

**Helge Bjerck**

**Institute of Biology**

**The genetic basis of variation in clutch size and onset of laying in a wild population of  
House Sparrows (*Passer domesticus*)**

## **Abstract**

Understanding the basis of genetic variation is a general goal in evolutionary biology. Towards this goal, the architectures of the standing genetic variation in two life history traits, clutch size and onset of laying, were investigated in an insular house sparrow population off the coast of Northern Norway. A newly developed 200K SNP-array was used for genome-wide association mapping (GWAS) and genome partitioning. No genome-wide significant or suggestive loci were detected for either trait. Estimated heritabilities were 0.10 and 0.06 for clutch size and onset of laying, respectively. While genome partitioning was able to show that clutch size was polygenic, differences between subpopulations confounded attempts to characterize lay date. This study highlights the dangers of performing GWAS across diverging subpopulations.

## **Introduction**

Clutch size and onset of laying are quantitative life history traits closely linked to fitness in birds. Indeed, as a measure of fecundity, clutch size comprises one of the primary components of fitness (Roff 2002), and offspring survival is closely related to onset of laying in many species (Verhulst & Nilsson 2008). However, across avian species, both of these traits vary considerably. This variation is found not only between species, but between and within populations as well (e.g. Lack 1947; Monaghan & Nager 1997; Lambrechts & Dias 1993; Noordwijk et al. 1981). Ever since Lack's seminal paper noting this variation, attempts to model the forces that structure it have been frequent in scientific literature (e.g. Cody 1966, Lack 1968, Winkler and Wallin 1987, Rowe et al. 1994).

We expect birds to maximize their clutch size while minimizing the cost to their residual reproductive value (Pianka & Parker 1975). This would include a trade-off with both egg size (Smith & Fretwell 1974), later clutches (Pianka & Parker 1975), and fledgling survival. Similarly, onset of laying is ideally timed such that hatching of chicks coincides with the

maximum availability of insect food (Arcese & Smith 1988). As a result, timing of laying is expected to be dependent on factors that accurately predict when this maximum will be. It has been shown experimentally that lay date can be artificially manipulated by providing extra food or by changing the photoperiod (Arcese & Smith 1988; Lambrechts et al. 1997). The optimality of a given phenotype is dependent on the host of factors that determine its fitness. However, these factors can be variable through time and space and organisms may find them difficult to predict by simple heuristics.

Indeed, it has been shown that differences in onset of laying between subpopulations can have variable fitness consequences through time (Ringsby et al. 1998; Ringsby et al. 2002). This can be one explanation for the existence of substantial genetic variation for these types of plastic traits, even though they are under apparent strong selection (Charmantier & Gienapp 2013).

The heritability of clutch size tends to be significant in avian species, ranging between 0.21 to 0.48, demonstrating that there is a genetic component to variation in clutch size (Postma & Noordwijk 2005). A significant heritability has also been shown for onset of laying, e.g. Noordwijk et al. 1981; Merila & Sheldon 2000.

It is clear that in order to properly understand the variation in these traits, it is necessary to integrate the knowledge about both their ultimate causation and their proximate causation, i.e. the work of evolutionary ecologists and the work of physiologists (see Tinbergen 1963, Visser et al. 2010). As such, mapping the standing genetic variation (SGV) within its ecological context represents an important step in bridging these disciplines.

Additionally, we are interested in how this genetic variation is structured. In general, we expect that SGV of traits under selection will have architectures consisting of many genes of small effect, i.e. a polygenic architecture, as selection will quickly fix genes of large-effect size. However, theoretical work has shown that the relative strength of other evolutionary

forces has an effect on the number, effect size, and clustering of the genes involved in the variation (e.g. Yeaman & Whitlock 2011). Specifically, when adaptive divergence of traits between populations is muffled by migration and drift, we expect these traits to have an oligogenic architecture, i.e. few genes of large effect size, and we expect that these genes will be clustered together (Yeaman & Whitlock 2011).

The goal of this study, then, was to analyze the genetic basis of variation in both clutch size and onset of laying in a wild House Sparrow population. We examined how much of the phenotypic variation is due to genetic effects, estimated whether the genetic basis is oligogenic or polygenic, and attempted to locate genes of significant effect size in the genome. This was accomplished using two complementary methodologies: a genome-wide association study (GWAS) and genome partitioning.

With the ongoing revolution in molecular genetics, the cost of performing higher power quantitative genetic analyses with molecular data is decreasing at an incredible rate. This allows for the relatively quick and easy construction of Single Nucleotide Polymorphism (SNP) arrays and the proliferation of genome-wide association studies in non-model organisms (Jensen et al. 2014). In turn, these have allowed for unprecedented opportunities to address this major question in biology: what is the exact nature of the causal connection between genotype and phenotype (Schielzeth & Husby 2014)?

In short, a GWAS is a method for associating phenotypic traits with genetic variants of known genomic location. It exploits linkage disequilibrium (LD) between genetic markers and the causal genes themselves. If genes coding for some quantitative trait are in LD with some genetic marker, then individuals who possess a particular allele (or alleles) at this genetic marker should score higher for the trait (Visscher et al. 2012). This requires that SGV for this trait exists in the study population and that a sufficient number of genetic markers are available (Schielzeth & Husby 2014). In the past, linkage mapping was generally done using crosses and

hence utilizing recent LD. However, SNP-typing and high density SNP-arrays enable far greater marker densities. As a result, it is now possible to utilize historical population-wide LD; this grants genome-wide association studies greater resolution and power than linkage mapping studies (Risch & Merikangas 1996, Mackay et al. 2009).

Genome partitioning is a multi-marker method that allows us to partition the additive genetic variance into each chromosome (Yang et al. 2011a; Yang et al. 2011b). If the trait of interest is polygenic, then we expect that causal loci are of small effect size and spread throughout the genome. Then, the variance in the trait explained by a genomic region should be roughly proportional to its size. If this expectation is not met, then there are likely genomic regions with one or more QTL's which, together, have a substantial effect size (Slate 2013, Yang et al. 2011b).

To date, the majority of gene mapping studies have focused on model organisms, generally humans or domesticated animals (Jensen et al. 2014). For many study systems, the challenges of genomic studies in the wild were insurmountable. Until recently, genetic studies in natural vertebrate populations have required extensive pedigree information, requiring long-term studies. Linkage mapping could then be carried out using information from relatives in pedigrees (e.g. Slate et al. 2002; Tarka et al. 2010; Johnston et al. 2010, see also review in Slate et al. 2010). However, the large sample sizes needed to draw robust conclusions from studies of genetic architecture may be difficult to generate in the wild and environmental heterogeneity can confound efforts to properly separate variance components. Nevertheless, studying organisms in the wild is important as it allows us to view fitness-related traits in their environmental context (Ellegren & Sheldon 2008). The relationship between phenotype and environment is often complex in unknown ways and laboratory studies have the potential to miss this complexity by not considering factors that affect the selection or expression of traits (Kruuk et al. 2014).

In the past few years, though, several studies of natural vertebrate populations have endeavored to utilize GWAS approaches to characterize variation in morphology and life history. For example, Johnston et al. (2014) successfully identified a locus connected to time to sexual maturation in wild salmon (*Salmo salar*) using a GWAS with 5568 SNP's. Similarly, long-term data from a wild population of Soay sheep (*Ovis aries*), where individuals were typed on a 50K SNP-chip, was used in a GWAS to identify a locus explaining variation in horn morphology (Johnston et al. 2011, 2013). Furthermore, Husby et al. discovered at least one locus associated with clutch size in a collared flycatcher population using a 50K SNP-chip (2015). Two studies on great tits using a 10K SNP-chip did not detect any genome-wide significant loci connected to clutch size or egg mass, but were able to show that these traits were polygenic through genome partitioning (Santure et al. 2013; 2015).

In light of this, we see that the insular house sparrow (*Passer domesticus*) population off the coast of Northern Norway offers a unique opportunity for studying quantitative genetics in the wild. Here, extensive individual-based data has been collected continuously on a number of islands since 1993, allowing for the generation of large sample sizes. Pedigree data has been collected for each individual, along with a slew of morphological and life history traits (e.g. Ringsby et al. 2002, 2009; Jensen et al. 2003, 2004, 2008; Husby et al. 2006; Pärn et al. 2009; Kvalnes et al. 2013). Each island represents a discrete sub-population within a metapopulation, where gene flow, genetic drift, and selection can all be measured (Holand et al. 2011; Jensen et al. 2013; Baalsrud et al. 2014). Further, much genomic groundwork has recently been done with this species using data collected in this project. A 10K SNP-chip was produced and was shown to be effective at, for instance, detecting sub-populations correlating to different islands in the study system (Hagen et al. 2013). The whole genome has been sequenced and assembled based on a linkage map (Elgvin et al. in prep; Hagen et al. ms). These resources were then used to create a custom 200K SNP-array (Hagen et al. in prep).

## **Materials & Methods**

### *Study Population*

The house sparrow is a small non-migratory passerine bird species. They are reproductively mature in their first breeding season after birth and generally attempt to reproduce in this season (Anderson 2006). They generally produce 1-3 clutches per season and the breeding season lasts from early May to mid August (Ringsby et al. 1998). Seasonal variation in clutch size has been documented, generally increasing initially and then decreasing through successive clutches in the season (Westneat et al. 2014, Seel 1968). The incubation period lasts around 11 days and the nestling period lasts around 15 days (Seel 1968).

Also, the House Sparrow is known for living commensally with humans (Anderson 2006). Our study system consists of a collection of 18 islands off the coast of Northern Norway in the region of Helgeland (Figure 1). Here, the sparrows are found almost exclusively on dairy farms and in residential areas (Husby et al. 2006). This predictability, along with the small size of the islands, allows us to track the populations with relative ease. As a result, this metapopulation has been tracked continuously since 1993.

Mark-recapture rates of adult birds show that 80-100% of the individuals in our study system are banded (Pärn et al. 2009, 2012, but see also Ringsby et al. 1998, 1999 and Holand et al. 2015, 2016). Furthermore, the house sparrow is a sedentary species; only a relatively small proportion of recruits are dispersers (Pärn et al. 2009, 2012), and the dispersal distances are short (Tufto et al. 2005; Pärn et al. 2009, 2012) relative to the extensive size of the study area (ca. 1600 km<sup>2</sup>). As a consequence, the majority of the individuals in the study system have been tracked from hatching to death, such that we generally have data on several clutches per individual (Ringsby et al. 1999, Pärn et al. 2009, Sæther et al. 1999, Billing et al. 2012, Kvalnes et al. 2013). Additionally, this means that accurate estimates of fitness components are

available. This data has been used to investigate the connection between fitness and a host of traits including dispersal rates (Pärn et al. 2009), brood sex ratio (Husby et al. 2006), egg mass (Kvalnes et al. 2013), and morphology (Jensen et al. 2004, 2008).

Specifically, it has been shown that the fitness consequences of onset of laying vary substantially between years (Ringsby et al. 2002, Ringsby et al. 1998). Significant differences between islands in onset of laying, along with autocorrelated daily weather conditions, produces variable selection between years with different weather conditions, as the fitness of a lay date is dependent on these weather conditions (Ringsby et al. 2002).

### *Phenotypic Data Collection*

Each year, fieldwork was carried out through the breeding season. About every 7-10 days, each island is thoroughly searched for active nest sites, such that each brood is visited at least 2-3 times. Clutch size is measured as the maximum number of eggs, nestlings, and/or dead nestlings recorded throughout the breeding attempt (Husby et al. 2006). Onset of laying for each female individual was defined as the date at which the first egg was laid in the female's first clutch of the breeding season. This date is determined, given the clutch size, simply by extrapolating backwards from either hatching date or the average laying rate of one egg per day (Kvalnes et al. 2013). This date is then transformed into a numerical variable by counting the number of days into the year, i.e. number of days since December 31<sup>st</sup>.

When nestlings are 8-12 days old, a blood sample is taken and each bird is uniquely banded with one numbered aluminum ring and three colored plastic rings. Adults and fledged juveniles are captured with mist nets throughout the summer and also in the autumn. At capture any unbanded birds are banded, they are measured for a host of morphological traits, and a blood sample is taken. All procedures follow an established field protocol and measurements are subsequently corrected by field worker (see Jensen et al. 2003). Observational data



supplements this, as the unique bands can be visible at a distance, meaning that recapture is not necessary in order to track the individuals in the population.

The blood samples collected from adults and nestlings are then used to genotype all individuals on 14 highly variable microsatellite markers. These are then compared in order to determine the genetic parentage of each brood (see e.g. Jensen et al. 2008; Billing et al. 2012; Baalsrud et al. 2014). Additionally, due to high nest site fidelity, it is generally possible to extrapolate maternity to all clutches in a nest within a breeding season, such that unhatched clutches can be assigned mothers (Kvalnes et al. 2013).

Clutches that did not hatch and had a clutch size less than three were not included in the analysis of clutch size as it is reasonable to suspect that the majority of these were incomplete clutches. Similarly, first lay dates after the 15<sup>th</sup> of July were removed from the analysis of lay date as these are likely second or third clutches where the first clutch was simply not observed (Kvalnes et al. 2013). See appendix for a supplementary table including the descriptive statistics of clutch size and onset of laying.

### *Genotyping*

The individuals in this study were genotyped on a custom Affymetrix Axiom 200k SNP-array. Analysis of linkage disequilibrium decay of the earlier 10K SNP-chip (Hagen et al. 2013) revealed that approximately 200,000 SNP's would be required to guarantee that a QTL would be in linkage disequilibrium with a marker, regardless of the QTL's location (Husby, Hagen, Jensen; personal communication). As a result, we expect that there are few genomic regions wherein a significant QTL could remain undetected.

The SNP's for the 200K array were identified through 10X whole-genome Illumina re-sequencing of 33 house sparrows from 15 populations spread throughout northern Scandinavia. These genomes were then aligned with the reference genome (Elgvin et al. ms) and SNP's with

little variation in their flanking sequences were chosen (Hagen et al. ms). In general, SNP's are evenly spaced with an average of 6 kbp between them, though about 15,000 SNP's are concentrated within a host of candidate genes (Hagen et al. ms).

The data for this study comes from the seven islands with the highest mark-recapture rates in our study system (Figure 1). On five of the islands (Nesøy, Gjerøy, Hestmannøy, Indre Kvarøy, and Aldra), sparrows from 1998 until present were genotyped. On the other two (Træna and Selvær), sparrows from 2003 until present were genotyped (Table 1). There is a discrepancy in the time series used as the “outer” islands of Træna and Selvær experienced a population crash around 2000 and did not fully recover until 2003 (Baalsrud et al. 2014). Altogether, 3247 adult individuals were genotyped from these islands. These constitute almost all adult house sparrows present on these islands during these periods.

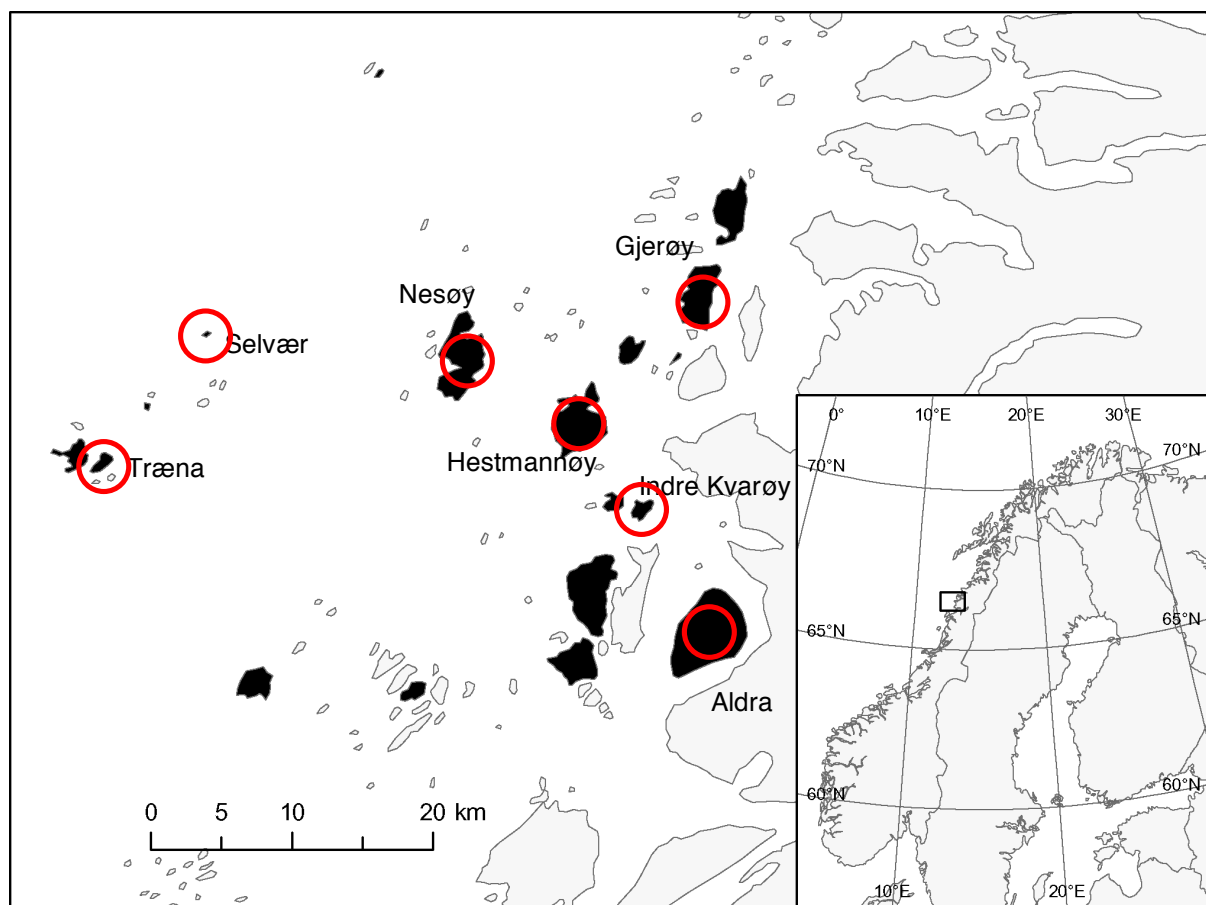


Figure 1: Map of islands included in the long-term house sparrow study off the coast of Northern Norway. Islands in black are island in the study system that have been continuously followed. Populations used in this study are circled in red.

Markers not among the 186,056 SNP's that were ranked as Polymorphic High Resolution, Affymetrix's highest quality class, were not used for any analyses. Also, markers with a call-rate less than 95% and a minor allele frequency less than 0.01 were removed. An identity-by-state matrix was constructed for the entire population based on the remaining markers where average kinship values of a pair are below the diagonal, numbers of SNP's typed for a pair are above the diagonal, and homozygosity of an individual is on the diagonal. Principal components analysis was then performed on this matrix and six clusters of individuals were delineated based on the first four principal components. Markers that were significantly out of Hardy-Weinberg equilibrium in all six clusters were removed, where the significance threshold was Bonferroni-corrected such that the genome-wide significance threshold equaled  $2.84 \times 10^{-7}$ . After this quality control, 184,409 SNP's were available for further analysis.

Similarly, 66 individuals were removed where IBS between any pair exceeded 0.95, 46 individuals were removed where call rate was less than 0.95, and 192 individuals were removed where heterozygosity at the Z chromosome exceeded or fell under the expected rate given the sex of the individual. Work is on-going to examine in greater detail why so many individuals appear to have been incorrectly sexed. Together, 2808 individuals passed this quality control, and of these, 620 had adequate clutch size data and 607 had adequate data on laying date (Table 1).

Table 1: Overview of the number of SNP-genotyped individuals coming from each island and the period during which data from these islands will be included. This gives a rough idea of the relative importance of each island in the analyses. CS=Clutch Size, LD=Lay Date.

<b>Island</b>	<b>Years</b>	<b># of individuals genotyped</b>	<b># of ind. w/ CS data</b>	<b># of ind. w/ LD data</b>
<b>Nesøy</b>	1998-2013	161	13	13
<b>Gjerøy</b>	1998-2013	657	123	123
<b>Hestmannøy</b>	1998-2013	1100	255	250
<b>Indre Kvarøy</b>	1998-2013	378	77	73
<b>Aldra</b>	1998-2013	211	30	27
<b>Træna</b>	2003-2013	466	65	65
<b>Selvær</b>	2003-2013	274	57	56
	Total =	3247	620	607

### *Model Building*

Preliminary model building for clutch size showed that the best mixed model with the smallest AIC value included the fixed effects island, age, lay date, clutch number, and a clutch number by lay date interaction, where lay date was mean centered within island (see also Westneat et al. 2009). Similar model building for onset of laying showed that the best model included island, year, and age effects, where the age of individuals was simply classified as first year or older as per Noordwijk, Balen, & Scharloo (1981). Both models included ID as random effects. These models were constructed using the R package “lme4” (Bates 2010).

While inclusion of tarsus length as a proxy for body size did decrease the AIC values for both traits, the decision was made to not include it in the final analyses as there were several individuals with missing morphological data. As a result, inclusion of tarsus length would necessitate a loss in sample size outweighing the benefits. Climatic variables were not considered, though early results from a parallel study indicate that these have little effect on laying date (Ringsby, Johansen, personal communication, 2016).

### *Genome-Wide Association Study (GWAS)*

The genome-wide association studies were done using two R packages: GenABEL and RepeatABEL. GenABEL includes functions that allow for the efficient handling of the large datasets inherent to SNP-data (Aulchenko et al. 2007). GRAMMAR-gamma, a function within GenABEL, can be used for generating p-values and effect sizes for each SNP sans repeated measurements (Svischeva et al. 2012). However, GRAMMAR-gamma only accepts one value per individual, the mean lifetime clutch size and the onset of laying for an individual's first year was used. RepeatABEL allows for the analysis of data with repeated measurements (Husby et al. 2015; Rönnegård et al. 2016). As we have data on several clutches for many individuals, the bulk of the analysis was done using RepeatABEL so as to maximize our power. RepeatABEL uses the GRM to control for relatedness between individuals.

P-values are corrected for genomic inflation in order to correct for population stratification. Significance and suggestive thresholds were Bonferroni-corrected such that the genome-wide significance level corresponds to  $p=2.71 \times 10^{-7}$  and the suggestive level corresponds to  $p=5.42 \times 10^{-6}$ .

### *Genome Partitioning*

Genome partitioning is performed using the software Genome-wide Complex Traits Analysis (GCTA) (Yang et al. 2011a, Yang et al. 2013). As GCTA only accepts one value per individual, the mean lifetime clutch size and the onset of laying for an individual's first year was used.

The methodology used here is similar to the one described in Yang et al. (2011b), where a genetic relationship matrix (GRM) is constructed for each genomic region. Here, the genomic regions used are autosomes. These GRM's are then adjusted in order to control for the imperfect LD between causal sites and tagged markers. Then, all GRM's are fitted

simultaneously in a mixed effects model and variance components are estimated using AI-REML (Gilmour, Thompson, & Cullis 1995). In the case of model converge problems, the chromosome with the fewest SNP's was successively removed until the model converged. As a result, the partitioning of mean lifetime clutch size included 19 chromosomes and the partitioning of lay date included 27 chromosomes. No fixed factors were used in these analysis as inclusion of additional covariates only led to convergence problems, necessitating the removal of more chromosomes.

### *Heritability*

Estimates of heritability of our traits come from three different sources. The three methods used in this study (GRAMMAR-gamma, RepeatABEL, and GCTA) all function through the estimation of variance components. They can therefore be used to calculate a trait's additive genetic variance and, in turn, the narrow-sense heritability. However, as GRAMMAR-gamma and GCTA only accept one value per individual, these two ignore within-individual variation. As a consequence, many of the fixed effects accounted for in the GWAS no longer have any meaning, e.g. age and year effects. As the fixed effects influence the phenotypic variance in the model, this must be accounted for when comparing heritability estimates (Wilson 2008).

For our purposes, the estimates of heritability from RepeatABEL are the most relevant. These best reflect the proportions of additive genetic and environmental variance actually viewed by our GWA studies. However, note that traditional estimates of heritability are derived from estimates of relatedness based on a pedigree. Here, relatedness is calculated using the identity-by-state at all markers. As a result, we aren't reliant on the accuracy of a pedigree, nor is noise introduced by the variance around the pedigree-predicted relatedness due to recombination and segregation (Berenos et al. 2014).

## Results

No loci were genome-wide significant at the Bonferroni-corrected significance level in the GWA studies on mean lifetime clutch size, clutch size, or onset of laying; nor were there any suggestive loci in any of these analyses (see appendix for a summary of the top ten SNP's in each analysis) (Figure 2).

Partitioning the variance by chromosome using GCTA generally corroborated these results. For clutch size, the correlation between chromosome size and proportion of variance explained was non-significant (Figure 3,  $r^2=0.06$ ,  $p=0.32$ ). However, this was largely driven by chromosome 20. After removing this from the analysis, there is a clear correlation between the size of the chromosome and the proportion of variance explained ( $r^2=0.28$ ,  $p=0.02$ ). These results suggest that clutch size is a polygenic trait. Moreover, as no SNP's on chromosome 20 seemed salient in the GWAS, the results seem to indicate that a relatively large proportion of loci of small effect size may be clustered on this chromosome.

For onset of laying, the correlation between chromosome size and proportion of variance explained was non-significant (Figure 3,  $r^2=0.01$ ,  $p=0.59$ ). Chromosome 4 explained far more than any other chromosome, although removing this chromosome did not improve the correlation (Figure 3). This outlier is due to the stratification in our data due to the between-island differences in lay date; these differences were not controlled for in the genome partitioning (see Appendix). These results suggest that onset of laying has an oligogenic architecture.

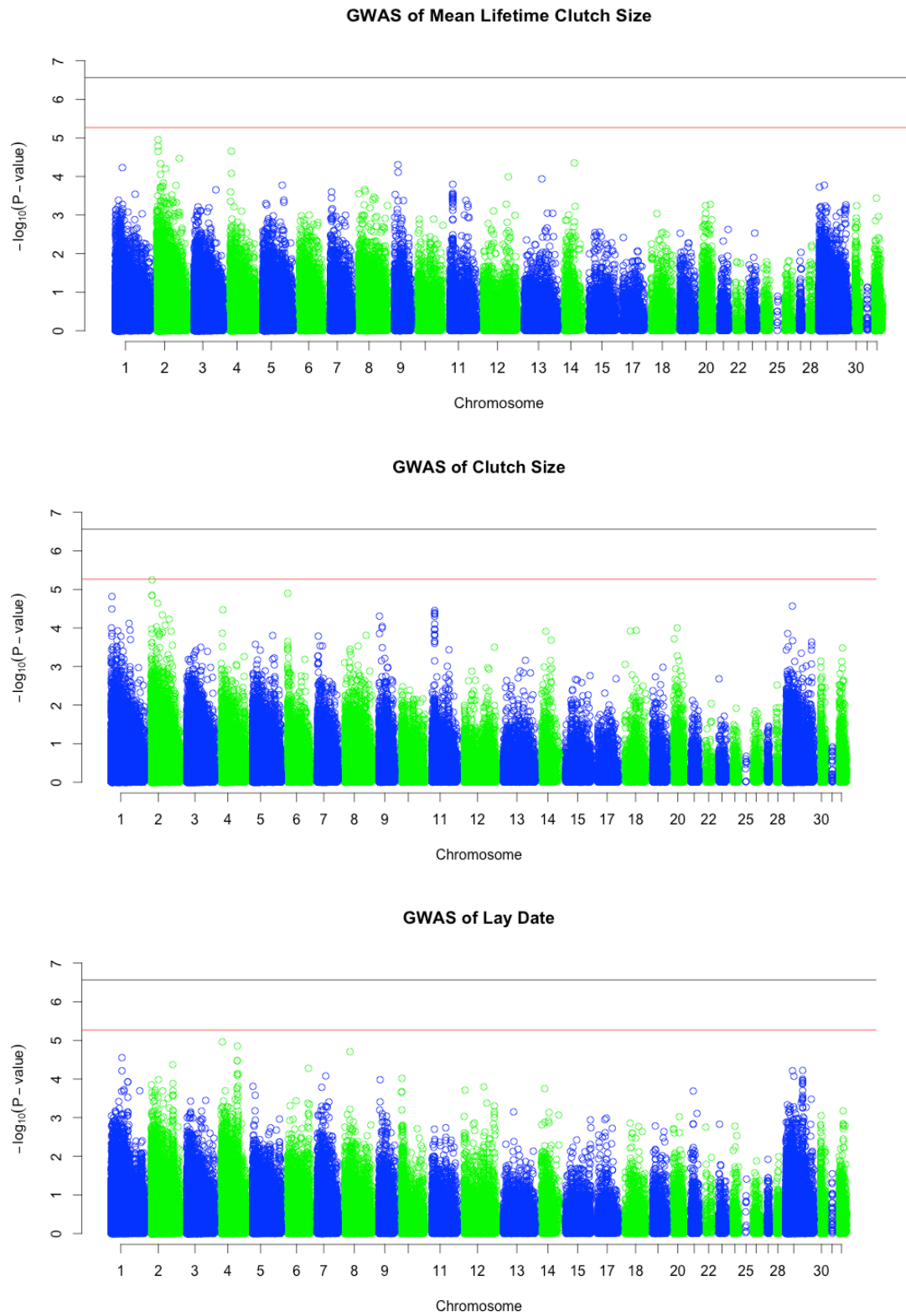


Figure 2: Manhattan plots of GWAS. Relative positions between markers are only correct within scaffolds.



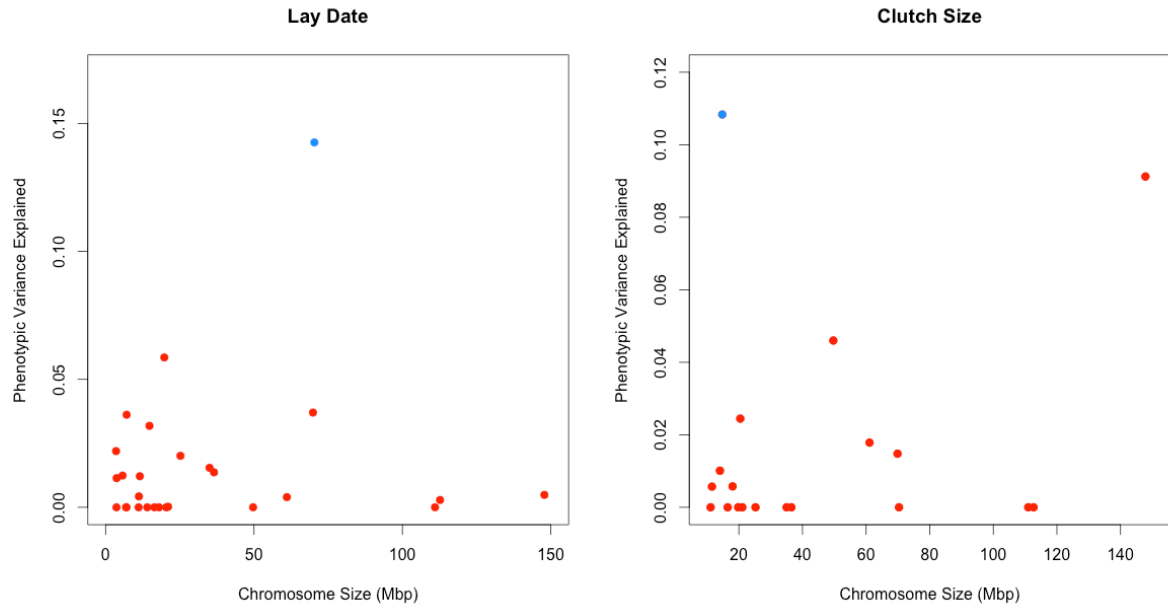


Figure 3: Chromosome partitioning of lay date and clutch size. The point in blue refers to chromosome 4 and chromosome 20 in the plot of lay date and clutch size, respectively

The heritability of mean lifetime clutch size was 0.14 in GRAMMAR-gamma and 0.20 in GCTA. The RepeatABEL estimates of heritability from the GWAS was 0.10 for clutch size and 0.07 for lay date. GCTA estimated the heritability of lay date as 0.30.

Table 2: Units of clutch size are number of eggs and units of lay date are number of days

### Variance Components of RepeatABEL GWAS

Source	Clutch Size	Lay Date
Additive Genetic Variance	0.120	22.500
Permanent Environmental Variance	0.090	19.500
Residual Variance	0.950	303.600

## Discussion

The genetic architectures of two life history traits, clutch size and onset of laying, were investigated using two complementary methods: GWAS and genome partitioning.

For clutch size, it's polygenic nature is shown by the absence of significant hits in our GWAS (Figure 3), despite our relatively high marker density. Polygenicity can also be seen in the correlation between chromosome size and variance explained, although, the collection of markers on chromosome 20 explain more variance than expected given the size of the chromosome. In combination, the GWAS and chromosome partitioning results indicate that the genes controlling the variation in clutch size, while many and of small-effect, are not spread evenly throughout the genome.

These findings are generally corroborated by previous research. In Rowe's modelling work, the primary factor determining clutch size for each female is her initial condition; females that begin the season with more energy will have a larger clutch size (Rowe 1994). As condition is expected to be highly polygenic, we expect clutch size to be highly polygenic as well. Also, genome-wide association studies on clutch size in two populations of great tit (*Parus major*) found no significant loci, and genome partitioning of variance showed that the amount of variance explained by a chromosome was correlated with its length (Santure et al. 2013; 2015). This implies that clutch size is a polygenic trait. Conversely, a GWAS on the collared flycatcher found one significant loci that explained about 28% of the additive genetic variance, indicating that relatively few genes may structure the genetic variation (Husby et al. 2015).

However, note that the great tit studies had a sample size of ~1000 and used ~6,000 SNP's (Santure et al. 2013; 2015) while the collared flycatcher study had a sample size of 313 and used ~37,000 SNP's (Husby et al. 2015). As a result, the great tit study may have missed loci not in linkage disequilibrium with their relatively few markers, and, indeed, they conclude

that they have low power to detect loci of a significant effect size (Santure et al. 2015). The collared flycatcher study, though it benefitted from using repeated values and a higher density of SNP's, suffers from a small sample size and the authors note that their estimated effect sizes are likely overestimated due to the Beavis effect (Husby et al. 2015). Then, the present study may have the best combination of sample size and number of markers out of these studies. As a consequence, we may have the best power out of available studies. However, a proper power analysis would be necessary in order to decisively conclude this.

For lay date, our results are inconclusive. The confounding effects of subpopulation structure (see Appendix) are too great to make robust conclusions. While no significant or suggestive SNP's were detected in the GWAS (Figure 2), the chromosome partitioning shows a clear oligogenic structure (Figure 3). These results contradict each other. However, this makes sense as the former analysis controlled for mean differences between islands and the latter did not.

There seems to be a dearth of similar investigations into the genetic basis of lay date, though there are several studies establishing that the trait has a significant heritability (e.g. McCleery et al. 2004, Cooke & Findlay 1982, Noordwijk et al. 1981) and a few studies have focused on connecting the phenology of lay date with genetic variation at a candidate gene, *Clock* (e.g. Liedvogel et al. 2012; Leder et al. 2006).

In general, the observed heritabilities of both our traits are lower than those seen in other studies (e.g. Sheldon et al. 2003; Noordwijk et al. 1981; Postma & Noordwijk 2005). There are a myriad number of methodological and biological potential reasons for this. This discrepancy might stem from established issues with heritability as a statistic, e.g. it is dependent on the fixed effects (Wilson 2008), and it is dependent on the amount of environmental variance (Postma 2014). It may also be because our study uses repeated values, such that within-individual variation is taken into account. Additionally, the house sparrow is

a multi-brooded species meaning that these traits may be more plastic than in other single-brooded species. However, the most salient differences between the present study and previous studies is that genomic relatedness is directly calculated with the markers rather than estimated through a pedigree. This has been shown to produce lower estimates of genetic variance (Berenos et al. 2014).

These findings call for follow-up studies. In order to make more general conclusions about the genetic architecture of these traits, studies with larger sample sizes will be needed (Mackay et al. 2009). Additionally, further research should focus on characterizing the environmental variables connected to variation in these life history traits. Currently, the majority of the variation remains unexplained, and this limits our ability to detect loci (Mackay et al. 2009). Specifically, the substantial differences between islands in onset of laying must be properly characterized. Perhaps a structured association would cast more light on this problem.

In conclusion, this study shows that it is difficult to gain a full understanding of the genetic basis of a complex trait in a wild population even with a large number of markers. The success of this approach remains highly dependent on access to quality ecological and phenotypic data. While the metapopulation structure can be seen as this study system's greatest strength, it clearly poses a host of methodological problems as well.

## **Bibliography**

- Anderson, T. R. (2006). Biology of the ubiquitous house sparrow.
- Arcese, P., & Smith, J. N. (1988). Effects of population density and supplemental food on reproduction in song sparrows. *The Journal of Animal Ecology*, 119-136.
- 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. E., Hagen, I. J., Myhre, A. M., Ringsby, T. H., Pärn, H., & 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.

Bérénos, C., Ellis, P. A., Pilkington, J. G., & Pemberton, J. M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular ecology*, 23(14), 3434-3451.

Billing, A. M., Lee, A. M., Skjelseth, S., BORG, Å. A., Hale, M. C., Slate, J. O. N., ... & Jensen, H. (2012). Evidence of inbreeding depression but not inbreeding avoidance in a natural house sparrow population. *Molecular ecology*, 21(6), 1487-1499.

Charmantier, A., & Gienapp, P. (2014). Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. *Evolutionary Applications*, 7(1), 15-28.

Cody, M. L. (1966). A general theory of clutch size. *Evolution*, 174-184.

Cooke, F., & Findlay, C. S. (1982). Polygenic variation and stabilizing selection in a wild population of lesser snow geese (*Anser caerulescens caerulescens*). *The American Naturalist*, 120(4), 543-547.

Elgvin, T.O., Trier, C.N., Tørresen, O.K., Hagen, I.J., Lien, S., Alexander, S.L., Nederbragt, A.J., Ravinet, M., Jensen, H., & Sætre, G. (in prep) Genomics of homoploid hybrid speciation

Ellegren, H., & Sheldon, B. C. (2008). Genetic basis of fitness differences in natural populations. *Nature*, 452(7184), 169-175.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., & Thompson, R. (2009). ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Hagen, I. J., Billing, A. M., Rønning, B., Pedersen, S. A., Pärn, H., Slate, J., & Jensen, H. (2013). The easy road to genome-wide medium density SNP screening in a non-model species: development and application of a 10 K SNP-chip for the house sparrow (*Passer domesticus*). *Molecular ecology resources*, 13(3), 429-439.

Hagen, I.J., Lien, S., Billing, A.M., Tarka, M., Elgvin, T.O., Trier, C.N., Sætre, G. & Jensen, H. (in prep). A genome wide, medium density linkage map for the house sparrow, *Passer domesticus*.

Hagen, I.J., Kemppainen, P., Billing, A.M., Lien, S., Elgvin, T.O., Trier, C.N., Sætre, G., Husby, A., & Jensen, H. (in prep). Effects of population demography on linkage disequilibrium in the house sparrow (*Passer domesticus*) genome.

- Holand, A. M., Jensen, H., Tufto, J., & Moe, R. (2011). Does selection or genetic drift explain geographic differentiation of morphological characters in house sparrows *Passer domesticus*?. *Genetics research*, 93(05), 367-379.
- Husby, A., Kawakami, T., Rönnegård, L., Smeds, L., Ellegren, H., & Qvarnström, A. (2015, May). Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. In *Proc. R. Soc. B* (Vol. 282, No. 1806, p. 20150156). The Royal Society.
- 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.
- Jensen, H., Moe, R., Hagen, I. J., Holand, A. M., Kekkonen, J., Tufto, J., & Sæther, B. E. (2013). Genetic variation and structure of house sparrow populations: is there an island effect?. *Molecular Ecology*, 22(7), 1792-1805.
- 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(6), 1275-1293.
- Jensen, H., Szulkin, M., Slate, J. (2014). Molecular quantitative genetics. Charmantier, A., Garant, D., & Kruuk, L. E. (Eds.). *Quantitative Genetics in the Wild* (pp. 209-227). Oxford University Press.
- Jensen, H., Sæther, B. E., Ringsby, T. H., Tufto, J., Griffith, S. C., & Ellegren, H. (2004). Lifetime reproductive success in relation to morphology in the house sparrow *Passer domesticus*. *Journal of Animal Ecology*, 73(4), 599-611.
- 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*). *Journal of evolutionary biology*, 16(6), 1296-1307.
- Johnston, S. E., Beraldi, D., McRae, A. F., Pemberton, J. M., & Slate, J. (2010). Horn type and horn length genes map to the same chromosomal region in Soay sheep. *Heredity*, 104(2), 196-205.
- Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, 502(7469), 93-95.

Johnston, S. E., McEwan, J., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., ... & Slate, J. O. N. (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(12), 2555-2566.

Johnston, S. E., Orell, P., Pritchard, V. L., Kent, M. P., Lien, S., Niemelä, E., ... & Primmer, C. R. (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular ecology*, 23(14), 3452-3468.

Kruuk, L. E. (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.

Kruuk, L. E., Charmantier, A., Garant, D. (2014). The study of quantitative genetics in wild populations. Charmantier, A., Garant, D., & Kruuk, L. E. (Eds.). *Quantitative Genetics in the Wild* (pp. 1-15). Oxford University Press.

Kruuk, L. E., Slate, J., & Wilson, A. J. (2008). New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annual Review of Ecology, Evolution, and Systematics*, 39, 525-548.

Kvalnes, T., Ringsby, T. H., Jensen, H., & Sæther, B. E. (2013). Correlates of egg size variation in a population of house sparrow *Passer domesticus*. *Oecologia*, 171(2), 391-402.

Lack, D. (1947). The significance of clutch size. *Ibis*, 89(2), 302-352.

Lack, D. (1968). *Ecological Adaptations for Breeding in Birds*. Chapman and Hall, London.

Lambrechts, M. M., Blondel, J., Maistre, M., & Perret, P. (1997). A single response mechanism is responsible for evolutionary adaptive variation in a bird's laying date. *Proceedings of the National Academy of Sciences*, 94(10), 5153-5155.

Lambrechts, M. M., & Dias, P. C. (1993). Differences in the onset of laying between island and mainland Mediterranean Blue Tits *Parus caeruleus*: phenotypic plasticity or genetic differences?. *Ibis*, 135(4), 451-455.

Leder, E. H., Danzmann, R. G., & Ferguson, M. M. (2006). The candidate gene, *Clock*, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *Journal of Heredity*, 97(1), 74-80.

- Liedvogel, M., Cornwallis, C. K., & Sheldon, B. C. (2012). Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. *Journal of evolutionary biology*, 25(5), 813-823.
- Mackay, T. F., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, 10(8), 565-577.
- McCleery, R. H., Pettifor, R. A., Armbruster, P., Meyer, K., Sheldon, B. C., & Perrins, C. M. (2004). Components of variance underlying fitness in a natural population of the great tit *Parus major*. *The American Naturalist*, 164(3), E62-E72.
- Merilä, J., & Sheldon, B. C. (2000). Lifetime reproductive success and heritability in nature. *The American Naturalist*, 155(3), 301-310.
- Monaghan, P., & Nager, R. G. (1997). Why don't birds lay more eggs?. *Trends in Ecology & Evolution*, 12(7), 270-274.
- Van Noordwijk, A. J., van Balen, J. V., & Scharloo, W. (1981). Genetic variation in the timing of reproduction in the great tit. *Oecologia*, 49(2), 158-166.
- Pärn, H., Jensen, H., Ringsby, T. H., & Sæther, B. E. (2009). Sex-specific fitness correlates of dispersal in a house sparrow metapopulation. *Journal of Animal Ecology*, 78(6), 1216-1225.
- Pärn, H., Ringsby, T. H., 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.
- Pianka, E. R., & Parker, W. S. (1975). Age-specific reproductive tactics. *American Naturalist*, 453-464.
- Postma, E., & van Noordwijk, A. J. (2005). Genetic variation for clutch size in natural populations of birds from a reaction norm perspective. *Ecology*, 86(9), 2344-2357.
- Postma E. (2014). Molecular quantitative genetics. Charmantier, A., Garant, D., & Kruuk, L. E. (Eds.). *Quantitative Genetics in the Wild* (pp. 16-33). Oxford University Press.
- Ringsby, T. H., Sæther, B. E., & Solberg, E. J. (1998). Factors affecting juvenile survival in house sparrow *Passer domesticus*. *Journal of Avian Biology*, 241-247.



Ringsby, T. H., Saether, B. E., Tufto, J., Jensen, H., & 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(15), 3963-3980.

Roff, D. A. (2002). Life history evolution (Vol. 7). Sunderland: Sinauer Associates.

Rowe, L., Ludwig, D., & Schluter, D. (1994). Time, condition, and the seasonal decline of avian clutch size. *American Naturalist*, 698-722.

Rönnegård, L., McFarlane, E. S., Husby, A., Kawakami, T., Ellegren, H., & Qvarnström, A. (2016). Increasing the power of genome wide association studies in natural populations using repeated measures-evaluation and implementation. *Methods in Ecology and Evolution*.

Santure, A. W., Cauwer, I., Robinson, M. R., Poissant, J., Sheldon, B. C., & Slate, J. (2013). Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. *Molecular ecology*, 22(15), 3949-3962.

Santure, A. W., Poissant, J., De Cauwer, I., Oers, K., Robinson, M. R., Quinn, J. L., ... & Slate, J. (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*.

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., & Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, 11(s1), 1-8.

Seel, D. C. (1968). Clutch Size, incubation, and hatching success in the house sparrow and tree sparrow *Passer spp.* at Oxford. *Ibis*, 110(3), 270-282.

Sheldon, B. C., Kruuk, L. E. B., & Merila, J. (2003). Natural selection and inheritance of breeding time and clutch size in the collared flycatcher. *Evolution*, 57(2), 406-420.

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.

Slate, J., Visscher, P. M., MacGregor, S., Stevens, D., Tate, M. L., & Pemberton, J. M. (2002). A genome scan for quantitative trait loci in a wild population of red deer (*Cervus elaphus*). *Genetics*, 162(4), 1863-1873.

Slate, J., Santure, A. W., Feulner, P. G., Brown, E. A., Ball, A. D., Johnston, S. E., & Gratten, J. (2010). Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics*, 26(6), 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.

Smith, Christopher C., and Stephen D. Fretwell. "The optimal balance between size and number of offspring." *American Naturalist* (1974): 499-506.

Summers-Smith, D. (1988). *The sparrows*. A&C Black.

Svishcheva, G. R., Axenovich, T. I., Belonogova, N. M., van Duijn, C. M., & Aulchenko, Y. S. (2012). Rapid variance components-based method for whole-genome association analysis. *Nature genetics*, 44(10), 1166-1170.

Tarka, M., Åkesson, M., Beraldi, D., Hernández-Sánchez, J., Hasselquist, D., Bensch, S., & Hansson, B. (2010). A strong quantitative trait locus for wing length on chromosome 2 in a wild population of great reed warblers. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1692), 2361-2369.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20(4), 410-433.

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. *The American Naturalist*, 165(1), E13-E26.

Verhulst, S., & Nilsson, J. Å. (2008). The timing of birds' breeding seasons: a review of experiments that manipulated timing of breeding. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1490), 399-410.

- Visser, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, 90(1), 7-24.
- Visser, M. E., Caro, S. P., Van Oers, K., Schaper, S. V., & Helm, B. (2010). Phenology, seasonal timing and circannual rhythms: towards a unified framework. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1555), 3113-3127.
- Westneat, D. F., Bókony, V., Burke, T., Chastel, O., Jensen, H., Kvalnes, T., ... & Stewart, I. R. (2014). Multiple aspects of plasticity in clutch size vary among populations of a globally distributed songbird. *Journal of Animal Ecology*.
- Westneat, D. F., Stewart, I. R., & Hatch, M. I. (2009). Complex interactions among temporal variables affect the plasticity of clutch size in a multi-brooded bird. *Ecology*, 90(5), 1162-1174.
- Wilson, A. J. (2008). Why  $h^2$  does not always equal  $VA/VP$ ?. *Journal of evolutionary biology*, 21(3), 647-650.
- Winkler, D. W., & Wallin, K. (1987). Offspring size and number: a life history model linking effort per offspring and total effort. *American Naturalist*, 708-720.
- Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., ... & Visscher, P. M. (2011b). Genome partitioning of genetic variation for complex traits using common SNPs. *Nature genetics*, 43(6), 519-525.
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011a). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1), 76-82.
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2013). Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Genome-wide association studies and genomic prediction*, 215-236.
- Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration–selection balance. *Evolution*, 65(7), 1897-1911.

## Appendix

### *Supplementary Note: Island Effect on Lay Date*

Genetic drift is very important in structuring the genetic variation in these small stochastic island populations (Jensen et al. 2013, Holand et al. 2011). As a result, we expect to see differences between islands regardless of similarities in the adaptive landscape. Indeed, there are marked differences in lay date between islands (e.g. Ringsby et al. 2002).

In order to investigate this, two different analyses were run for onset of laying: one including island as a fixed effect and one without. These two analyses presented drastically different results. Removing island as a fixed effect tripled the additive genetic variance and a marginally increases the permanent environmental variance. Heritability increased to 0.16 from 0.06.

Also, as seen in the Results, no SNP's were genome-wide significant or suggestive when including island as a fixed effect. However, without controlling for the island effect, all of the top six most significant SNP's were located on chromosome 4 and four of these were genome-wide suggestive. The most significant of these was located ~31 Mbp from the rest, which seemed to be clustered around a peak (Figure 5). At this peak, linkage disequilibrium between SNP's was very high, such that these SNP's seemed to be detecting the same effect in the population and were functionally the same. As a result, these four suggestive SNP's were only detecting two loci.

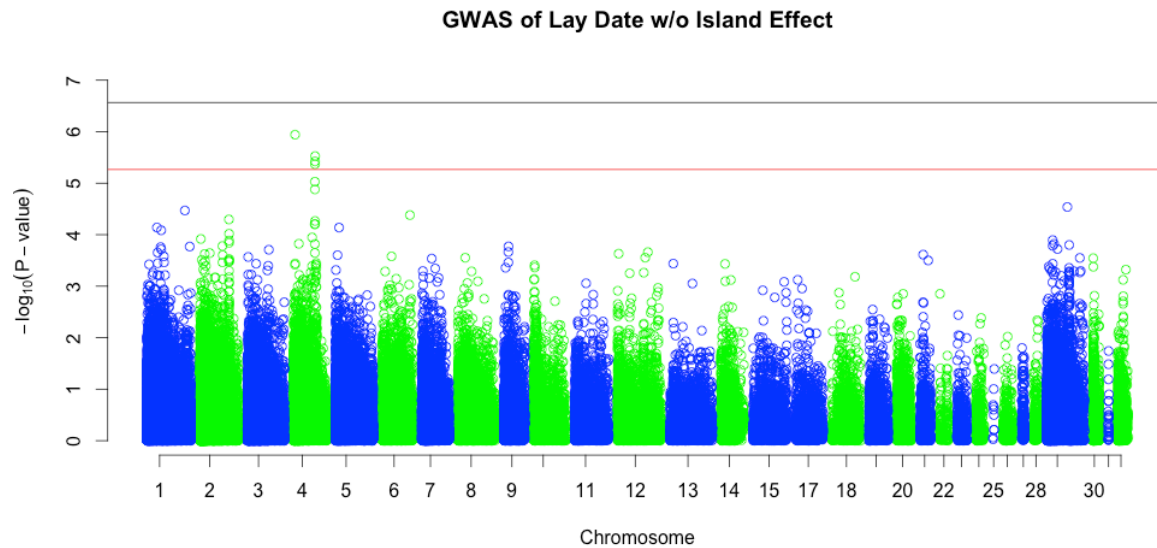


Figure 5: Manhattan plot of GWAS of onset of laying without taking into account differences between island. Relative positions of markers are only correct within scaffolds.

These two loci explained a surprisingly large proportion of the variance in laying date, each accounting for approximately 20% of the additive genetic variation, though only 2-3% of the phenotypic variation. In combination with the non-significant correlation between chromosome size and genetic variance explained, these results support an oligogenic model for this trait.

When we plot the p-values from each analysis against each other (Figure 6), we see that including island systematically increases p-values. Additionally, the most significant SNP's all lie exactly where they would be predicted to be given a linear regression between the two sets of p-values. This indicates that the significance of these SNP's can be almost wholly explained by differences between islands.

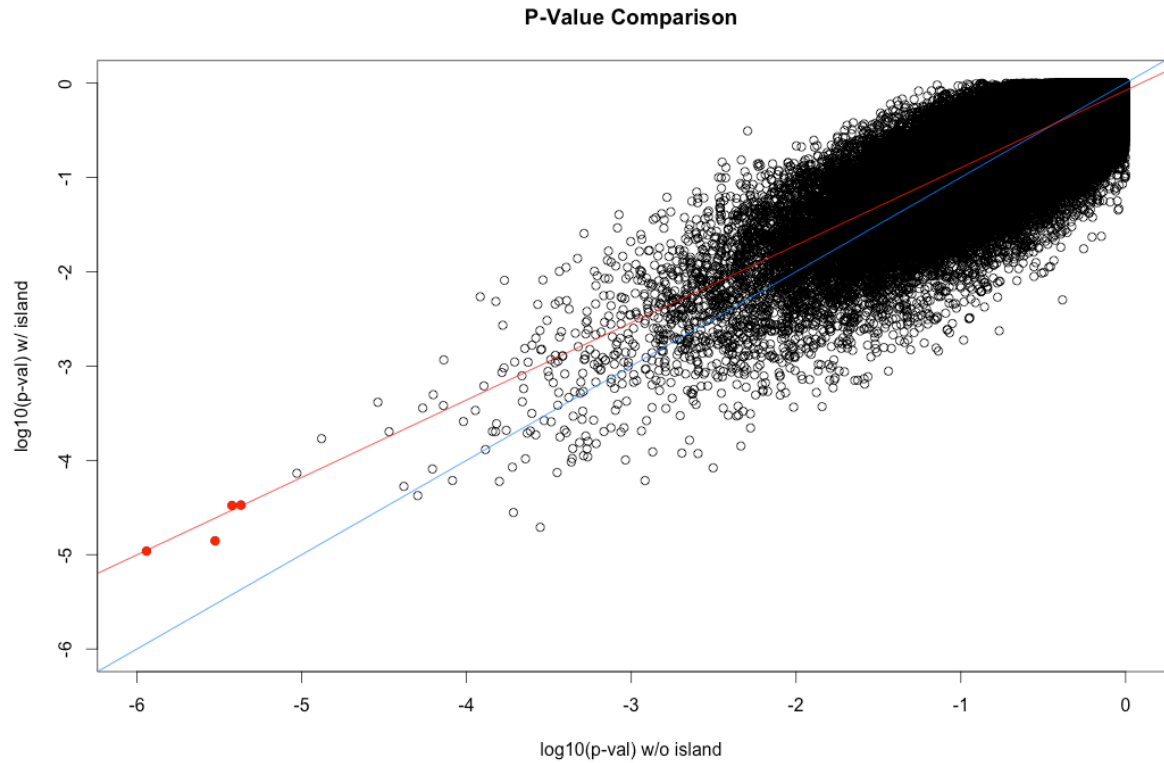


Figure 6: Plot comparing the p-values of the GWA studies of lay date excluding the island effect (abscissa) and including the island effect (ordinate). Here, the SNP's in red are the four suggestive SNP's, all on chromosome 4. The blue line represents equivalence, and the red line represents the linear regression. The linear regression did not include the SNP's in red, such that the effect is not driven by those outliers.

Though RepeatABEL uses the GRM to control for relatedness in the population (Rönnegård et al. 2016), the stratified relatedness produced by the differences between islands pose a problem for this analysis. As the differences between islands were revealed to be primarily genetic, then controlling out this variation may control out exactly the variation we are interested in. On the other hand, the differences in allele frequencies at these loci may be due to random genetic drift within each population. If the variation is best explained at the population level, then our sample size is actually equal to the number of populations and all our individuals merely reflect pseudoreplication. A different approach would be required in order to separate the differences in allele frequencies between populations due to mere random genetic drift and the differences that actually cause differences in lay date.

In any case, BLAST results of these suggestive SNP's against the collared flycatcher and zebra finch genome did not lead to any genes that easily lend themselves to interpretation. While homologues can be found for most of these SNP's in these two genomes, they are generally not near known genes. However, SNPa217703, the most significant SNP for lay date, comes within ~37 kbp of a gene related to Fragile X Syndrome, called AFF1. In the chicken genome, there were no significant hits.

## Supplementary Tables

Table 1: Descriptive statistics of traits of interest

	<i>Clutch Size</i>	<i>Onset of Laying (Days from Dec 31<sup>st</sup>)</i>
<i>Mean</i>	<i>4.85</i>	<i>141.8</i>
<i>Median</i>	<i>5</i>	<i>138</i>
<i>Within Individual Variance</i>	<i>0.19</i>	<i>59.8</i>
<i>Between Individual Variance</i>	<i>1.04</i>	<i>328.7</i>
<i>Total Variance</i>	<i>1.23</i>	<i>387.6</i>
<i># of Individuals</i>	<i>620</i>	<i>607</i>
<i># of Observations</i>	<i>1726</i>	<i>931</i>

## Summary of GRAMMAR-gamma GWAS of Clutch Size, Top 10 SNP's

SNP Names	Chromosome	Major Allele	Minor Allele	Effect Size	SE	P-value
SNPa462400	2	C	A	-0.224	0.051	0.00001
SNPa462413	2	A	C	-0.220	0.051	0.00002
SNPa434592	4	T	C	0.286	0.067	0.00002
SNPa462414	2	C	T	0.215	0.051	0.00002
SNPa44085	2	G	A	-0.221	0.053	0.00003
SNPa513360	14	A	G	0.245	0.060	0.00004
SNPa461196	2	T	C	-0.297	0.073	0.00005
SNPa285571	9	T	G	-0.412	0.102	0.00005
SNPa300393	1	G	A	0.238	0.059	0.0001
SNPa241736	2	G	A	-0.212	0.053	0.0001



### Summary of RepeatABEL GWAS of Lay Date, Top 10 SNP's

SNP Names	Chromosome	Major Allele	Minor Allele	Effect Size	SE	P-value
SNPa217703	4	C	A	-4.556	1.026	0.00001
SNPa123579	4	A	C	4.890	1.115	0.00001
SNPa374248	8	T	C	5.013	1.162	0.00002
SNPa152625	1	T	C	4.007	0.947	0.00003
SNPa123552	4	A	G	4.641	1.107	0.00003
SNPa123542	4	A	C	4.615	1.102	0.00003
SNPa44244	2	A	G	-3.930	0.950	0.00004
SNPa148320	6	A	G	-3.888	0.952	0.0001
SNPa61161	29	A	G	4.273	1.054	0.0001
SNPa251119	29	C	T	-5.546	1.370	0.0001

### Summary of RepeatABEL GWAS of Clutch Size, Top 10 SNP's

SNP Names	Chromosome	Major Allele	Minor Allele	Effect Size	SE	P-value
SNPa462400	2	C	A	-0.217	0.047	0.00001
SNPa493087	6	A	G	0.208	0.047	0.00001
SNPa462414	2	C	T	0.204	0.046	0.00001
SNPa462413	2	A	C	-0.208	0.047	0.00001
SNPa432427	1	A	G	-0.222	0.051	0.00002
SNPa400068	2	G	T	0.269	0.062	0.00002
SNPa57977	29	C	T	0.272	0.064	0.00003
SNPa432398	1	G	A	-0.204	0.048	0.00003
SNPa479653	4	T	G	-0.277	0.066	0.00003
SNPa480691	11	A	G	0.262	0.062	0.00004