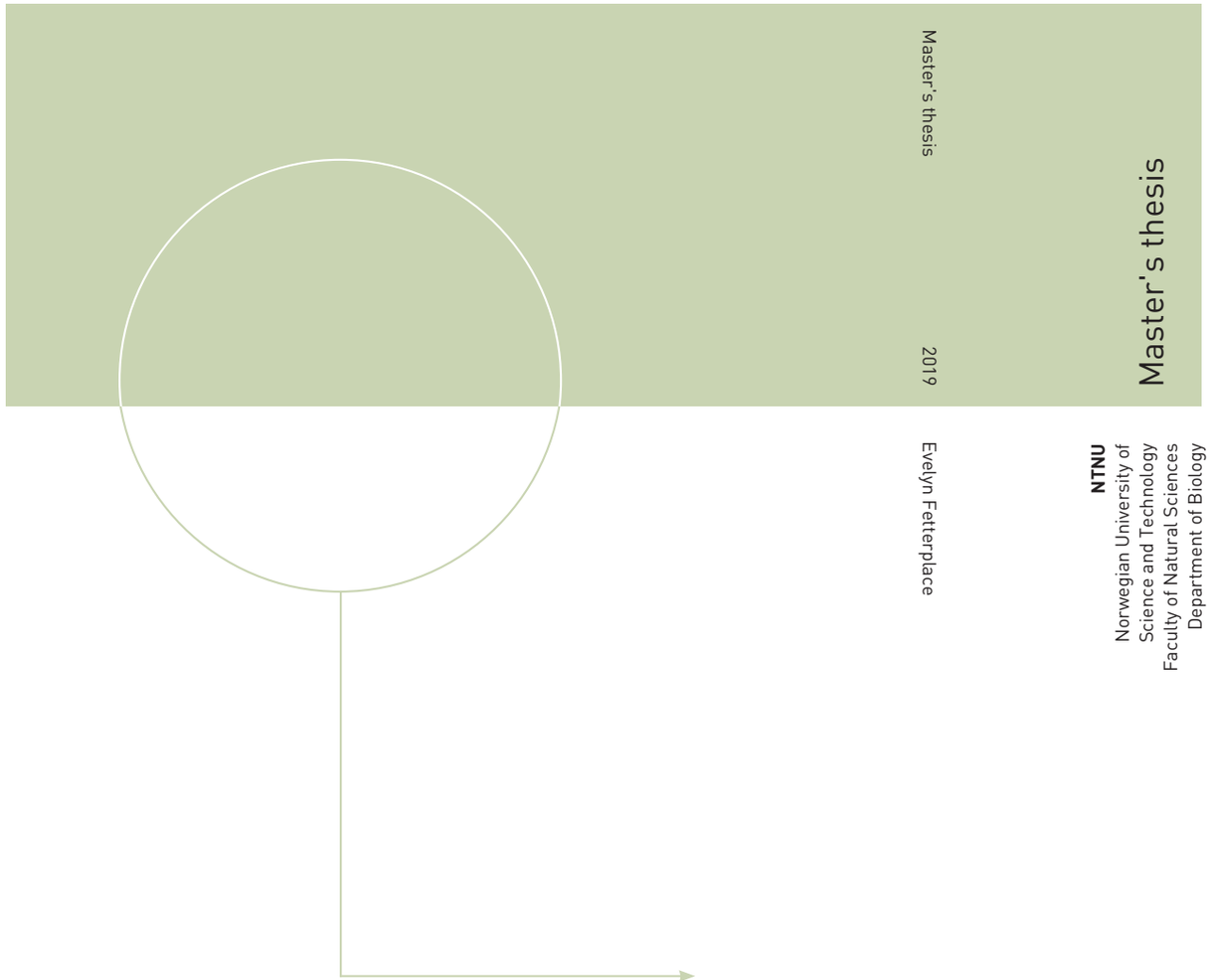

Evelyn Fetterplace

The Genetic Basis for Dispersal in a House Sparrow Metapopulation

May 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU

Norwegian University of
Science and Technology

The Genetic Basis for Dispersal in a House Sparrow Metapopulation

Evelyn Fetterplace

Submission date: May 2019

Supervisor: Professor Henrik Jensen

Co-supervisor: Dr. Alina Niskanen
Sarah Lundregan
Dilan Saatoglu

Norwegian University of Science and Technology
Department of Biology

Foreword

I would like to thanks my supervisor, Professor Henrik Jensen, and co-supervisors, Dr. Alina Niskanen, Sarah Lundregan and Dilan Saatoglu for guiding and teaching me throughout the course of my study. Thank you for your constant support and encouragement.

I am also grateful to Dr. Bernt Rønning for his coordination of the house sparrow fieldwork and to all other participants who made field season an enjoyable experience.

The work presented here is built upon the efforts of those involved in the house sparrow study system over the past three decades. Completion of this thesis would not have been possible without their dedication to collection of such high-quality data. I would also like to thank the Helgeland residents, whose ongoing cooperation has contributed to the success of the study system. I am grateful for the opportunity that this thesis has given me to participate in the ongoing work of the House Sparrow Project at CBD, NTNU.

Evelyn Fetterplace
Trondheim, May 2019

Abstract

Dispersal is an ecologically important trait that shows phenotypic variation in many populations, and which may play an increasingly important role in population dynamics as the climate changes. Dispersal and traits associated with dispersal, have been shown to be highly heritable in many species, including some bird species. For these reasons, dispersal is a suitable and interesting trait for investigating genetic architecture.

Knowledge of the underlying genetic architecture is required to understand the mechanisms driving phenotypic change in dispersal and to interpret their involvement in eco-evolutionary cycles. Previous studies have revealed several genes which may influence dispersal, but studies on dispersal as a phenotypic trait in itself, rather than traits associated with dispersal, are few, and the number of causal loci and their locations in the genome is largely unknown.

In this study, phenotypic, pedigree and genome wide SNP data, from an insular house sparrow metapopulation off the coast of northern Norway, was used to explore the genetic basis of dispersal. Rather than investigating specific traits that may be associated with dispersal in the house sparrow, the dispersal phenotype was defined as an individual that left their natal island, and successfully established on a new island. Heritability for dispersal was estimated using animal models in MCMCglmm, and dispersal was found to be a highly heritable trait. Genome partitioning analyses did not find a significant, positive relationship between chromosome size and proportion of variance explained. In addition, chromosome 12 explained a disproportionate amount of the variance in dispersal. GWA analysis was used to search for causal loci and revealed one locus of significant effect. No genes that have been associated with dispersal were found near significant or suggestive loci.

This work illustrates both the difficulties and advantages of performing association studies in natural populations and investigates the genetic architecture of dispersal in house sparrows.

Table of Contents

1. Introduction

The genetic architecture of ecologically important traits

The genetic architecture of a trait generally describes the genetic effects that make up and influence a phenotypic trait (Hansen 2006). This includes the number of genes affecting a trait and the magnitude of their effect, but also the form of variance in genes and the way this variation is passed on i.e. through additive, dominance or epistatic effects (Hansen 2006, Johnston et al. 2013). Understanding the link between genotype and phenotype is a key goal of evolutionary genetics (Johnston et al. 2013, Charmantier et al. 2014). The rise of genomics has given us a much greater ability to understand the genetic architecture behind phenotypic variation in important ecological traits (Rodrigues-Verdugo et al. 2017), and the relative importance of genes in the variation of traits, within the mix of genetic and environmental causes (Charmantier et al. 2014).

Adaptive evolutionary change in a population requires there to be heritable genetic variation in the underlying traits, and selection on those traits (Falconer and Mackay 1996, Lynch and Walsh 1998, Ronce 2007, Danchin et al. 2011). The speed of evolution also depends on the form of the genetic basis for the trait, and possible covariance with other traits. Hence, examining the genetic architecture is essential for understanding the mechanisms for variation in the phenotype and for predicting the potential for trait evolution (Orr 2005, Richards et al. 2010, Mackay et al. 2012). As a supplement to studies into life history parameters, ecological dynamics, and selection pressures, information on the genetic architecture also gives insight into the eco-evolutionary dynamics of a trait (Clutton-Brock and Sheldon 2010, Johnston et al. 2013, Hendry et al. 2017). An understanding of the genetic architecture of a trait does not, of course, provide a conclusion on the importance of a trait for the fitness of individuals or the potential for adaptation, this requires further experimentation (Barrett and Hoekstra 2011). However, additive genetic variance is a crucial element, in that it provides variation on which selection might act, so this element of the genetic architecture of a trait is a critical piece of the puzzle.

In the past decade, quantitative genetic studies on populations in their natural environment have increased considerably. This is due partly to the inclination of evolutionary biologists to answer their questions in a realistic setting and made possible by advances in molecular genetics and statistical methods, as well as better availability of suitable data sets (Charmantier et al. 2014).

Although many key questions in the field of quantitative genetics can be investigated without the use of natural populations, some questions are more precisely answered through the study of populations in their natural environment (Charmantier et al. 2014). These include questions about the effect of climate change on evolutionary dynamics, and predictions of evolutionary responses to selection pressures. Placing a study within the relevant environmental conditions allows genetic variation to be assessed relative to these other causes of variation, something that has increasingly been seen as important as we understand more about how environmental conditions can affect both the selection processes, and the expression of genetic variance (Charmantier and Garant 2005).

Dispersal

Dispersal is a complex trait made up of, and influenced by, many different factors, such as the propensity to disperse, distance and direction of movement ability, and settlement choice. In addition, many physiological, morphological and behavioural traits have been found to determine the possibility for dispersal (Clobert et al. 2012, Saastamoinen et al. 2018). Dispersal is often very sensitive to environmental cues and consequently, ecological factors like competition, population structure and resource abundance play an important role in creating the conditions for dispersal (Bowler and Benton 2005, Matthysen 2005, Pärn et al. 2012). Dispersal is an ecologically important trait in part because it has significant influence on the genetic variation in populations, because it is a crucial mechanism for gene flow and allele frequency changes within and between populations (Johnson and Gaines 1990, Tufto et al. 2005). This can be particularly important in small populations, because one or a few individuals dispersing in can change the gene pool more dramatically (Lenormand 2002). It also contributes to the distribution of genetic diversity, by increasing the proportion of total diversity that is contained within, rather than between, populations. This distribution of diversity is particularly central to the dynamics and evolution of spatially structured populations. Species can maintain genetic cohesion across space and maintain global persistence even in the presence of local extinction (Ronce 2007, Kahilainen et al. 2018).

Due to the effects of environmental change and increasing fragmentation of habitats, dispersal can be considered increasingly crucial, particularly for the survival of small populations (Hanski & Gaggiotti 2004, Legrand et al. 2017, Ronce 2007, Travis et al. 2010). Therefore, the ability of populations to disperse is a critical aspect to consider for the future of populations in the face of

changing environmental condtitions that may lead to population decrease and reduced genetic diversity. If populations can adapt towards greater dispersal capacity, this might be a way for species to thrive in a rapidly changing world. Whether dispersal will increase or decrease in response to habitat fragmentation will depend on factors such as the proportion of empty habitats to colonize, the degree of environmental heterogeneity, and importantly, the amount of genetic variation in dispersal traits (Cheptou et al. 2017, Cote et al. 2017). Additionally, the speed of evolutionary change involves a combination of standing genetic variation, new mutations and genetic covariances among traits (Reznick and Ghalambor 2001, Etterson 2004, Becks et al. 2010, Hendry 2013). Once we understand these parameters, we can then begin to determine whether dispersal as an adaptive mechanism will be sufficient for the evolutionary rescue of populations in fragmented habitats. Understanding dispersal and its evolution is therefore crucial to improve the management of natural populations (Ronce 2007, Dyck and Baguette 2005).

Many studies have looked at varying dispersal related traits, such as flight metabolism and wing length, particularly in birds and insects (Niitepõld et a. 2009, Saastamoinen et al. 2018). In the Pea aphid *(Acyrthosiphon pisum)* it was found that male wing polymorphisms are determined by a single locus on the X chromosome (Roff 1986, Caillaud et al. 2002,) Whereas in *Drosophila melanogastar* studies of variation affecting locomotion behaviour have given a list of candidate genes for dispersal, which could be potential regulators of dispersal in other organisms as well (Jordan et al. 2007, 2012). The foraging gene was one of the first *Drosophila* genes shown to influence locomotion behaviour, causing adults with the dominant 'rover' allele (for^R) to have a higher dispersal distance (Edelsparre et al. 2014). The Pgi gene in the Glanville fritillary butterfly (*Melitaea cinxia*) has also shown allelic variation associated with dispersal rate. Individuals with a specific allele of this gene, that is responsible for a metabolic enzyme associated with cellular energetics (Mattila & Hanski, 2014), have a higher flight metabolic rate and a higher dispersal propensity in the field (Haag et al. 2005; Niitepõld et al. 2009, 2011).

The genetic architecture of dispersal traits needs far more investigation for us to understand the eco-evolutionary dynamics of dispersal (Rodriquez-Verdugo et al. 2017). In many bird species, heritability estimates greater than zero have been found for propensity to leave the natal site (Saastamoinen et al. 2018), with some very high estimates such as 0.95 for western bluebirds, Sialia mexicana, based on animal model analysis (Duckworth and Kruuk 2009). However, estimates for traits involved in the entire process of dispersal, including departure, transfer and settlement are rare (Saastamoinen et al. 2018), and this study is unique in that it takes dispersal

as a trait in itself and investigates whether the propensity of house sparrows to leave their natal island and reach a new island has a genetic basis.

Estimating heritability

One of the major methods of understanding the evolutionary potential of a trait, is to estimate the heritability of the trait in a population. Additive genetic variance is the component of genetic variance that is independent of interactions with other genes, and with the environment, and is considered the major basis of evolutionary responses to selection (Lande 1979). Additive genetic variance can be used to yield an estimate of heritability; called narrow sense heritability (Wilson et al. 2010). Heritability is measured as the proportion of the phenotypic variation (V_p) in a trait that is due to additive variance in genetic factors (V_a) (i.e. V_a/V_p) (Visscher et al. 2008, Wilson 2008), and is one way of quantifying the relative importance of V_a on phenotypic variance.

Models that are used to estimate additive genetic variance are based on the assumption in quantitative genetics that complex traits are controlled by many genes spread over the genome (the infinitesimal model). In this case individuals that are closely related, and therefore share more genes, will be more phenotypically similar as well (Wilson et al. 2010, Charmantier et al. 2014). So, using population pedigrees and phenotypic covariance between individuals, researchers can estimate the additive genetic variance, or the relative importance of genes for the phenotypic variation in a population (de Villemereuil et al. 2013, Charmantier et al. 2014). Heritability was traditionally measured using parent-offspring regressions, and full-sib or half-sib designs, but these can be difficult to execute in natural populations due to the required experimental design and can be sensitive to dominance and environmental effects, or common environments between parents and offspring (de Villemereuil 2012, de Villemereuil et al. 2013).

In recent times the use of mixed models, particularly the animal model, has increased for estimating variance components in wild populations (Lynch and Walsh 1998, de Villemereuil et al. 2013). Animal models are generalised linear mixed models (GLMM) that incorporate information from detailed pedigrees in a way that simpler techniques cannot, considering the covariance between as many pairs as possible rather than only sibling relationships or parent-offspring relationships (Charmantier et al. 2014, Wilson et al. 2010). The animal model is in fact particularly suited to wild populations in many ways, because it can deal with the complexity that we expect to see in nature (Wilson et al. 2010), including unbalanced designs, missing trait data

and pedigree links, and varying environmental conditions (Lynch and Walsh 1998, Charmantier et al. 2014, Wilson et al. 2010, de Villemereuil et al. 2013). It also allows better exploration of non-genetic influences on phenotype such as the effects of sex or age (Wilson et al. 2010). The use of the animal model has created lower estimates of heritability in general, probably partly due to removing some environmental sources of covariance between individuals (Charmantier et al. 2014). Animal models have also been found to yield more accurate heritability estimates than parent-offspring regressions and sibling designs, particularly with small data sets or for binary traits (de Villemereuil et al. 2013). The animal model is also effective for incorporating repeated measures of individuals over their lifetime and including random effects (Kruuk 2004). This allows the utilisation of all available measurements, and it also allows the quantification of permanent between-individual differences apart from those due to additive genetic variance (Kruuk 2004). Many common statistical tools in quantitative genetics assume Gaussian distribution of traits and residuals, and hence can have problems with phenotypic traits that do not follow these assumptions of normality, such as binary traits like dispersal (de Villemereuil 2018). However, GLMMs include the core assumptions of the infinitesimal model but can also handle a wider range of characteristics of the phenotypic trait of interest (Bolker et al. 2009, de Villemereuil 2018).

Recent advances, especially in genotyping on SNP panels, have also allowed more widespread use of genome wide scanning methods of investigating genetic architecture, such as genome partitioning, and genome-wide association studies (GWAS) (Stapley et al. 2010, Ellegren 2014), in which the underlying causal loci of heritable traits can be mapped. In genome partitioning, the phenotypic variance explained by each chromosome is regressed on the size of the chromosome (Kemppainen and Husby 2018b). Studies showing that large chromosomes explain more variation in traits than small chromosomes (Yang et al. 2011) are one way in which the theory of quantitative genetics have been supported, because they indicate that many loci of small effect distributed across the genome influence a quantitative trait (Kemppainen and Husby 2018b). This is because larger chromosomes generally contain more genes, so they are expected to explain a larger proportion of the variation in the trait. Divergence from this pattern can indicate the presence of genes or loci of large effect (Kemppainen and Husby 2018b, Robinson et al. 2013). Genome wide association (GWA) analyses use genotype information on a dense set of genetic markers across a genome, to capture a significant proportion of the variation. Using data on thousands of single nucleotide polymorphism (SNP) genotypes, and phenotype variation data on the genotyped individuals, multiple association analyses are

undertaken to detect associations between genetic variants and trait variations (McCarthy et al. 2008). This allows researchers to better understand the nature of the genetic variation in a trait and find candidate genes for further investigation.

Evidence suggests that there is genetic variation for dispersal in natural populations and, based on evidence from other complex life history traits, it seems most likely that this variation would be caused by many genes of small effect (Tiffin and Ross-Ibarra 2014, Saastamoinen et al. 2018). Genome wide association studies are necessary to confirm this (Saastamoinen et al. 2018). Knowing the number and location of genes that contribute to dispersal variation helps us to understand more about the evolution of traits in natural populations. Traits that are controlled by only one gene or are oligogenic may provide greater evolutionary potential if selection acts on them and can be more likely to become fixed in a population. Polygenic traits are far more complex, because they may involve genes that influence several different fitness traits, and they can be more influenced by environment. This can allow more variants to be maintained in a population and complicate the evolutionary trajectory (Remington 2015). Conversely polygenicity could also make adaptation to unpredictable environments easier if there are many genetic possibilities by which a specific beneficial change in phenotype could be accomplished (Remington 2015). GWA approaches have been successfully utilised in many recent studies on a wide range of traits in natural populations (Johnston et al. 2011, Johnston et al. 2013, Husby et al. 2015, Santure et al. 2015, Barson et al. 2015). For example, Husby et al. 2015 found evidence for association between a genome region and phenotypic variance in clutch size in the collared flycatcher (*Ficedula albicollis*).

The house sparrow study system

The house sparrow (*Passer domesticus)* is an extensively studied model species, and the natural metapopulation off the Helgeland coast in Northern Norway is a uniquely suited population through which to study evolutionary changes and the genetic basis for phenotypic traits such as dispersal (Figure 1). House sparrow metapopulation consists of 18 subpopulations which cover an area more than $16km^2$ (Pärn et al. 2012, Baalsrud et al. 2014) and have been monitored since 1993. This study system has provided a large data set, including genetic, morphological and life history data (Jensen et al. 2003, Jensen et al. 2004, Pärn et al. 2009). One of the major benefits of such a long-term study is the thorough understanding of the ecology of the house sparrow metapopulation. A large data set is important for detecting associations

between genetic variants and phenotypic covariance, but it is also essential to understand the potential environmental effects, mating systems and behavioural ecology in order to anticipate the factors that may be important in quantitative genetic models (Charmantier et al. 2014).

House sparrows are sexually dimorphic, passerine birds, with a lifespan of up to 9 years in Northern Norway (Jensen et al. 2004). The clear population boundaries, and the high resighting rate of individuals (mean of 74%) allows high accuracy in estimating population sizes, individual survival, and inter-island dispersal (Tufto et al. 2005, Pärn et al. 2012, Holand et al. 2016). House sparrows are often associated with human settlements (Anderson 2006) and are predominantly found at agricultural or residential sites in the study system. This improves sampling efficiency and allows individuals to be monitored from hatching over consecutive years until they die (Jensen et al. 2004, Jensen et al. 2008, Pärn et al. 2009, Billing et al. 2012). The area of the site used in this study is also greater than the dispersal distance of most individuals (Pärn et al. 2009). Establishing the heritability of dispersal in natural populations can be very difficult, because individuals can often leave the study area without being detected. It can also be difficult to detect all individuals in the population in general. However, the large size of the study system, the ease of access to individuals, and the ability to track birds over many years, means that almost all dispersers can be detected, and we can obtain accurate information about kinship relationships within the populations.

Dispersal rate in the house sparrow metapopulation has been estimated at 22.5% (Saatoglu et al. 2019 *in prep*). Dispersers of both sexes in this system have been shown to have higher survival rates than residents (Altwegg et al. 2000), and dispersal probability was related to some environmental variables. Inbreeding is also quite high in the system, and shows negative fitness effects (Jensen et al. 2007, Niskanen et al. 2019 *submitted*). These details indicate that possible reasons for dispersal in the population include avoidance of environmental disadvantages, and the costs of inbreeding (Pärn et al. 2009). Dispersal rates among the house sparrows in this system also differ between islands and years (Pärn et al. 2012), so there are important environmental factors to consider that influence the phenotypic variation between individuals. In addition, differences in dispersal patterns between the sexes have been found in the Helgeland house sparrow population (Altwegg et al. 2000, Skjelseth et al. 2007), suggesting that to understand the evolution of dispersal rates, we must consider differences in selection pressures between the sexes.

Overview of the study

Here I present the investigation into the heritability and genetic architecture of dispersal in the Helgeland metapopulation of house sparrows, using a custom 200K Affymetrix Axiom SNP array. First, an animal model containing the metapopulation pedigree was used to estimate V_a and h^2 of dispersal. This was followed by genome partitioning to examine the phenotypic variance explained by each chromosome. Finally, a GWA analysis of the whole dataset was undertaken to determine whether the genome-wide approach would detect regions and candidate genes associated with dispersal for future functional studies.

2. Methods

2.1 Data collection

Data collection was held each year in the summer (breeding season), from May to August (Husby et al. 2006) and in autumn between September and October. During fieldwork, when chicks reached 8-13 days old, they were taken from their nest and ringed with one unique identity number and three colour combination bands. Un-ringed fledglings and adults were captured with mist-nets and ringed in the same way as the chicks (Ringsby et al. 2002). Natal island for the chicks is recorded at this time. Each individual's hatch year is considered as either; the first year recorded for nestlings or fledged juveniles captured in summer or autumn, or the year prior to first year recorded for birds first ringed as adults (Jensen et al. 2008). The identity numbers and colour combination bands allow dispersal and survival of birds over the breeding season to be recorded ecologically, based on recapture or resighting in subsequent years. A range of phenotypic traits are also measured for all captured birds, such as bill morphology, wing length and body mass, and small blood samples (ca. 25 µl) are collected from under the wing (Ringsby et al. 2002, Jensen et al. 2004, 2008; Pärn et al. 2009, 2012, Holand et al. 2016, Kvalnes et al. 2018). My participation in sampling to date was in May 2017.

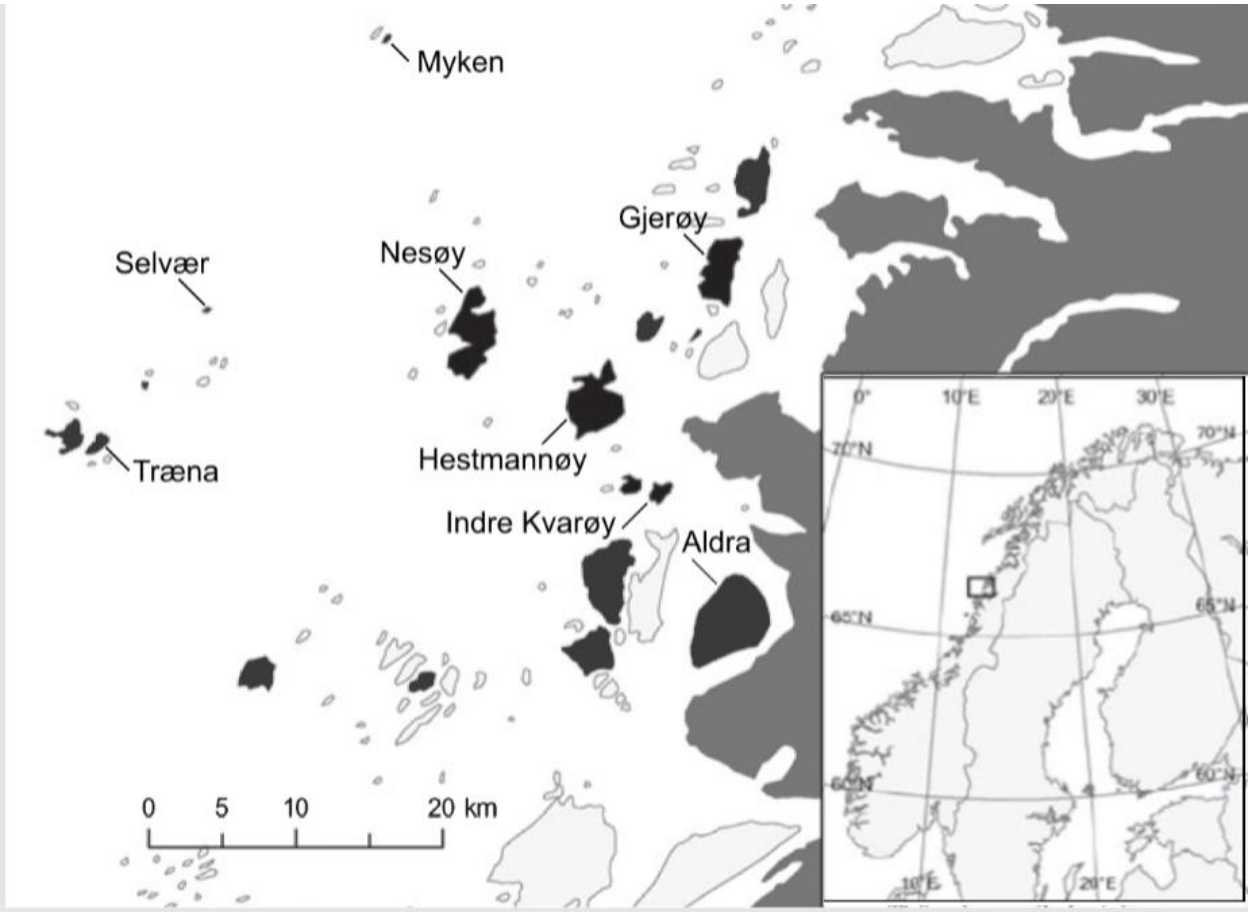

Figure 1. Islands included in the house sparrow metapopulation study system, Northern Norway (66°N, 13°E). Islands shaded black have been continuously followed since monitoring began. The eight populations used in this study are labeled with the island name. (Image adapted from Lundregan et al. 2018).

In this study I used dispersal data that includes only adult birds recorded on 8 of the study islands (Aldra, Gjerøy, Hestmannøy, Indre Kvarøy, Myken, Nesøy, Selvær and Træna), as these are the 8 islands that have been SNP-genotyped (Table 1, see chapter 2.2). Information about dispersal was available for 2745 SNP-genotyped individuals. Dispersal data was obtained by combining accurate ecological natal island information (where the information was obtained from nestlings and fledglings captured during the summer season) and also genetic assignment that detects natal island for those individuals that do not have this information (fledglings captured during the autumn season and/or un-ringed adults) (Kuismin et al. 2019 *submitted*, Saatoglu et al. 2019 *in prep.*).

Table 1: Period for which all recorded adult individuals on each island were genotyped on the 200K SNP microarray (Years), and number of individuals per island (Ni) after quality control for dispersal. Island refers to the birth (natal) island of the individuals.

Island	Years	Ni
Aldra (38)	1998 – 2013	179
Gjerøy (26)	1998 – 2013	524
Hestmannøy (27)	1998 – 2013	940
Indre Kvarøy (28)	1998 – 2013	331
Myken (22)	2004 – 2013	67
Nesøy (20)	1998 – 2013	126
Selvær (24)	2003 – 2013	275
Træna (23)	2003 – 2013	274
TOTAL	16	2716

2.2 Genotyping and Quality Control of the Dispersal Dataset

Blood samples from 3253 adult house sparrows were genotyped using a custom Affymetrix Axiom 200k SNP array (Lundregan et al. 2018). The array was developed based on the reference genome for house sparrow (Elgvin et al. 2017). Almost all house sparrows present on 8 islands during the years 1998/2003/2004-2013 were included. Approximately 90% of the adult population on each island were sampled each year, meaning that sample sizes are highly correlated with actual population sizes (Niskanen et al. 2019 *submitted*).

Successful genotypes were obtained for 3219 individuals from the Helgeland metapopulation. Quality control on this data set has previously been completed, to remove individuals with incorrect sex coding, too high identity by state (IBS>0.9), low call rate (<0.95%) and low minor allele frequency (<0.01), and the final data set consisted of 3116 house sparrow individuals (1580 females and 1536 males) (Niskanen et al. 2019 *submitted*).

Of these 3116 individuals, the natal islands of 2745 adults were revealed by using genetic assignment method and correction with a SNP-based pedigree, which gives more power in order to perform further analysis (Kuisman et al. 2019 *submitted*, Saatoglu et al. 2019 *in prep*).

A small group of birds (n = 29) from the 2745 data set were removed completely. These birds were born on a group of three islands (Træna, Selvær and Myken) denoted 'island 88' before 2004, but their adult island was also either Træna, Selvær or Myken, so determining whether they were dispersers was impossible. This created the final data set of 2716 birds (Table 1).

Some individuals (n=120) that were born outside the 8 SNP-genotyped islands (but dispersed into them as adults) were given a 'natal island' from inside the 8 islands i.e. their adult island was used as their natal island. Although it is possible some bias could be introduced in this way, the expectation is that this will not cause significant bias and it increases the power by including many extra dispersers.

Of the 200,000 SNP markers on the array the 184 804 markers ranked as PolyHigh resolution by CIGENE were used in further analysis, and quality control for individuals with dispersal was performed in the GenABEL R package (Aulchenko et al. 2007).

In total 183 088 markers and 2716 individuals (1325 female, 1391 male) passed the quality check.

2.3 Estimating Heritability of Dispersal

I estimated the genetic component of the phenotypic variation in dispersal using univariate animal models in the R package MCMCglmm, which fits generalised linear mixed models using Markov chain Monte Carlo techniques (Hadfield 2010). Using the R package *sequoia*, a metapopulation level pedigree has been constructed for the 3116 adult house sparrows in the study system, using a set of 605 highly informative SNPs (Niskanen et al. 2019). This pedigree was used in MCMCglmm for estimating variance components.

Bayesian methods, such as the Markov chain Monte Carlo (MCMC) method, are more easily adapted to non-Gaussian distributions, than frequentist methods such as restricted maximum likelihood (REML) or penalised quasi-likelihood (PQL), and are considered to provide more

accurate approximations of variance components, particularly for binary traits, as they mitigate downward bias of estimates (Hadfield 2010, de Villemereuil et al. 2013, de Villemereuil 2018). Bayesian estimates are known to have high sensitivity to prior distribution when sample size and variance are low (de Villemereuil et al. 2013). However highly complex models, for example those with many random effects, may only converge at all when using MCMC-based methods (de Villemereuil 2018).

The priors for the MCMCglmm model were specified by using a prior recommended for binary data (de Villemereuil 2012). The X^2 with 1 degree of freedom prior does put more weight in 0 than in 1, and for this reason may not be the perfect prior. However, the probabilistic weight of the X^2 with 1 degree of freedom prior is more spread between 0 and 1 than other priors, such as Inverse Gamma.

The effect of natal island and year of birth were estimated by adding these parameters to the model as random factors, because these are factors known to influence rates of dispersal among the Helgeland metapopulation of house sparrows *(*Pärn et al. 2012). To define a prior in MCMCglmm a list is used, where the R argument stands for the prior on the residual variance, and the G argument is for the random effect variance. When there are 3 random effects in a model (in this case, animal, island and year), 3 priors are required to be defined in G (de Villemereuil 2012). The prior was therefore adjusted to include 3 priors in the G argument list (Appendix III; Figure IV).

The posterior distributions of the estimates of the additive genetic (animal), year, island and residual (units) effects for dispersal were fitted with the default parameters (120,000 iterations, thinning interval of 100, and burn-in period of 10,000), and the plots showed some autocorrelation (see Appendix III; Figure I). Therefore, models were re-run with longer iterations (de Villemereuil 2012). The final model was run with 1e+06 iterations, a burn-in period of 10000, and a thinning interval of 100.

Heritability of dispersal was estimated using the posterior mode of each variance component (Table 2). The mode is used because the posterior distribution is not symmetrical. Heritability on the latent (model) scale $\left(h^2_{latent}\right)$ was estimated as an intra-class correlation coefficient, using the formula $h^2_{latent} = V_a/(V_a + V_{island} + V_{year} + V_r + \pi^2/3)$ and the parameter estimates are shown in Table 2. When random effects are added to the model they must be considered in the calculation

for total phenotypic variance, which is achieved by adding them to the denominator of the heritability formula (de Villemereuil 2012). The parameter $\pi^2/3$ is included in the formula denominator because the categorical model used has a logit link function, so we need to include the 'variance' of the link transformation in the total variance (Nakagawa and Schielzeth 2010). In addition, because we are using binary data (with 0 denoting non-dispersers and 1 denoting dispersers), the residual variance cannot be calculated and is set to an arbitrary value of 1 in MCMCglmm (Nakagawa and Schielzeth 2010). The heritability estimate was also made on the data scale, using the formula $h^2_{obs} = V_a / V_p$ because heritability on the expected data scale cannot be computed as an intra-class correlation coefficient in the same way that it can on the latent scale (Appendix I: Addendum III). The function "QGparams " from the package "QGglmm" was used to obtain variance component estimates on the data scale and to estimate h^2obs (de Villemereuil et al. 2016).

2.4 Genome Partitioning

The proportion of variance explained by each chromosome was estimated using a command line software tool, called GCTA (Yang et al. 2011). GCTA was first used to create relationship matrices (GRMs) for each chromosome. Average information Restricted Maximum Likelihood (AI REML) models with multiple GRMs fitted as random effects (GCTA option-mgrm) were used for chromosome partitioning (Yang et al. 2011). The -mgrm option in GCTA allows you to fit the GRM for each chromosome in a single model, and hence find the genetic variance due to each chromosome. Rather than testing the association of a particular SNP with a phenotypic trait, the GCTA analysis estimates the variance explained by all the SNPs on each chromosome (Yang et al. 2011).

Initially there were some problems with model non convergence, where more than half of the model components were constrained, which would cause the results to be unreliable. This was addressed by successively excluding the smallest chromosomes (Kemppainen and Husby 2018b). Chromosome length (Mbp) was taken from the house sparrow reference genome assembly, INSDC accession number MBAE00000000.1 (Elgvin et al. 2017). Five chromosomes (21 (5.71 Mb), 22 (3.67 Mb), 25 (0.48Mb), 27 (3.74 Mb) and 28 (3.53 Mb)) were excluded.

A corrected *p*-value was obtained through 'HC-resampling' to address the *p*-value inflation that can occur due to heteroscedasticity and censoring. If not accounted for, this can result in a

roughly 30% false positive rate for chromosome partitioning in bird genomes (Kemppainen and Husby 2018). Finally, the proportion of variance explained by each chromosome was plotted against chromosome size (Mbp).

2.5 GWAS

GWAS for dispersal was implemented using the R packages GenABEL (Aulchenko et al. 2007) and RepeatABEL (Rönnegård et al. 2016). The RepeatABEL package is a relatively new development that allows more effective use of GenABEL for natural populations.

GWA analysis is carried out by fitting a linear regression at every marker position on the genome (Rönnegård et al. 2016). RepeatABEL allows variables such as year and birth island to be included as random effects, along with the polygenic ID effects. Association analysis in RepeatABEL fits fixed SNP effects in a linear mixed model that can include random polygenic effects and permanent environmental effects, in order to correct for repeated measures and population structure (Rönnegård et al 2016). The model in this study included natal island and hatch year as random factors, in order to account for the fact that individuals may have common environmental effects as well as a common genetic background (Flint and Eskin 2012). The genomic relatedness matrix (GRM) also included as a polygenic random factor to account for relatedness between individuals.

GWAS in RepeatABEL generally assumes a Gaussian distribution, so it is important to consider whether this method is effective for binary data. Rönnegård et al. (2016) tested the package for its applicability to binary data and found that it is suitable, however as the proportion of successes for the trait (i.e. number of dispersers) decreases, so does the power to detect causal SNPs. If a trait has extreme binary proportions (i.e. <5% or >95%) it is recommended to treat the results with caution, but as the proportion of dispersers to non-dispersers in this study is 29% the method is a suitable one for this study.

The X chromosome tends to lack reported associations with traits, partly because of the design of genotyping arrays and other technical issues causing it to be removed from analyses. It is also less well understood how the X chromosome might need to be handled during quality control and analysis (Wise et al. 2013). For GWA analysis in the current study, autosomal SNPs only were used.

Multiple testing correction was carried out using a Bonferroni correction (Haynes 2013). Correction for multiple testing is necessary in GWAS, and the Bonferroni method is a simple approach that uses all single nucleotide polymorphisms across the genome. This is a highly conservative approach however and can limit the number of SNPs that are found to be genome wide significant (Duggal et al. 2008).

The top significant SNP in the GWAS was investigated for its proximity to gene coding regions in the house sparrow genome, and genes associated with dispersal or dispersal related traits. the top 3 SNPs on chromosome 12 were also investigated. The annotated house sparrow genome (Elgvin et al. 2017) was used to determine whether these SNPs were in intronic parts of the genome, or which genes of known function were flanking the SNP.

Unless otherwise stated, all statistical analyses were performed using R version 3.5.1 (R Core Team, 2018).

3. Results

3.1 Heritability Estimation

Additive genetic variance (V_a) estimated from analysis with MCMCglmm was 6.46 (95% C.I = 4.23-9.74). The heritability estimate indicates that dispersal is a highly heritable trait in house sparrow (h^2_{latent}= 0.53, HPD interval = 0.41-0.63, h^2_{obs}= 0.34) (Table 2).

Examining the posterior distribution of models using different priors found that the Inverse Gamma prior and the Parameter expanded F distribution prior did not fit the data correctly (Appendix III; Figures II and III), and the X^2 with 1 degree of freedom gave the best fit (Appendix III; Figure IV*).* Heritability estimates (on the latent scale) for models using each prior indicated that the heritability estimate was not particularly prior sensitive (Appendix II; Table II), so the choice of prior may not have been critical here, however the posterior distribution indicates that the X^2 with 1 degree of freedom prior was the best choice.

Table 2. Estimates of the variance components of dispersal on the latent model and expected data scales. Trait heritability on the latent scale was calculated using these variance components and the formula $V_a / V_a + V_{pe} + V_r + \pi^2/3$ where V_a is additive genetic variance, V_{pe} is permanent environmental variance (included by the use of natal island and birth year as random factors), and V_r is residual variance. Trait heritability on the expected data scale was calculated using the variance components and the formula V_a / V_p .

Scale	V_a [95% CI]	V_{pe}	V_r *	h^2
Latent	6.46 [4.23-9.74]	0.99 (year) 0.44 (island)	1	0.53
Expected data	0.083	0.246	-	0.34

*residual variance cannot be calculated for binary data; so is fixed at 1(de Villemereuil et al. 2012).

3.2 Genome Partitioning

In genome partitioning analysis using GCTA, the regression between chromosome size and proportion of variance in dispersal explained was not significant after HC-correction (p=0.27). This may be partly because many chromosomes appear to explain no variation in dispersal despite being large (Table 3).

Chromosome 12 (coloured red in Figure 2) appeared to explain a higher proportion of the variation for dispersal than expected from the general relationship between variance explained and size, and chromosome 5 also appears to explain a larger proportion of variance in dispersal than other chromosomes (Figure 2).

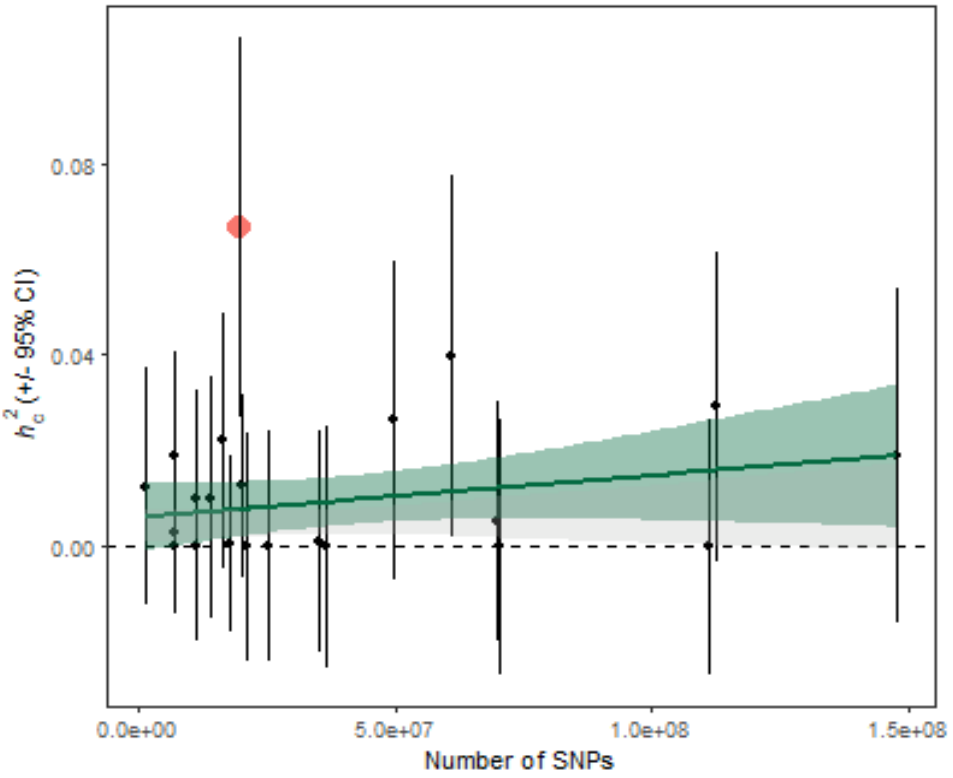

Figure 2. Relationship between explained variation in dispersal (+-SE) and chromosome size (number of SNPs). The green shaded area shows 95% C.I. for regression between number of SNPs and the proportion of variance explained by each chromosome (h_c^2), with black bars indicating 95% C.I. for each h_c^2 estimate. Grey shaded area indicates 95% quantiles formed by HC-resampling forming the null distribution for HC-corrected p-value.

Table 3. Genome partitioning output. Shown are the chromosome size (bp), the effect size and the standard error. High h^2 for chromosome 12 indicates that this is the significant (coloured red) chromosome in Figure 2.

Chromosome	**Size**	$\boldsymbol{h^2}$	**SE**	**p_out**
1	112670000	0.029210	0.016508	0.865353698
2	147840000	0.018851	0.017782	0.999799035
3	110990000	0.000001	0.013635	0.999799035
4	70350000	0.000001	0.013661	0.999799035
5	61080000	0.039725	0.019221	0.195136656
6	35010000	0.001079	0.011836	0.999799035
7	36520000	0.000001	0.012772	0.999799035
8	49690000	0.026366	0.016857	0.837419614
9	25220000	0.000001	0.012271	0.999799035
10	21080000	0.000001	0.012133	0.999799035
11	20430000	0.012800	0.009752	0.999799035
12	19790000	0.066468	0.020233	0.001607717
13	18020000	0.000767	0.009399	0.999799035
14	16470000	0.022259	0.013597	0.930667203
15	14040000	0.010231	0.012829	0.999799035
17	11240000	0.000001	0.009881	0.999799035
18	11530000	0.010138	0.011531	0.999799035
19	11120000	0.000001	0.007983	0.999799035
20	1478000	0.012595	0.012596	0.999799035
23	7030000	0.018849	0.011080	0.986736334
24	7080000	0.000001	0.007213	0.999799035
26	6900000	0.003097	0.007512	0.999799035
29	69870000	0.005278	0.012778	0.999799035

3.3 GWAS

GWA analyses in RepeatABEL revealed a single genome-wide significant SNP, SNPa105045 on chromosome 15 (*p*= 1.25e-07, position = 2695766, effect size = 0.10± 0.02) (Table 4). Other suggestive 'peaks' can be seen in the GWA analysis (Figure 4) where areas more influential in dispersal could be found, one of which (on chromosome 12) corresponds to the results shown in genome partitioning.

Table 4: Summary statistics for the top ten SNPs associated with dispersal in the RepeatABEL GWAS. Top SNP is significant at the genome-wide local significance level of 2.7 x 10^-7. For each SNP the table shows its name, chromosome, position (bp), the reference allele A1, effect allele A2, estimated effect size of A2 with standard error, and *p*-value.

SNP	Chr.	Position	A1	A2	Effect ± SE	*p*-value
SNPa105045	15	2695766	G	A	0.104 ± 0.020	1.25e-07
SNPa16999	12	14555542	T	G	-0.052 ± 0.012	1.53e-05
SNPa14838	12	10813153	G	A	-0.061 ± 0.014	1.64e-05
SNPa344637	2	22728732	C	T	0.051 ± 0.012	1.74e-05
SNPa235914	1	23895981	C	T	0.063 ± 0.015	1.97e-05
SNPa480081	11	2315811	G	A	0.051 ± 0.012	2.24e-05
SNPa397538	23	4055552	G	A	0.135 ± 0.032	3.23e-05
SNPa89783	11	7242384	C	A	0.069 ± 0.017	3.41e-05
SNPa14833	12	10808310	T	C	-0.063 ± 0.015	3.76e-05
SNPa17197	12	14885979	T	C	-0.052 ± 0.013	4.03e-05

Lamda = 1.01 (SE = 2.75e-05), so it was not necessary to correct the estimate for inflation (Figure 3).

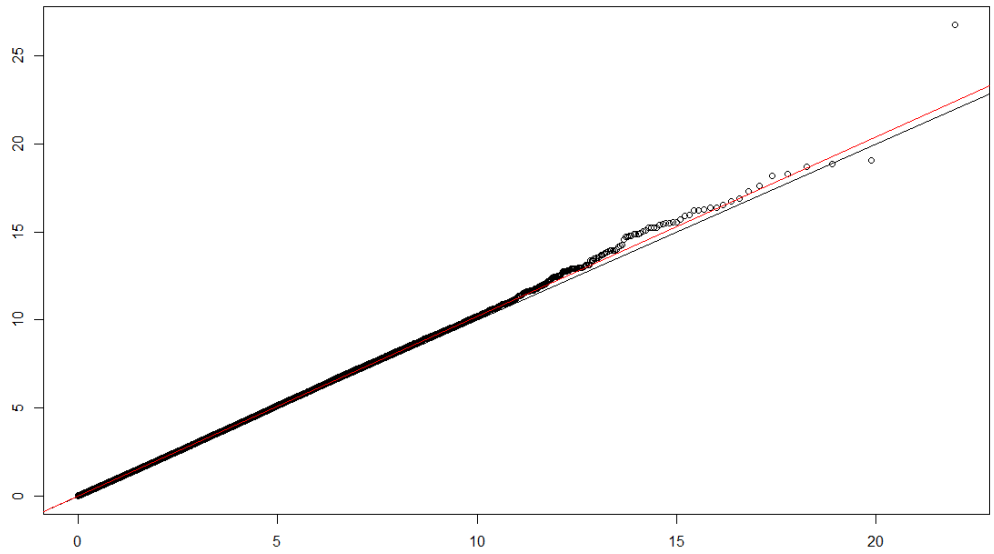

Figure 3. X^2-X^2 plot for GWA scan. Black line of slope 1: expected under no inflation; Red line: fitted slope.

The significant locus is near two gene coding regions in the annotated house sparrow genome (Appendix II; Table I). The closest gene of known function downstream of SNPa105045 is IV00_00043860 (11 Kbp away), and upstream is IV00_00043864 (15 Kbp away).

Gene IV00_00043860 is analogous to UPB1 that is associated with amino acid biosynthesis in Sumatran orangutan (*Pongo abelii)*. Gene IV00_00043864 is analogous to ADORA2A, that is associated with G protein-coupled receptor activity in *Homo sapiens*.

Genes surrounding the top 3 SNPs on chromosome 12 were also investigated. The SNP closest to genome-wide significance on chromosome 12 was SNPa16999. The closest gene of known function downstream is IV00_00039264 (205 Kbp away). Gene IV00_00039264 is analogous to SUCLG2, that is associated with carbon metabolism and the citric acid cycle in the common pigeon (*Columbia Livia).* The closest gene of known function upstream is IV00_00039280 (214 Kbp away). Gene IV00_00039280 is analogous to FAM19A4, which modulates injury-induced and chemical pain hypersensitivity in the crab-eating macaque (*Macaca fascicularis).*

No genes were found that have been associated with dispersal or dispersal related traits (Appendix II: Table I).

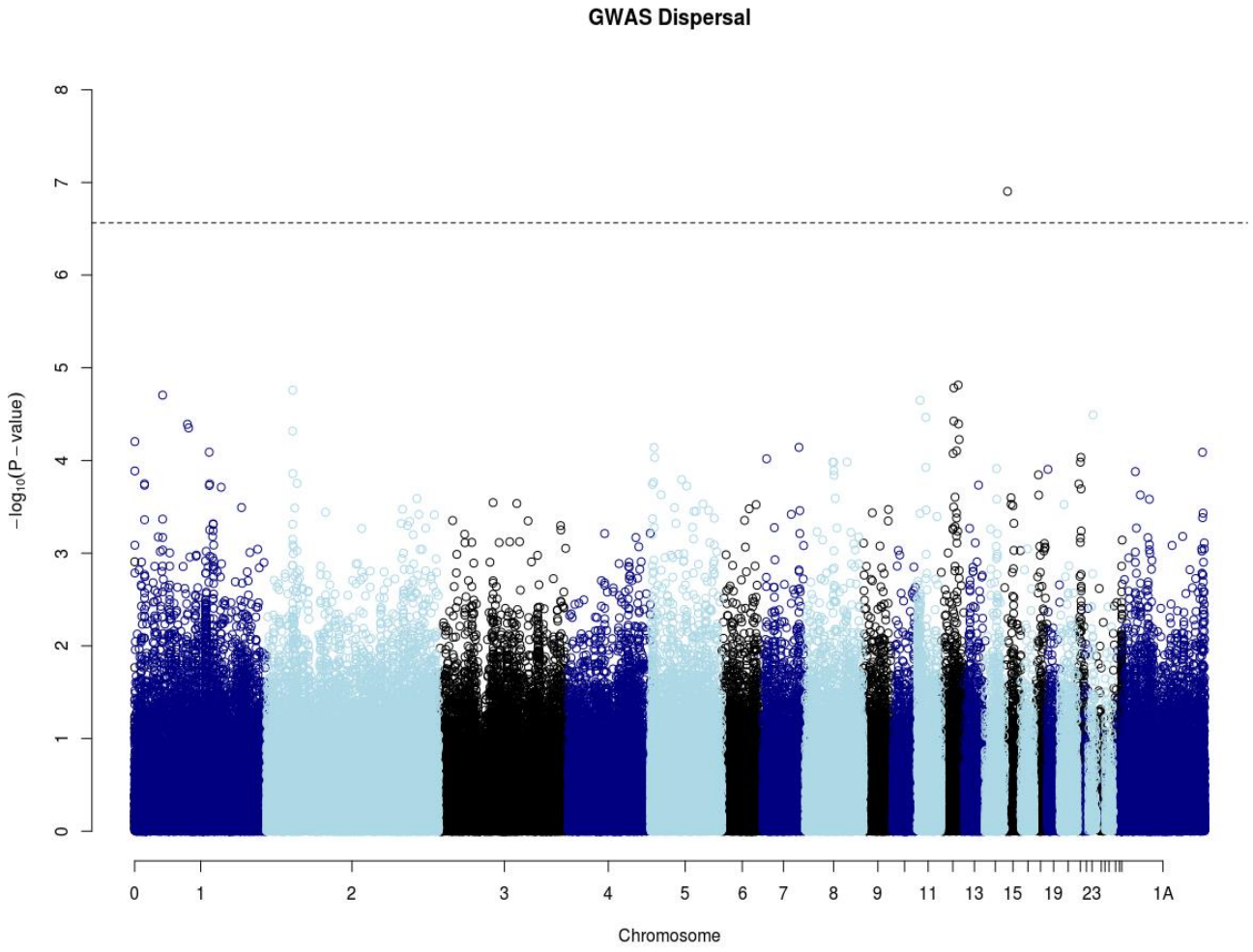

Figure 4. Manhatten plot of RepeatABEL GWA scan for dispersal (N=2716 measurements) on 183 088 SNPs. Position of markers on the X axis corresponds to their bp position on their chromosome. Genome-wide local significance level after correction is 2.7e-7 (dotted line).

4. Discussion

This study has provided evidence to support the genetic basis for dispersal in the Helgeland house sparrow metapopulation. The animal model implemented in MCMCglmm indicated that dispersal has substantial additive genetic variance ($V_a = 6.46, \text{C.I} = 4.23 - 9.74$), with some variance also attributable to year ($0.99,\ [0.91 - 1.08]$) and island ($0.44,\ [0.07 - 1.99]$) (corresponding values on the data scale can be found in Table 2). It also indicated that dispersal is highly heritable in the study population ($\text{h}^2_{\text{latent}} = 0.53$), which is in concordance with other studies showing that dispersal and dispersal traits can be highly heritable in wild populations.

The heritable basis for dispersal

The most common studies that have found evidence for heritability of dispersal traits are in plants and insects, usually on seed or pollen dispersal structures or wing morphology (Roff and Fairbairn 2001, Ronce 2007). In vertebrates, indirect evidence is accumulating in the form of within-family (sibling and parent–offspring) resemblance in dispersal behavior (Massot and Clobert 2000, Massot et al. 2003, Doligez and Pärt 2008, Sharp et al. 2008). Some studies in vertebrates have also shown high heritability estimates for personality traits that could be related to dispersal, such as exploratory behaviour in great tits, *Parus major*, which have ranged between 0.1 and 0.78 (Dingemanse et al. 2002, Drent et al. 2003), and boldness in dumpling squid, *Euprymna tasmanica,* where h^2 estimates range from 0.2–0.8 (Sinn et al.2006). In zebrafish estimates of heritability of boldness were found to be quite high ($\text{h}^2 = 0.76$) (Ariyomo et al. 2013). This trait was described as the propensity of an individual to take risks and can be conceivably linked to a higher dispersal propensity. They also found a high heritability for aggressiveness ($\text{h}^2 = 0.36$).

Most previous studies have focused on the heritability of dispersal related traits (Saastamoinen et al. 2018), such as boldness (Ariyomo et al. 2013) and flight metabolic rate (Niitepõld et al. 2009, Mattila and Hanski 2014,), but this study looks at the propensity to disperse from the natal patch as a trait in its own right. When behavioural or morphological traits are used as proxies for dispersal, it is often assumed that the trait has a direct functional link to dispersal propensity, for example that bolder individuals will move further due to curiosity and lack of fear (Duckworth et al. 2015). However, it is also likely that such traits may affect the dynamics within a population,

and that these dynamics cause variation in dispersal tendencies (Bowler and Benton 2005). Therefore, the correlation between dispersal and a particular trait may be tangled up with the environmental drivers, rather than there being a causal link between them (Duckworth et al. 2015). This could mean that an estimate of the heritability of a trait associated with dispersal provides less clear information about the evolutionary potential of dispersal itself. Considering dispersal as a trait also has the benefit of including all the stages of dispersal, because individuals are evaluated as dispersers when they are captured or sighted on an island that is not the island they were born on. Consequently 'dispersal' as a trait accounts for emigration from the natal island, inter-island movement, and immigration into a new island, as well as successful establishment. Hence, rather than providing information about a specific trait that may be useful in one or all of these stages of dispersal, this study has effectively investigated the heritability of *successful* dispersal into a new local population and this has not been considered a lot in previous literature (Saastamoinen et al. 2018). This study does not take into account all measures of 'success' as it is not limited only to dispersers that manage to reproduce or provide recruits. Although there is some evidence that at least male dispersers have a lower lifetime production of recruits in this population (Pärn et al. 2009), there is also some evidence that survival of dispersers is higher than among the residents of the new adult island, and also higher than the adults who remained on their natal island (Altwegg et al. 2000). More investigation is required to determine whether dispersers have an advantage or disadvantage in fitness in the Helgeland metapopulation, however higher survival is often linked to higher fitness due to the availability of more seasons in which to breed, and so it is likely that increased survival of dispersers could have knock-on effects on fitness.

There are cases where heritability of dispersal as a trait in its own right have been estimated, for example in Doligez et al. (2009) where dispersal for the collared flycatcher was defined as an observed change of patch between the years of birth and first breeding, or between breeding years. Rather than using a continuous variable such as dispersal distance, they chose the binary definition with the expectation that it would minimise methodological problems. In a patchy population such as the one used in that study, and the current study population, dispersal distance can easily be constrained by the site layout, and so individual dispersal distance and settlement decisions will be highly influenced by environmental constraint (Doligez et al. 2004, Doligez and Pärt 2008, Doligez et al. 2009). Previous studies on the collared flycatcher population have also shown the binary variable of dispersal to respond to many social and environmental factors (Doligez et al. 2002, 2004, Doncaster et al. 1997). High heritability

estimates have also been found in western bluebirds, *Sialia mexicana*, for propensity to leave the natal site ($h^2 = 0.95$) based on animal model analysis. Dispersal as a binary trait can create difficulties in the data as well. Assigning a dispersal status to individuals usually requires two successive observations, because they must be detected on their birth island and on an adult island. This means that dispersal status for those that are not recaptured or captured for the first time as adults can be uncertain or incorrectly assigned (Doligez et al. 2011). One of the ways that this problem has been mitigated in the current study is through genetically assigning individuals to birth islands (Saatoglu et al. *in prep*), so any individuals captured as adults have an accurate dispersal status.

An important issue to consider in the models is that effects of common environment, such as that shared by chicks in the same nest, can make up a significant proportion of variance (Kruuk 2004). In collared flycatcher chicks, 49% of variance in body condition is accounted for by differences between nest-boxes (Merilä et al. 2001), and similar results have been found in long-tailed tits (MacColl & Hatchwell 2003), and blue tits (Charmantier et al. 2004). Although my study included effects of natal island and birth year, nest or mother identity could have been included as random effects in the model, in order to estimate the component of variance due to differences in the trait between offspring of different mothers that is additional to additive genetic effects. These are factors that could influence the condition of individuals and therefore their dispersal propensity. Future work could include genetic correlation analysis of fitness indications (size, etc.) and dispersal, through bi-variate or multi-variate animal models, to see if dispersal is related to any of these conditions that might be influenced by mother care, and nest quality.

Estimating heritability

Heritability as a parameter can be very useful and informative, because it provides a simple measure of the importance of heritable genetic variation (additive genetic variance) for individual differences, and it allows this information on trait variation to be compared within and across populations (Visscher et al. 2008). It also allows predictions about the response to selection (Visscher et al. 2008). To obtain very accurate estimates of heritability however, hundreds or thousands of observations are needed, because the accuracy of a heritability estimate is affected by bias based on technical issues such as sampling error (Visscher et al. 2008). Therefore, it is important to have a large sample size and high-quality pedigree structure, as

well as to account for confounding effects (Visscher et al. 2008). One of the main challenges of heritability estimation is its basis on the assumption that there are infinite unlinked loci which have a small impact on the phenotypic variance (de Villemereuil et al. 2013). This assumption ignores other genetic effects and constraints such as linkage (de Villemereuil et al. 2013), and it is worth considering whether a heritability estimate with these genetic factors unaccounted for is truly accurate, or even informative. Another important problem is that wild populations usually have imperfect individual detection rate, and this can bias the estimate of heritability (de Villemereuil et al. 2013). The detection error rate in the Helgeland metapopulation has been estimated at 4.5% (Saatoglu et al. 2019 *in prep*), which indicates that the accuracy of estimates from this system should not be greatly affected by this issue. Future work into the genetic basis for dispersal could benefit from investigating non-additive genetic effects, such as dominance genetic variance and epistatic variance, and how these might change estimates of heritability.

This study has been done on a very large sample size, particularly compared to many other studies of wild populations (Saastamoinen et al. 2018, Doligez et al. 2011), so its ability to estimate heritability accurately should be quite high. However, problems associated with each method of estimating heritability are also important to consider in order to determine if the heritability estimate in this study is reliable. The use of generalized linear mixed models (GLMMs) to analyze non-Gaussian traits has been increasing (Wilson et al. 2011, Morrissey et al. 2012, de Villemereuil et al. 2013) because they provide many benefits over traditional methods, such as parent-offspring regressions and sibling designs. GLMMs are often more effective for non-Gaussian traits and are also particularly useful for considering the various environmental variables in natural environments. Within GLMMs however there are also a variety of ways to estimate variance components of quantitative traits. Frequentist mixed models (REML, PQ), and Bayesian mixed model methods (MCMC) contain many differences that can affect their suitability for estimating the heritability of a certain trait. Bayesian method have been referenced as the better option for non-Gaussian traits, although there are various considerations to be taken into account (Appendix I: Addendum IV). Doligez et al. (2011) used MCMCglmm to fit an animal model on dispersal for the Collared Flycatcher and estimated the heritability of natal dispersal to be 0.39 (0.31-0.47), a similar result to the one found in this study, using the same method. Whilst we should be cautious about drawing conclusions from it, heritability can be a very useful measure of trait variation, and it provides many new opportunities that assist with understanding phenotypic variation and the relationship between genes and the environment (Visscher et al. 2008).

An important aspect of MCMCglmm that can influence the accuracy of heritability estimates is the issues of autocorrelation and convergence. Testing of various model lengths, and thinning intervals allowed the model with the highest effective sample size, and acceptable levels of autocorrelation to be chosen (Appendix I: Addendum II), and this should allow confidence in the heritability estimate.

Heritability estimates were done on both the latent scale and the data scale. Heritability estimates on the expected data scale are expected to be significantly lower than the latent trait scale (Villemereuil et al. 2018, Appendix I: Addendum III). Due to the noise in GLMMs, generated by the error processes and over-dispersion variance in the latent scale, phenotypic variances on the expected data scale can be very large (Villemereuil et al. 2018). Consequently, heritability estimates are often quite small. However, heritability estimates in this study were still very high even on the data scale (h^2obs= 0.34), and so considering the reliability of these estimates, particularly knowing the prior sensitivity in MCMC models is important (de Villemereuil et al. 2013) (Appendix, Addendum I).

Genome partitioning and genome-wide association studies

From genome partitioning analysis, it is not possible to conclude that dispersal is a polygenic trait, because the slope of the regression was not significantly positive (p=0.27). With a polygenic trait, you would expect to see a significant positive relationship between chromosome effect size and the number of SNPs on a chromosome. It is unlikely that dispersal is a trait regulated by only one or a few genes however, and the slope of the regression does show a general positive trend. This trend however is quite weak, and multiple chromosomes with differing sizes explain the same amount of variance. However, if you consider the mean estimates per chromosome (Table 3), many large chromosomes don't explain any variation at all. It may be possible then that there are multiple genes that affect dispersal as we might expect, but they are not distributed evenly across the genome and thus do not follow the pattern of correlation between chromosome size and amount of variance explained (Kemppainen and Husby 2018). One of the reasons that there was no significant positive relationship indicated, may be the binary nature of the trait of interest. Binary traits tend to have a large error, and this could influence the significance of the positive association (Kemppainen and Husby 2018b).

The genes found around the significant SNP (SNPa105045) on chromosome 15 have not been shown to have any association with dispersal or dispersal related traits. However, other suggestive peak regions in the GWAS indicated that there may be alternative areas of the house sparrow genome which explain some dispersal variation. In particular, a number of SNPs in chromosome 12 reach closest to genome wide significance in addition to SNPa105045. Probably caused by these SNPs, chromosome 12 was also shown to have a disproportionate effect on dispersal variance for its size. GWAS results are indicative of dispersal as a complex polygenic trait that is influenced by many genes spread across the genome that each have a small effect on variation in the trait (Slate et al. 2005, Slate et al. 2010 Stapley et al. 2010. In fact, studies have indicated that most complex traits in wild populations might have this kind of genetic architecture (Santure et al. 2015). Based on GWA studies of diverse populations, in fact there has not been a great amount of evidence for any large effect QTL. In Husby et al. (2015), only one genome-wide SNP for clutch size was detected, and it explained only 3.9% of the variance.

Although the power of this study is high for a study on a wild animal population, it is still quite low in comparison to GWA studies in humans, so it is possible the sample size is still too low to detect the small effects genes are having on trait variance. As it is unlikely that dispersal is regulated by a few genes of large effect, the power may need to be significantly higher to detect causal genes in this case. The significance threshold used is also very strict, as it is not based on sample size but is divided by the number of tests (183033).

SNP array and sample sizes

The 200K SNP array used in this study has a high marker density, and sample size in comparison to many association studies in wild populations (Santure et al. 2015, Chaves et al. 2016, Johnston et al. 2011, Schielzeth and Husby 2014). For this reason, overestimation of effect sizes should be minimal (Slate 2013), and the power to detect important genetic variants high. However, variants may still be missed when marker density and sample size are high, especially if we are looking for rare variants with low minor allele frequency MAF (<0.05), and small effect (Wilkening et al. 2009). Since it is likely that the genes involved in dispersal are of small effect, this is something to consider in the case of my study. Large effect genes that have been identified through GWA studies in wild populations tend to relate to near Mendelian traits under strong selection, or those involved in adaptation to trophic niches such as bill morphology

in Darwin's finches (Johnston et al. 2011, Johnston et al. 2013, Lamichhaney et al. 2015, Chaves et al. 2016, Lamichhaney et al. 2016). In addition, you can find so-called 'missing heritability' problems, such as the case with human height. This is a highly polygenic trait, for which hundreds of variants across the genome have been identified, and yet these explain only around 20% of the variation in human height (Marouli et al., 2017).

One of the issues with current quantitative genetic models, because they were developed for captive populations, is that they assume that phenotypic variance is consistent among all individuals, because detection of individuals is perfect (Cam 2009). This is obviously unlikely in natural populations (Cam 2009,), and can cause biased estimates based on the type of individuals that are more often captured, and a flawed understanding of the demographic parameters of interest (Doligez et al. 2011). One major benefit of this study system therefore is the unique coverage of the population, in which we can be fairly sure that a large percentage of the population is covered. Although not foolproof, this is a good way of minimising these problems. Although success rate can be low for detecting outlier loci of quantitative traits, and those detected often have a small effect, genome wide studies are still important for examining the genetic basis of ecologically important traits in natural populations (Slate et al. 2010, Stapley et al. 2010). Examination of such traits in their natural context can help to untangle complicated environment-phenotype-genotype interactions, and to understand the mechanisms behind adaptive evolution.

Implications of results for populations

Evolutionary studies are motivated by the importance of being able to predict the effect of natural or sexual selection on traits, particularly whether they can create a permanent phenotypic change (Kruuk 2004). Estimating the genetic basis of quantitative traits is intended as a tool to better be able to make and understand these predictions (Kruuk 2004). With these indications that dispersal is a heritable trait in the Helgeland house sparrow metapopulation, and that there are relatively many genes of relatively small effect across the genome that influence the variation in this trait, a logical next step then is to consider what implications this has for the Helgeland metapopulation, and other populations.

There are indications that dispersal strategies can evolve in the wild, in response to selective pressures (Kokko and López-Sepulcre 2006, Duckworth and Badyaev 2007, Thomas et al.

2001). In crested tits (Lens and Dhont 1994), fragmentation of important habitat was found to delay natal dispersal. It also increased dispersal distance in nuthatches but decreased the probability of natal dispersal for them (Mattysen et al. 1995). Whilst there is no evidence that dispersal is a recently evolved trait in the house sparrow, changes in dispersal patterns and frequency may be brought on by environmental change. Potential reasons for house sparrow dispersal include; changing environmental conditions, e.g. between 1995-2010 in Troms, Northern Norway, average temperature in May increased by around 2.5 °C, representing an increase of 0.19 °C per year (Barrett, 2011). In the Glanville fritillary butterfly (*Melitaea cinxia)* metapopulation of the Åland islands, heritable genetic variation for a metabolic enzyme associated with dispersal has been found, along with environmental drivers of dispersal between local populations within the metapopulation (Niitepõld & Saastamoinen, 2017). It has also been shown that changing climate is altering the metapopulation dynamics through increased dispersal, with potential negative effects for the survival of the metapopulation (Kahilainen et al. 2018), and that habitat fragmentation has changed allele frequencies in the population in the past (Fountain et al. 2016). Determining the relative importance of individual phenotype and environmental or population dynamics effects on dispersal is of course difficult (Duckworth et al. 2015) but knowing that there is additive genetic variance for dispersal means we can understand the possible evolutionary outcomes of such pressures. If one imagines there are only environmental causes for dispersal, and at certain times there are no environmental triggers for dispersal, the dispersal rates would be low. In a metapopulation, this could create circumstances in which the population might not be able to survive because extinct patches are not recolonised and sink populations are not ‘rescued’ by immigrants (Hanski and Gaggiotti 2004). Because dispersal seems to be a heritable trait in the Helgeland metapopulation however, there is possibility that dispersal may also evolve in response to environmental changes. If dispersal propensity also corresponds with higher fitness, then dispersal frequency in this population is likely to increase. If it corresponds with lower fitness, the likelihood is that dispersal frequency would decrease. This correlation with fitness may also change over time, particularly when the fitness of the individual is related to environmental variables. For example, in a metapopulation such as this one, individuals that disperse may have higher fitness, but as fragmentation of habitats becomes more significant the distances between local populations could become larger and the costs associated with dispersing greater. Then the fitness of dispersers might decrease, causing the evolution of a lower dispersal propensity. Unfortunately, considering that habitat fragmentation is increasing, this particular outcome could cause difficulties for the survival of the metapopulation if local populations are not able to be

supplemented. Alternatively, as the climate changes many species need to adjust their home ranges to maintain favourable conditions. Dispersers in the house sparrow metapopulation may have higher fitness due to their ability to find more suitable habitat and dispersal propensity may increase.

5. Conclusion

Results presented here indicate that dispersal is highly heritable, and most likely governed by polygenic genetic architecture. A significant locus was identified which may explain a small proportion of the variation in dispersal. Additionally, chromosome 12 seemed to explain a disproportionate amount of variation in dispersal, and this chromosome holds a number of SNPs that were the next closest to significance. However, none of the investigated SNPs were close to any genes of known association with dispersal or dispersal related traits.

Uncovering loci for quantitative traits in natural populations is difficult and that fact is highlighted by this study. In order to detect large effect variants, it may be necessary to select traits with high heritability and high variability in a population, and that are under strong selection.Future work could aim to determine the extent to which variants identified here affect dispersal alone, or if environmental causes have a greater influence on dispersal variability.

Despite difficulties encountered in uncovering genes for ecologically important traits in natural populations, animal model and GWA studies can, on occasion, provide valuable insights into eco-evolutionary dynamics in such populations. Consequently, association studies are likely to remain a valuable investigative tool in genetic and eco-evolutionary research into the future.

References

ALTWEGG, R., RINGSBY, T.H., SAETHER, B.E. (2000). "Phenotypic correlates and consequences of dispersal in a metapopulation of house sparrows Passer domesticus." *Journal of Animal Ecology* 69(5): 762-770.

ANDERSON, T. (2007) (Ed.). “Biology of the Ubiquitous House Sparrow: From Genes to Populations”. *Oxford University Press*.
ARIYOMO, T. O., CARTER, M., Watt, P.J. (2013). "Heritability of Boldness and Aggressiveness in the Zebrafish". *Behavior Genetics* 43(2): 161-167.

AULCHENKO, Y.S., RIPKE, S., ISAACS, A., VAN DUIJN, C.M. (2007). "GenABEL: an R library for genome-wide association analysis." *Bioinformatics* 23(10): 1294-1296.

BAALSRUD, H. T., SÆTHER, B., HAGEN, I. J., MYHRE, A. M., RINGSBY, T. H., PÄRN, H. AND JENSEN, H. (2014). "Effects of population characteristics and structure on estimates of effective population size in a house sparrow metapopulation." *Molecular Ecology* 23(11): 2653-2668.

BARRETT, R. D. H. AND H. E. HOEKSTRA (2011). "Molecular spandrels: tests of adaptation at the genetic level." *Nature Reviews Genetics* 12: 767.

BARSON, N. J., et al. (2015). "Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon." *Nature* 528: 405.

BECKS, L., ELLNER, S.P., JONES, L.E., HAIRSTON, N.G.. (2010). "Reduction of adaptive genetic diversity radically alters eco-evolutionary community dynamics." *Ecology Letters* 13(8): 989-997.

BILLING, A. M., LEE, A. M., SKJELSETH, S., BORG, Å. A., HALE, M. C., SLATE, J., PÄRN, H., RINGSBY, T. H., SÆTHER, B.E. & JENSEN, H. (2012). “Evidence of inbreeding depression but not inbreeding avoidance in a natural 618 house sparrow population”. *Molecular Ecology*, 21:1487-1499.

BOLKER, B. M., BROOKS. M.E., CLARK, C.J., GEANGE, S.W., POULSEN, J.R., STEVENS, M.H., WHITE, J.S.,(2009). "Generalized linear mixed models: a practical guide for ecology and evolution." *Trends in Ecology & Evolution* 24(3): 127-135.

BOWLER,D.E., BENTON,T.G .(2005). “Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics”. *Biological Reviews* 80: 205–225.

BRESLOW, N. E. AND X. LIN (1995). "Bias Correction in Generalised Linear Mixed Models with a Single Component of Dispersion." *Biometrika* 82(1): 81-91.

CAILLAUD, M. C.,BOUTIN, M.,BRAENDLE, C. & SIMON, J. C. (2002). “A sex-linked locus controls wing polymorphism in males of the pea aphid, Acyrthosiphon pisum”. *Heredity* 89: 346–352.

CAM, E. (2009). “Contribution of capture-mark-recapture modeling to studies of evolution by natural selection. Modeling Demographic Processes In Marked Populations”, (eds D.L. Thomson, E.G. Cooch & M.J. Conroy): 83–129. Springer, Boston, Massachusetts, USA.

CHARMANTIER, A., KRUUK, L.E.B., BLONDEL, J., LAMBRECHTS, M.M. (2004). "Testing for microevolution in body size in three blue tit populations." *Journal of Evolutionary Biology* 17(4): 732-743.

CHARMANTIER, A. GARANT, D. (2005). "Environmental quality and evolutionary potential: lessons from wild populations." *Proceedings. Biological Sciences* 272(1571): 1415-1425.

CHARMANTIER, A., GARANT, D., KRUUK, L.E.B. (2014). “Quantitative Genetics in the Wild”, Oxford University Press.

CHAVES, J. A., COOPER, E.A., HENDRY, A.P., PODOS, J., DE LEÓN, L.F., RAEYMAEKERS, J.A., MACMILLAN, W.O., UY, J.A. (2016). "Genomic variation at the tips of the adaptive radiation of Darwin's finches." *Molecular Ecology* 25(21): 5282-5295

CHEPTOU, P.O., HARGREAVES, A.L., BONTE, D., JACQUEMYN, H. (2017). "Adaptation to fragmentation: evolutionary dynamics driven by human influences." *Philosophical Transactions of the Royal Society B: Biological Sciences* 372:1712.

CLOBERT, J., BAGUETTE, M., BENTON, T.G. & BULLOCK, J.M. (2012). “Dispersal Ecology and Evolution”. Oxford University Press, Oxford, UK.

CLUTTON-BROCK, T., SHELDON, B.C. (2010). “Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology”. *Trends in Ecology & Evolution* 25: 562-573.

COTE, J., BESTION, E., JACOB, S., TRAVIS, J., LEGRAND, D., BAGUETTE, M. (2017). "Evolution of dispersal strategies and dispersal syndromes in fragmented landscapes." Ecography 40(1): 56-73.

DANCHIN, E., CHARMANTIER, A., CHAMPAGNE, F. A., MESOUDI, A., PUJOL, B., BLANCHET, S. (2011). “Beyond DNA: integrating inclusive inheritance into an extended theory of evolution”. *Nature Reviews Genetics* 12, 475–486.

DE VILLEMEREUIL, P. (2012). “Tutorial Estimation of a biological trait heritability using the animal model”, How to use the MCMCglmm R package.

DE VILLEMEREUIL, P., GIMENEZ, O., DOLIGEZ, B.(2013). “Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits”. *Methods in Ecology and Evolution*, 4(3):260–275.

DE VILLEMEREUIL, P., SCHIELZETH, H., NAKAGAWA, S., MORRISSEY, M. (2016). "General Methods for Evolutionary Quantitative Genetic Inference from Generalized Mixed Models." *Genetics* 204(3): 1281-1294.

DE VILLEMEREUIL, P. (2018). “Quantitative genetic methods depending on the nature of the phenotypic trait”, *Annals of the New York Academy of Sciences*, 1422(1).

DINGEMANSE, N., BOTH, C., DRENT, P.J., VAN OERS, K., VAN NOORDWIJK, A.J. (2002). “Repeatability and heritability of exploratory behaviour in great tits from the wild”. *Anim Behav,* 64(6):929–938.

DOLIGEZ, B., DANCHIN, E., CLOBERT, Jl. (2002). "Public Information and Breeding Habitat Selection in a Wild Bird Population." *Science* 297(5584): 1168-1170.
DOLIGEZ, B., PÄRT, T. (2008) “Estimating fitness consequences of dispersal: a road to ‘know-where’? Non-random dispersal and the underestimation of dispersers’ fitness”. *Journal of Animal Ecology* 77:1199–121.

DOLIGEZ, B., GUSTAFSSON, L.,PÄRT, T. (2009). “Heritability' of dispersal propensity in a patchy population", *Proceedings. Biological sciences,* 276(1668): 2829-2836.

DOLIGEZ, B., DANIEL, G., WARIN, P. ET AL. (2011). “Estimation and comparison of heritability and parent-offspring resemblance in dispersal probability from capture-recapture data using different methods: The Collared Flycatcher as a case study”, *Journal of Ornithology*, 152(2): 539-554.

DONCASTER, C., CLOBERT, J., DOLIGEZ, B., GUSTAFSSON, L., DANCHIN, E. (1997). “Balanced Dispersal Between Spatially Varying Local Populations: An Alternative To The Source-Sink Model”. *The American Naturalist*, 150(4):425-45.

DRENT, P.J., VAN OERS, K., VAN NOORDWIJK, A.J. (2003). “Realized heritability of personalities in the great tit (Parus major)”. *Proc R Soc Lond B Biol Sci*, 270(1510):45–51.

DUCKWORTH, R.A., BADYAEV, A.V. (2007). "Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird." *Proceedings of the National Academy of Sciences of the United States of America,* 104(38): 15017-15022.

DUCKWORTH, R. A., KRUUK, L. E. B. (2009). “Evolution of genetic integration between dispersal and colonization ability in a bird”, *Evolution* 63, 968–977.

DUCKWORTH, R. A., BELLONI, V., ANDERSON, S.R. (2015). "Cycles of species replacement emerge from locally induced maternal effects on offspring behavior in a passerine bird." *Science,* 347(6224): 875-877.

DUGGAL, P., GILLANDERS, E., HOLMES, T.N., BAILEY-WILSON, J.E. (2008). "Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies." *BMC Genomics,* 9:516-516.

EDELSPARRE, A.H., LIM, J.H., ANWARI, M., FITZPATRICK, M.J. (2014). “Alleles underlying larval foraging behaviour influence adult dispersal in nature”, *Ecology Letters,* 17, 333–339.
ELGVIN, T.O., TRIER, C.N., TØRRESEN, O.K., et al. (2017). “The genomic mosaicism of hybrid speciation”, *Science Advances*, 3(6).

ELLEGREN, H. (2014). "Genome sequencing and population genomics in non-model organisms." *Trends Ecol Evol* 29(1): 51-63.

ETTERSSON, J. R. (2004). "Evolutionary potential of *chamaecrista fasciculata* in relation to climate change", *Evolution,* 58(7): 1459-1471.

FALCONER, D.S, MACKAY, T.F.C. (1996). *“*Introduction to Quantitative Genetics”, Fourth Edition. Longman Group Limited, Essex.

FLINT, J., ESKIN, E. (2012). “Genome-wide association studies in mice”, *Nature Reviews Genetics*, 13:807–817.

FOUNTAIN, T., NIEMINEN, M., SIRÉN, J., WONG, S. C., LEHTONEN, R., & HANSKI, I. (2016). “Predictable allele frequency changes due to habitat fragmentation in the Glanville fritillary butterfly”. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10), 2678–2683.

GIENAPP, P., FIOR, S., GUILLAUME, F., et al. (2017). "Genomic Quantitative Genetics to Study Evolution in the Wild." *Trends in Ecology & Evolution*, 32(12): 897-908.

HAAG, C.R., RIEK, M., HOTTINGER, J.W., PAJUNEN, V.I., EBERT, D. (2005). “Genetic diversity and genetic differentiation in Daphnia metapopulations with subpopulations of known age.” *Genetics* 170:1809–1820.

HADFIELD, J. D., RICHARDSON, D.S., BURKE, T. (2006). "Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework." *Molecular Ecology,* 15(12): 3715-3730.

HADFIELD, J.D. (2010). “MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package”. *Journal of Statistical Software*, 33(2):1-22.

HADFIELD, J.D.(2016). “MCMCglmm course notes”. http://cran.r-project.org/web/packages/MCMCglmm/index.html

HANSEN, T.F. (2006). "The Evolution of Genetic Architecture." *Annual Review of Ecology, Evolution, and Systematics*, 37(1): 123-157.

HANSKI, I., GAGGIOTTI, O.E. (2004). “Ecology, genetics, and evolution of metapopulations”, San Diego:Academic Press, 3-22.

HAYNES, W. (2013). “Bonferroni Correction”. *Encyclopedia of Systems Biology*. W. Dubitzky, O. Wolkenhauer, K.-H. Cho and H. Yokota. New York, NY, Springer New York: 154-154

HENDRY, A.P. (2013). "Key questions in the genetics and genomics of eco-evolutionary dynamics." *Heredity,* 111(6): 456-466.

HENDRY, A.P. (2017).“Eco-evolutionary Dynamics”, Princeton University Press, Princeton and Oxford.

HEYWOOD, J.S. (2005). "An Exact Form of the Breeder's Equation for the Evolution of a Quantitative Trait under Natural Selection." *Evolution,* 59(11): 2287-2298.

HOFFMAN, A.A., WILLI, Y. (2008). "Detecting genetic responses to environmental change." *Nat Rev Genet,* 9(6): 421-432.
Holand, H. et al. Spatial variation in senescence rates in a bird metapopulation. Oecologia 181, 865–871 (2016).

HUSBY, A., SÆTHER, B.E., JENSEN, H., Ringsby, T.H. (2006). "Causes and consequences of adaptive seasonal sex ratio variation in house sparrows." *Journal of Animal Ecology*, 75(5): 1128-1139.

HUSBY, A., KAWAKAMI, T., RÖNNEGÅRD, L., et al. (2015). "Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait." *Proceedings of the Royal Society B: Biological Sciences,* 282(1806).

JENSEN, H., SÆTHER, B.E., RINGSBY, T.H., TUFTO, J., GRIFFITH, S.C., ELLEGREN, H. (2003). "Sexual variation in heritability and genetic correlations of morphological traits in house sparrow (Passer domesticus)." *J Evol Biol,* 16(6): 1296-1307.

JENSEN, H., SÆTHER, B.E., RINGSBY, T.H., TUFTO, J., GRIFFITH, S.C., ELLEGREN, H. (2004). “Lifetime reproductive 694 success in relation to morphology in the house sparrow Passer domesticus”. *Journal of Animal Ecology*, 695(73):599-611.

JENSEN, H., BREMSET, E.M., RINGSBY, T.H., SÆTHER, B.-E. (2007). “Multilocus heterozygosity and inbreeding depression in an insular house sparrow metapopulation”. *Molecular Ecology,* 16: 4066–4078.

JENSEN, H., STEINSLAND, I., RINGSBY, T.H., SÆTHER, B.-E. (2008). “Evolutionary Dynamics of a Sexual Ornament in the House Sparrow (Passer domesticus): The Role of Indirect Selection within and between Sexes”. *Evolution*, 62:1275-1293.

JOHNSON, M.L., GAINES, M.S. (1990). "Evolution of Dispersal: Theoretical Models and Empirical Tests Using Birds and Mammals." *Annual Review of Ecology and Systematics* 21(1):449-480.

JOHNSTON, S.E., McEWAN, J.C., PICKERING, N.K., et al. (2011). “Genome wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population”, *Molecular Ecology,* 20(11):2555-2566.

JOHNSTON, S.E., GRATTEN, J., BERENOS, C., et al. (2013), 'Life history trade offs at a single locus maintain sexually selected genetic variation', *Nature* 502:93-95.

JORDAN, K.W., CARBONE, M.A., YAMAMOTO, A., MORGAN, T.J., MACKAY, T.F. (2007). "Quantitative genomics of locomotor behavior in Drosophila melanogaster". *Genome Biology* 8(8):172.

JORDAN, K.W., CRAVER, K.L., MAGWIRE, M.M., CUBILLA, C.E., MACKAY, T.F.C., ANHOLT, R.R.H. (2012). "Genome-wide association for sensitivity to chronic oxidative stress in Drosophila melanogaster". *PLoS One*, 7.

KAHILAINEN, A., VAN NOUHUYS, S., SCHULZ, T., & SAASTAMOINEN, M. (2018). "Metapopulation dynamics in a changing climate: Increasing spatial synchrony in weather conditions drives metapopulation synchrony of a butterfly inhabiting a fragmented landscape". *Global change biology*, *24*(9), 4316–4329.

KEMPPAINEN, P., HUSBY, A. (2018). "Inference of genetic architecture from chromosome partitioning analyses is sensitive to genome variation, sample size, heritability and effect size distribution". *Molecular Energy Resources,* 18(4).

KEMPPAINEN, P., HUSBY, A. (2018b). "Accounting for heteroscedasticity and censoring in chromosome partitioning analyses." *Evolution Letters,* 2(6): 599-609.

KOKKO, H., LÓPEZ-SEPULCRE, A. (2006). "From Individual Dispersal to Species Ranges: Perspectives for a Changing World." *Science,* 313(5788): 789-791.

KRUUK, L.E.B. (2004). "Estimating genetic parameters in natural populations using the "animal model"." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 359(1446): 873-890.

KUISMIN, M., SAATOGLU, D., NISKANEN, A.K., JENSEN, H., SILLANPÄÄ, M.J. (2019). "Genetic assignment of individuals to source populations using network estimation tools". *Submitted.*

KVALNES, T., JENSEN, H., RØBERG, A., HOLAND, H. (2018). "Offspring fitness and the optimal propagule size in a fluctuating environment." *Journal of Avian Biology,* 49(7).

LAMICHHANEY, S., BERGLUND, J., et al. (2015). "Evolution of Darwin's finches and their beaks revealed by genome sequencing." *Nature,* 518: 371.

LAMICHHANEY, S., HAN, F., et al. (2016). "A beak size locus in Darwin's finches facilitated character displacement during a drought." *Science,* 352(6284): 470-474.

LANDE, R. (1979). "Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry." *Evolution,* 33(1): 402-416.

LEGRAND, D., COTE, J., et al. (2017). "Eco-evolutionary dynamics in fragmented landscapes." *Ecography* 40(1):9-25.

LENORMAND, T. (2002). "Gene flow and the limit to natural selection". *Trends Ecol Evol*, 17: 183-189.

LENS, L., DHONDT, A.A. (1994). "Effects of habitat fragmentation on the timing of Crested Tit Parus cristatus natal dispersal." *Ibis,* 136(2):147-152.

LUNDREGAN, S. L. HAGEN, I., et al. (2018). "Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis". *Mol. Ecol.* 27:3498–3514.

LYNCH, M., WALSH, B. (1998). "Genetics and Analysis of Quantitative Traits". Sinauer, Sunderland.

MACCOLL, A.D.C., HATCHWELL, B.J. (2003). "Sharing of caring: nestling provisioning behaviour of long-tailed tit, Aegithalos caudatus, parents and helpers." *Animal Behaviour* 66(5): 955-964.

MACKAY, T.F.C., RICHARDS, S., STONE, E.A., BARBADILLA, A., AYROLES, J.F., ZHU, D., et al. (2012). "The Drosophila melanogaster Genetic Reference Panel". *Nature*, *482*(7384), 173-178.

MAROULI, E., GRAFF, M., MEDINA-GOMEZ, C., LO, K. S., WOOD, A. R., KJAER, T. R., et al. (2017). "Rare and low-frequency coding variants alter human adult height". *Nature*, 542(7640).

MASSOT, M., CLOBERT, J. (2000). "Processes at the origin of similarities in dispersal behaviour among siblings". *J Evol Biol,* 13:707–719.

MASSOT, M., HUEY, R.B., TSUJ,I J., VAN BERKUM, F.H. (2003). "Genetic, prenatal, and postnatal correlates of dispersal in hatchling fence lizards (Sceloporus occidentalis)". *Behav Ecol* 14:650–655.

MATTHYSEN, E., ADRIAENSEN, F., DHONT, A.A. (1995). "Dispersal Distances of Nuthatches, Sitta europaea, in a Highly Fragmented Forest Habitat." *Oikos,* 72(3): 375-381.

MATTHYSEN, E. (2005). "Density-dependent dispersal in birds and mammals". *Ecography,* 28:403–416.

MATTILA, A.L.K., HANSKI, I. (2014). "Heritability of flight and resting metabolic rates in the Glanville fritillary butterfly". *Journal of Evolutionary Biology,* 27:1733–1743

MCCARTHY, M. I., ABECASIS, G.R., CARDON, L.R., et al. (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges." *Nature Reviews Genetics,* 9: 356.

MERILÄ, J., KRUUK L.E.B., SHELDON, B.C.,. (2001). "Natural selection on the genetical component of variance in body condition in a wild bird population." *Journal of Evolutionary Biology,* 14(6): 918-929.

NAKAGAWA, S., SCHIELZETH, H. (2010). "Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists". *Biological Reviews*, 85(4):935–956.

NIITEPÕLD K., SMITH, A.D., OSBORNE, J.L., REYNOLDS, D.R., CARRECK, N.L., MARTIN, A. P., MARDEN, J. H., OVASKAINEN, O., HANSKI, I. (2009). "Flight metabolic rate and Pgi genotype influence butterfly dispersal rate in the field". *Ecology,* 90:2223–2232.

NIITEPÕLD, K., MATTILA, A.L.K., HARRISON, P.J., HANSKI, I. (2011). "Flight metabolic rate has contrasting effects on dispersal in the two sexes of the Glanville fritillary butterfly". *Oecologia* 165: 847–854.

NIITEPÕLD, K., SAASTAMOINEN, M. (2017). "A candidate gene in an ecological model species: phosphoglucose isomerase (*Pgi*) in the Glanville fritillary butterfly (*Melitaea cinxia*)". *Annales Zoologici Fennici*, 54: 259-273.

NISKANEN et al. (2019), "Consistent scaling of inbreeding depression in space and time." *Submitted*.

ORR, H.A. (2005). "The genetic theory of adaptation: a brief history". *Nature Reviews Genetics* 6:119–127.

PÄRN, H., JENSEN, H., RINGSBY, T., SÆTHER, B. (2009). "Sex-Specific Fitness Correlates of Dispersal in a House Sparrow Metapopulation". *Journal of Animal Ecology, 78*(6):1216-1225.

PÄRN, H., RINGSBY, T., JENSEN, H., SÆTHER, B.E. (2012). "Spatial heterogeneity in the effects of climate and density-dependence on dispersal in a house sparrow metapopulation." *Proceedings of the Royal Society B: Biological Sciences* 279(1726):144-152.

REMINGTON, D.L. (2015). "Alleles versus mutations: Understanding the evolution of genetic architecture requires a molecular perspective on allelic origins." *Evolution,* 69(12): 3025-3038.

REZNICK, D.N., GHALAMBOR, C.K. (2001). "The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution." *Genetica,* 112-113:183-198.

RICHARDS, C.L., BOSSDORF, O., PIGLIUCCI, M. (2010). “What role does heritable epigenetic variation play in phenotypic evolution?” *BioScience,* 60:232–237.

RINGSBY, T. H., SÆTHER, B. , TUFTO, J. , JENSEN, H. AND SOLBERG, E. J. (2002). "Asynchronous spatiotemporal demography of a house sparrow metapopulation in a correlated environment." *Ecology* 83(2): 561-569.

ROBINSON, M.R., SANTURE, A.W., DECAUWER, I., SHELDON, B.C. & SLATE, J. (2013). “Partitioning of genetic variation across the genome using multimarker methods in a wild bird population”. *Molecular Ecology*, 22:3963-3980.

ROFF, D.A. (1986). "The evolution of wing dimorphism in insects." *Evolution,* 40(5):1009-1020.

ROFF, D.A., FAIRBAIRN, D.J. (2001). “The genetic basis of dispersal and migration, and its consequences for the evolution of correlated traits”. In: Clobert J, Danchin E, Dhondt AA, Nichols JD (eds) *Dispersal*. Oxford University Press, Oxford, pp 191–202.

RONCE, O., OLIVIERI, I., CLOBERT, J., DANCHIN, E. (2001). “Perspectives on the Study of Dispersal Evolution. Dispersal”. J. Clobert, E. Danchin, A. A. Dhondt and J. D. Nichols, *Dispersal,* Oxford University Press: 341-357.

RONCE, O. (2007). "How Does It Feel to Be Like a Rolling Stone? Ten Questions About Dispersal Evolution." *Annual Review of Ecology, Evolution, and Systematics* 38(1): 231-253.

RÖNNEGÅRD, L., MCFARLANE, S.E., HUSBY, A., et al. (2016). "Increasing the power of genome wide association studies in natural populations using repeated measures - evaluation and implementation." *Methods in ecology and evolution,* 7(7): 792-799.

R CORE TEAM. (2017). “R: A language and environment for statistical computing”. [Online] R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at; http://www.R-project.org/

RODRÍGUEZ-VERDUGO, A., BUCKLEY, J., STAPLEY, J. (2017). "The genomic basis of eco-evolutionary dynamics." *Molecular Ecology,* 26(6):1456-1464.

RUDMAN, S.M., BARBOUR, M.A., CSILLERY, K., et al. (2018). "What genomic data can reveal about eco-evolutionary dynamics." *Nat Ecol Evol,* 2(1): 9-15.

SAASTAMOINEN, M., BOCEDI, G., COTE, J., et al. (2018). "Genetics of dispersal." *Biological Reviews,* 93(1): 574-599.

SAATOGLU D., NISKANEN A.K., KUISMIN M., et al. (2019). “Dispersal in a house sparrow metapopulation - identifying “cryptic” dispersers using genetic assignment.” *in prep.*

SANTURE, A.W., POISSANT, J., DE CAUWER I., et al. (2015). "Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations." *Molecular Ecology,* 24(24): 6148-6162.

SCHIELZETH, H., HUSBY, A. (2014). "Challenges and prospects in genome-wide quantitative trait loci mapping of standing genetic variation in natural populations." *Annals of the New York Academy of Sciences,* 1320(1): 35-57.

SHARP, S.P., BAKER, M.B., HADFIELD, J.D., SIMEONI, M., HATCHWELL, B.J. (2008) “Natal dispersal and recruitment in a cooperatively breeding bird”. *Oikos,* 117:1371–1379.

SKJELSETH, S., RINGSBY, T. H., TUFTO, J., JENSEN, H., & SAETHER, B. E. (2007). "Dispersal of introduced house sparrows Passer domesticus: an experiment." *Proceedings. Biological sciences,* 274(1619): 1763-1771.

SLATE, J.O.N. (2005). “Quantitative trait locus mapping in natural populations: progress, caveats and future directions”. *Molecular Ecology*, 14:363-379

SLATE, J., SANTURE, A.W., FEULNER, P.G.D., et al. (2010). “Genome mapping in intensively studied wild vertebrate populations”. *Trends in Genetics*, 26:275-284.

SLATE, J. (2013). "From beavis to beak color: a simulation study to examine how much qtl mapping can reveal about the genetic architecture of quantitative traits." *Evolution* 67(5): 1251-1262.

SORENSEN, D., GIANOLA, D. (2002). “Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics”. *Springer Science and Business Media*, New York.

STAPLEY, J., REGER, J., FEULNER, P.G., et al. (2010). “Adaptation genomics: The next generation”. *Trends in Ecology and Evolution*, 25(12):705-12.

THOMAS, C. D., BODSWORTH, E.J., WILSON, R.J., et al. (2001). "Ecological and evolutionary processes at expanding range margins." *Nature,* 411(6837): 577-581.

TIFFIN, P., ROSS-IBARRA, J. (2014). "Advances and limits of using population genetics to understand local adaptation." *Trends Ecol Evol,* 29(12): 673-680.

TRAVIS, J.M.J., MÜNKEMÜLLER, T., BURTON, O.J. (2010). "Mutation surfing and the evolution of dispersal during range expansions." *J Evol Biol,* 23(12): 2656-2667.

TUFTO, J., RINGSBY, T. H., DHONDT, A. A., ADRIAENSEN, F., MATTHYSEN, E. (2005). A’ parametric model for estimation of dispersal patterns applied to five passerine spatially structured populations”. *Am. Nat*. 165:13-26.

VISSCHER P. M., WRAY, N.R., HILL, W.G. (2008). "Heritability in the genomics era — concepts and misconceptions." *Nature Reviews Genetics,* 9: 255.

WILKENING, S., CHEN, B., BERMEJO, J.L., CANZIAN, F. (2009). "Is there still a need for candidate gene approaches in the era of genome-wide association studies?" *Genomics,* 93(5): 415-419.

WILSON, A. J., RÉALE, D., CLEMENTS, M.N., et al. (2010). "An ecologist’s guide to the animal model." *Journal of Animal Ecology,* 79(1): 13-26.

WILSON, A. J., MORRISSEY, M.B., ADAMS, J.M., et al. (2011). "Indirect genetics effects and evolutionary constraint: an analysis of social dominance in red deer, Cervus elaphus." *Journal of Evolutionary Biology,* 24(4): 772-783.

WISE, A.L., GYI, L., MANOLIO, T.A. (2013). “eXclusion: Toward Integrating the X Chromosome in Genome-wide Association Analyses”. *American Journal of Human Genetics*, 92: 643-647.

YANG,J., LEE, S.H., GODDARD, M.E., VISSCHER, P.Ml. (2011). "GCTA: a tool for genome-wide complex trait analysis." *Am J Hum Genet,* 88(1): 76-82.

Appendix

I: Addendum

Prior distributions in MCMCglmm

When estimating the heritability of a binary trait the residual variance (V_r) is not able to be calculated from the data, and so it must be fixed. In MCMC it is fixed to 1 (Hadfield 2010). Because it is also necessary to add a variance component for the link variance (in this case the logit link variance), there is no longer symmetry in the heritability calculation.
$h^2 = V_a / V_a + V_r$ is then $h^2 = V_a / V_a + 1 + \pi^2/3$.

When the residual variance is fixed to 1, the variance of the additive genetic effect must remain small. So, using a prior distribution on V_a with a long tail, such as the Inverse Gamma distribution (V=1, nu=0.002 in MCMCglmm), means that most probabilistic weight is placed on $h^2 = 1$. Various studies have recommended the use of an X^2 *distribution* with 1 degree of freedom prior for binary data (Hadfield 2010, de Villemereuil et al. 2013), mostly because the dispersion of estimates is larger with other common priors and they result in low precision. One of the main issues with this prior is a downward bias of heritability estimate when using a small sample size e.g. n= < 1000 (de Villemereuil et al. 2013), but in general the strength of a prior tends to fade with increasing sample size and with a sufficient sample size the prior issue can become unimportant (de Villemereuil 2012). Due to the large sample size in this study it can be confidently assumed as the best option for prior distribution.

To increase the confidence in the prior distribution In this study, a variety of prior distributions were tested to determine which would provide the best fit for this data (Appendix III: Figures III, IV and V). I also compared the heritability estimates obtained by MCMCglmm using each prior distribution (Appendix II: Table II). The priors tested were; 1) the *Inverse Gamma distribution* (V=1, nu=0.002), 2) *Expanded F distribution* (V=1, nu=1, alpha.mu=0, alpha.V=10), and 3) X^2 *distribution* with 1 degree of freedom (m (V=1, nu=1000, alpha.mu=0, alpha.V=1). The results showed little prior sensitivity in the heritability estimate, but the posterior distribution of variance components indicates that the X^2 with 1 degree of freedom is the best fit for the data.

Autocorrelation in MCMCglmm

It is important to check the convergence and autocorrelation when using the MCMC algorithm. The burn-in period in MCMCglmm allows you to avoid the influences of the starting values and wait for model convergence before saving the iteration values (Sorensen and Gianola 2002), but it is impossible to know how long this period should be before testing it (de Villemereuil 2012). The thinning interval is used to avoid autocorrelation, and also to lighten memory usage (Doligez et al. 2011). Successive iterations of MCMC tend to be correlated with each other, because the nature of the algorithm means that the estimation of a new value for a parameter at each stage is based on the current estimated value of the parameter from the previous iteration (de Villemereuil 2012). To get a large enough effective sample size, you must have a large uncorrelated sample where all iterations are independent, and an effective sample size of at least 1000 is recommended (de Villemereuil 2012). To ensure they are independent you can choose to save only one iteration value in every 10 iterations for example (de Villemereuil 2012). The number chosen should depend on the total number of iterations in the model, to make sure that the sample size is large enough, and will be different for each model.
The ideal chain length (i.e the total number of iterations) will also change depending on a variety of factors, and there is no simple way to determine how long it should be (Sorensen and Gianola 2002). If there are not enough iterations, the saved values may not be representative of the whole distribution, so to an extent it is desirable to run the model for as long as possible.

Various model lengths and burn-in periods were tested for this study and the chosen model, with 1 million iterations, using a burn-in period of 10,000 iterations, and a thinning interval of 100 (i.e. 10,000 iterations were used in total) (Appendix III: Figure IV) was found to provide the best convergence and autocorrelation results. A model with 5 million iterations was tested, but although the autocorrelation and convergence was better the sample size was much lower, and the heritability estimates did not vary greatly (Appendix II: Table III). Since the time and memory needed for a model this much longer are significantly higher, and many models were run the shorter model was chosen for further analyses.

Heritability on the data scale and the latent scale

Another interesting aspect of using GLMMs to estimate heritability is that they provide estimates of quantitative genetic parameters on the latent model scale. The latent scale is convenient for statistical analysis, but obtaining estimates on the scale upon which the traits were measured can also be desirable and helpful for making clear biological inferences from the estimates (de Villemereuil et al. 2016). Because the assumption of the infinitesimal model of quantitative genetics results in a normally distributed genetic component, models require that some component is normally distributed (de Villemereuil et al. 2018). GLMM's create this normally distributed trait, called the latent trait, upon which to estimate variance components (de Villemereuil et al. 2018) GLMMs include ways of dealing with the difficulties that arise in estimating additive genetic variance in non-normal traits, such as the fact that they are inherently non-additive on the scale on which they are measured and they have many complex sources of variation (de Villemereuil et al. 2016). The latent scale in a GLMM assumes additive effects on trait expression, and uses a link function to relate expected values for a trait to the latent scale (de Villemereuil et al. 2016). Current methods for making evolutionary predictions, such as the Lande equation (Lande 1979), and the breeder's equation (Heywood 2005) rely on parameter estimates coming from normally distributed traits, so for traits that are non-normally distributed, which are traits most commonly used in GLMMs, it is necessary to make estimates on a latent scale. The latent scale in a GLMM meets the assumptions of these equations for predictions of evolution, so in general it is preferable to use these estimates when investigating evolutionary potential (de Villemereuil et al. 2016). However researchers also naturally seek to make inferences on the scale upon which the data is observed, because this is the scale on which we see selection acting (de Villemereuil et al. 2016). The major difference between estimates on the latent and data scale is that the latent scale creates a Gaussian distribution, where as in this case the distribution on the data scale is binomial. In addition, absolute values range from negative to positive, which is impossible on the data scale. Therefore both the sign and range of values differ on the latent scale from the observed and expected data scales (de Villemereuil et al. 2016). So the interpretations of heritability based on the expected phenotype and the observed phenotype will be different and deciding on which scale to estimate heritability can depend on what biological question is being asked (de Villemereuil et al. 2016).

Bayesian vs. Frequentist methods of estimating variance components

The development of Bayesian approaches to quantitative genetic variance component estimation, an approach that combines prior information and observed data to draw statistical inference (Visscher et al. 2008), is expected to provide more efficient analysis and heritability estimates (Hadfield et al. 2006). Various comparisons of heritability estimates using animal models with frequentist and bayesian methods have indicated that MCMC is a more effective method when looking at binary traits (de Villemereuil et al. 2013), although it is not always completely clear which method is the best, and understanding their behaviour particularly for binary data needs more investigation (Charmantier et al. 2014). De Villemereuil et al. 2013, found through their comparison studies that parent-offspring regressions and frequentists animal models had much higher dispersion of estimates compared to the animal model using MCMC. With small sample sizes MCMC tends to lose its advantage and become much more biased, and does not estimate heritability as accurately when the heritability is low (de Villemereuil et al. 2013), but this was not considered a significant problem for this study as the sample size used here is considered large for a wild population, and heritability was consistently high. De Villemereuil et al. 2013 recommend MCMC as the preferred method, particularly for binary data, because it has limited inaccuracy and high precision.

II: Supplementary Tables

Table I. Flanking genes for SNPs associated with a significance 'peak' on RepeatABEL GWAS. Shown are the markers (SNPs) and location (chromosome), the two flanking genes (if SNP is intergenic) or containing gene (if SNP is intronic). Functions of the genes based on UniProt are listed. Only marker SNPa105045 was significantly associated with dispersal, but other listed SNPs are the top SNPs found on each observed 'peak' on GWAS Manhatten plot (Figure 4).

SNP	CHR	Flanking genes	Distance (kbp)	Analogous Gene	Function
SNPa105045	15	IV00_00043860 IV00_00043864	11 15	UPB1 in Pongo abelii ADORA2A in homo sapiens	Amino acid biosynthesis G protein-coupled receptor activity.
SNPa16999	12	ID=IV00_00039264 ID=IV00_00039280	205 214	SUCLG2 in Columbia livia FAM19A4 in Macaca fascicularis	Protein coding gene. Carbon metabolism and citric acid cycle Modulates injury-induced and chemical pain hypersensitivity
SNPa14838	12	ID=IV00_00039019	Intronic	QRICH1: Glutamine-rich protein 1 (Bos taurus)	Regulation of cell morphogenesis, and transcription of RNA polymerase II.
SNPa17197	12	ID=IV00_00039291	Intronic	UBA3: NEDD8-activating enzyme E1 catalytic subunit (Homo sapiens)	Down-regulates steroid receptor activity. Necessary for cell cycle progression.
SNPa344637	2	ID=IV00_00000830 ID=IV00_00000857	180 589	FAM92A1: Protein FAM92A1 (Bos taurus) Runx1t1: Protein CBFA2T1 (Mus musculus)	Protein folding. Calcium ion binding Acts as a negative regulator of adipogenesis. Fat cell differentiation.

Table II. Heritability estimates obtained in MCMCglmm using three different prior distributions. Heritability estimates are on the latent scale. All models are run with 1e+06 iterations, a burn-in period of 10000, and a thinning interval of 100.

Prior distribution	h^2_{latent} **estimate**
Inverse Gamma	0.59
Expanded F distribution	0.57
X^2 **with 1 degree of freedom**	0.53

Table III. Comparison of model parameters to assess best model for use in this study. Shown are the parameters for each model (nitt = total number of iterations, burn = burn-in period, thin = thinning interval), the effective sample size and the autocorrelation for each factor. Autocorrelation values should be around 0.1 or lower ideally. Models are done with slightly different data than the final model used in the study, so they are meant to compare to each other not to the final model.

Model parameters	**Effective sample size**	**Autocorrelation values**	**Analysis/** h^2_{latent} **estimate**
Nitt: 1e+06 Burn: 10000 Thin: 100	2401	Year: 0.8928722 Island: 0.001714489 Units: 0.80016605	$h^2_{latent} = 0.51$
Nitt: 1e+06 Burn: 10000 Thin: 5000	198	Year: 0.108928322 Island: 0.04650537 Units: -0.12995491	$h^2_{latent} = 0.47$
Nitt: 5e+06 Burn: 10000 Thin: 5000	990	Year: -0.02693241 Island: 0.035357695 Units: 0.01839180	$h^2_{latent} = 0.49$

III: Supplementary Figures

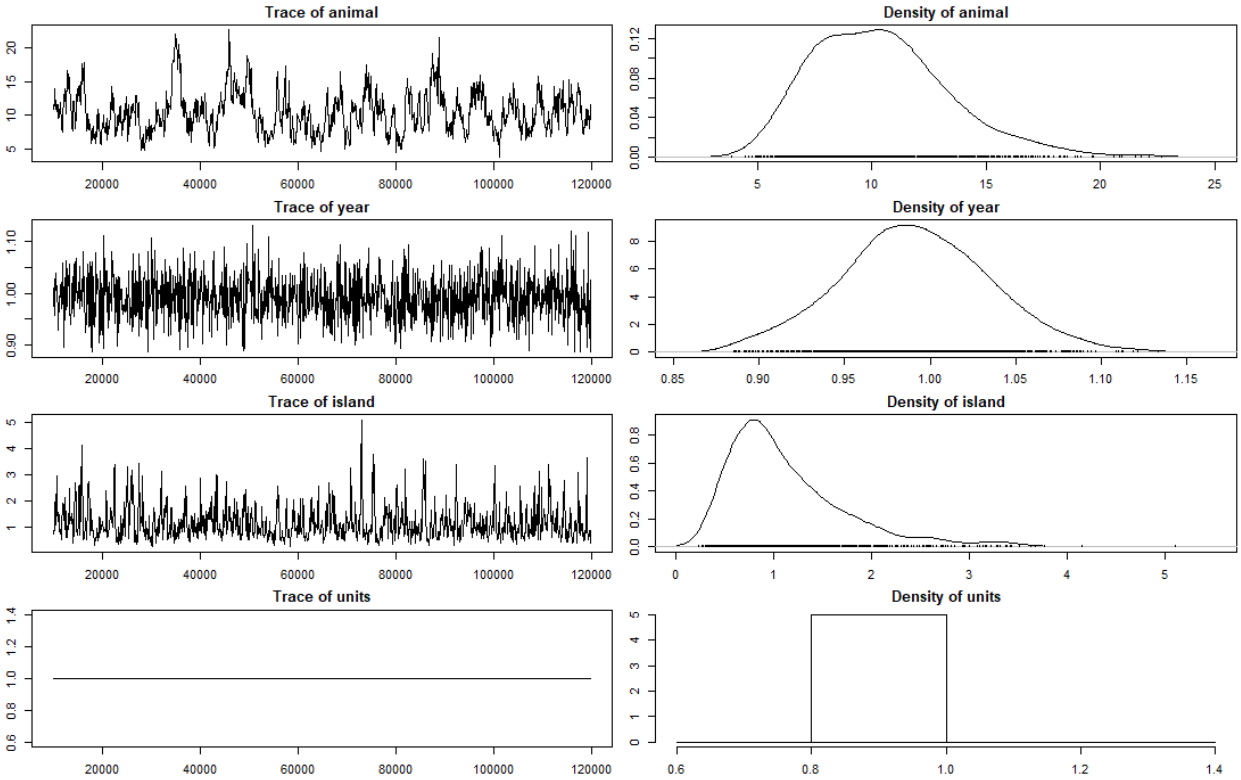

Figure I. MCMCglmm model with default parameters.
modelbinary3.1.1 <- MCMCglmm(disp ~ 1, random = ~animal + year + island, family = "categorical", prior = prior, pedigree = pedigree2, data = sparrow2 , nitt = 120000, burnin = 10000, thin = 100).

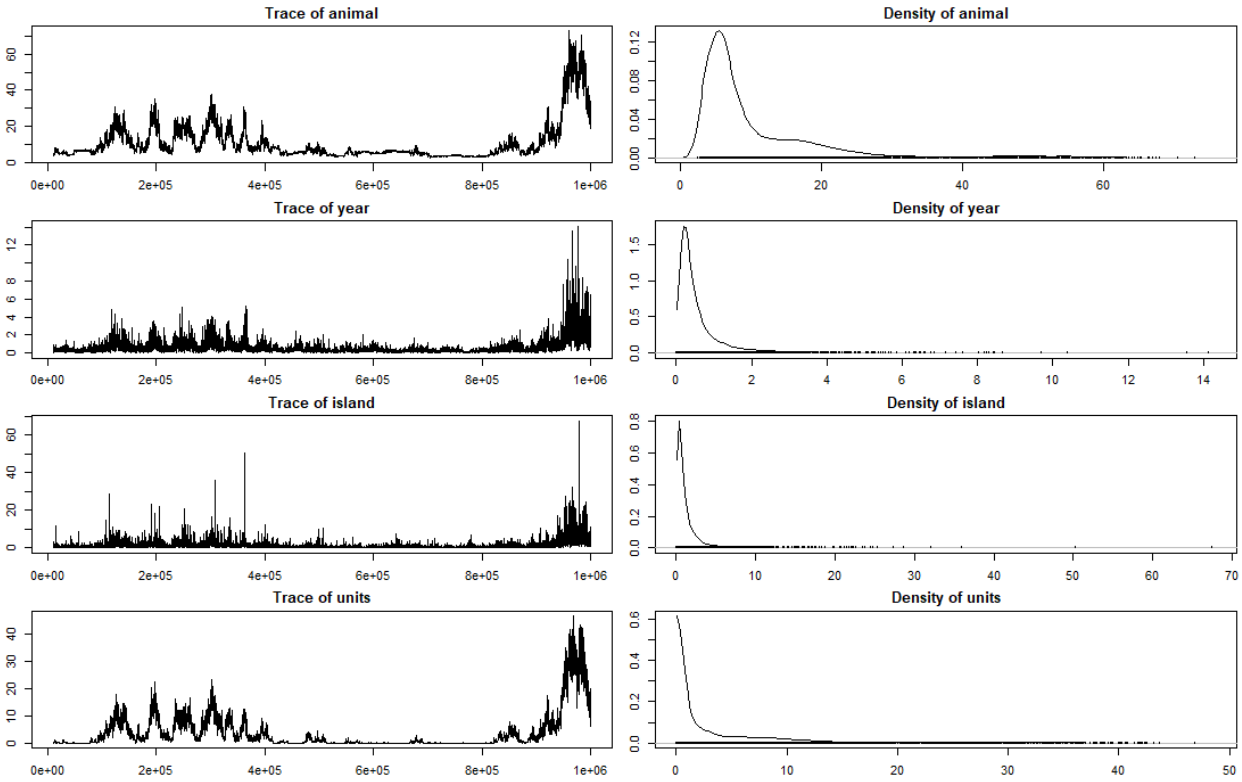

Figure II: Trace of the variances using an Inverse Gamma distribution prior
modelbinary5.3 **<-** MCMCglmm(disp ~ 1, random = ~animal + year + island, family = "categorical", prior = priorInvG, pedigree = pedigree2, data = sparrow4 , nitt = 1e+06, burnin = 10000, thin = 100)
Prior: *priorInvG* <- list(R = list(V=1, nu=0.002), G = list(G1 = list(V=1, nu=0.002), G2 = list(V=1, nu=0.002), G3=list(V=1, nu=0.002)))

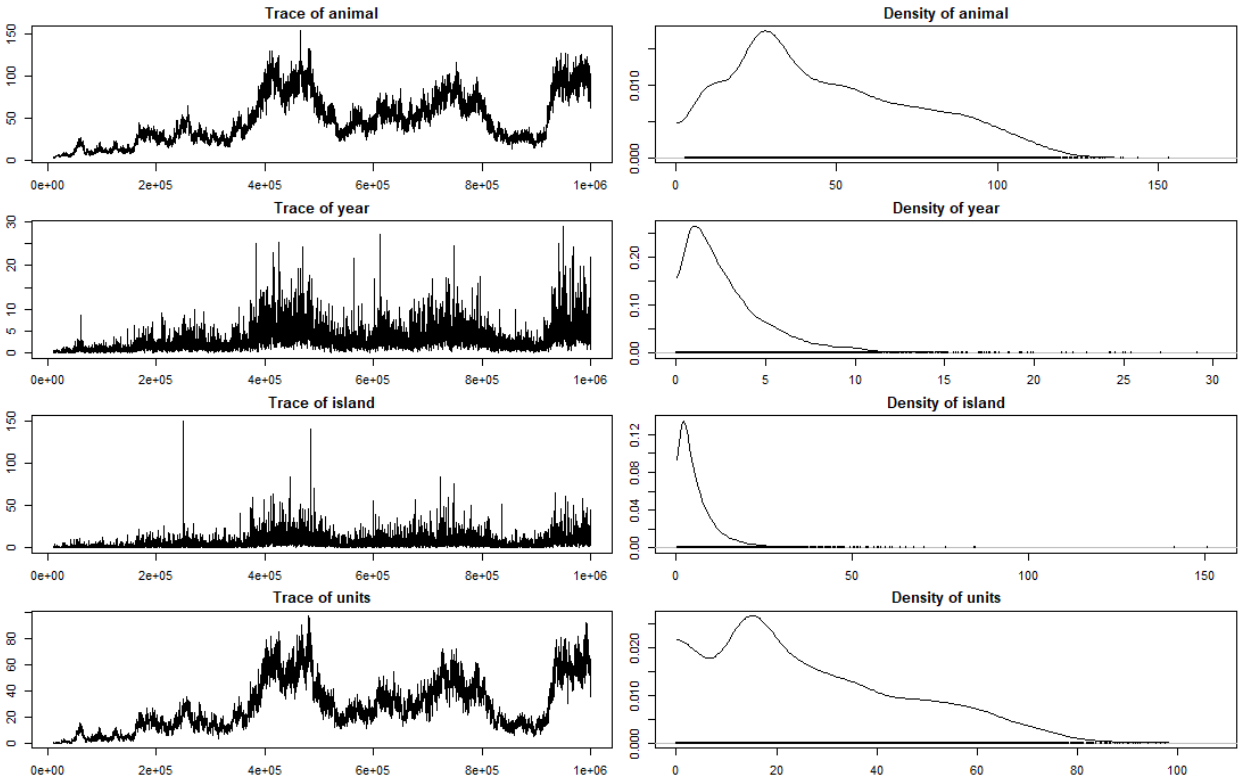

Figure III: Trace of the variances using a parameter expanded F distribution prior.
modelbinary5.4 <- MCMCglmm(disp ~ 1, random = ~animal + year + island, family = "categorical", prior = priorExpF, pedigree = pedigree2, data = sparrow4 , nitt = 1e+06, burnin = 10000, thin = 100)
Prior: priorExpF <- list(G = list(G1 = list(V = 1, nu = 1, alpha.mu = 0, alpha.V = 10), G2 = list(V = 1, nu = 1, alpha.mu = 0, alpha.V = 10), G3=list(V=1, nu=1, alpha.mu=0,alpha.V=10)), R = list(V = 1, nu = 0.002))

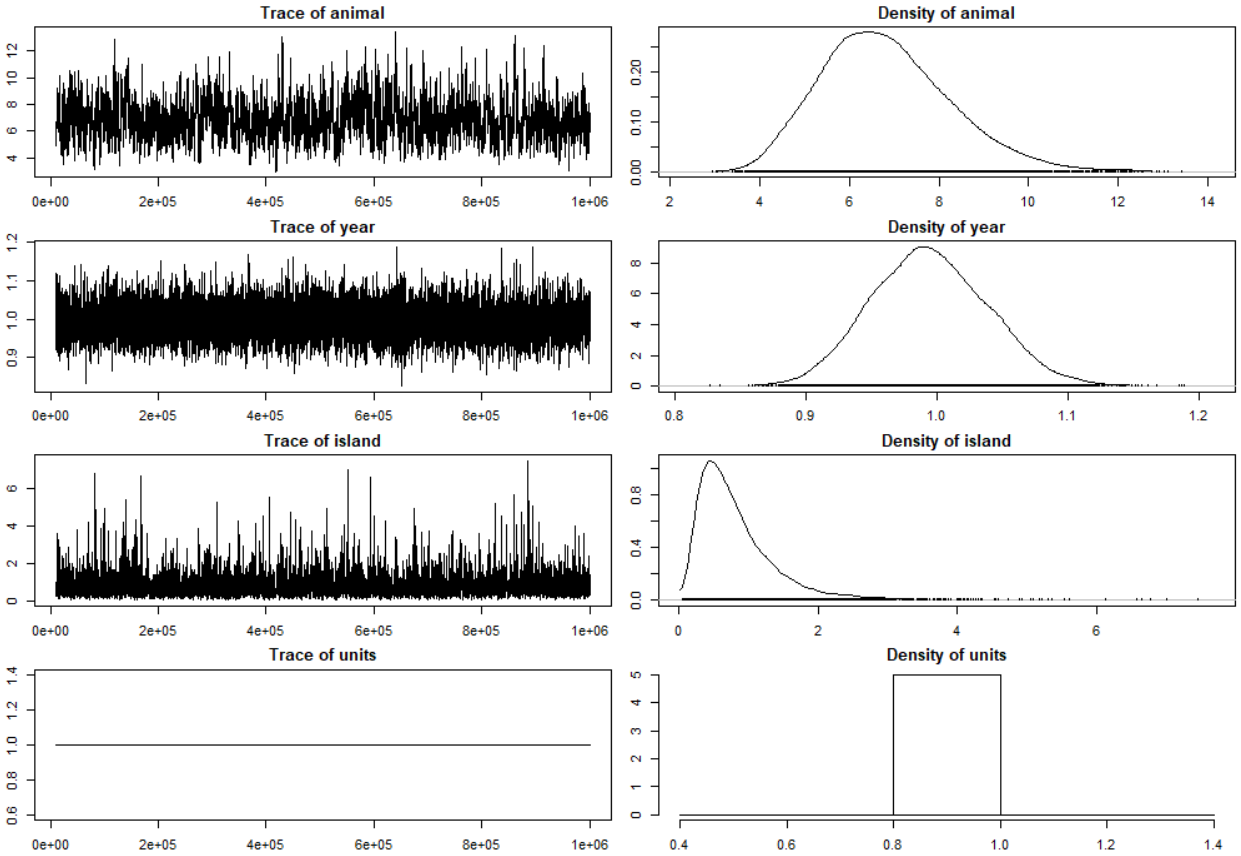

Figure IV. Trace of the variances using X2with 1 degree of freedom prior, and data used for heritability estimate. modelbinary3.1.7 <- MCMCglmm(disp ~ 1, random = ~animal + year + island, family = "categorical", prior = prior, pedigree = pedigree2, data = sparrow4 , nitt = 1e+06, burnin = 10000, thin = 100)
prior <- list(R = list(V = 1, fix = 1), G = list(G1 = list(V = 1, nu = 1000, alpha.mu = 0, alpha.V = 1), G2 = list(V = 1, nu = 1000, aplha.mu = 0, alpha.v = 1), G3 = list(V = 1, nu = 1000, alpha.mu = 0, alpha.V = 1)))