Inferences of Genetic Architecture of Bill Morphology in House Sparrow using a High Density SNP Array Point to a Polygenic Basis

Sarah L. Lundregan[1*], Ingerid J. Hagen[1,2], Jostein Gohli[1,3], Alina K. Niskanen[1,4], Petri Kemppainen[1,3], Thor Harald Ringsby[1], Thomas Kvalnes[1], Henrik Pärn[1], Bernt Rønning[1], Håkon Holand[1], Peter S. Ranke[1], Anna S. Båtnes[1], Linn-Karina Selvik[1], Sigbjørn Lien[5], Bernt-Erik Sæther[1], Arild Husby[1, 3, 6§], and Henrik Jensen[1§]

1: Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, Norway

2: Norwegian Institute for Nature Research, P.O. Box 5685 Torgard, NO-7485 Trondheim, Norway

3: Organismal and Evolutionary Biology Research Programme, University of Helsinki, Finland

4: Department of Ecology and Genetics, University of Oulu, Oulu, Finland

5: Centre for Integrative Genetics, Department of Animal and Aquacultural Sciences, Faculty of Life Sciences, Norwegian University of Life Sciences, Ås, Norway

6: Department of Ecology and Genetics, EBC, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden

* Author for correspondence: Sarah.Lundregan@ntnu.no

§ Joint senior authors

Key-words: Ecological genetics, GWAS, beak morphology, bill morphology, heritability, polygenic

Abstract

Understanding the genetic architecture of quantitative traits can provide insights into the mechanisms driving phenotypic evolution. Bill morphology is an ecologically important and phenotypically variable trait, which is highly heritable and closely linked to individual fitness. Thus, bill morphology traits are suitable candidates for gene mapping analyses. Previous studies have revealed several genes that may influence bill morphology, but the similarity of gene and allele effects between species and populations is unknown. Here, we develop a custom 200K SNP array and use it to examine the genetic basis of bill morphology in 1857 house sparrow individuals from a large-scale, island metapopulation off the coast of Northern Norway. We found high genomic heritabilities for bill depth and length, which were comparable with previous pedigree estimates. Candidate gene and genome wide association analyses yielded six significant loci, four of which have previously been associated with craniofacial development. Three of these loci are involved in bone morphogenic protein (BMP) signalling, suggesting a role for BMP genes in regulating bill morphology. However, these loci individually explain a small amount of variance. In combination with results from genome partitioning analyses this indicates that bill morphology is a polygenic trait. Any studies of eco-evolutionary processes in bill morphology are therefore dependent on methods that can accommodate polygenic inheritance of the phenotype and molecular-scale evolution of genetic architecture.

Introduction

Knowledge of the genetic architecture of ecologically important traits in natural populations provides insights into their potential for phenotypic evolution (Lande 1979, Clutton-Brock and Sheldon, 2010). Identification of causal loci improves understanding of the genetic drivers and selective processes leading to adaptation, which for quantitative traits may not always proceed according to conventional models of selection and evolution (Pritchard and Di Rienzo, 2010). In birds, bill morphology is an example of an ecologically important trait, which has substantial phenotypic variation, is typically highly heritable and closely linked to individual fitness (James and Zach, 1979; Merilä *et al*., 2001; Jensen *et al*., 2008; Bosse *et al*., 2017). Bill form and function varies both between and within species in response to a range of ecological factors, including availability of differing food sources and interspecific competition (Abzhanov *et al*., 2006; Riyahi *et al*., 2013; Soons *et al*., 2015). Subtle differences in bill morphology can have important consequences for variation in individual fitness by affecting relative foraging efficiency (Temeles *et al*., 2009), defining diet (Soons *et al*., 2015), altering survival probability (Boag and Grant, 1981; Grant and Grant, 2002) or affecting nestling provisioning rate (Forstmeier *et al*., 2001; Ringsby *et al*., 2009).

Additive genetic variance, the fuel for evolutionary responses to selection (Lande, 1979), is commonly standardised by the phenotypic variance to yield heritability (Falconer and Mackay, 1996). Recent advances in genomics have made estimation of genomic heritabilities feasible (Stapley *et al*., 2010; Gienapp *et al*., 2017). Genomic heritability has the advantage that it can be used where the pedigree is unknown or incomplete and it accounts for cryptic relatedness and variation in identity by descent (IBD) around the pedigree expectation (Visscher *et al*., 2008). Currently, heritability is often estimated using genome-wide marker data and animal models (Kruuk, 2004; Gienapp *et al*., 2017). In addition to the heritability of

a trait, inheritance type (i.e. Mendelian, oligogenic, polygenic) can be determined using genome partitioning and may have important consequences for trait evolutionary dynamics. For instance, for polygenic inheritance, which is the expectation for quantitative traits like bill morphology (Grant and Grant, 1994; Tiffin and Ross-Ibarra, 2014; Santure *et al*., 2015, Bosse *et al*., 2017), adaptation likely occurs by modest allele frequency shifts across many loci and these loci are more likely to differ between populations (Pritchard and Di Rienzo, 2010).

If a trait is heritable, gene mapping may be performed to identify putatively causal loci. Recent technological advances have reduced costs and paved the way for use of genome-wide association study (GWAS) approaches in ecological model species, in particular via large scale genotyping on SNP panels (Stapley *et al*., 2010). With a few notable exceptions (e.g. Johnston *et al*., 2011; Johnston *et al*., 2014; Barson *et al*., 2015; Chaves *et al*., 2016), success rate is low for detection of outlier loci using GWAS on quantitative traits in natural populations and loci detected are often of small effect (e.g. Husby *et al*., 2015; Silva *et al*., 2017; Bosse *et al*., 2017). Such studies are nonetheless important, despite challenges such as environmental heterogeneity and population structure, as they allow us to examine the genetic basis of ecologically important traits in their natural context (Slate *et al*., 2010; Stapley *et al*., 2010). Although quantitative traits are expected to often be under polygenic control, with signals of selection distributed across many loci (Tiffin and Ross-Ibarra, 2014), several large effect genes contributing to bill morphology have been identified in avian species (e.g. Abzhanov *et al*., 2004; Abzhanov *et al*., 2006; Brugmann *et al*., 2010; Lamichhaney *et al*., 2015; Lamichhaney *et al*., 2016; Bosse *et al*., 2017). As polygenic inheritance increases the likelihood that effect sizes of adaptive loci will differ between populations, examining the effects of these genes using a regional candidate gene approach may increase statistical power

to determine their importance for phenotypic variation across different groups (Wilkening *et al*., 2009).

The house sparrow (*Passer domesticus*) is a ubiquitous and extensively studied ecological model species ideally suited to gene mapping studies on bill morphology (Anderson, 2006). A natural house sparrow metapopulation exists at the Helgeland coast in Northern Norway and has been monitored since 1993, allowing establishment of a large sample dataset that includes genetic and pedigree information, alongside morphological and life history data including measurements of bill depth and length (e.g. Jensen *et al*., 2003; Jensen *et al*., 2004; Pärn *et al*., 2009; Ringsby *et al*., 2009). Bill depth and length measures have high repeatability within the Helgeland system, both traits are highly heritable (e.g. Jensen *et al*., 2003; Jensen *et al*., 2008) and bill length has been related to components of individual fitness (Jensen *et al*., 2004; Jensen *et al*., 2008). Larger bills have also been found to increase nestling feeding rate and, consequently, number of successful recruits (Ringsby *et al*., 2009).

Here we examine the genetic architecture of bill morphology in the Helgeland metapopulation of house sparrows using a new, custom 200K Affymetrix Axiom SNP array. First, genomic heritability was estimated to determine whether the SNP markers captured the additive genetic variance previously estimated for bill morphology in the house sparrow. Subsequently, genome partitioning analysis was performed to examine inheritance type and identify outlier chromosomes. The 200K array is enriched for markers in the region of a number of genes, including nine candidate genes that have previously been related to bill morphology in other bird species: ALX1, BMP4, Calm1, TGFbrII, Dkk2, FZD1, FGF8, Shh, and FGF19 (Table 1). Candidate gene association tests were performed to examine whether any of these genes explained variation in bill morphology in the house sparrow. This was

followed by GWA analysis of the whole dataset to determine whether the genome-wide approach would support the findings from candidate gene analyses and if additional regions and candidate genes for future studies could be detected. Significant markers in the Helgeland dataset were subsequently tested for enrichment in an independent dataset from a second set of house sparrow populations situated 75-295 km South of the Helgeland system (hereafter called the 'Southern populations') to determine whether significant associations in the Helgeland dataset were replicated across populations.

Materials and Methods

Study Population and Data Collection

This study utilises data from eight insular subpopulations from a long-term study of a house sparrow metapopulation, in an archipelago off the Helgeland coast in Northern Norway (66°30'N, 12°30'E; Fig. 1). The data includes adult birds recorded between 1998-2013 (Tables 2 and S1). Only adult measurements taken within breeding season (May – August) were used to reduce seasonal intra-individual variation in bill morphology (Anderson, 2006). Each individual's hatch year was determined as either 1) the first year recorded for nestlings or fledged juveniles captured in the summer and autumn, or 2) the year prior to first year recorded for birds first ringed as adults (Jensen *et al*., 2008). Birds were captured using mist nets, blood samples were collected from the brachial vein, and slide callipers were used to measure bill depth (measured directly in front of the nares) and bill length (from the tip of the bill to the base of the skull) to the nearest 0.01mm (Jensen *et al*., 2004; Jensen *et al*., 2008). All birds were ringed with a metal ring with a unique individual ID number and a unique combination of three plastic colour rings. This was done to allow individual identification and give information on e.g. age at subsequent captures. Phenotypic measurements taken by different fieldworkers were adjusted to T.H.R measurements by adding mean differences when found significant ($P < 0.05$) using paired T-tests (Kvalnes *et al*., 2017). To give a single value for each individual, measurements of adult birds (second calendar year (2CY) and older) were adjusted to May in 2CY using a linear mixed effects model with age, age^2 and month as fixed factors, random intercepts for year, cohort and ID number, and a random slope for the effect of age within ID number to account for any between-individual variation. If age and month effects were significant ($P < 0.05$) in likelihood ratio tests, predicted values from the model were used to adjust measurements to May in 2CY before calculation of mean

146 phenotypic values (Kvalnes *et al*., 2017). Unless otherwise stated, all statistical analyses were
147 performed using R version 3.3.3 (R Core Team, 2017).

148

149 *Re-sequencing and SNP array development*

150 The newly developed custom house sparrow Affymetrix 200K SNP array was based on the
151 re-sequencing of 33 individual house sparrows from 13 different locations in Norway (29
152 individuals) and two locations in Finland (four individuals). See Supplementary Material
153 (Tables S2-S4; Figure S1) for more information on geographic locations for the re-sequenced
154 individuals, development of the 200K SNP array, and genotyping of individuals. The 200 000
155 SNPs on the array are distributed across 29 of the chromosomes in the house sparrow
156 genome, with mean and median distances between SNPs being shorter than 5000 bp (Table
157 S4). It is expected that this array will provide an important genomic tool for an avian species
158 with a nearly global distribution and European origin (Anderson, 2006). The array may be
159 used to examine population structure, dispersal, inbreeding, demography, and the genetic
160 basis for phenotypes relevant for physiological, ecological, and evolutionary processes.

161

162 *Quality Control of the Bill Morphology Dataset*

163 Of the 3219 individuals from the Helgeland metapopulation successfully genotyped on the
164 200K array, 1958 had phenotypic data for bill depth and bill length. For the purpose of this
165 study on bill morphology only the 185 587 markers ranked as PolyHigh (polymorphic
166 markers with well-defined clusters, at least two individuals possess the minor allele) or
167 MonoHigh (monomorphic markers with a single well-defined cluster) resolution by CIGENE
168 were used in further analysis. Mendelian error checking was performed in PLINK (version

1.9, Purcell *et al*., 2007) using previously constructed microsatellite-based pedigrees (Jensen *et al*., 2008; Billing *et al*., 2012). First, parental links including more than 5% Mendelian errors were removed from the pedigree, then loci with more than 10% Mendelian errors were excluded from the dataset. Quality control for individuals with bill data was performed in GenABEL (Aulchenko *et al*., 2007). Individuals with incorrect sex coding (38) and too high identity by state (IBS > 0.9; 27) were removed, as were markers with low call rate (<95%; 197) and low minor allele frequency (<0.01; 57). If IBS was over threshold for a pair, the individual with lower overall call rate was excluded. In total 183 109 markers and 1857 individuals (986 female, 871 male) passed the quality check (Table 2).

PCA and Model Selection

Principal components analysis (PCA) was run on the 2CY May-adjusted bill depth and bill length data. The first ("bill size") and second ("bill shape") principal components were highly correlated with bill length and bill depth respectively (Table S5) so the main analyses were carried out only on the measured values. Results for these principal components are presented in the Supplementary Material.

Preliminary exploration of phenotypic data was carried out using AICc model selection (Burnham, 2002), to identify factors and covariates that may influence bill morphology. For the repeated measures data, a 7^{+} age category was created to reduce inaccuracy caused by low numbers of individuals in the oldest age classes. Best models for bill depth and length, for both the repeated measures and 2CY May adjusted data, are displayed in Table S6. In the current dataset, body mass showed significant correlation with both bill depth ($r = 0.18$, $P < 0.001$) and bill length ($r = 0.17$, $P < 0.001$). Association analyses

were therefore performed both with and without inclusion of body mass as a model covariate, to determine whether significant markers influenced bill morphology independent of or relative to body mass.

Heritability Estimation

Genomic heritability for bill depth and length was calculated from variance component estimates given by the rGLS function in RepeatABEL (Rönnegård *et al*., 2016). The indep function in PLINK 1.9 with recommended parameters 50 5 2 was used to produce a list of markers in approximate linkage equilibrium (LE) for creation of the genetic relatedness matrix (GRM) because linkage disequilibrium (LD) between markers can confound relatedness estimates (Santure *et al*., 2010; Lopes *et al*., 2013; Eu-ahsunthornwattana *et al*., 2014; Gienapp *et al*., 2017). The GRM was then calculated using identity by state (IBS) at all autosomal markers in LE (Speed *et al*., 2012). As RepeatABEL permits inclusion of repeated individual measurements, temporal effects such as age and month can be included when estimating variance components, which may improve accuracy of h^2 estimates by properly accounting for within individual variation.

Genome Partitioning

Genome partitioning allows determination of the inheritance type of a trait by partitioning proportion of additive genetic variance explained by chromosome (Yang *et al*., 2011a, Yang *et al*., 2011b). Polygenic traits are expected to be represented by positive relationship between chromosome effect size and chromosome length, as larger chromosomes are likely to contain more protein coding genes. This positive linear relationship may be disrupted or inverted

where large effect QTL are present (Kemppainen and Husby, 2018). GCTA (Yang *et al*., 2011a) was used to perform genome partitioning to determine mode of inheritance. Chromosome specific GRMs were computed for all autosomes, then average information Restricted Maximum Likelihood (AI REML) models with multiple GRMs fitted as random effects (GCTA option -mgrm) were used to estimate the proportion of variance in each trait explained by each chromosome. GCTA does not support repeated observations, so the May 2CY-adjusted phenotypic data was used. Problems with model non-convergence were addressed by successively excluding the smallest chromosomes. Chromosome length (Mbp) was taken from the house sparrow reference genome assembly, INSDC accession number MBAE00000000.1 (Elgvin *et al*., 2017). Three chromosomes were excluded for bill depth and eight for bill length.

Typically, the relationship between variance explained and chromosome length is tested by a one-sided ordinary least squares (OLS) regression. However, due to heteroscedasticity (variance increases with number of SNPs on each chromosome) and censoring (negative values replaced by an arbitrary small positive value of 10^{-6}) this leads to *P* value inflation, resulting in ~30% false positives under the null hypothesis for typical chromosome partitioning analyses in bird genomes (Kemppainen & Husby, 2018). To account for this, a corrected *P* value was attained from 'HC-resampling' in which a null distribution for the original (uncorrected) *P* value from OLS regression is formed by resampling chromosomal heritability (h^2_c) estimates from a normal distribution with mean 0 and standard deviation equal to the SE_h (the standard error estimate for each h^2_c attained from GCTA), with 10^6 replicates. This leads to uniform distribution of *P* values under the null hypothesis of no association between phenotype and genotype (i.e. when $h^2 = 0$; Kemppainen & Husby, 2018). Because SNP numbers on the smallest chromosomes were highly variable

(Table S8), number of SNPs on each chromosome was used as a proxy for number of genes on a chromosome (n_g). When all chromosomes are included, number of SNPs on a chromosome is highly correlated with chromosome length (r = 0.998, $P < 0.001$).

Association Analyses

Three different GWA approaches were used, as each offered unique advantages with respect to our dataset: single marker association analyses were implemented in the R package RepeatABEL (Rönnegård *et al*., 2016), which permits inclusion of individual ID as a random factor to account for within individual variation. Bayesian sparse linear mixed models (BSLMM, Zhou *et al*., 2012) were implemented in GEMMA (Zhou and Stephens, 2012). BSLMM jointly model all markers and are well suited to GWA analyses on quantitative traits, as they provide estimates of the proportion of phenotypic variance explained by large effect (PGE) vs polygenic markers (PVE) allowing inference of their relative contributions. Finally, a multi-marker method, quantitative trait cluster association testing (QTCAT, Klasen *et al*., 2016), was implemented. QTCAT may increase power in GWAS on polygenic traits by searching for clusters of markers significantly associated with a given trait, mitigating the need to correct for population structure and genetic background by accounting for correlation between markers whilst simultaneously associating them with the phenotype.

A family-wise error rate (FWER) of 0.05 was used for all association analyses. The GRM used in these analyses was calculated using IBS at genome-wide SNPs in approximate LE. Proportion of phenotypic variance explained by significant markers was estimated by dividing the additive genetic variance explained by the marker by the total phenotypic variance (V_p) for the associated trait: $h^2_{SNP} = \frac{2pq(a^2)}{V_p}$, where p and q are allele frequencies, and

a is the marker effect size (Falconer and Mackay, 1996). For RepeatABEL analyses, multiple testing correction was carried out using the FWER approximation method (Halle *et al*., 2016). Prior to calculation of significance thresholds, one of each pair of SNPs with correlation r > 0.999 was removed. Full order and order three versions of the method were then used to determine the local significance level (α_{loc}) for candidate gene and GWA models respectively (see Supplementary Material).

Candidate gene association studies have increased statistical power due to lower number of markers tested (Tabor *et al*., 2002), which increases the likelihood of detecting trait associated loci with low minor allele frequency (MAF) or low effect size (Mackay *et al*., 2009). Candidate gene analysis was carried out in RepeatABEL on genes outlined in Table 1, which have previously been associated with bill morphology. The implemented models are outlined in Table S6. SNPs within 55 Kbp upstream or downstream of candidate genes for bill morphology were selected for analysis (Table 1), based on the distance at which LD decays to background levels in the house sparrow in Norway (Elgvin *et al*., 2017). Candidate gene studies do not allow identification of novel trait associated genes (Tabor *et al*., 2002), therefore GWAS for each trait were also conducted in RepeatABEL using autosomal only SNPs, as analysis of sex chromosomes requires special handling during quality control and association analysis (Wise *et al*., 2013). For both the GWA and candidate gene analyses in RepeatABEL, genome-wide markers in approximate LE (N = 57 202 SNPs) were used to estimate λ for each trait. This is particularly important in candidate gene analyses because low numbers of markers are insufficient to estimate overdispersion of test statistics caused by population structure (Devlin *et al*., 2001). For BSLMM analyses phenotypic data adjusted to May in 2CY was used. Hatch year and island were not included as fixed factors in the models as BSLMM only accepts continuous covariates. A conservative posterior inclusion probability

(PIP) cut-off of 0.1 was used, as suggested in Chaves *et al*. (2016), Comeault *et al*. (2014) and Riesch *et al*. (2017). Since BSLMM does not accept missing genotypes, these were imputed using LinkImpute (Money *et al*., 2015). QTCAT analyses also utilised the imputed dataset alongside phenotypic data adjusted to May in 2CY. The best AICc models used in the analysis are outlined in Table S6.

For all GWA analyses, the annotated house sparrow genome (Elgvin *et al*., 2017) was used to determine whether significant SNPs were exonic, intronic or flanked by a gene that had previously been related to craniofacial development or to one of our candidate genes. The annotated collared flycatcher (*Ficedula albicolis*) FicAlb1.5 (Kawakami *et al*., 2014b), and chicken (*Gallus gallus*) Gallus_gallus-5.0 genome assemblies were also used to determine position of significant SNPs in relation to known genes in the flycatcher and chicken. BLASTN searches with normal sensitivity settings were performed via Ensembl (Yates *et al*., 2016) using a sequence derived by alignment against the house sparrow genome, which spanned 1000 bp either side of the significant SNP.

Result Verification

Verification of significant results from RepeatABEL, BSLMM and QTCAT analyses was carried out using RepeatABEL in an independent dataset from a second set of house sparrow populations (N individuals = 710, N records = 1343) on three islands, Vega (65°40′N, 11°55′E), Leka (65°05′N, 11°40′E), and Lauvøya, (63°55’N, 9°55’E) situated 75-295 km South of the Helgeland system (Table S1). Analyses were carried out as in the Helgeland dataset and the GRM was created using SNPs in approximate LE. For significantly associated candidate genes, the same set of SNPs as tested in the Helgeland metapopulation was used.

For genome-wide significant markers, SNPs 55 Kbp either side of the marker were tested. Models used were the same as for RepeatABEL analyses in the Helgeland metapopulation (Table S6).

Results

Heritability Estimation and Genome Partitioning

Heritability estimates for the two bill morphology traits from analyses in RepeatABEL were relatively high: 0.35 for bill depth, and 0.38 for bill length (Table 3). In genome partitioning analysis, after HC-correction the regression between h^2_c and n_g was close to being significant (at $\alpha = 0.05$) for bill depth ($P = 0.074$) and was significant for bill length ($P = 0.024$), supporting the hypothesis that bill morphology is polygenic. Since on average a positive regression is expected between h^2_c and n_g under the null hypothesis of $h^2 = 0$, statistics from OLS regression testing for this relationship are not meaningful. However, in Figure 2, in addition to the HC-corrected regression line (including confidence intervals) the 95 % quantiles formed by the regression lines from OLS regression for each trait are shown. Some chromosomes appeared to explain a higher (chromosome 7) or lower (chromosome 6) proportion of the variation for all bill morphology traits than expected from the general relationship between variance explained and size (Fig. 2). Chromosomes 5 and 18, explained a larger proportion of the variation in bill depth, but a smaller proportion of the variation in bill length than expected (Fig. 2, Table S8). Furthermore, chromosomes 3 and 11 explained a disproportionately large proportion of the variation in bill length but not in bill depth (Fig. 2, Table S8).

Candidate Gene Association Analyses

None of the SNPs in or near candidate genes were significantly associated with bill length or bill depth in this study population at the local significance threshold for all candidate gene SNPs tested (Table S9, $\alpha_{loc} = 0.0001$, based on the 924 SNPs in or within 55 Kbp of candidate

genes for bill morphology and which passed quality control). Furthermore, no SNPs in or near candidate genes were significantly associated with bill depth or bill length at respective single-marker set significance levels.

GWAS

GWA analyses in RepeatABEL revealed a single significant SNP on chromosome 3, SNPa77348 (β = 0.054 ± 0.010, P = 1.227^{-7}, α_{loc} 3.140^{-7}) which explained 1.6% of the variance in bill depth (Table 4; Fig. 3). The significant locus is in a gene-free region of the annotated house sparrow genome (Fig. S4), situated between cysteine-rich motor neuron 1 protein (CRIM1, 1.07 Mbp away) and a gene encoding an unknown protein (125 Kbp away). CCAAT/enhancer binding protein zeta (CEBPZ) is the closest gene of known function to SNPa77348 (150 Kbp away). The candidate SNP is located at position 6.33 Mbp on chromosome 3 in the flycatcher genome (E-value = <0.000, alignment = 87.94%), and the closest genes in the flycatcher are CEBPZ and a lincRNA (ENSFALG00000018842). Several such lincRNAs lie between the candidate SNP and CRIM1 in both the flycatcher and chicken genomes.

GWA analyses using BSLMM found a single region significantly associated with bill depth, and three SNPs significantly associated with bill length when body mass was included as a model covariate (Table 5, Fig. 4). These SNPs were estimated to explain between 0.5 and 0.8% of the variance in respective traits (Table 5). The proportion of variance explained by polygenic contribution to the traits (PVE) was greater than the proportion of genetic variance explained by large effect markers (PGE) (Table 5). Marker effect sizes (β) from RepeatABEL and BSLMM were correlated (r = 0.590, P <0.001 for bill depth; r = 0.558, P < 0.001 for bill

length). The majority of the flanking genes in the house sparrow genome for SNPs detected in the BSLMM analyses have previously been linked to craniofacial development or abnormalities, or to skeletal development (Table 6; Fig. 5). The marker with the highest PIP for bill depth was SNPa77348 on chromosome 3 (γ = 0.446, β = 0.038), this is the same marker that showed significant association with bill depth in the RepeatABEL analysis (Table 4). A second significant marker in this region, SNPa77303, was also detected for bill depth (γ = 0.140, β = 0.036). SNPa262222 (γ = 0.276, β =-0.061) was significantly associated with bill length and was within the intron of Glis family zinc finger 1 (Glis1). SNPa263551 was also associated with bill length (γ = 0.154, β = 0.059), flanking genes were InaD-like protein (INADL) and Nuclear factor 1A (NFIA) in the house sparrow, flycatcher and chicken genomes. SNPa524686 associated with bill length (γ = 0.214, β = 0.063) is flanked by Nyctalopin (NYX) and Monoamine oxidase A (MAOA) in the house sparrow genome. In the chicken and flycatcher genomes, Calcium/Calmodulin dependent serine protein kinase (CASK) is situated between MAOA and NYX on chromosome 1, CASK has previously been linked to craniofacial dysmorphogenesis (Laverty and Wilson, 1998). Two additional SNPs were significantly associated with bill morphology when body mass was not included as a model covariate. SNPa500795 was then significantly associated with bill depth (γ = 0.141, β = -0.060), and the marker is within the intron of bone morphogenic protein kinase 2 (Bmpk2) in the house sparrow genome. Furthermore, SNPa707, was significantly associated with bill length when body mass was not included as a model covariate (γ = 0.108, β = -0.060). This marker is within the intron of Cbl Proto-Oncogene B (CBLB) in the house sparrow genome.

Analysis using the multi-marker GWA method, QTCAT, did not identify any significant associations at a FWER of 0.05. Selection frequency did not exceed 0.18 for any

quantitative trait cluster (QTC), and lowest *P* value observed for any SNP was 0.494. Excluding island and hatch year as fixed factors did not alter the QTCAT results.

Result Verification

Of the six GWAS significant SNP regions (RepeatABEL or BSLMM) tested for association with bill depth or length in the Southern populations, only one showed borderline significance (Table S13; Fig. S6): SNPa670, 39.5 Kbp away from SNPa707, showed significant association with bill length at the single marker set level (P = 0.0015, α_{loc} = 0.0023). This association was, however, not significant at the combined candidate gene threshold (α_{loc} = 0.0003, based on the 275 SNPs in or within 55 Kbp of markers that were significantly associated with bill morphology in GWA analyses in the Helgeland metapopulation).

392 **Discussion**

393 Using genome-wide SNP data and RepeatABEL variance component estimates, bill
394 morphology was shown to be heritable in this study population (Table 3). Heritability
395 estimates for the depth and length of bills (0.35 and 0.38, respectively) were high and broadly
396 comparable with previous findings, both in the house sparrow and in other bird species. In
397 Silva *et al*. (2017), heritabilities for bill depth and length were 0.32 and 0.39, based on a
398 subset of individuals from the Helgeland metapopulation of sparrows typed on a10K SNP
399 array (Hagen *et al*., 2013). In Jensen *et al*. (2003), pedigree based heritabilities for bill depth
400 and length in a sub-set of the same metapopulation of house sparrows were 0.32 and 0.48
401 respectively. The higher value for bill length is consistent with the tendency for pedigree-
402 based estimates to be higher than genomic heritabilities (Berenos *et al*., 2014). Estimates in
403 other avian species vary, but are generally also high (e.g. Keller *et al*., 2001; Merilä *et al*.,
404 2001).

405 Genome partitioning in GCTA revealed close to significant positive relationship (after
406 HC-correction) between chromosome length and proportion of variance explained for bill
407 depth, and significant positive relationship for bill length (Fig. 2). This result provides support
408 for the general notion that bill morphology is polygenic in nature (Boag, 1983; Grant and
409 Grant, 1994). Accordingly, most complex traits in free-living populations may be influenced
410 by many genes of small effect (Pritchard and Di Rienzo, 2010; Santure *et al*., 2015). Disparity
411 was observed between chromosomes that contributed disproportionately to a trait and whether
412 significant SNPs for the trait were located on these chromosomes in GWA analyses (Table 5,
413 Table S8). For example, although chromosome 5 explained much of the variance in bill depth
414 (but not bill length; Fig. 2, Table S8) and this chromosome is where three of our nine
415 candidate genes for bill morphology are located (Calm1, FGF19 and BMP4), no significant

hits were detected on this chromosome. This is, however, not unexpected for polygenic traits, where disproportionately high contribution to phenotypic variance by a chromosome may be caused by clustering of many small effect QTL, rather than presence of a large effect QTL (Yeaman, 2013; Schielzeth and Husby, 2014; Kemppainen and Husby, 2018). Significant markers detected in GWA analyses in the present study explained only a small proportion (0.5 - 1.6%) of the variance in their respective traits (Tables 4 and 5). This is generally indicative of polygenic genetic architecture. The high PVE:PGE ratio for all traits in our BSLMM analyses (Table 5) also supports the hypothesis that bill morphology is polygenic. Consequently, our GWAS results largely corroborate those from genome partitioning.

Of the nine candidate genes tested in association analyses in RepeatABEL, none were significantly associated with bill morphology at the local significance level calculated using all candidate SNPs. Furthermore, no large effect genes were discovered in our GWA analyses. Previous studies on bill morphology in Darwin's finches identified many genes of small effect alongside two large effect genes: ALX1 (Lamichhaney *et al*., 2015; Lamichhaney *et al*., 2016) and HMGA2 (Chaves *et al*., 2016; Lamichhaney *et al*., 2016). That none of the candidate loci associated with bill morphology in previous studies were associated with bill morphology here is somewhat surprising and illustrates the problems often associated with replicating candidate gene-trait associations in subsequent studies. Detection can be difficult, even with a high-density marker set as used in this study. Population structure may be partially responsible: if different markers are in LD with a causative variant in different populations then this will reduce power to detect the association across populations (Knief *et al*., 2017). Low effective population size (N_e) in subpopulations could pose a problem if alleles are fixed in smaller populations due to drift and this covaries with environmental effects on bill morphology (Conner and Hartl, 2004). Crucially, adaptation in quantitative

traits likely occurs by small allele frequency shifts at many loci (Tiffin and Ross-Ibarra, 2014). If a trait is sufficiently polygenic, similar adaptive responses may be achieved by modest frequency shifts at different combinations of loci (Pritchard and Di Rienzo, 2010), and the loci in question may differ between populations.

Is it instead possible that our study simply missed signals from variants near previously identified genes due to lack of statistical power (Cohen, 1998)? Our 200K SNP array has higher marker density compared to nearly all similar association studies on ecologically important traits in natural vertebrate populations (e.g. Johnston *et al*., 2011; Johnston *et al*., 2013; Kawakami *et al*., 2014a; Santure *et al*., 2015; Chaves *et al*., 2016; Huisman *et al*., 2016, but see Bosse *et al*., 2017), with average distance between markers of approximately 5000 bp, affording high power to detect causal variants. Nearly 92% of the SNPs on our custom 200K array passed quality control, and more than 180 000 SNPs could be used to assess the genetic architecture of bill morphology. GWAS sample size in the Helgeland metapopulation (N = 1857 individuals, N = 4239 records) is also larger than in the majority of studies outlined in Schielzeth and Husby (2014), which should reduce overestimation of effect sizes due to the Beavis effect (Slate, 2013). GWA studies in natural populations (e.g. Santure *et al*., 2013; Husby *et al*., 2015; Silva *et al*., 2017; Kardos *et al*., 2016) have often found little evidence for large effect QTL, especially where LD between markers is low. For traits with polygenic genetic architecture, trait linked variants are more likely to have lower effect size and be missed in GWAS even where marker density and sample size are high (Wilkening *et al*., 2009). There is mounting evidence that adaptive evolution may often occur according to a "molecular-based model" via simultaneous selection on standing variation, coupled with cumulative effects of many mutations and intragenic recombination (Remington, 2015). This as opposed to the established Fisher-Orr model

(Fisher 1930; Orr 1998), where adaptive evolution is expected to occur by sequential fixation of mutations toward the fitness optimum. This "molecular perspective" indicates that gene alleles influencing quantitative traits are likely to vary at many causative sites, which occur in different combinations between individuals. Single-marker GWA approaches likely lack power to detect the variation associated with such complex alleles (Remington, 2015). Accordingly, under this model for genetic architecture, joint-modelling of markers by the BSLMM approach used in the present study may account for its greater success.

Where large effect genes have been identified in natural populations they often influence near Mendelian traits under strong selection (Johnston *et al*., 2011; Johnston *et al*., 2013), examine divergently selected populations (Johnston *et al*., 2014; Barson *et al*., 2015; Bosse *et al*., 2017), or focus on closely related (sub)species (Lamichhaney *et al*., 2015; Chaves *et al*., 2016; Lamichhaney *et al*., 2016). In populations that are panmictic or where there is substantial gene flow, stabilising selection may lower the frequency of large effect alleles (Remington, 2015). In Bosse *et al*. (2017) differentiation between great tit populations was used in EigenGWAS (Chen *et al*., 2016) to identify outlier genomic regions under divergent selection. These were then related to bill length using post-hoc analyses. The exploitation of population divergence, alongside higher marker density (N =485 122 SNPs), may explain why none of the genes described in Bosse *et al*. (2017) were associated with house sparrow bill length in the present study. Similarly, in Lamichhaney *et al*. (2015; 2016) F_{ST} outlier approaches were used to identify large effect genes related to bill shape and bill size respectively, these genes were then linked to diversification of bill morphologies using phylogenetic analyses. Differences in bill morphology between Darwin's finch species have also been exploited in studies on gene expression (Abzhanov *et al*., 2004; Abzhanov *et al*., 2006) and epimutations (Skinner *et al*., 2014) to reveal genes important for bill phenotype. In

association studies on divergent populations or related (sub)species, approaches that take advantage of marked differentiation may be more likely to reveal large-effect loci.

For all significant markers in our GWA analyses, at least one flanking gene (or the containing gene where the SNP was intronic) had plausible association with craniofacial morphology (Table 6). Perhaps most noteworthy is the collective contribution of BMP related genes to variation in bill depth (Table 6; for a more extensive discussion of the functions of flanking genes, see Supplementary Material). However, these significant loci explained only a small proportion of the variance in their associated traits. Several genome-wide significant variants detected in this study were more than 100 Kbp away from flanking genes that have, nonetheless, previously been implicated in skeletal or craniofacial development. Neutral intergenic markers can indicate causative genes in GWAS via LD with a causal variant (Stephan *et al*., 2006) and LD range in a population indicates the average maximum distance from a gene the neutral locus may be (Backström *et al*., 2006). Previous studies on LD in natural bird populations show that the distance at which LD decreases to background level is short (Backström *et al*., 2006; Kawakami *et al*., 2014a) and inversely related to N_e (Li and Merila, 2010). LD range in the Helgeland metapopulation of house sparrows is longest in Aldra, an inbred subpopulation (Billing *et al*., 2012), and shortest in Hestmannøy, the largest subpopulation (Hagen *et al*., in prep; Baalsrud *et al*., 2014). LD also varies by chromosome, but on average falls to background levels after approximately 55 Kbp (Elgvin *et al*., 2017). SNPs more distant than this from their flanking genes are less likely to be in LD with causal variants in a gene or promoter. Consequently, significant markers that were situated more than 55 Kbp from their flanking genes in the present study may represent spurious trait associations, underscoring the challenges of gene mapping for polygenic traits. Although several of these genes have previously been connected to craniofacial morphology (Table 6),

over-interpretation should be avoided, as ability to construct meaningful narrative may not always imply result validity (Pavlids *et al*., 2012). Alternatively, distal SNPs may be linked to a causal variant in an enhancer, or other type of intergenic regulatory element (Fig. S12). Although enhancers may skip several genes to regulate a more distal gene, up to 60% of enhancers interact with the nearest promoter (Yao *et al*., 2015). Flanking genes to significant markers were investigated here for simplicity and because regulatory elements have not yet been well annotated in any avian genome. A study by Chan *et al*. (2010) revealed that adaptive evolution of the pelvic region in sticklebacks was governed by deletion mutations in the *Pel* enhancer region for Pitx1, underscoring how major morphological changes can occur due to variation in intergenic DNA regions. It is worth noting that the genetic architecture in association studies is complex, as the total historical variation in the sampled population is included. LD between a given SNP and a causal variant is determined by the history of mutation, recombination and coalescence of lineages, and SNP-trait associations may not decrease monotonically with distance (Remington, 2015; Li *et al*., 2018).

Population structure and kinship are commonly included in GWA models to factor out spurious SNP-trait associations that occur due to shared ancestry. However, this may also factor out relevant variation if molecular differences between populations correlate with phenotypic differences in the examined trait (Remington, 2015). In the present metapopulation of house sparrows, differences in bill morphology between islands existed (Fig. 6), so controlling for island may factor out genuine associations. This may offer a partial explanation as to why more significant variants were detected by BSLMM analyses (Table 5) than by the RepeatABEL method (Table 4), as BSLMM only allows for continuous covariates so island could not be included as a fixed factor. However, removing island as a fixed factor in RepeatABEL GWAS did not yield any additional significant hits (results not shown).

Finally, our study system presented the opportunity to perform tests of repeatability on results from the main Helgeland system in a second set of genetically distinct island populations, which is uncommon in GWA studies. Result verification in these Southern populations yielded only one borderline significant result for SNPa707 (Table S13; Fig. S6). Low replicability in the Southern populations may indicate spurious SNP-trait associations in the main analysis. Conversely, low replicability may be explained by smaller sample size in the Southern populations, or differing LD patterns from the Helgeland population. Replicability is often low in association studies, where significant loci frequently differ between populations or (sub)species. Lack of signals at candidate regions in the Southern populations, despite our large-scale dataset, reinforces the message that loci responsible for differences in phenotype may routinely differ between populations for quantitative traits with polygenic genetic architecture (Pritchard and Di Rienzo, 2010). This may particularly be true if adaptive evolution commonly occurs according to the "molecular-based model" (Remington, 2015).

Our results support previous findings in many bird species that bill morphology is highly heritable. Six significant loci were identified for bill depth or length using genome-wide association analysis, each explaining only a small proportion of the variation in bill morphology. Two of these loci were flanked by genes involved in BMP signalling. Additionally, an outlier chromosome in genome partitioning analysis contained BMP4 and the co-expressed Calm1. These results indicate that BMP related genes may collectively contribute to bill morphology variation in house sparrows. This work illustrates the difficulties involved in discovering loci for quantitative traits in natural populations: despite the high marker density of our 200K SNP array and our large-scale dataset, previously identified large effect genes for bill morphology were not associated with bill length or depth here. This, alongside lack of strong signals for marker regions tested in the Southern

560 populations, suggests bill morphology is polygenic in nature in the house sparrow and likely
561 in other bird species. Hence, methods that can accommodate polygenic inheritance of the
562 phenotype and molecular-scale evolution of genetic architecture are needed to study any eco-
563 evolutionary processes in bill morphology.

564

565 **Acknowledgements**

566 We thank L.E. Andersen, O.R. Davidsen, I. Dorrestein, T. Kolaas, L.K. Larsen, P.A.
567 Martinussen, S.J. Monkan, M. Mørkved, K. Nygård, S.L. Sommerli, Ø.Ø. Tennfjord, T.R.
568 Østerås and other fieldworkers and students for their contributions to fieldwork, J. Kekkonen
569 and J. Brommer for allowing access to Finnish sparrow DNA samples for resequencing, and
570 R. Røsbak for her assistance with laboratory analyses. Thanks is also extended to M. Langaas,
571 Ø. Bakke and K. Halle for providing the scripts for multiple testing correction. We would also
572 like to thank L. Andersson for pre-publication information about involvement of ALX1 in
573 Darwin's finch bill morphology and for access to information about ALX1, S. Johnston,
574 whose comments greatly improved the paper, as well as two anonymous reviewers and
575 Subject Editor Dr. S. Rogers for valuable comments on previous versions of the paper.
576 Finally, we extend our gratitude to the residents of the study system, whose ongoing
577 cooperation has contributed to its success. This study was supported by grants from the
578 Norwegian Research Council (programmes STORFORSK, Strategic University Program in
579 Conservation Biology, project 221956, and projects 214553 and 239974 to A.H.), the
580 Norwegian Directorate for Nature Management, the EU-commission (project METABIRD),
581 and the Academy of Finland (project 295204 to A.N.). This work was also partly supported
582 by the Research Council of Norway through its Centres of Excellence funding scheme

583 (project 223257). The fieldwork was carried out in accordance with permits from the Ringing
584 Centre at Stavanger Museum, Norway. Whole genome re-sequencing of house sparrows was
585 carried out at Genome Quebec, McGill University, Canada. Genotyping on the custom house
586 sparrow Affymetrix Axiom 200K SNP array was carried out at CIGENE, Norwegian
587 University of Life Sciences, Norway.

References

Abzhanov, A., Kuo, W. P., Hartmann, C., Grant, B. R., Grant, P. R. & Tabin, C. J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature,* 442**,** 563-567.

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R. & Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science,* 305**,** 1462-1465.

Abzhanov, A. & Tabin, C. J. (2004). Shh and Fgf8 act synergistically to drive cartilage outgrowth during cranial development. *Developmental Biology,* 273**,** 134-148.

Aulchenko Y.S., Ripke S., Isaacs A., van Duijn C.M. (2007). GenABEL: an R package for genome-wide association analysis. *Bioinformatics*, 23(10), 1294-6.

Ameye, L. & Young, M. F. (2002). Mice deficient in small leucine-rich proteoglycans: novel in vivo models for osteoporosis, osteoarthritis, Ehlers-Danlos syndrome, muscular dystrophy, and corneal diseases. *Glycobiology,* 12**,** 107r-116r.

Anderson, T. R. (2006). *Biology of the Ubiquitous House Sparrow: From Genes to Populations,* Oxford, Oxford University Press.

Backström, N., Qvarnström, A., Gustafsson, L. & Ellegren, H. (2006). Levels of linkage disequilibrium in a wild bird population. *Biology Letters,* 2**,** 435-438.

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, 2653-2668.

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T. F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., Erkinaro, J., Lien, S. & Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature,* 528**,** 405-408.

Bech-Hansen, N. T., Naylor, M. J., Maybaum, T. A., Sparkes, R. L., Koop, B., Birch, D. G., Bergen, A. A. B., Prinsen, C. F. M., Polomeno, R. C., Gal, A., Drack, A. V., Musarella, M. A., Jacobson, S. G., Young, R. S. L. & Weleber, R. G. (2000). Mutations in NYX, encoding the leucine-rich proteoglycan nyctalopin, cause X-linked complete congenital stationary night blindness. *Nature Genetics,* 26**,** 319-323.

Berenos, 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**,** 3434-3451.

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 house sparrow population. *Molecular Ecology*, 21, 1487-1499.

Boag, P. T. (1983). The heritability of external morphology in Darwin's ground finches (Geospiza) on Isla Daphne Major, Galapagos. *Evolution,* 37**,** 877-894.

Boag, P. T. & Grant, P. R. (1981). Intense natural selection in a population of Darwin's Finches (Geospizinae) in the Galapagos. *Science,* 214**,** 82-5.

Bosse, M., Spurgin, L.G., Laine, V.N., Cole, E.F., Firth, J.A., Gienapp, P., Gosler, A.G., Mcmahon, K., Poissant, J., Verhagen, I., Groenen, M., Van-Oers, K., Sheldon, B.C., Visser, M.E., Slate, J. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, 358(6361), 365-368.

Brugmann, S. A., Powder, K. E., Young, N. M., Goodnough, L. H., Hahn, S. M., James, A. W., Helms, J. A. & Lovett, M. (2010). Comparative gene expression analysis of avian embryonic facial structures reveals new candidates for human craniofacial disorders. *Human Molecular Genetics,* 19**,** 920-930.

Brunner, H. G., Nelen, M., Breakefield, X. O., Ropers, H. H. & Van Oost, B. A. (1993). Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science,* 262**,** 578-580.

Burnham, K. P. (2002). *Model selection and multimodel inference: a practical information-theoretic approach,* New York, Springer.

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., Southwick, A. M., Absher, D. M., Grimwood, J., Schmutz, J., Myers, R. M., Petrov, D., Jónsson, B., Schluter, D., Bell, M. A. & Kingsley, D. M. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science (New York, N.Y.),* 327**,** 302-305.

Chaves, J. A., Cooper, E. A., Hendry, A. P., Podos, J., De León, L. F., Raeymaekers, J. A. M., Macmillan, W. O. & Uy, J. A. C. (2016). Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology***,** 25(21), 5282–5295.

Chen, G. B., Lee, S. H., Zhu, Z. X., Benyamin, B. & Robinson, M. R. (2016). EigenGWAS: Finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity*, 117, 51–61.

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.
Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition. Hillsdale, N.J., Lawrence Erlbaum Associates.
Comeault, A. A., Soria-Carrasco, V., Gompert, Z., Farkas, T. E., Buerkle, C. A., Parchman, T. L. & Nosil, P. (2014). Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in Timema cristinae. *The American Naturalist,* 183**,** 711-727.
Conner, J. K. & Hartl, D. L. (2004). *A primer of ecological genetics,* Sunderland, Mass., Sinauer Associates.
Devlin, B., Roeder, K. & Wasserman, L. (2001). Genomic Control, a New Approach to Genetic-Based Association Studies. *Theoretical Population Biology,* 60**,** 155-166.
Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., Ravinet, M., Jensen, H. & Sætre, G.-P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances,* 3**(**6), e1602996.
Eu-Ahsunthornwattana, J., Miller, E. N., Fakiola, M., Wellcome Trust Case Control, C., Jeronimo, S. M. B., Blackwell, J. M. & Cordell, H. J. (2014). Comparison of Methods to Account for Relatedness in Genome-Wide Association Studies with Family-Based Data. *PLoS Genetics,* 10**,** e1004445.
Falconer, D & Mackay, T. (1996). *Introduction to Quantitative Genetics, Ed 4*. Longmans Green, Harlow, Essex, UK.
Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press, Oxford (Variorum ed. 1992.Oxford Univ. Press, New York).
Forstmeier, W., Leisler, B. & Kempenaers, B. (2001). Bill morphology reflects female independence from male parental help. *Proceedings of the Royal Society B: Biological Sciences,* 268**,** 1583-1588.
Gienapp, P., Fior, S,. Guillaume, F., Lasky, J, R., Sork, V, L. & Csilléry, K. (2017). Genomic Quantitative Genetics to Study Evolution in the Wild. *Trends in Ecology and Evolution*, 32(12), 897-908.
Grant, P. & Grant, B. (1994). Phenotypic and genetic-effects of hybridization in Darwins Finches. *Evolution,* 48(2), 297-316.
Grant, P. R. & Grant, B. R. (2002). Unpredictable Evolution in a 30-Year Study of Darwin's Finches. *Science,* 296**,** 707-711.
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**,** 429-439.
Halle, K. K., Bakke, Ø., Djurovic, S., Bye, A., Ryeung, E., Wisløff, U., Andreassen, O. A. & Langaas, M. (2016). Efficient and powerful familywise error control in genome-wide association studies using generalized linear models. *Cornell University Library*, arXiv:1603.05938.
Haworth, K. E., Wilson, J. M., Grevellec, A., Cobourne, M. T., Healy, C., Helms, J. A., Sharpe, P. T. & Tucker, A. S. (2007). Sonic hedgehog in the pharyngeal endoderm controls arch pattern via regulation of Fgf8 in head ectoderm. *Developmental Biology,* 303**,** 244-258.
Huisman, J., Kruuk, L. E., Ellis, P. A., Clutton-Brock, T. & Pemberton, J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America*, 113(13), 3585-3590.
Husby, A., Kawakami, T., Rönnegård, L., Smeds, L., Ellegren, H. & Qvarnström, A. (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). https://doi.org/10.1098/rspb.2015.0156.
James, N. M. S. & Zach, R. (1979). Heritability of Some Morphological Characters in a Song Sparrow Population. *Evolution,* 33**,** 460-467.
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**,** 1296-307.
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**,** 599-611.
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.
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**,** 93-95.

Johnston, S. E., Mcewan, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., Pemberton, J. M. & 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**,** 2555-2566.

Johnston, S. E., Orell, P., Pritchard, V. L., Kent, M. P., Lien, S., Niemelä, E., Erkinaro, J. & 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**,** 3452-3468.

Kardos, M., Husby, A., McFarlane, S. E., Qvarnström, A. & Ellegren, H. (2016). Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resource*s, 16, 727-741.

Kawakami, T., Backström, N., Burri, R., Husby, A., Olason, P., Rice, A. M., Ålund, M., Qvarnström, A. & Ellegren, H. (2014a). Estimation of linkage disequilibrium and interspecific gene flow in Ficedula flycatchers by a newly developed 50k single-nucleotide polymorphism array. *Molecular Ecology Resources,* 14**,** 1248-1260.

Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C. F., Olason, P. & Ellegren, H. (2014b). A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology,* 23**,** 4035-4058.

Kearns, A. E., Donohue, M. M., Sanyal, B. & Demay, M. B. (2001). Cloning and characterization of a novel protein kinase that impairs osteoblast differentiation in vitro. *Journal of Biologial Chemistry,* 276**,** 42213-42218.

Keller, L. F., Grant, P. R., Grant, B. R. & Petren, K. (2001). Heritability of morphological traits in Darwin/'s Finches: misidentified paternity and maternal effects. *Heredity,* 87**,** 325-336.

Kemppainen, P. & Husby, A. (2018). Inference of genetic architecture from chromosome partitioning analyses depends on genome variation, sample size, heritability and effect size distribution. *Molecular Ecology Resources*, xx, xxx-xxx. https://doi.org/10.1111/1755-0998.12774.

Kim, Y. S., Lewandoski, M., Perantoni, A. O., Kurebayashi, S., Nakanishi, G. & Jetten, A. M. (2002). Identification of Glis1, a novel Gli-related, Kruppel-like zinc finger protein containing transactivation and repressor functions. *Journal of Biological Chemistry,* 277**,** 30901-30913.

Klasen, J. R., Barbez, E., Meier, L., Meinshausen, N., Bühlmann, P., Koornneef, M., Busch, W. & Schneeberger, K. (2016). A multi-marker association method for genome-wide association studies without the need for population structure correction. *Nature Communications,* 7**,** 13299.

Knief, U., Schielzeth, H., Backström, N., Hemmrich-Stanisak, G., Wittig, M., Franke, A., Griffith, S. C., Ellegren, H., Kempenaers, B. & Forstmeier, W. (2017), Association mapping of morphological traits in wild and captive zebra finches: reliable within, but not between populations. *Molecular Ecology*, 26, 1285–1305.

Kolle, G., Jansen, A., Yamada, T. & Little, M. (2003). In ovo electroporation of Crim1 in the developing chick spinal cord. *Developmental Dynamics,* 226**,** 107-111.

Kruuk, L.E. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical transactions of the Royal Society of London. Series B*, 29(359), 873-890.

Kumar, M., Ray, P. & Chapman, S. C. (2012). FGF and BMP signaling are required for specifying pre-chondrogenic identity in neural crest derived mesenchyme and initiating the chondrogenic program. *Developmental Dynamics,* 241**,** 1091-1103.

Kvalnes, T., Ringsby, T. H., Jensen, H., Hagen, I. J., Rønning, B., Pärn, H., Holand, H., Engen, S. & Sæther, B. E. (2017). Reversal of response to artificial selection on body size in a wild passerine bird. *Evolution***,** 71(8), 2062-2079.

Lamichhaney, S., Berglund, J., Almen, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerova, M., Rubin, C.-J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T. & Andersson, L. (2015). Evolution of Darwin/'s finches and their beaks revealed by genome sequencing. *Nature,* 518**,** 371-375.

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almén, M. S., Webster, M. T., Grant, B. R., Grant, P. R. & Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science,* 352**,** 470-474.

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

Laverty, H. G. & and Wilson, J. B. (1998). Murine CASK is disrupted in a sex-linked cleft palate mouse mutant. *Genomics*, 53(1), 29-41.

Li, M. H. & Merila, J. (2010). Extensive linkage disequilibrium in a wild bird population. *Heredity,* 104**,** 600-610.

Li, Z., Kemppainen, P., Rastas, P. & Merilä, J. (2018). Linkage disequilibrium clustering□based approach for association mapping with tightly linked genomewide data. *Molecular Ecology Resources*, https://doi.org/10.1111/1755-0998.12893.

Lopes, M. S., Silva, F. F., Harlizius, B., Duijvesteijn, N., Lopes, P. S., Guimarães, S. E. & Knol, E. F. (2013). Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. *BMC Genetics,* 14**,** 1-11.

Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics,* 10**,** 565-577.

Mallarino, R., Grant, P. R., Grant, B. R., Herrel, A., Kuo, W. P. & Abzhanov, A. (2011). Two developmental modules establish 3D beak-shape variation in Darwin's finches. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 4057-4062.

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**,** 918-929.

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G. Y. & Myles, S. (2015). LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3 (Bethesda),* 5**,** 2383-90.

Musialik, E., Bujko, M., Kober, P., Grygorowicz, M. A., Libura, M., Przestrzelska, M., Juszczyński, P., Borg, K., Florek, I., Jakóbczyk, M., Baranowska, A. & Siedlecki, J. A. (2014). Comparison of promoter DNA methylation and expression levels of genes encoding CCAAT/enhancer binding proteins in AML patients. *Leukemia Research,* 38**,** 850-856.

Ogata, T., Muroya, K., Matsuo, N., Shinohara, O., Yorifuji, T., Nishi, Y., Hasegawa, Y., Horikawa, R. & Tachibana, K. (2001). Turner syndrome and Xp deletions: clinical and molecular studies in 47 patients. *Journal of Clinical Endocrinology & Metabolism,* 86**,** 5498-5508.

Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, 52, 935–949.

Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., Rexroad III, C. E. & Moen, T. (2015). The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources*, 15(3), 662-672.

Pavlidis, P., Jensen, J. D., Stephan, W. & Stamatakis, A. (2012). A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution*, 29(10), 3237-3248.

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**,** 1216-1225.

Pritchard, J. K. & Di Rienzo, A. (2010). Adaptation – not by sweeps alone. *Nature Reviews Genetics,* 11**,** 665-667.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W. & Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics,* 81(3) 559-575.

Rao, A., O'donnell, S., Bain, N., Meldrum, C., Shorter, D. & Goel, H. (2014). An intragenic deletion of the NFIA gene in a patient with a hypoplastic corpus callosum, craniofacial abnormalities and urinary tract defects. *European Journal of Medical Genetics,* 57**,** 65-70.

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/.

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.

Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A. A., Farkas, T. E., Lucek, K., Hellen, E., Soria-Carrasco, V., Dennis, S. R., De Carvalho, C. F., Safran, R. J., Sandoval, C. P., Feder, J., Gries, R., Crespi, B. J., Gries, G., Gompert, Z. & Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution,* 1**,** https://doi.org/10.1038/s41559-017-0082.

Riyahi, S., Hammer, Ø., Arbabi, T., Sánchez, A., Roselaar, C. S., Aliabadian, M. & Sætre, G.-P. (2013). Beak and skull shapes of human commensal and non-commensal house sparrows Passer domesticus. *BMC Evolutionary Biology,* 13**,** 1-8

Ringsby, T., Berge, T., Sæther, B.-E. & Jensen, H. (2009). Reproductive success and individual variation in feeding frequency of House Sparrows (Passer domesticus). *Journal of Ornithology,* 150**,** 469-481.

Rönnegård, L., Mcfarlane, S. E., 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,* 7**,** 792-799.

Santure, A. W., De 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**,** 3949-3962.

Santure, A. W., Poissant, J., De Cauwer, I., Van Oers, K., Robinson, M. R., Quinn, J. L., Groenen, M. A. M., Visser, M. E., Sheldon, B. C. & Slate, J. (2015). Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. *Molecular Ecology,* 24**,** 6148-6162.

Santure, A. W., Stapley, J., Ball, A. D., Birkhead, T. R., Burke, T. & Slate, J. O. N. (2010). On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology,* 19**,** 1439-1451.

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**,** 35-57.

Shirakawa, K., Maeda, S., Gotoh, T., Hayashi, M., Shinomiya, K., Ehata, S., Nishimura, R., Mori, M., Onozaki, K., Hayashi, H., Uematsu, S., Akira, S., Ogata, E., Miyazono, K. & Imamura, T. (2006). CCAAT/Enhancer-Binding Protein Homologous Protein (CHOP) Regulates Osteoblast Differentiation. *Molecular and Cellular Biology*, 26, 6105-6116

Silva, C. N. S., Mcfarlane, S. E., Hagen, I. J., Ronnegard, L., Billing, A. M., Kvalnes, T., Kemppainen, P., Ronning, B., Ringsby, T. H., Sæther, B. E., Qvarnstrom, A., Ellegren, H., Jensen, H. & Husby, A. (2017). Insights into the genetic architecture of morphological traits in two passerine bird species. *Heredity,* 119**,** 197-205.

Simovich, M. J., Bland, S. D., Peiffer, D. A., Gunderson, K. L., Cheung, S. W., Yatsenko, S. A. & Shinawi, M. (2008). Delineation of the proximal 3q microdeletion syndrome. *American Journal of Medical Genetics Part A,* 146A**,** 1729-1735.

Skinner, M. K., Guerrero-Bosagna, C., Haque, M. M., Nilsson, E. E., Koop, J. A. H., Knutie, S. A. & Clayton, D. H. (2014). Epigenetics and the evolution of Darwin's Finches. *Genome Biology and Evolution*, 6(8), 1972–1989.

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**,** 1251-1262.

Slate, J., Santure, A. W., Feulner, P. G. D., Brown, E. A., Ball, A. D., Johnston, S. E. & Gratten, J. (2010). Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics,* 26**,** 275-284.

Soons, J., Genbrugge, A., Podos, J., Adriaens, D., Aerts, P., Dirckx, J. & Herrel, A. (2015). Is Beak Morphology in Darwin's Finches Tuned to Loading Demands? *PLoS ONE,* 10**.** https://doi.org/10.1371/journal.pone.0129479.

Speed, D., Hemani, G., Johnson, Michael R. & Balding, David J. (2012). Improved Heritability Estimation from Genome-wide SNPs. *American Journal of Human Genetics,* 91**,** 1011-1021.

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A. D., Beckerman, A. P. & Slate, J. (2010). Adaptation genomics: the next generation. *Trends in Ecology & Evolution,* 25**,** 705-712.

Stephan, W., Song, Y. S. & Langley, C. H. (2006). The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics,* 172(4)**,** 2647-2663.

Tabor, H. K., Risch, N. J. & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics,* 3**,** 391-397.

Temeles, E. J., Koulouris, C. R., Sander, S. E. & Kress, W. J. (2009). Effect of flower shape and size on foraging performance and trade-offs in a tropical hummingbird. *Ecology,* 90**,** 1147-1161.

Tiffin, P. & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution,* 29**,** 673-680.

Vaccaro, P., Brannetti, B., Montecchi-Palazzi, L., Philipp, S., Helmer Citterich, M., Cesareni, G. & Dente, L. (2001). Distinct binding specificity of the multiple PDZ domains of INADL, a human protein with homology to INAD from Drosophila melanogaster. *Journal of Biological Chemistry,* 276**,** 42122-42130.

Visscher, P.M., Hill, W.G. & Wray, N.R. (2008). Heritability in the genomics era – concepts and misconceptions. *Nature Reviews Genetics*, 9, 255-266.

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**,** 415-419.

Wilkinson, L., Kolle, G., Wen, D., Piper, M., Scott, J. & Little, M. (2003). CRIM1 Regulates the Rate of Processing and Delivery of Bone Morphogenetic Proteins to the Cell Surface. *Journal of Biological Chemistry,* 278**,** 34181-34188.

Wise, Anastasia L., Gyi, L. & Manolio, Teri A. (2013). eXclusion: toward integrating the X chromosome in genome-wide association analyses. *American Journal of Human Genetics,* 92**,** 643-647.

Wu, P., Jiang, T.-X., Suksaweang, S., Widelitz, R. B. & Chuong, C.M. (2004). Molecular Shaping of the Beak. *Science (New York, N.Y.),* 305**,** 1465-1466.

Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. (2011a). GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics,* 88**,** 76-82.

Yang, J., Weedon, M. N., Purcell, S., Lettre, G., Estrada, K., Willer, C. J., Smith, A. V., Ingelsson, E., O'connell, J. R., Mangino, M., Mägi, R., Madden, P. A., Heath, A. C., Nyholt, D. R., Martin, N. G., Montgomery, G. W., Frayling, T. M., Hirschhorn, J. N., Mccarthy, M. I., Goddard, M. E., Visscher, P. M. & The, G. C. (2011b). Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics,* 19**,** 807-812.

Yao, L., Berman, B. P. & Farnham, P. J. (2015). Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. *Critical Reviews in Biochemistry and Molecular Biology,* 50**,** 550-573.

Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F. J., Maurel, T., Mclaren, W., Murphy, D. N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S. J., Cunningham, F., Aken, B. L., Zerbino, D. R. & Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Research,* 44**,** D710-D716.

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adapted loci. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), E1743–E1751.

Zhou, X., Carbonetto, P. & Stephens, M. (2012). Polygenic Modeling with Bayesian Sparse Linear Mixed Models. https://doi.org/10.1371/journal.pgen.1003264

Zhou, X. & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics,* 44**,** 821-824.

Data Accessibility

Phenotype and genetic data will be deposited in Dryad on acceptance of the manuscript.

Author Contributions

H.J., S.L.L., A.H. and I.J.H. designed the study. H.J., T.H.R., H.P.A., T.K., P.S.R., H.H., I.J.H., B.R. and A.H. collected field data. I.J.H and H.J. developed the SNP array. A.S.B. and L.-K.S. prepared the samples for SNP-genotyping, A.H. and A.N. cleaned the SNP genotype data. S.L.L and J.G. analysed the data with guidance from H.J. and A.H. The article was written by S.L.L. with input from all authors.

Table 1: Candidate genes enriched on our custom Affymetrix 200K SNP array that have previously been related to avian bill morphology, residing chromosome (Chr.), number of SNPs within 55 Kbp of the gene after quality control, reference(s) and study species. Gene roles were taken from associated references.

Gene	Name	Chr.	SNPs	Ref.	Species	Role
ALX1	ALX homeobox-1	1A	167	(Lamichhaney *et al*., 2015)	*Geospiza*	Associated with blunt bills
BMP4	Bone morphogenic protein 4	5	56	(Abzhanov *et al*., 2006, Abzhanov *et al*., 2004)	*Geospiza*	Differentially expressed in the developing bill of species with divergent bill morphologies
Calm1	Calmodulin	3	146			
TGFbrII	Transforming growth factor-beta receptor type-2	7	22	(Mallarino *et al*., 2011)	*Geospiza*	
Dkk2	Dickkopf Wnt signalling pathway inhibitor 2	4	160	(Brugmann *et al*., 2010)	*Anas platyrhynchos, Gallus gallus, Coturnix spp.*	Expression of these Wnt signalling pathway members is upregulated in species with broad bills and downregulated in those with narrow bills
FZD1	Frizzled-1	2	173			
FGF8	Fibroblast growth factor 8	6	67	(Abzhanov and Tabin, 2004, Wu *et al*., 2004)	*Anas platyrhynchos, Gallus gallus, Geospiza*	FGF8 interacts with other proteins in the frontonasal ectoderm, including Shh, to induce expression of BMP4 at the proper domain on the developing bill
Shh	Sonic hedgehog	2	123			
FGF19	Fibroblast growth factor 19	5	10	(Haworth *et al*., 2007, Kumar *et al*., 2012)	*Gallus gallus*	Strong expression in pharyngeal endoderm

Table 2: Period for which all recorded adult individuals on each island were genotyped on the 200K SNP microarray, number of SNP genotyped individuals that passed quality control on each island (N_i) and number of records per island (N_r).

Island	Years	N_i	N_r
Aldra	1998 – 2013	146	493
Gjerøy	1998 – 2013	402	907
Hestmannøy	1998 – 2013	717	1709
Indre Kvarøy	1998 – 2013	251	556
Myken	2004 – 2013	38	54
Nesøy	1998 – 2013	98	278
Selvær	2003 – 2013	107	124
Træna	2003 - 2013	98	118
TOTAL	16	1857	4239

Table 3: Variance components of RepeatABEL GWAS for the measured bill morphology traits depth and length. Trait heritabilities (h^2) were calculated using these variance components and the formula $\mathbf{h^2 = V_a/(V_a+V_{pe}+V_r)}$, where V_a is additive genetic variance, V_{pe} is permanent environmental variance, and V_r is residual variance.

Trait	V_a [95% CI]	V_{pe}	V_r	h^2
Bill Depth	0.030 [0.027, 0.034]	0.020	0.035	0.35
Bill Length	0.112 [0.100, 0.125]	0.077	0.108	0.38

Table 4: Summary statistics for the top five SNPs associated with bill depth in the RepeatABEL GWAS. Top SNP is significant at the genome-wide local significance level of 3.14^{-7}. For each SNP the table shows its name, chromosome, position (bp), the reference allele A1, effect allele A2, minor allele frequency (MAF), estimated effect size of A2 with standard error, adjusted *P* value, and Hardy-Weinberg (HWE) *P* value. Effect size and *P* values for significant SNPs when body mass was not included as a covariate shown in grey. Proportion of variance in bill depth accounted for by each marker (h^2_{SNP}) is also given.

SNP	Chr.	Position	A1	A2	MAF	Effect ± SE	*P* value	HWE *P* value	h^2_{SNP}
SNPa77348	**3**	**21541818**	**C**	**T**	**0.39**	**0.054 ± 0.010** 0.055 ± 0.010	**1.227^{-7}** 1.151^{-7}	**0.407**	**0.016**
SNPa77303	3	21474218	T	G	0.36	0.051 ± 0.011 0.052 ± 0.011	1.495^{-6} 1.476^{-6}	0.013	0.014
SNPa196812	7	21744692	G	A	0.15	0.067 ± 0.014 0.066 ± 0.015	3.073^{-6} 3.396^{-6}	0.073	0.013
SNPa500795	4	1474181	A	G	0.12	-0.070 ± 0.015 -0.070 ± 0.015	3.169^{-6} 3.164^{-6}	0.541	0.012
SNPa390338	29	32346307	G	T	0.16	0.063 ± 0.014 0.058 ± 0.014	6.449^{-6} 1.393^{-5}	0.723	0.012

Table 5: Significant markers and the traits they were associated with in BSLMM analysis on bill depth and bill length**.** Proportion of variance explained by both sparse and random terms (PVE), and proportion of genetic variance explained by the sparse terms (PGE) are given for each trait. Summary statistics for significant SNPs are given: residing chromosome (Chr.) position, minor allele frequency (MAF), and Hardy-Weinberg equilibrium (HWE) *P* value. Shown outputs from BSLMM are Posterior Inclusion Probability (PIP) and effect size (β). PIP and β when body mass was not included as a covariate in the model are shown in grey. Proportion of variance in associated trait accounted for by each marker (h^2_{SNP}) is also given.

Trait	PVE	PGE	SNP	Chr.	Position	MAF	HWE *P* value	PIP	β	h^2_{SNP}
Bill Depth	0.481	0.333	SNPa77348	3	21541818	0.39	0.408	0.446 0.393	0.038 0.044	0.008
			SNPa77303	3	21474218	0.36	0.012	0.140	0.036	0.007
			SNPa500795	4	1474181	0.12	0.513	0.141	-0.060	0.009
Bill Length	0.546	0.283	SNPa262222	8	4914956	0.45	0.083	0.276 0.224	-0.061 -0.058	0.005
			SNPa524686	1	106229088	0.35	0.025	0.214 0.147	0.063 0.059	0.005
			SNPa263551	8	2436193	0.42	1.000	0.154 0.220	0.059 0.060	0.005
			SNPa707	1	17578765	0.27	0.314	0.108	-0.060	0.005

964 **Table 6:** Flanking genes for significant markers from RepeatABEL and BSLMM analyses.
965 Shown are the markers and the analysis they were significant in, the bill morphology traits
966 each significant marker was associated with, and the two flanking genes (if SNP is intergenic)
967 or containing gene (if SNP is intronic). Functions of the genes based on NCBI and literature
968 search are listed. One marker, SNPa77348, was significantly associated with bill depth in both
969 the RepeatABEL and BSLMM analysis. CASK is situated between MAOA and NYX in the
970 chicken genome, there is currently no position information for this gene in the house sparrow
971 genome.

SNP (analysis)	**Associated Trait**	**Flanking Gene**	**Distance (Kbp)**	**Function**	**Reference(s)**
SNPa77348 (RepeatABEL, BSLMM)	Bill depth	CRIM1	1072	Inhibits BMP4 receptor activation and signalling	(Kolle *et al*., 2003; Wilkinson *et al*., 2003)
		CEBPZ	150	Transcriptional process involving heat shock factors, regulates osteogenesis via BMP signalling	(Musialik *et al*., 2014; Shirakawa et al., 2006)
SNPa500795 (BSLMM, no body mass)	Bill depth	Bmpk2	Intronic	BMP2 induced osteoblast differentiation	(Kearns *et al*., 2001)
SNPa262222 (BSLMM)	Bill length	Glis1	Intronic	Developmental gene expression control, high expression in craniofacial region	(Kim *et al*., 2002)
SNPa263551 (BSLMM)	Bill length	INADL	128	Organisation of multimeric complexes at plasma membrane	(Vaccaro *et al*., 2001)
		NFIA	197	DNA replication, craniofacial abnormality, anteverted nares	(Rao *et al*., 2014)
SNPa524686 (BSLMM)	Bill length	NYX	171	Collagen fibrilogenesis, retinal disease	(Ameye and Young, 2002; Bech-Hansen *et al*., 2000)
		CASK	*Gallus gallus*	Linked to craniofacial dysmorphogenesis	(Laverty and Wilson, 1998)
		MAOA	441	Lymphogenic, oxidative deamination of neurotransmitters. Agression, Turner syndrome	(Ogata *et al*., 2001; Brunner *et al*., 1993)
SNPa707 (BSLMM, no body mass)	Bill length	CBLB	Intronic	Linked to craniofacial defects in humans, anteverted nares and flattened nasal bridge	(Simovich *et al*., 2008)

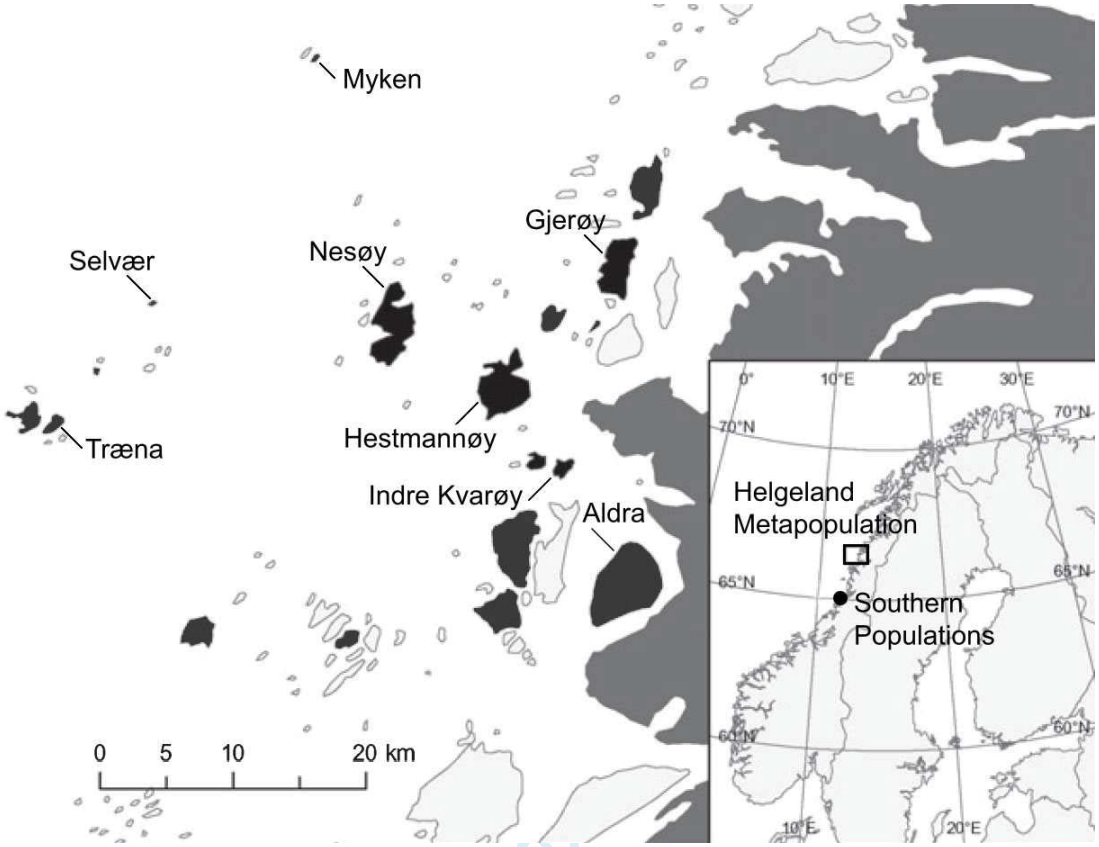


Figure 1: Islands included in the house sparrow metapopulation system, Helgeland, Northern Norway (66°30'N, 12°30'E). Islands shaded in black have been continuously followed since monitoring began. The eight islands used in main analysis in this study are indicated. The locator map shows the latitude of the Helgeland metapopulation and the Southern populations: Vega (65°40′N, 11°55′E), Leka (65°05′N, 11°40′E), and Lauvøya, (63°55'N, 9°55'E).

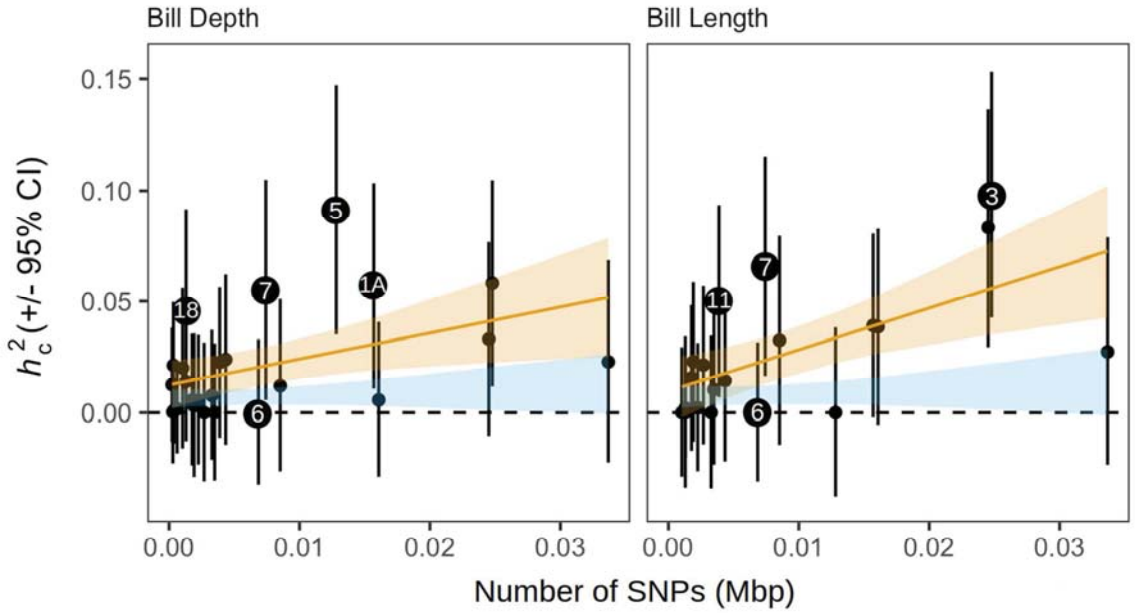


Figure 2: Relationship between explained variation in bill morphology (± SE) and chromosome size for **A)** bill depth, and **B)** bill length. Numbers indicate chromosome number. The yellow shaded area shows 95% C.I. for OLS regression between number of SNPs and the proportion of variance explained by each chromosome (h^2_c), with black bars indicating 95% C.I. for each h^2c estimate. Blue shaded area indicates 95% quantiles formed by HC-resampling forming the null distribution for HC-corrected *P* value.

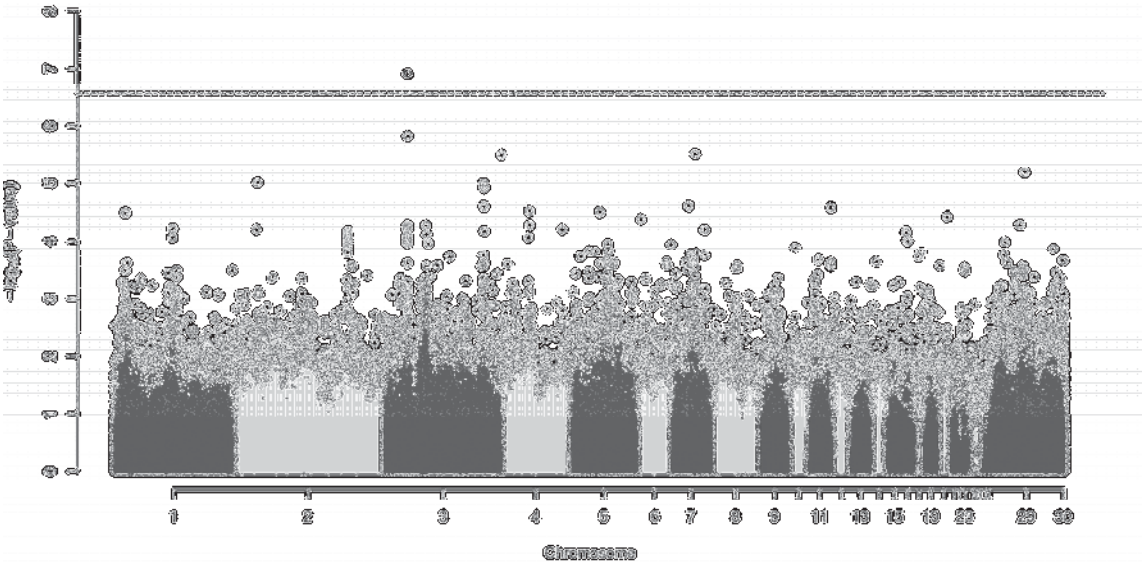
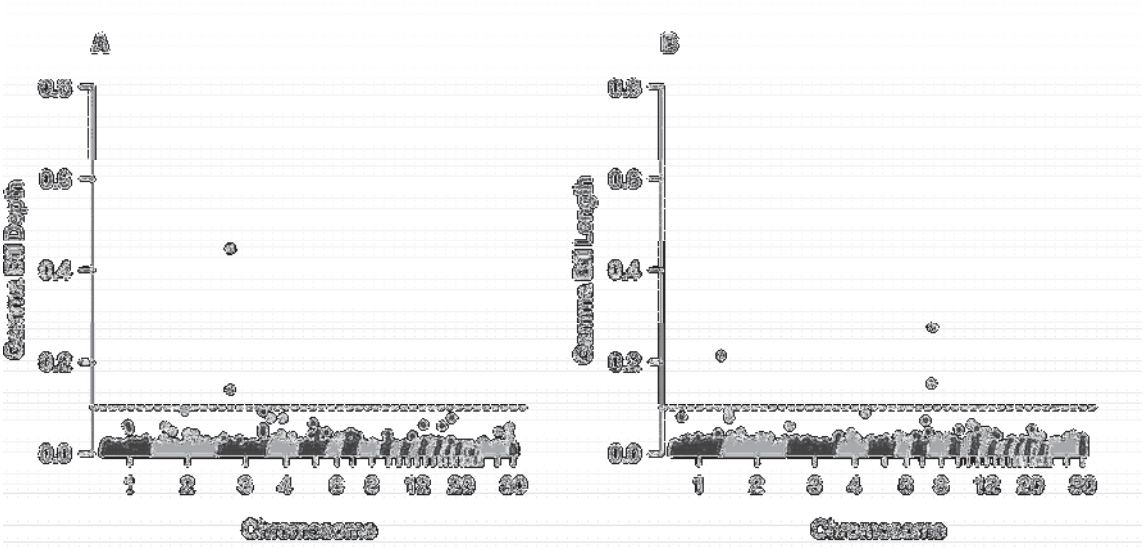


Figure 3: Manhattan plot of RepeatABEL GWA scan for bill depth (N individuals = 1857; N records = 4239) on 181 454 autosomal SNPs. No SNPs on chromosome 16 or Z are included and neither are markers without a position (those with a zero value for chromosome). 'Chromosome' 30 is a linkage group with no chromosome name. Position of markers on the X axis correspond to their bp position on their chromosome. Local significance level when a FWER of 0.05 is controlled for is 3.14^{-7} (dotted line).



999

1000 **Figure 4:** Manhattan plot of BSLMM scans for **A)** bill depth and **B)** bill length. Position of
1001 markers on the X axis corresponds to their bp position on their chromosome. SNPa77348 and
1002 SNPa77303 in the same gene region exceeded the PIP (γ) cut off of 0.1 for bill depth when
1003 body mass was included as covariate in the model. For bill length, three markers had a PIP >
1004 0.1 when body mass was included as a covariate in the model.

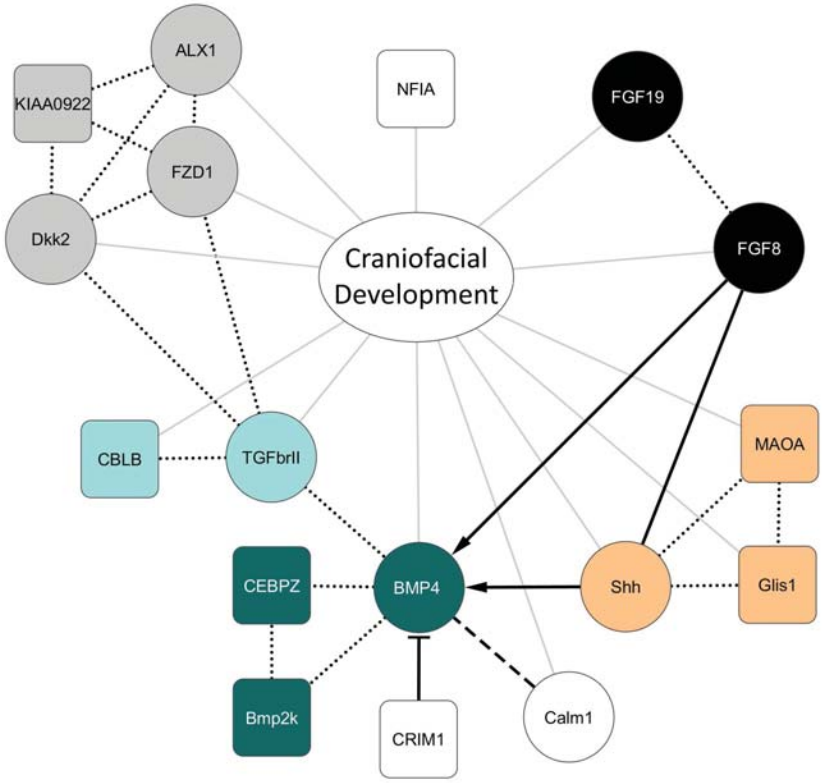
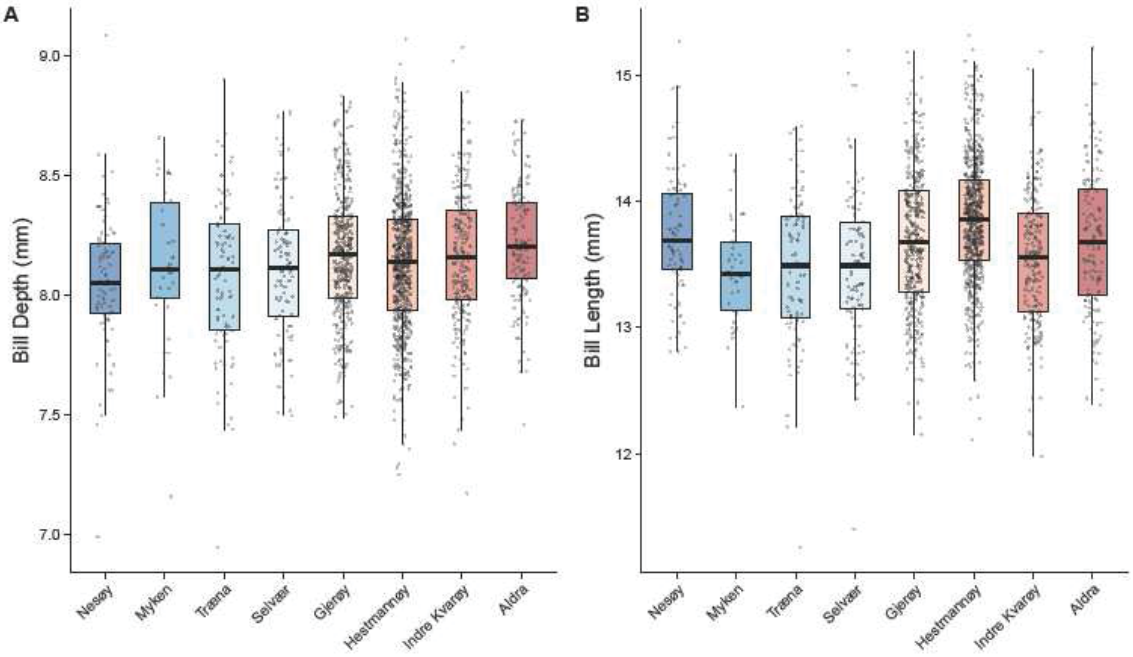


Figure 5: Pathway diagram displaying links between candidate genes (circular nodes) for bill morphology (craniofacial development) used in this study. For details about the functions of these candidate genes in relation to bill morphology, see Table 1. Square nodes represent flanking genes for significant SNPs associated with bill morphology in GWA analyses, and that have previously been linked to craniofacial development or to one of the candidate genes evaluated here. Full grey edges indicate link to craniofacial development, full black edges indicate direct protein-protein interaction, bars and arrows indicate negative and positive regulation, dashed edges indicate co-expression and dotted edges indicate involvement in the same signalling pathway. Nodes are coloured according to main signalling pathway: FZD1, Dkk2, ALX1 and KIAA0922 are all involved in the Wnt pathway, TGFbrII and CBLB are part of the TGF pathway, BMP4, Bmp2k and CEBPZ belong to the BMP pathway, FGF8 and FGF19 are part of the FGF pathway, and Shh, Glis1 and MAOA all belong to the Shh pathway. White nodes indicate candidate genes that do not share a main signalling pathway with another candidate, or for which main signalling pathway is unknown. For literature supporting these links, see Table S11.

1022



1023
1024

1025 **Figure 6:** Individual bill measurements by hatch island using the May 2CY adjusted
1026 phenotypic data for **A)** bill depth and **B)** bill length. Data median, 25th and 75th percentiles and
1027 1.5 interquartile range are shown.

1028