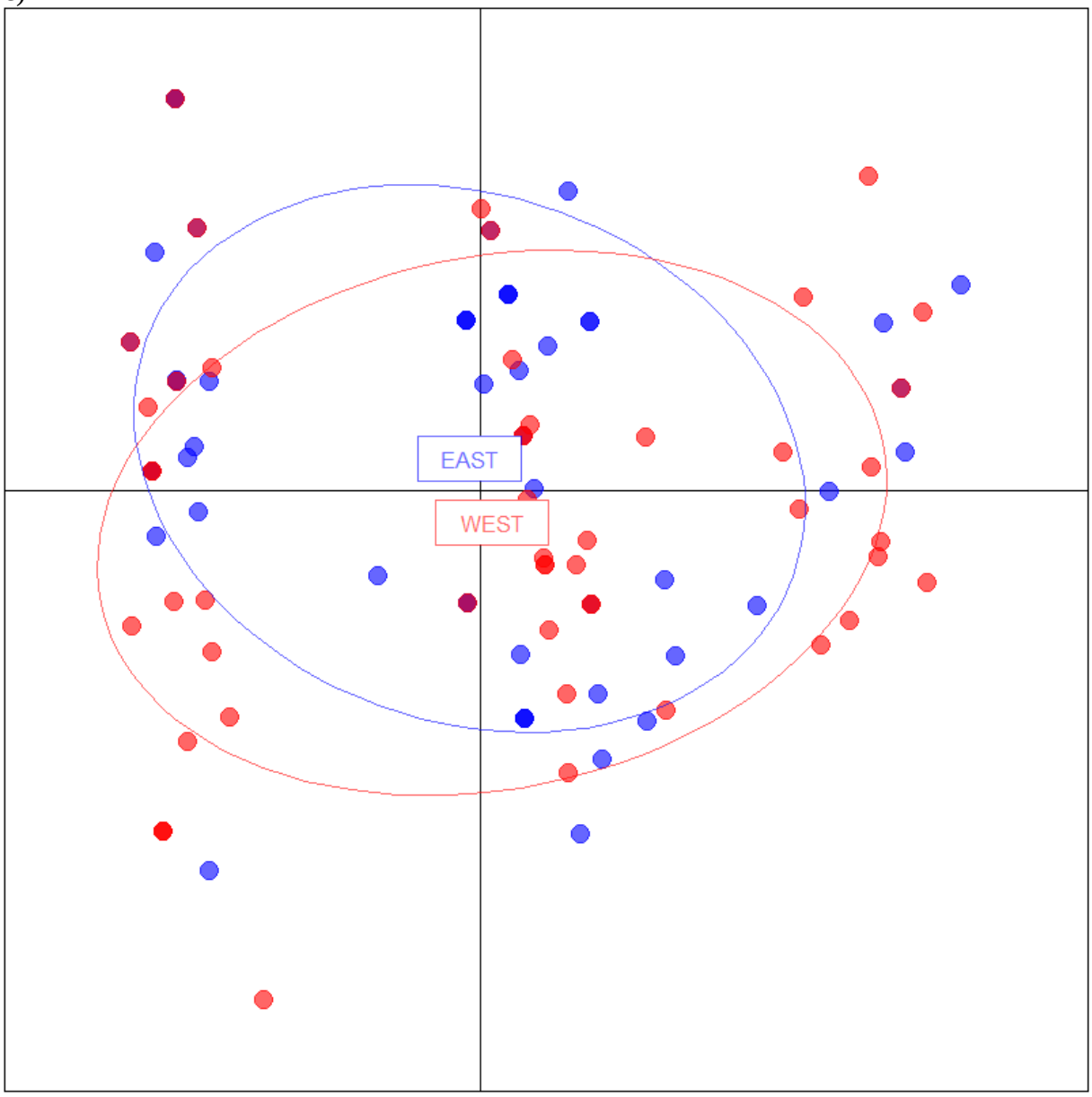
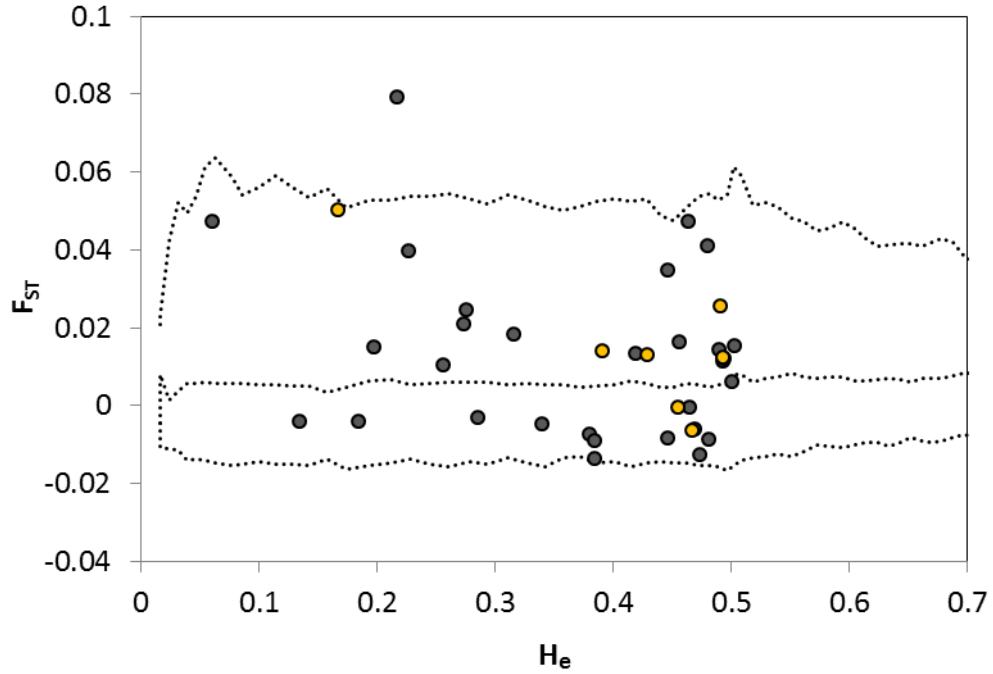
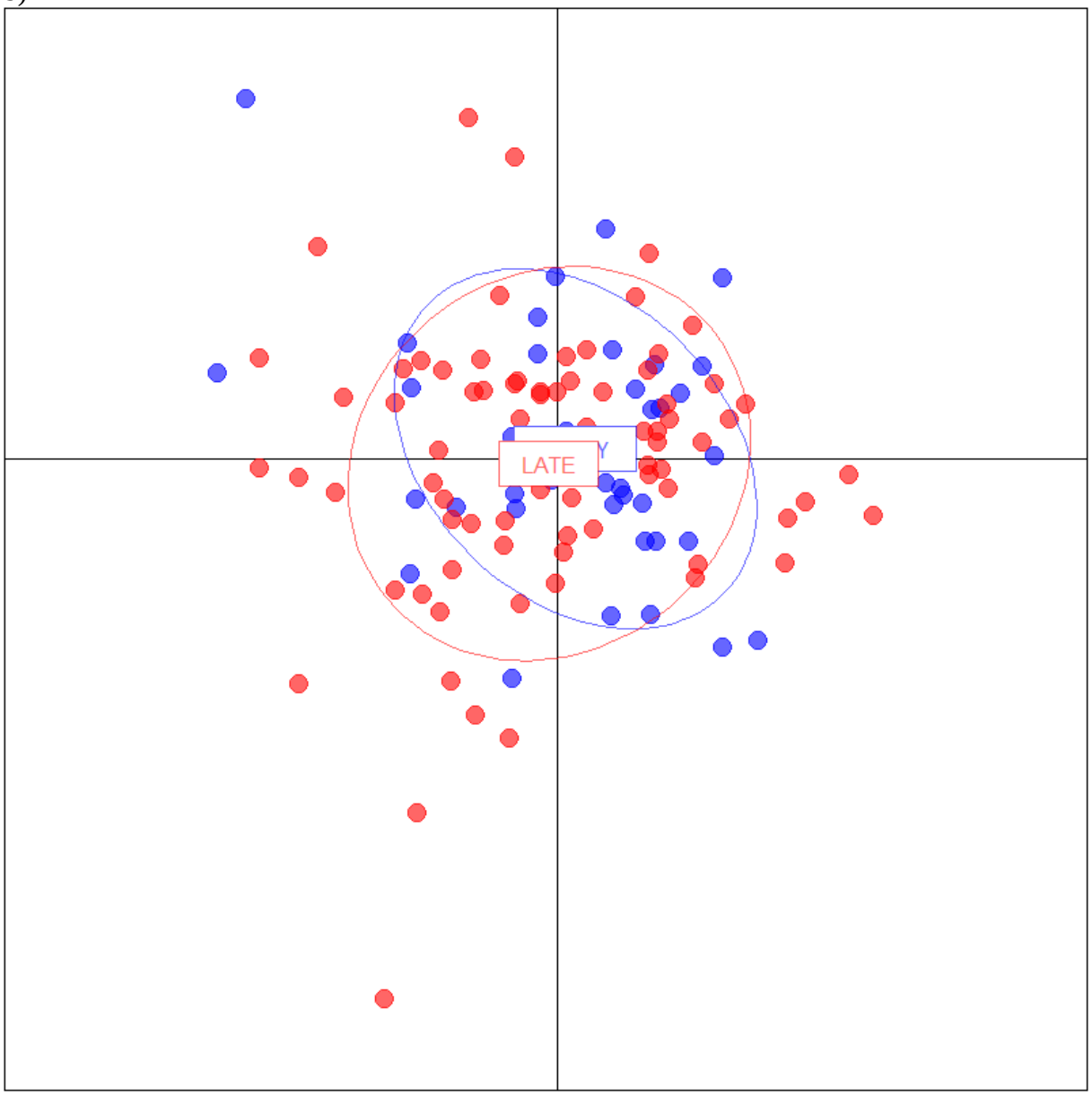


Figure 5
a)



b)



c)