NTNU

Norwegian University of
Science and Technology

Genetic Basis of Variation in Bill Morphology in a Free-Living House Sparrow Metapopulation

Sarah Lundregan

MSc in Biology
Submission date: May 2017
Supervisor: Henrik Jensen, IBI
Co-supervisor: Arild Husby, IBI
Ingerid Hagen, IBI

Norwegian University of Science and Technology
Department of Biology

Foreword

I would first like to extend my thanks to my supervisor, Associate Professor Henrik Jensen, and co-supervisors, Dr. Ingerid Hagen and Associate Professor Arild Husby, for their invaluable input and guidance throughout the course of my study. Your advice has been thought provoking and our discussions have been enjoyable and informative. I would like to thank Professor Mette Langaas, Professor Øyvind Bakke and Kari Krizak Halle (PhD) for their advice on multiple testing correction and for providing the script used for this here. I am also grateful to Dr. Bernt Rønning for his coordination of the house sparrow fieldwork and for my training and to all other participants who made field season an enjoyable experience.

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

Finally, I would like to extend my sincere gratitude to my husband, Hans Kvernsjøli, for his support and encouragement throughout the past two years, I couldn't have done it without you!

Sarah Lundregan

Trondheim, May 2017

Abstract

Bill morphology is an ecologically important trait, which shows substantial phenotypic variation, is highly heritable and is under selection in many bird species. These characteristics underscore the suitability of bill morphology traits for gene mapping analysis and increase the probability that evolution of these traits will occur at a rate conducive to generation of ecological and evolutionary (eco-evo) feedback. Knowledge of the underlying genetic architecture of bill morphology is required to understand the mechanisms driving phenotypic change in these traits and to interpret their involvement in eco-evo cycles. Previous studies have revealed several genes which may influence bill morphology but total number of causal loci, their locations in the genome and magnitude of their individual effects is unknown. Here, the genetic basis of four bill morphology traits (bill depth, bill shape, bill length and bill size) was explored using phenotypic and genome-wide SNP data from a large-scale dataset from an insular house sparrow metapopulation off the coast of Northern Norway. Genomic heritabilities for the above bill morphology measures were estimated using variance component methods, proportion of variance explained was high and broadly in line with previously defined estimates. Chromosome partitioning analyses found significant, positive relationship between chromosome size and proportion of variance explained for the four examined traits, indicating a polygenic basis for bill morphology. Candidate gene and GWA methods were used to search for causal loci. No large effect was observed for any bill morphology trait in candidate gene regions, although weak association was detected for ALX1 and FGF8 for bill shape and size respectively. GWA analysis revealed a significant locus of small effect size on bill depth, situated approximately 1 Mbp away from CRIM1, which exerts upstream control over BMP4. Increase in bill depth with hatch year was observed in the Helgeland metapopulation, as was increase in the frequency of the allele conferring deeper bills at the significant marker. Bill depth may be under positive selection in this metapopulation or covariation of genetic change and environmental fluctuations may be driving its increase. This work illustrates both the difficulties and advantages of performing association studies in natural populations and attempts to determine which approaches may be most likely to result in detection of causal loci.

Table of Contents

Foreword i

Abstract iii

1: Introduction 1

2: Methods 9

2.1: Study Population 9

2.2: Data collection 10

2.3: Genotyping and Quality Control 10

2.4: Principal Components Analysis 11

2.5: Model Selection 12

2.6: Heritability 13

2.7: Genome Partitioning 14

2.8: Single Marker Association Analyses 14

2.8.1: Candidate Genes 15

2.8.2: GWAS 16

2.9: Quantitative Trait Cluster Association Test 16

2.10: Pathway Determination 17

2.11 Result Verification 17

3: Results 18

3.1: Heritability 18

3.2: Genome Partitioning 18

3.3: Candidate Gene Association Analyses 19

3.4: GWAS 21

3.5: QTCAT 22

3.6: Result Verification 22

4: Discussion 23

5: Conclusion 32

References 33

Appendix I

I: Addendum I

Measures Data Adjustment I

Population Structure I

Multiple Testing Correction II

Pathway Determination IV

II: Supplementary Tables V

III: Supplementary Figures VIII

1: Introduction

Recent years have seen increased interest in the study of the interaction between ecological and evolutionary (eco-evo) dynamics, which recognises that adaptive evolution can occur so rapidly that it may influence ecological processes and that ecological changes (eg. in population density or age structure) may lead to selection and affect rate of adaptation (Lowe *et al.*, 2017, Hendry, 2016). Eco-evo feedbacks are bidirectional interactions which may occur at any level of organisation; from genes, individuals and populations to communities and ecosystems. Due to hierarchical organisation, changes at one level may affect others and the link between genes and ecosystems is mediated by population parameters (Hendry, 2012). Population level research is therefore invaluable in gaining mechanistic understanding of eco-evo interactions (Lowe *et al.*, 2017). An excellent example of such feedback in action is how fluctuating environment and density dependence may affect the rate of evolution through stabilising selection toward mean phenotypic values due to differential selection at different carrying capacities (Sæther and Engen, 2015). To date, several studies have quantified the importance of change in heritable traits for population growth and dynamics. For example, Pelletier *et al.* (2007) demonstrated that distribution of body sizes in a population of Soay sheep markedly influenced population growth and in Kinnison *et al.* (2008) it was shown that rapid adaption of invasive Chinook salmon to local habitats influenced fitness and survival and altered population growth rates relative to those of non-adapted counterparts. A study on Yellowstone wolves used an integral projection model to show that environmental change was expected to generate eco-evolutionary change, and that changes in the mean environment are likely to affect this species to a greater extent that changes in extent of variability (Coulson *et al.*, 2011).

In birds, bill morphology is an example of an ecologically important trait, which has been shown to have substantial phenotypic variation, be highly heritable and be under selection in many bird species (James and Zach, 1979, Merilä *et al.*, 2001, Jensen *et al.*, 2008). From Darwin's finches and other passerines to waders and birds of prey, bill form and function vary 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). Species specific and individual differences are the result of trade-offs between various environmental factors, cranial skeletal structure and related musculature (Bock, 1966, van der Meij and Bout, 2004) as well as genetic architecture and its degree of plasticity. Some birds, for example Darwin's finches (Grant and Grant, 2014) and Hawaiian honeycreepers (Tokita *et al.*, 2017), have greater scope for variability whereas raptor bill morphology is constrained by evolutionary integration with braincase and body size (Symonds and Tattersall, 2010). Ultimately, individual

genetic makeup coupled with plastic response toward environmental conditions will dictate phenotype. Bill morphology variation can have important fitness consequences in nature: Subtle differences between individuals can affect their relative foraging efficiency (Temeles *et al.*, 2009), define dietary niche (Soons *et al.*, 2015) or make the difference between survival and starvation (Boag and Grant, 1981). Bill size has also been positively related to nestling provisioning rate (Ringsby *et al.*, 2009, Forstmeier *et al.*, 2001). As a result, birds with particular bill morphologies may experience both increased offspring survival and lifetime reproductive success. A core principle of eco-evo dynamics is that evolutionary change must be rapid enough for selected changes in phenotype to feedback on ecological dynamics on a contemporary time scale; heritability of a trait coupled with the strength of selection on said trait determines how closely rate of evolutionary change can track environmental changes (DeLong *et al.*, 2016). A landmark paper by Grant and Grant (2002) demonstrated that bill morphology of a Darwin's finch species (*Geospiza scandens*) changed significantly several times over the 30 year study period. Calculation of the ratio of phenotypic to population size change showed feedback may occur within 8 generations for the heritable traits bill length and depth in this species (DeLong *et al.*, 2016) and that phenotypic changes (ie. adaptive phenotypic evolution) were equally as important as changes in island productivity for persistence of the species during the study timeframe (Grant and Grant, 2002). Approaches like those outlined below that seek to determine the genetic architecture of traits implicated in eco-evo dynamics may help to define mechanisms behind these processes in nature. Although phenotypes are central to eco-evo dynamics, through selection acting on phenotype and phenotype affecting ecology, evolution only occurs when allele frequencies in the genes underlying these phenotypes change. Understanding eco-evo dynamics at the genetic level therefore requires knowledge of the genetic architecture of the phenotypic traits driving these relationships.

Heritability is a crucial parameter in determining evolutionary response to selection and the extent to which a trait is genetically determined can be evaluated by calculating its environmental and genetic component (Speed *et al.*, 2012). Nowadays heritability is usually calculated using animal (mixed-effect) models, whereas more traditional methods include parent-offspring regression or directly measuring response to selection (Wilson, 2008). As well as calculating heritability, it is of interest to determine the inheritance type of a trait. Partitioning proportion of additive genetic variance explained for a given trait by chromosome facilitates this (Yang *et al.*, 2011a, Yang *et al.*, 2011b). Significant regression with a positive slope of chromosome effect size on chromosome length is expected to be observed for polygenic inheritance, whereas inversion or disruption of this positive linear relationship may occur where large effect quantitative trait loci (QTL) are present (Robinson *et al.*, 2013, Santure *et al.*, 2015). As quantitative traits, bill

morphology measures are expected to be under polygenic control, with signals of selection distributed across many loci (Tiffin and Ross-Ibarra, 2014). Several genes of large effect for bill morphology have, however, been identified in some avian species (eg. Abzhanov *et al.*, 2006, Brugmann *et al.*, 2010, Mallarino *et al.*, 2011).

Where a trait is heritable gene mapping may be performed, which involves identification of the genomic positions of genes coding for the trait (Conner and Hartl, 2004). This was previously accomplished by linkage mapping studies, which first utilised experimental crosses or pedigree information to chart recombination events and hence linkage disequilibrium (LD) between markers and create the genetic linkage map, a time consuming and costly process. (Slate, 2005, Schielzeth and Husby, 2014). Once a linkage map is constructed, it is possible to identify quantitative trait loci (QTL) that connect phenotypic variation among individuals with both phenotype and marker genotype data to single or multiple causal genes (Pardo-Diaz *et al.*, 2015). Traditionally, such linkage mapping studies were performed on model organisms or humans, for whom data was more easily obtainable and testing was comparably accessible. However, success in identifying trait associated genes was often limited, requiring fine mapping by use of an ever-increasing number of crosses (Mackay *et al.*, 2009). Families with detailed family history were used instead of controlled crosses in humans, to perform linkage mapping via estimation of IBD probabilities (Schielzeth and Husby, 2014). Some success in identifying QTL was nevertheless demonstrated in humans (Almasy and Blangero, 2009), several plant species (Bradshaw *et al.*, 1998, Lin and Ritland, 1997) and in *Drosophilla* (Leips and Mackay, 2000), where genes detected were predominantly involved in traits with relatively simple Mendelian inheritance (Conner and Hartl, 2004). Other early mapping studies were based on candidate genes, selected because of *a priori* hypotheses about association with the trait of interest. This reduced cost and time intensiveness of mapping, as smaller regions of the genome were tested, but reproducibility was low and this method does not allow identification of novel genes associated with a trait (Tabor *et al.*, 2002). Recent advances in next generation and high throughput sequencing technologies have paved the way for expansion of genomics research to wild populations, in particular via large scale genotyping on single nucleotide polymorphism (SNP) panels (Stapley *et al.*, 2010, Jensen *et al.*, 2014). These panels can be utilised in genome wide association studies (GWAS), which exploit historical LD between causal genes and genetic markers to uncover trait associations. Because of this, GWAS is sometimes called 'linkage disequilibrium mapping' (Mackay, 2004). As LD decays exponentially with increasing genetic distance, causal genes are more likely to be near to markers which score highly for a trait (Wray *et al.*, 2013). Marker densities of more than 500k are possible with high density SNP arrays, affording them greater power than linkage mapping studies in line with

increasing resolution. (Ha *et al.*, 2014). Many methods for detecting QTL are available, including family based association tests centred on transmission of alleles within pedigrees, linear mixed model (LMM) approaches or non-linear methods like Maximum Quasi-Likelihood (Euahsunthornwattana *et al.*, 2014). LMM association approaches, like those implemented in GenABEL (Aulchenko *et al.*, 2007) offer higher power than family based methods, are able to control genomic inflation (artificial differences in allele frequencies due to population structure or cryptic relatedness, which cause inflated *P* values) to an appropriate level and are computationally efficient. Extensions of the LMM approach to permit repeated measurements, as in RepeatABEL (Rönnegård *et al.*, 2016), allow within-individual variation to be fully accounted for and may produce more accurate estimates.

This era of genome-wide association studies calls into question whether candidate gene approaches are still necessary or desirable, but candidate gene studies may result in increased statistical power to detect extant associations due to lower number of markers tested (Tabor *et al.*, 2002). Power depends on sample size and LD between SNPs in the region, combined with minor allele frequency (MAF) and effect size of the associated marker (Mackay *et al.*, 2009). Trait associated genetic polymorphisms with low MAF or in regions of low LD are unlikely to be detected by GWAS but may be via association testing of candidate genes. The regional candidate gene approach is also useful to confirm previously demonstrated trait associations and, with deep resequencing of candidate regions, to determine causative variants (Wilkening *et al.*, 2009). Confirmation of the effects of candidate genes in separate species or populations reveals the extent of generality of gene functions across different groups and is important for wider understanding of genetic architecture. This approach was successfully used in Haag *et al.* (2005) where variation at the *pgi* locus had a direct effect on dispersal rate and thereby on metapopulation dynamics in the Glanville fritillary butterfly. Follow up studies confirmed the importance of this gene for flight, dispersal, and population growth (Hanski and Saccheri, 2006, Niitepõld *et al.*, 2009), linking genetic and ecological dynamics in this species. Novel multi-marker methods, which can increase power in association studies, may be another way to detect QTL which are missed by traditional single marker association methods. One such method is quantitative trait cluster association testing (QTCAT), which searches for clusters of markers significantly associated with a given trait. For complex traits under polygenic control, multi-marker association methods are expected to be superior to single-marker techniques as they mitigate the need for population structure correction by accounting for correlation between markers at the same time as associating them with the phenotype (Klasen *et al.*, 2016).

Challenges regarding adequate sample size, unpredictable LD patterns and ability to define sufficient high quality markers are amplified in wild populations of non-model species, due to environmental heterogeneity, population structure, and lack of previous genetic studies, consequently such work is rare (Jensen *et al.*, 2014). However, studies in natural populations are important, as they allow us to examine the genetic basis of ecologically important traits in their natural context (Slate *et al.*, 2010). Environment-phenotype interactions are often complex; large scale studies in natural populations allow year and age effects over time to be assessed via repeated individual measures (Clutton-Brock and Sheldon, 2010). They also aid evaluation of the degree to which results from laboratory studies in model species can be extrapolated to natural populations and non-model species, and are invaluable to inform management decisions (Stapley *et al.*, 2010). Recently, GWA approaches have been successfully utilised in several studies in natural vertebrate populations (eg. Johnston *et al.*, 2013, Husby *et al.*, 2015, Santure *et al.*, 2015, Johnston *et al.*, 2011, Barson *et al.*, 2015, Johnston *et al.*, 2014). As association mapping relies on historical recombination events it does not require multigenerational genetic data, however, long term monitoring systems in natural populations are well suited to association study as they monitor phenotypic and population data over several years and allow age effects and changes over time to be linked to individual genotype (Clutton-Brock and Sheldon, 2010). Ideally a species should also possess qualities such as pervasiveness and interesting adaptive history which make it applicable for wider ecological study (Ellegren, 2014, Slate *et al.*, 2010), as in the 'ecological models' the collared flycatcher (Ellegren *et al.*, 2012) and threespine stickleback (Jones *et al.*, 2012).

A recent study of Darwin's finch species used whole-genome re-sequencing of 120 representative individuals to indicate a polygenetic basis for changes in bill morphology and to identify a key haplotype, ALX1-b, associated with blunt bills (Lamichhaney *et al.*, 2015). Other avian studies have revealed several more genes affecting bill morphology: Bone morphogenic protein 4 (BMP4) and calmodulin (Calm1) are differentially expressed in *Geospiza* species with divergent bill morphologies and regulate bill development in chicken embryos (Abzhanov *et al.*, 2006, Abzhanov *et al.*, 2004). Transforming growth factor-beta receptor type-2 (TGFbrII) has been shown to be differentially expressed in developing premaxillary bone of species with differing bill morphologies (Mallarino *et al.*, 2011). Expression of Wnt signalling pathway members, including Dickkopf Wnt signalling pathway inhibitor 2 (Dkk2) and frizzled-1 (FZD1), is upregulated in species with broad bills and downregulated in those with narrow bills (Brugmann *et al.*, 2010). Fibroblast growth factor 8 (FGF8) interacts with other proteins in the frontonasal ectodermal zone, including sonic hedgehog (Shh), to induce expression of BMP4 at the proper domain on the developing bill, potentially influencing its morphology (Abzhanov and Tabin, 2004, Wu *et al.*,

2004). The fibroblast growth factor signalling pathway is important during neural crest (NC) migration and pharyngeal endoderm formation, pharyngeal endoderm is important for regulation of craniofacial morphogenesis, especially bill formation, in developing chicks (Haworth *et al.*, 2007). After NC migration both FGF8 and fibroblast growth factor 19 (FGF19) maintain strong expression in the pharyngeal endoderm (Kumar *et al.*, 2012). Functional studies like those outlined above often involve developmental knockout or knockdown approaches and are the benchmark criterion by which association between genes and traits is confirmed (Barrett and Hoekstra, 2011).

The house sparrow (*Passer domesticus*) is an ideal candidate for gene mapping studies of traits potentially involved in eco-evo dynamics, as much is known about its biology and ecology (Anderson, 2006). A natural house sparrow metapopulation exists in Helgeland, northern Norway and has been studied extensively since 1993, allowing establishment of a large sample dataset. The insular metapopulation covers approximately 1600 km^2 and is characterized by relatively low dispersal between islands, allowing examination of selection, genetic drift, and gene flow (Holand *et al.*, 2011, Jensen *et al.*, 2013, Pärn *et al.*, 2012, Jensen *et al.*, 2008). Genetic and pedigree information is available for the majority of individuals, alongside morphological and life history data including measurements of bill depth and length (eg. Jensen *et al.*, 2003, Jensen *et al.*, 2004, Pärn *et al.*, 2009, Ringsby *et al.*, 2009). Within this system, bill size has been found to correlate with rate of feeding of offspring (Ringsby *et al.*, 2009). The same study concluded that parental investment, represented by nestling feeding rate, may increase number of successful recruits as well as bolstering chances of surviving to recruitment stage. Jensen *et al.* (2004, 2008) showed that bill morphology in adult house sparrows is related to survival and reproductive success, and hence that it is under positive selection. It has been shown that bill depth and length measures are highly repeatable within the Helgeland system and that the traits are also highly heritable (Jensen *et al.*, 2003, Jensen *et al.*, 2008). Evidence of a large, additive genetic component to phenotypic variation in bill depth and length indicates suitability of the traits for gene mapping analysis. Combined with the fact that bill morphology is under selection (Jensen *et al.*, 2008), this increases the chances that evolution will occur at a rate conducive to creation of eco-evolutionary feedback (DeLong *et al.*, 2016). Genes which have previously been related to bill morphology (see above) are therefore of particular interest for association studies in the Helgeland metapopulation.

Genetic data for study individuals is available from 1993-2014 and a 10k SNP microarray was recently developed and successfully used to detect population stratification relating to island divisions in the Helgeland system (Hagen *et al.*, 2013). This information was used to develop a linkage map which was utilised to assemble the house sparrow reference genome into chromosomes (Elgvin *et al.*, in press.). Subsequently, a 200k Affymetrix Axiom SNP array (Hagen

et al., in prep.-b) was developed. The array includes SNPs within and in the genomic vicinity to candidate genes for a diverse range of phenotypic traits in model species and in humans, totalling 16,827 SNPs on the array. The remaining 183,173 SNPs are distributed evenly across the house sparrow genome. SNPs within or in the vicinity of nine genes which, based on the above outlined literature, may be related to bill morphology were included on the array: ALX1, BMP4, Calm1, Dkk2, FGF8, FGF19, FZD1, Shh and TGFbrII. Functions of these candidate gene were predicted using online databases UniProt (Uniprot, 2017) and Ensembl (Yates *et al.*, 2016). Relationships between these candidate genes and proteins which directly interact with or regulate them are shown in Fig. 1.

Here, the genetic architecture of four bill morphology traits: bill depth, bill shape, bill length and bill size, was investigated using data from the Helgeland metapopulation of house sparrows. First, heritability estimates were examined to determine whether the high density, genome-wide panel of SNPs captured the additive genetic variance previously estimated for bill morphology in house sparrows. If this is the case it would suggest that at least some typed SNPs are in LD with genes affecting the traits. Subsequently, chromosome partitioning analysis was performed to examine on which chromosomes the genes causing variation in these traits reside. As SNPs within and close to a number of candidate genes for bill morphology have been typed in this study population, a candidate gene approach was then used to examine whether any of these genes explained variation in bill morphology. This was followed by GWA analysis of the whole dataset to determine whether: 1) this genome-wide approach would support the findings from candidate gene analyses and 2) to determine whether additional regions and candidate genes for future studies could be detected. GWAS was performed using single marker methods, GRAMMAR-gamma for single measurements and RepeatABEL for repeated individual measures, as well as with the multi-marker method QTCAT, which may have greater power than the previous methods to detect links between polygenic traits and associated gene regions. Significant markers in the Helgeland dataset were tested for enrichment in an independent dataset from a second set of house sparrow populations, using the RepeatABEL method, to determine whether significant associations in the Helgeland dataset were maintained in these Southern populations. Finally, a pathway map was created for a potential causal gene to illustrate its conceivable mechanisms of action and possibly indicate new candidate genes for bill morphology.

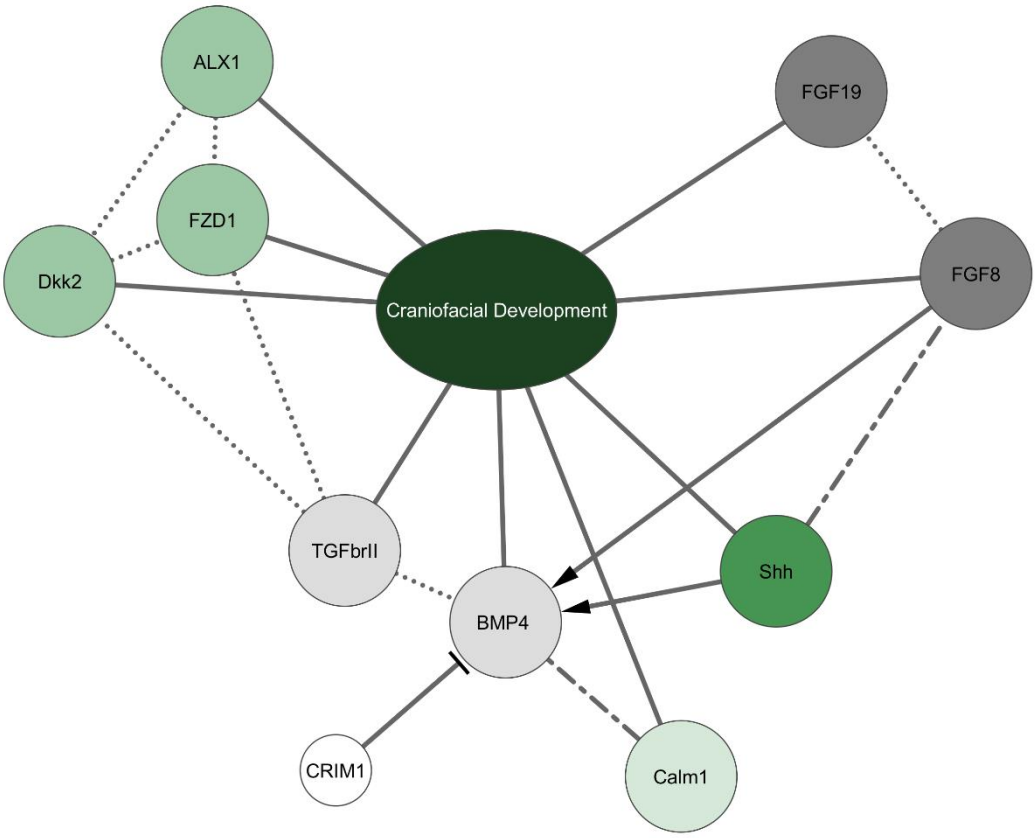

Figure 1: Pathway diagram displaying links between candidate genes for bill morphology (craniofacial development) used in this study. Full edges indicate direct protein-protein interaction, dotted edges indicate involvement in the same signalling pathway, dashed edges indicate co-expression, bars and arrows indicate negative and positive regulation respectively. Nodes are coloured according to main signalling pathway; FZD1, Dkk2 and ALX1 are all involved in the Wnt pathway, TGFbrII and BMP4 belong to the BMP signalling pathway, FGF8 and FGF19 are part of the Fgf signalling pathway. Calm1 and Shh belong to the calmodulin and sonic hedgehog signalling pathways, which are linked to but not part of the signalling pathways for the other genes. Note that CRIM1 was not among the 9 candidate genes, but was included in this figure as it falls within ≈ 1 Mbp of a significant SNP found in RepeatABEL GWAS on bill depth and it negatively regulates BMP4.

2: Methods

2.1: Study Population

This study utilises data from a long-term study of a house sparrow metapopulation, consisting of 18 insular subpopulations in an archipelago at the Helgeland coast in Northern Norway (Fig. 2). The house sparrow is a non-migratory passerine bird, which is sexually dimorphic and with a lifespan of up to 9 years in Northern Norway (Jensen *et al.*, 2004). The species is sedentary and exhibits low dispersal rates, in the insular Helgeland system only around 10% of fledglings that recruit into the breeding population are dispersers (Pärn *et al.*, 2009, Pärn *et al.*, 2012). Up to 20% of fledglings survive to recruitment (Sæther *et al.*, 1999). House sparrows are often associated with human settlements (Anderson, 2006) and are predominantly found at agricultural or residential sites in the study system. This improves sampling efficiency and, in conjunction with low dispersal rates, allows individuals to be monitored from hatching over consecutive years until they die (Billing *et al.*, 2012, Jensen *et al.*, 2004, Jensen *et al.*, 2008, Pärn *et al.*, 2009). This enables repeated morphological measurements to be taken, which capture within-individual variation in bill morphology (Davis, 1954, Greenberg *et al.*, 2013) (Appendix I; Table I).

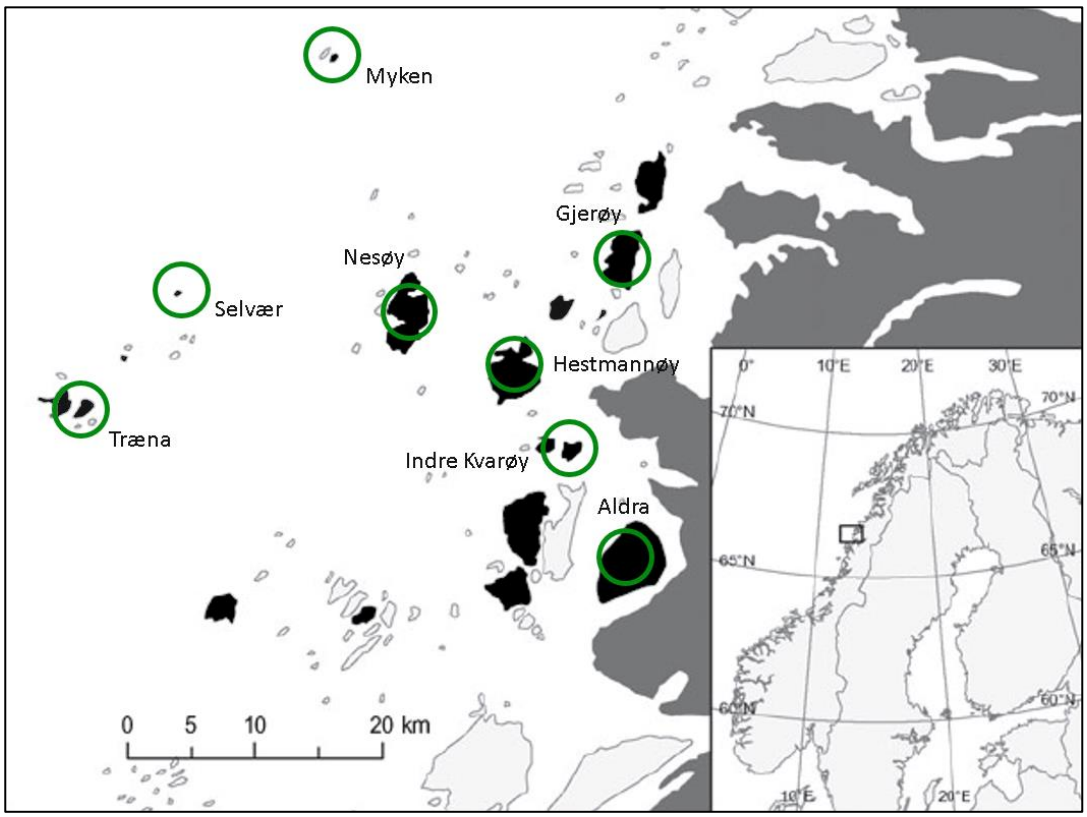

Figure 2: Islands included in the house sparrow metapopulation study system, Northern Norway (66°N, 13°E). Islands shaded black have been continuously followed since monitoring began. The eight populations used in this study are circled in green.

2.2: Data collection

Sampling has been ongoing on from 1993 to present in all insular subpopulations in the metapopulation. However, due to financial constraints, data used in this study includes only adult birds recorded on 8 of the study islands during 6 (5 islands), 11 (2 islands) or 10 (1 island) years (Table 1). My participation in sampling thus far was from May to July, 2016. Adult house sparrows were captured using mist nets for blood sampling and morphological measurements, including measurements for bill depth and bill length. Slide callipers were used to measure these two bill morphology traits to the nearest 0.01mm (Jensen *et al.*, 2004, Jensen *et al.*, 2008). During the breeding season, May – August (Husby *et al.*, 2006), nestlings were collected from nests between the age of 8-13 days, banded for identification with three coloured plastic rings and one metal ring displaying identification number, then morphological measurements and blood samples taken before replacement into nests. Any un-banded adult or fledged birds captured in mist nests were banded as above. Only adult measures from each individual were utilised in this study (Jensen *et al.*, 2004, Jensen *et al.*, 2008) and measurements taken outside of the breeding season were not used to reduce seasonal intra-individual variation in bill morphology (Anderson, 2006). A linear regression method was used to adjust measurements collected by different fieldworkers to an internal standard and to control for age and seasonal effects in analyses where one measure per individual was needed (Jensen *et al.*, 2008). See Appendix I; Measures Data Adjustment.

Table 1: Period for which all recorded adult individuals on each island were genotyped on the 200k SNP microarray, number of SNP genotyped individuals which passed quality control on each island and number of measurements per island.

Island	Years	Individuals	Measurements
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**

2.3: Genotyping and Quality Control

Study individuals (Table 1) were genotyped on a custom Affymetrix Axiom 200k SNP array (Hagen *et al.*, in prep.-b). The array was developed based on the reference genome for *P. domesticus* (Elgvin *et al.* in press.) and whole genome sequencing of 33 individuals from 15 populations across Norway and Finland, with at least one male and one female from each population. Linkage

disequilibrium (LD) decay analysis based on data from the 10k SNP array indicated that 200k SNPs would be sufficient to ensure LD between a potential QTL and a marker (Hagen *et al.*, 2013). Furthermore, preliminary analysis of LD based on 200k SNP genotype data from the 8 study populations suggest that LD decays to background levels between 15-20kb (Hagen *et al.*, in prep.-a). Between 1993 and 2016, 14,100 individuals (nestlings, fledged juveniles, and adults) were recorded on one of the 8 included islands in the Helgeland metapopulation (Nesøy, Myken, Træna, Selvær, Gjerøy, Hestmannøy, Indre Kvarøy and Aldra) (Fig. 2). 3150 individuals who were adults on these study islands between 1998-2013 were genotyped on the 200k array; of these individuals 1958 had phenotypic data for bill depth and bill length. Of the 200,000 SNP markers on the array, 184,804 were categorised as PolyHigh Resolution, Affymetrix's highest quality class which requires good cluster resolution and at least two copies of the minor allele. Only markers ranked as this category were used in further quality control analyses.

Quality control was performed using the GenABEL R package (Aulchenko *et al.*, 2007) The dataset was subset to include only individuals with bill morphology data (1958 individuals) prior to quality control to ensure the final dataset adhered to quality control parameters. 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 1).

2.4: Principal Components Analysis

Principal components analysis using the prcomp base function in R was run on the age and month adjusted (Appendix 1; Measures Data Adjustment), centred bill depth and bill length measures to derive two principal components: PC1 and PC2 (Table 2). As prcomp utilises single variance decomposition, data was mean-centred (equivalent to eigen-decomposition performed on the covariance matrix) to ensure the first principle component described the direction of maximum variance rather than the data mean (centring ensures PC1 runs parallel to the main direction of the data cloud). Here, PC1 is analogous to bill size, as depth and length have a direct relationship; PC2 is analogous to bill shape, as depth and length have an inverse relationship. No significant difference between the sexes was observed ($P > 0.1$). This process was done separately for single and repeated measures data.

Table 2: Loadings for each principal component and the proportion of variance (± SD) in bill morphology they explain. PC1 is analogous to bill size and PC2 to bill shape. Bill size explains most variation in bill morphology between individuals.

Data		age1billD	age1billL	Proportion of Variance	SD
Single Measures	**PC1**	0.200	0.980	0.820	0.549
	PC2	0.980	-0.200	0.180	0.257
Repeated Measures	**PC1**	0.186	0.983	0.821	0.596
	PC2	0.983	-0.186	0.179	0.278

2.5: Model Selection

Preliminary exploration of phenotypic data was carried out to identify factors and covariates that may influence bill morphology. This was performed separately for single and repeated measures data, as packages used for genetic association testing required different types of phenotypic input. For single measures data, where repeated measurements for each individual were adjusted to age 1 in May using a predictive mixed model approach (Appendix I; Measures Data Adjustment), sex, island and hatch year showed effect on bill depth and PC2 (hereafter referred to as bill shape). Difference between sexes for bill length and PC1 (hereafter referred to as bill size) was small, but sex was included as a factor in model selection nevertheless, as sex has previously been found to affect all bill morphology traits (Jensen *et al.*, 2003, Jensen *et al.*, 2008). The best linear model with smallest AICc (Burnham, 2002) value for bill depth and bill shape included sex, island and hatch year as fixed factors. For bill length and bill size the best model did not include sex as a fixed factor (Appendix II; Table III). Preliminary exploration of the repeated measures data identified sex, island, age, and measurement month as potentially affecting bill morphology. Before further analysis, repeated measures data was pruned to remove outliers and to include only measurements from May to August. A 7+ age category was also created to mitigate effects on accuracy caused by low numbers of individuals in the upper age classes. Models used in AICc model selection were constructed using the R package, lme4 (Bates *et al.*, 2015). The best model for bill depth, bill size and bill shape included sex, island and month as fixed factors, age as a covariate, and ID as a random factor. The best model for bill length was similar, except age was not included as covariate (Appendix II; Table III). Interaction effects were not included in any model selection, as prediction of how interactions will affect genetic associations is complex, especially where range of variables does not naturally encompass zero, running the risk of controlling away any genetic effects (Aschard, 2016).

Genes influencing bill morphology may influence the bill alone, alternatively increases in bill dimensions may occur as a result of larger body size, usually strongly correlated with higher body

mass (Chaves *et al.*, 2016). In the current dataset, both bill depth ($r = 0.18$, $P < 0.001$) and bill length ($r = 0.17$, $P < 0.001$) showed significant correlation with body mass. Inclusion of body mass as a covariate in any gene mapping analyses would thus result in tests for associations between bill morphology and marker after variation in bill morphology explained by variation in body mass was controlled for (ie. tests would be for bill morphology relative to body mass). Genes which affect bill morphology through overall skeletal development are, however, of interest, as increased fitness due to greater bill dimensions (Forstmeier *et al.*, 2001, Ringsby *et al.*, 2009) may often be gained irrespective of the causal genetics behind altered bill morphology. Body mass was not included as a covariate in any model for this reason.

2.6: Heritability

All heritability estimates derived in this study make use of variance component estimation methods. The polygenic function used in conjunction with the GRAMMAR-gamma method in GenABEL automatically outputs a value for narrow-sense heritability (Aulchenko *et al.*, 2007). Heritability was calculated manually for RepeatABEL and GCTA methods using variance component estimates and the formula $h^2 = V_a/V_a+V_{pe}+V_r$, where V_a is additive genetic variance, V_{pe} is permanent environmental variance and V_r is residual variance. Here, the kinship matrix, which specifies the covariance structure for the mixed models, is calculated using IBS at all markers in linkage equilibrium (LE) after excluding markers within and close to candidate genes. Utilisation of IBS gives several advantages over use of a pedigree to determine relatedness including elimination of inaccuracy due to recombination-segregation induced noise, incompleteness of pedigree or cryptic relatedness (Speed *et al.*, 2012). Of the three methods used to derive heritability estimates, only RepeatABEL permits inclusion of repeated individual measures. Allowing repeated measures means variables like age and month can be included, fixed effects alter h^2 estimates by removing V_{pe} and this should be kept in mind when evaluating trait heritability estimates (Wilson, 2008). Inclusion of repeated measures improves accuracy of heritability estimates by taking into account within-individual phenotypic variation, which often reduces V_a and h^2 in line with decreasing repeatability of measurements. Observed reductions in V_a when using repeated measures confirm increased power to correct for circumstances which obscure estimation of V_a, such as high LD, environmental covariances or selection (Åkesson *et al.*, 2008). The above observations indicate that estimates of heritability produced by RepeatABEL will be most accurate.

2.7: Genome Partitioning

A command line software tool, GCTA (Yang *et al.*, 2011a), was used to conduct genome-wide complex trait analysis. GCTA allows detection of 'missing heritability' which is often due to many SNPs with small effect that GWAS cannot detect. Contributions of SNPs to variation in bill morphology should be correlated between individuals of similar genotype. GCTA fits chromosomal genomic relationship matrices (GRMs) as random effects to calculate the variance explained by all SNPs on a chromosome rather than testing association of specific SNPs to a trait. Here, average information Restricted Maximum Likelihood (AI REML) models (Gilmour *et al.*, 1995) with multiple GRMs fitted as random effects were used to give the proportion of variance in each of the four bill morphology traits explained by each chromosome. As the GCTA software does not support repeated observations, phenotype data adjusted to age one in May was utilised. Problems with model non-convergence were addressed by successively removing the smallest chromosomes. Controlling for covariates exacerbated model non-convergence meaning more chromosomes had to be removed for the model to converge, as did inclusion of principle components to define population structure, therefore these components were not used in analyses. Proportion of variance explained by each chromosome was subsequently plotted against chromosome size (Mbp) to visualise and determine genetic architecture of bill morphology. Polygenic traits are expected to be represented by positive relationship between chromosome effect size and chromosome length, this positive linear relationship may be disrupted or inverted where large effect QTL are present (Robinson *et al.*, 2013). Hence, linear regressions were fitted between variance explained by each chromosome and chromosome length, using the stats package in R (R Core Team, 2017), to test significance of these relationships (Santure *et al.*, 2015).

2.8: Single Marker Association Analyses

Prior to association analysis, identity by state (IBS) was calculated for individuals with bill morphology data, using all passed autosomal markers minus SNPs in and close to candidate genes for bill morphology. The indep function in PLINK 1.9 with recommended parameters 50 5 2 was used to produce a list of markers in approximate LE (Purcell *et al.*, 2007). This list of markers was used to compute the IBS (GRM) matrix, both for GenABEL and RepeatABEL analyses, as LD between markers can confound relatedness estimates (Eu-ahsunthornwattana *et al.*, 2014, Lopes *et al.*, 2013, Santure *et al.*, 2010) and inclusion of candidate markers when calculating the GRM can cause loss of power (Yang *et al.*, 2014). The reverse distance matrix, 0.5-f, was then computed, classic multidimensional scaling (Gower, 1966, Mardia, 1978) performed and k-means clustering (Hartigan and Wong, 1979) implemented to define first principal components of variation in the

distance matrix for visual determination of population structure and identification of outliers (Appendix III; Fig. I). The same reduced set of markers utilised for GRM calculation was used to estimate the inflation factor (λ) for use in genomic control (Appendix II; Table IV). *P*-values were subsequently corrected for this inflation factor, if it was greater than 1.00, to account for population stratification (Devlin *et al.*, 2001, Hinrichs *et al.*, 2009). Significance thresholds were calculated using a custom method (pers. comm. Kari Halle, Mette Langaas, 2017), based on the order 3 method outlined in Halle *et al.*, (2016) and adapted for use on traits with Gaussian distribution (Appendix I; Multiple Testing Correction). Prior to calculation of significance thresholds all SNPs with correlation > 0.999 were removed, order 3 and full order methods were then used to determine the local alpha level for GWA and candidate gene association models respectively (Appendix I; Multiple Testing Correction).

2.8.1: Candidate Genes

SNPs within 55 Kbp upstream or downstream of candidate genes for bill morphology were selected for analysis. An adapted candidate gene approach, which utilised the IBS matrix calculated using all SNPs in linkage equilibrium rather than a relatedness matrix calculated using gene-specific SNPs, was employed to determine the extent to which SNP variation on or close to these functionally relevant genes can explain variation in bill morphology. Models were first fitted using the variance covariance matrix produced using the polygenic function combined with the GRAMMAR-Gamma approach from the R package GenABEL (Aulchenko *et al.*, 2007). As GRAMMAR-Gamma only accepts one value per individual, single measures data with bill morphology measurements adjusted to age one in May was used here (Appendix I; Measures Data Adjustment). A second R package, RepeatABEL (Rönnegård *et al.*, 2016), was also used to estimate candidate gene associations. This package permits phenotypic data with replicates, maximising power to detect genetic associations with a trait. The GRM for these models was fitted using the GenABEL method, implemented in RepeatABEL. Association analysis in RepeatABEL involves estimation of the covariance matrix (V) using an animal model fitted via a hierarchical generalized linear model (HGLM) as a first step, followed by the rGLS model which fits each marker as a fixed effect, with correlation structure (GRM) and permanent environmental effects (ID) fitted as random effects.

2.8.2: GWAS

GWAS for each trait were conducted using GenABEL and RepeatABEL, as for the candidate gene association tests. Autosomal SNPs only were used for GWAS, as analysis of sex chromosomes requires special handling during quality control and association analysis (Wise *et al.*, 2013). The annotated house sparrow genome (Elgvin *et al.*, in press.) was used in conjunction with a custom code (Pers. comm. Henrik Jensen, 2017) to determine whether significant SNPs were in exonic or intronic parts of genes or within 55 Kbp of an annotated gene. The annotated collared flycatcher genome assembly FicAlb1.5 (Kawakami *et al.*, 2014b) was also used to determine position of significant SNPs in relation to known genes in *Ficedula albicollis*. BLAST search was performed via Ensembl (Yates *et al.*, 2016) using a sequence derived by alignment against the house sparrow genome, which spanned 1000bp either side of the significant SNP.

2.9: Quantitative Trait Cluster Association Test

Genome wide multi-marker association tests were carried out using the R package QTCAT (Klasen *et al.*, 2016). Linear mixed models (as form the basis of the GRAMMAR-Gamma and RepeatABEL methods) correct for genetic background and population structure simultaneously by using the GRM to model genetic covariance between individuals. Estimation of the random effect assumes infinitesimal genetic background, contributed to by many loci with small effect. Use of random effects in this manner can cause true associations to be masked by genetic background, as the hypothesis tested is restricted to detection of locus effect on phenotype independent of population structure or genetic background. QTCAT overcomes the need for population structure correction by simultaneously considering correlations between markers whilst making multi-marker associations. The method also mitigates the need to correct for genetic background, as all markers are simultaneously associated to the phenotype. Both these features may increase power to detect causal loci by allowing testing of the unrestricted hypothesis; whether a specific locus has a significant effect on the trait of interest.

As QTCAT does not accept missing genotypes (Klasen *et al.*, 2016), the Java programme LinkImpute (Money *et al.*, 2015) was used prior to implementation of QTCAT to impute missing genotypes. LinkImpute accepts files in PLINK's ped and map format. Here input files were subset beforehand using PLINK (Purcell *et al.*, 2007) to contain only individuals and markers which passed GenABEL quality control. LinkImpute is based on a *k*-nearest neighbour imputation method (LD-*k*NNi), a major advantage of which is that it is not reliant on phasing or ordered markers. LD-*k*NNi takes into account LD between SNPs when choosing nearest neighbours, only

SNPs in high LD with the SNP to be imputed are used to determine nearest neighbour and imputation weightings. In control tests, LinkImpute performed well, demonstrating higher accuracy and faster runtime than existing methods (Money *et al.*, 2015). Fixed factors used were the same as for GenABEL analysis: sex, island and hatch year were included as fixed factors for bill depth and bill shape; island and hatch year were included as fixed factors for bill length and bill size.

2.10: Pathway Determination

Literature search and online methods iHop (Hoffmann and Valencia, 2004) and STRING (Szklarczyk *et al.*, 2015) were utilised to determine potential mechanisms of action for genes implicated in controlling bill morphology in GWAS and QTCAT. Cytoscape 3.2.1 (Shannon *et al.*, 2003) was used to produce pathway diagrams (Appendix I; Pathway Determination, Appendix III; Fig. V).

2.11 Result Verification

Significant results were verified in an independent dataset from a second set of house sparrow populations on three islands, Leka, Vega and Lauvøya, situated approximately 150 km South of the Helgeland system, off the coast of mid-Norway. Quality control was carried out as in the Helgeland dataset and the GRM for use in RepeatABEL was created using SNPs in approximate LE and with no markers from the candidate set to be tested. For significantly associated candidate genes, the same sets of SNPs as tested in the Helgeland metapopulation were used. For our GWAS significant marker, SNPs 55 Kbp either side of the marker were tested. Best models selected using AICc model selection were the same as for RepeatABEL analyses in the Helgeland metapopulation system (Appendix II; Table III).

Unless otherwise stated, all statistical analyses were performed using R version 3.3.3 (R Core Team, 2017).

3: Results

3.1: Heritability

All four bill morphology traits showed relatively high heritabilities, ranging from 0.35 in the RepeatABEL estimate for bill depth to 0.55 in the GRAMMAR-gamma estimate for bill size. Heritabilities of the four studied bill morphology traits from GRAMMAR-gamma, RepeatABEL and GCTA are summarised in Appendix II; Table V. RepeatABEL estimates are most precise and are utilised as the definitive trait heritability values in this study. Heritability estimates derived using this method for bill depth, bill shape, bill length and bill size are 0.35, 0.38, 0.38, and 0.40 respectively, calculated from GWAS variance components (Table 3).

Table 1: Variance components of RepeatABEL GWAS for bill morphology traits: depth, shape, length, and size, alongside trait heritabilities.

Source of Variance:	Additive Genetic	Permanent Environment	Residual	h^2
Bill Depth	0.030	0.020	0.035	0.35
Bill Shape	0.030	0.014	0.034	0.38
Bill Length	0.112	0.077	0.108	0.38
Bill Size	0.125	0.078	0.107	0.40

3.2: Genome Partitioning

Chromosome-wise partitioning of variance using GCTA showed regressions between chromosome size and proportion of the variance explained were significant for all bill morphology traits: bill depth (Fig. 3A, $R^2 = 0.17$, $F(1, 22) = 4.59$, $P = 0.04$), bill shape (Fig. 3B, $R^2 = 0.27$, $F(1,18) = 7.93$, $P = 0.01$), bill length (Fig. 3C, $R^2 = 0.36$, $F(1,17) = 9.66$, $P = 0.006$), and bill size (Fig. 3D, $R^2 = 0.23$, $F(1, 25) = 7.68$, $P = 0.01$). Consequently, this gives substantial evidence toward bill morphology being polygenic in nature in this study population. However, despite the positive relationships between chromosome effect size and chromosome size, some chromosomes, like chromosome 7, appear to explain a relatively large proportion of the variation for all bill morphology traits (Fig. 3). Furthermore, other chromosomes, like chromosomes 5 and 18, appear to explain a disproportionately large amount of the variation in bill depth and bill shape, but not of the variation in bill length and bill size (Fig. 3A, 3B). Similarly, chromosomes 3 and 11 appear to explain a disproportionately large amount of the variation only for bill length and bill size Fig. 3C, 3D.

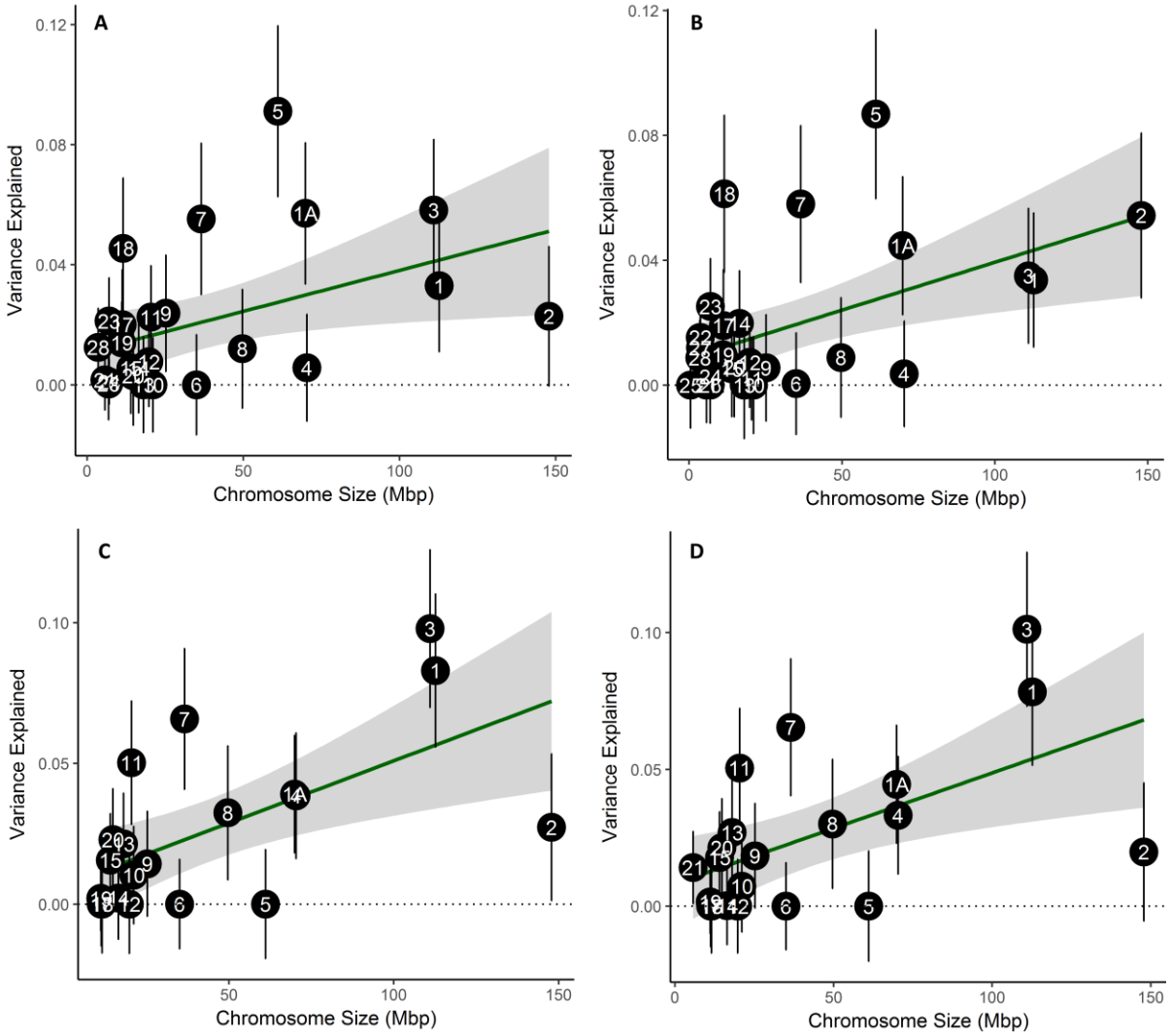

Figure 3: Relationship between explained variation in bill morphology (± SE) and chromosome size (Mbp). **A)** Bill depth, **B)** bill shape, **C)** bill length, **D)** bill size. 1857 individuals and 183,109 SNPs were used to partition V_a in bill morphology traits across chromosomes.

3.3: Candidate Gene Association Analyses

Following adjustment for genomic inflation (estimated using autosomal SNPs in approximate LE and without inclusion of SNPs from candidate genes for bill morphology) one SNP within 55 Kbp of ALX1, SNPa174071, passed the single-gene significance threshold for bill shape (P = 0.0009, α_{loc} = 0.001) using the GRAMMAR-gamma method in GenABEL (Fig. 4). Appendix II; Table VI gives summary statistics for the top 5 SNPs from this scan, SNPs are in significant LD (Fig.4, LD between all SNP combinations >0.94). The association was, however, non-significant at the threshold for all candidate genes combined (α_{loc} = 0.00009, based on the 768 SNPs in or within 55 Kbp of candidate genes for bill morphology and which passed quality control). No other SNPs within or near candidate genes showed significant association with bill morphology under the

GRAMMAR-gamma method. Moreover, no significant associations were detected between SNPs in or within 55 Kbp of ALX1 and any of the bill morphology traits using the RepeatABEL method ($P > 0.005$, $\alpha_{loc} = 0.001$).

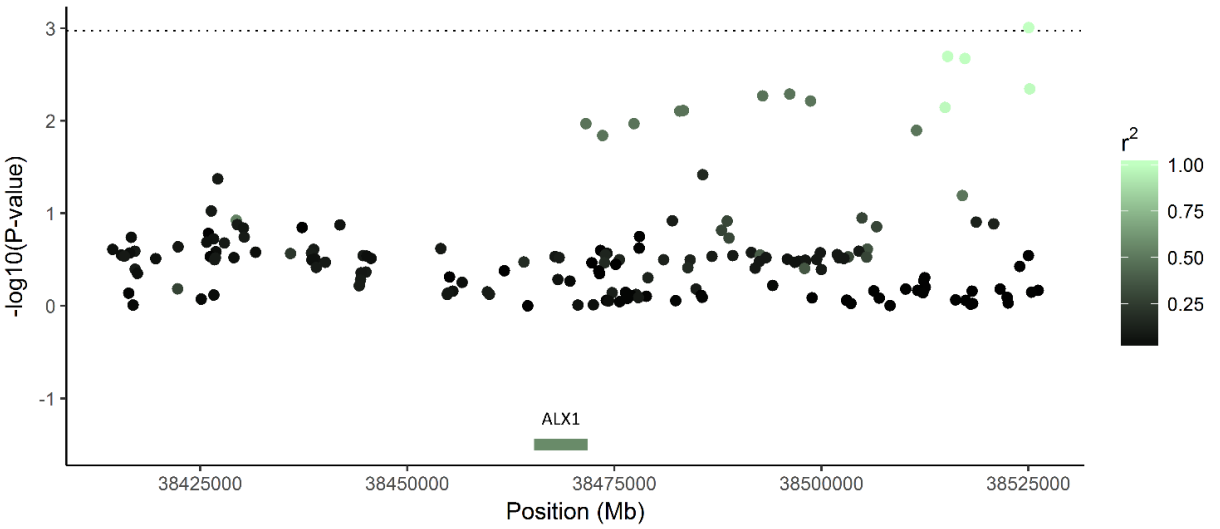

Figure 4: Manhattan-linkage disequilibrium plot for the 55 Kbp flanking regions either side of ALX1. Result derived from the GRAMMAR-gamma association test for ALX1 SNPs on bill shape. SNPs are coloured after degree of LD with reference marker, SNPa174071 ($P = 0.0009$). The dashed line represents the single-gene local alpha level for ALX1 in relation to bill shape ($\alpha_{loc} = 0.001$).

Using the RepeatABEL method, a single SNP within 55 Kbp of FGF8, SNPa276021, surpassed the single-gene significance threshold for bill size ($P = 0.0007$, $\alpha_{loc} = 0.0009$) (Fig. 5) after adjustment for genomic inflation factor of 1.01 (Appendix II; Table IV). Appendix II; Table VII gives summary statistics for the top 5 SNPs from this scan. Unlike linkage patterns for ALX1, the top 5 SNPs in the FGF8 candidate gene analysis were in very low LD (Fig. 5). This association was non-significant at the combined candidate gene threshold for bill size ($\alpha_{loc} = 0.00009$) and was not detected using the GRAMMAR-gamma method in GenABEL. FGF8 was the only candidate gene tested to show association with bill morphology when using the RepeatABEL method.

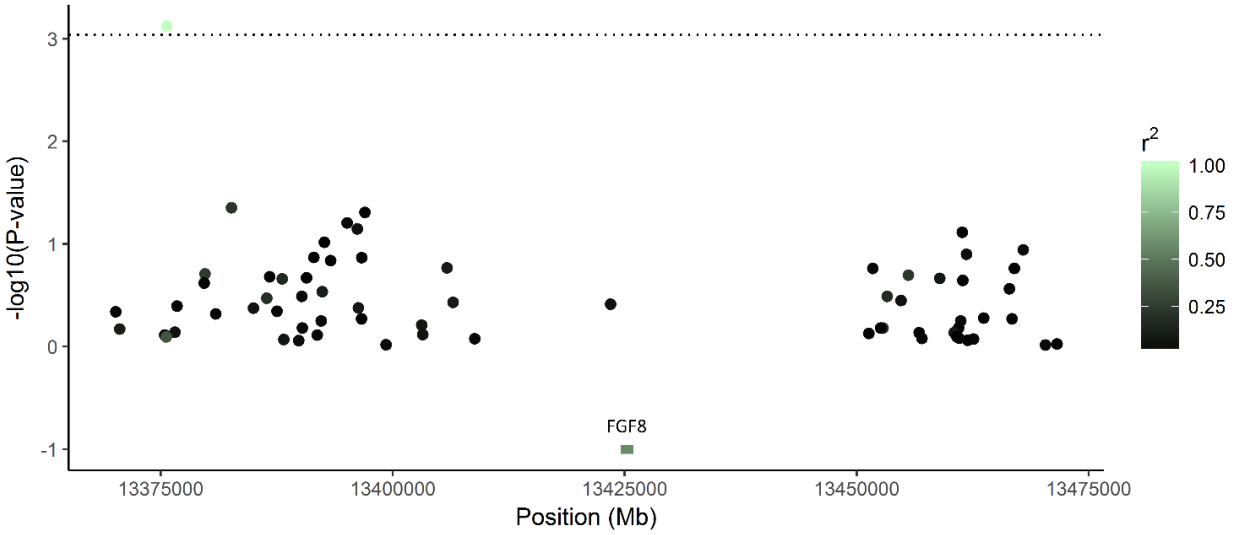

Figure 5: Manhattan-linkage disequilibrium plot for the 55 Kbp flanking regions either side of FGF8. Results are derived from the RepeatABEL association test for FGF8 SNPs on bill size. SNPs are coloured after degree of LD with reference marker, SNPa276021 (P = 0.0007). Single-gene local alpha level for FGF8 in relation to bill size represented by dashed line (α_{loc} = 0.0009).

3.4: GWAS

No loci were genome-wide significant at trait-specific local alpha levels (Appendix; Table IV) for any of the four studied bill morphology measures when using the GRAMMAR-gamma method in GenABEL. However, analysis in RepeatABEL revealed a single genome-wide significant SNP, SNPa77348, for bill depth on chromosome 3 (Fig. 6, Appendix II; Table VIII). This marker is neither within an exon or intron, nor within 55 Kbp of an annotated gene on the house sparrow or flycatcher genomes. The significant locus is located in a gene-free region of the annotated house sparrow genome, situated between a gene similar to 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 is the closest upstream gene of known function to SNPa77348 (CEBPZ, 150 Kbp away). Genes within approximately 1 Mbp of the suggestive SNP are listed in Appendix II; Table X. The candidate SNP is located at 6.33 Mbp on the flycatcher genome (E-value = 0.0, alignment = 87.94%), and the closest upstream gene in the flycatcher is CEBPZ. The closest downstream gene in the flycatcher is for a lincRNA, ENSFALG00000018842, which is transcribed upstream of CRIM1, the closest downstream gene with known function. Several such lincRNA variants lie between the candidate SNP and CRIM1 in the flycatcher genome.

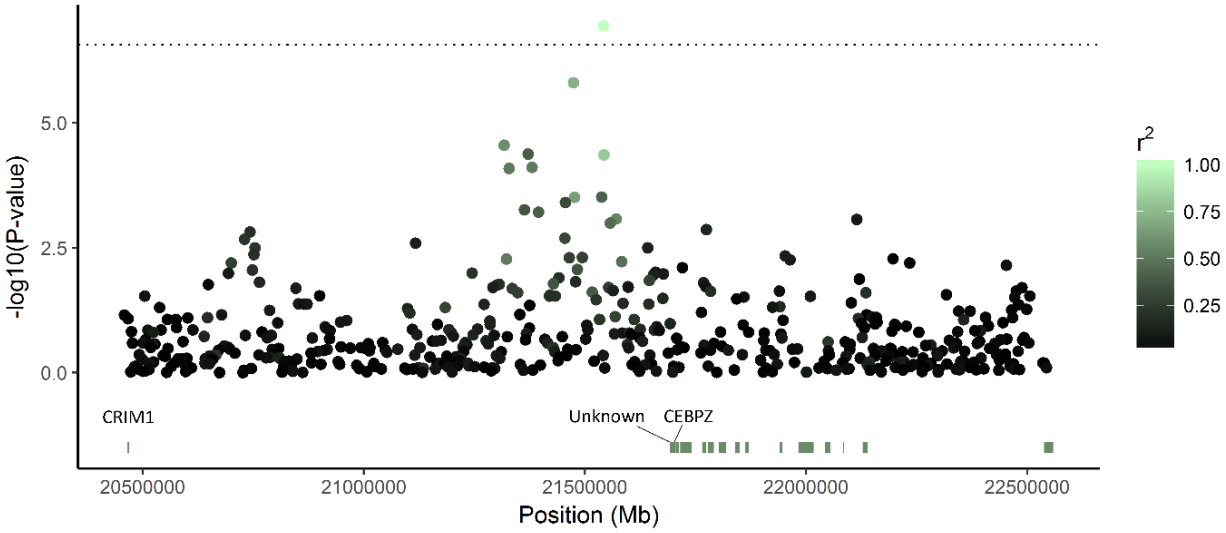

Figure 6: Manhattan-linkage disequilibrium plot for the 1 Mbp flanking regions either side of top SNP, SNPa77348. Results are derived from the RepeatABEL GWAS on bill depth. SNPs are coloured after degree of LD with the reference marker. Genes within 1 Mbp of the significant variant are displayed, and closest genes with known function are annotated.

3.5: QTCAT

Analysis using the multi-marker GWA method, QTCAT, yielded no significant novel or confirmatory associations for any bill morphology trait 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 0.012. Excluding island and hatch year as fixed factors did not alter the results.

3.6: Result Verification

If island is included as a fixed factor in analyses, no significant associations are detected in the Southern dataset for candidate genes FGF8 and ALX1, nor for the 55 Kbp region either side of SNPa77348. Where island was not included as a fixed factor, one marker, SNPa77370, showed significant association with bill shape at single marker set significance threshold ($P = 0.0031$, $a_{loc} = 0.0032$) when examining the 55 Kbp region either side of SNPa77348 (Appendix; Table IX, Fig. IV). This is in contrast to results from the Helgeland system, where the significant marker was associated with bill depth, but bill depth and shape are very highly correlated ($r = 0.93$, $P = <0.001$) and bill depth explains 98% of the variation in bill shape in the PCA (Table 2). If the significance level is calculated based on all 258 markers tested in the Southern dataset, α_{loc} is 0.0004 when the same FWER of 0.05 is controlled for. The association between SNPa77370 and bill depth, in the model without island as fixed factor, is not significant at this alpha level.

4: Discussion

The current analysis of bill morphology variation in the Helgeland metapopulation of free living house sparrows provides an interesting comparison to independent findings in other species. Heritability estimates for the depth, shape, length, and size of bills (0.35, 0.38, 0.38, 0.40) are high and are broadly comparable with previous findings, both in the house sparrow and in other bird species. For example, in Jensen *et al.* (2003) heritabilities for bill length and depth, determined using a pedigree reconstructed from microsatellite data in a sub-set of the same metapopulation of house sparrows, were 0.48 and 0.32 respectively. Pedigree-based estimates in a wild population of ground finch (*Geospiza fortis*) were 0.79 and 0.56 for bill length and depth, and were 0.66 and 0.83 for principal components 1 and 2 of bill morphology, corresponding to bill size and shape as in this study (Keller *et al.*, 2001). Analysis of a large collared flycatcher dataset gave heritability estimates for bill length and depth as 0.44 and 0.52 (Merilä *et al.*, 2001). Here, definitive values for trait heritability were taken from RepeatABEL analyses and were calculated from the variance components included in the model, after removing any phenotypic variance explained by fixed factors and covariates. Although this does not take into account dominance or epistatic variance, V_a estimates are likely to be accurate and confirm that there is a substantial genetic component to inheritance of bill morphology (Rönnegård *et al.*, 2016).

GCTA analysis identified genetic architecture of bill morphology as polygenic. Visual examination of the relationships between variance explained and chromosome length (Fig. 3) also revealed that, whilst some chromosomes (eg. chromosome 7) are important contributors to variance explained for all traits, others (eg. chromosomes 5 and 18) are important only for certain traits. These results give some substantiation to significant results in our single marker association tests. For example, chromosome 1A falls outside the 95% CI for bill shape and this is where ALX1 is situated. Chromosome 1A is also the location of HMGA2 a locus controlling bill size in Darwin's finches (Chaves *et al.*, 2016, Lamichhaney *et al.*, 2016). The SNP array used here was, unfortunately, not enriched for SNPs in this gene region so HMGA2 was not among candidate genes tested. Chromosome 5, the strongest outlier for bill depth and bill shape, is where three of our nine candidate genes for bill morphology are located: Calm1, FGF19 and BMP4. However, there were no significant associations on chromosome 5 in single marker GWA analysis, nor were there in analyses using QTCAT. In contrast, chromosome 3 where the GWAS significant marker, SNPa77348, is located did not fall outside of the 95% CI for bill depth. This may be because this marker had a small effect size, which was not large enough to cause deviation from the linear relationship between chromosome size and variance explained. Bill morphology measures are

generally agreed to be polygenic in nature (Boag, 1983, Grant and Grant, 1994). An analysis of the genetic architecture of quantitative traits in two free-living populations of great tits (*Parus major*) found that the majority of complex traits in wild populations may be influenced by many genes of small effect (Santure *et al.*, 2015). Traits tested did not include bill morphology measures but included body mass, which is significantly correlated with bill dimensions in our dataset. Furthermore, recent studies which utilise genome-wide screens confirm existence of many genes with small influence on bill morphology, alongside two large effect loci: HMGA2 for bill size (Chaves *et al.*, 2016, Lamichhaney *et al.*, 2016) and ALX1 for bill shape (Lamichhaney *et al.*, 2016, Lamichhaney *et al.*, 2015). In this study, polygenic inheritance of bill morphology was suggested by candidate gene and GWA tests, in line with results from genome partitioning and with current consensus on the genetic architecture of these traits. In candidate gene association tests, markers significant at the single gene level were detected near ALX1 (SNPa174071, 0.05% of variance explained) and FGF8 (SNPa276021, 0.07% of variance explained) for bill shape and bill size respectively. Repeated measures GWAS identified a single significant marker in an intergenic region, SNPa77348, which explained 0.05% of the variance in bill depth (Appendix II; Table VIII). Markers like these which explain <1% of the variance in a trait are considered to be of low effect size (Goddard *et al.*, 2016). This may either be because ALX1 and FGF8 are relatively small effect genes in this population closely linked to their respective significant markers, or significant markers could be weakly linked to genes further away with larger effects. Results here agree with expectations for quantitative traits, which are often explained by many loci with small effects, coupled with few loci of median to large effect (Mackay *et al.*, 2009).

Weak association was detected at the single gene significance level for candidate genes ALX1 ($P = 0.0009$) and FGF8 ($P = 0.0007$) for bill shape and bill size respectively. That ALX1 was associated specifically with bill shape is interesting, as it is this measure of bill morphology with which ALX1 has previously been linked (Lamichhaney *et al.*, 2015, Lamichhaney *et al.*, 2016). Positive values for principal component 2, bill shape, correspond to shorter, blunter bills and negative values to longer, more pointed bills (Table 2). Strong LD structure was observed between the top five markers for the ALX1 scan on bill shape (Fig. 4), increasing credibility of the association as selective sweeps increase LD between neutral loci close to a causative variant (Stephan *et al.*, 2006). The weak signal for ALX1 was detected using GRAMMAR-gamma but not RepeatABEL and could suggest spurious association, as repeated measurements fully account for within-individual variation, reducing chance of type 1 error and lowering incidence of type 2 error (Rönnegård *et al.*, 2016). The second candidate gene displaying single-gene significant association, FGF8, is involved in production of elongated bill structure during facial development when it is

expressed in tandem with Shh (Abzhanov and Tabin, 2004). Artificial overexpression of FGF8 alone in chick embryos led to reduced chondrogenesis and significant reduction in bill size (Abzhanov and Tabin, 2004). Confidence in the relatively weak association observed here is improved by its detection in RepeatABEL, which likely has higher power than GenABEL (Rönnegård *et al.*, 2016). Conversely, low LD between the associated polymorphism and surrounding markers (Fig.5) is not in line with anticipated increase in LD in regions close to causative variants if these are under selection (Stephan *et al.*, 2006) and may suggest spurious association. Weak associations demonstrated for both the above candidate genes suggest that, if the associations are genuine, they contribute only little to the variance in the bill morphology phenotypes they affect.

A single significant marker was detected in RepeatABEL GWAS on bill depth, no significant associations were found for any other bill morphology trait, and none were detected using the GRAMMAR-gamma method. Our 200k SNP chip has higher marker density compared to similar association studies on ecologically important traits in outbred populations (Santure *et al.*, 2015, Chaves *et al.*, 2016, Johnston *et al.*, 2011), with average distance between markers of approximately 6000 bp, affording higher power to detect causal variants. Sample size in the Helgeland metapopulation (N = 1857 individuals, N = 4239 measurements) is 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 wild populations, humans and livestock have often found little evidence for large effect QTL, especially where LD between markers is low. This is demonstrated by a recent study in a wild great tit population, where no large effect markers for clutch size or egg mass were identified (Santure *et al.*, 2013) and in Husby *et al.* (2015), where only one genome-wide significant SNP explaining 3.9% of the variance in clutch size was detected. Evaluation of the highly polygenic human height also provides an excellent example; GWAS studies have identified 697 variants which together only explain around 20% of the heritability in height (Marouli *et al.*, 2017). In Silva *et al.* (in press.), where GWAS was carried out using house sparrows from a subset of islands in the Helgeland metapopulation genotyped on the 10k array described in Hagen *et al.* (2013) one marker, SNP11485 on chromosome 20, was significantly related to bill length. The same marker on the 200k SNP array was not associated with bill length in the present study, disparity may be a result of fewer individuals from a smaller subset of islands being typed on the 10k array. Rare variants with low MAF (<0.05) and low effect size may be missed in GWAS even where marker density and sample size are high (Wilkening *et al.*, 2009). In light of results suggesting polygenic inheritance for bill morphology traits and for the correlated trait body mass in this house sparrow population, it is probable that power to detect

such variants was low here. Where large effect variants have previously been identified by GWAS in wild populations they related to near Mendelian traits under strong sexual selection like the horn type and size in Soay sheep (Johnston *et al.*, 2013, Johnston *et al.*, 2011), traits that captured lifetime fitness differences between alternate phenotypes as with age at maturity in Atlantic salmon (Johnston *et al.*, 2014, Barson *et al.*, 2015), or those involved in bill morphology dependent adaptation to trophic niches and subsequent speciation events as in Darwin's finches (Chaves *et al.*, 2016, Lamichhaney *et al.*, 2015, Lamichhaney *et al.*, 2016). Traits with such characteristics may be more likely to yield large effect loci in a GWA study.

Here, GWAS significant marker for bill depth, SNPa77348 on chromosome 3, was situated in a gene-free region 1.07 Mbp away from CRIM1, and close to a protein of unknown function and a gene analogous to CEBPZ. CEBPZ plays an important role in response to environmental stimuli via a transcriptional process involving heat-shock factors (Musialik *et al.*, 2014), but has not been connected to craniofacial or skeletal development in any species. CRIM1 inhibits BMP receptor activation and subsequent BMP signalling by direct binding with BMP4 and BMP7 (Kolle *et al.*, 2003, Wilkinson *et al.*, 2003). This illustrates a plausible mechanism through which CRIM1 could influence bill morphology via upstream action, as BMP4 expression has previously been linked with bill phenotype (Abzhanov *et al.*, 2006, Abzhanov *et al.*, 2004, Wu *et al.*, 2004). LD between a significant neutral locus and a causative variant is one means by which significant markers like ours that do not code for missense mutations can indicate causative genes in GWAS (Stephan *et al.*, 2006). LD range in a population dictates how far from a gene a trait the associated variant may be (Backström *et al.*, 2006). Previous studies on LD in wild bird populations show that its range is very dependent on effective population size (N_e), as is expected from the theory and is widely observed in the animal breeding community (Charlesworth and Charlesworth, 2010). Li and Merila (2010) reported high levels of LD which spanned several mega bases in the Siberian jay which has a small N_e, whereas studies of linkage disequilibrium on the Z-chromosome in the migratory collared flycatcher reported LD which fell to background level at 500 Kbp (Backström *et al.*, 2006) or at distances as close as 17 Kbp (Kawakami *et al.*, 2014a). LD range in the Helgeland metapopulation of house sparrows is longest in Aldra, an inbred subpopulation, and shortest in Hestmannøy, the largest subpopulation. LD varies by chromosome, but on average falls to background levels after 15-20 Kbp (Hagen *et al.*, in prep. -a). Although LD may be up to 10-fold higher in genomic islands of differentiation (Kawakami *et al.*, 2014a) SNPa77348 is likely too distant from CRIM1 to be linked with a causal variant within the gene. The distance is also too great for the significant variant to be in or in LD with a promotor region for CRIM1 (Cho, 2012).

Intergenic, non-coding SNPs linked with a phenotype often imply regulatory function (Lamichhaney *et al.*, 2016). Our significant marker is approx. 150Kb away from a protein of unknown function, which one could speculate is involved in regulation of CRIM1. Alternatively, our significant SNP could be in or in LD with a marker in an enhancer element for CRIM1, regulatory proteins bind to enhancers which then function via DNA looping and may be more than 1 Mbp from the promoter they act upon (Cho, 2012). A study by Chan *et al.* (2010) revealed adaptive evolution of the pelvic region in sticklebacks was governed by regulatory changes in Pitx1, caused by recurrent deletion mutations in its *Pel* enhancer region. Their findings underscore how major morphological changes can occur due to variation in intergenic DNA regions, relatively far from the actual gene affecting the trait. In the current study, BLAST against the flycatcher genome assembly revealed several lincRNAs close to the significant variant, SNPa77348, which are transcribed upstream of CRIM1. Analogous sequences in the house sparrow genome may influence BMP4 levels via upstream control of CRIM1 and the significant variant found here may represent, or be in LD with, a causal variant in one of these lincRNAs (Hrdlickova *et al.*, 2014). Brodie *et al.* (2016) suggest that, although SNPs are more likely to be relevant when they are closer to genes, causative genes may be found up to 2 Mbp away. This could occur if SNPs are markers for large structural variations like copy number variants, inversions, or balanced translocations. These structural variations can affect gene expression, but the latter two cannot be detected using a SNP microarray and detection of copy number variants is unlikely unless a microarray is designed for this purpose (McCarroll, 2008, Shaffer *et al.*, 2007). Finally, synonymous mutations in an exon (also not relevant to our significant marker) can affect speed of protein folding, altering structure and accordingly, function (Hunt *et al.*, 2009).

In this dataset, linear regression of age 1 May adjusted bill depth on hatch year, the trait with which marker SNPa77348 was associated, explains 1% of the variation ($R^2 = 0.01$, $F(2, 1780) = 7.9$, $P = 0.0004$) and indicates that bill depth is increasing by 0.004 mm each year in this house sparrow metapopulation ($\beta = 0.004 \pm 0.001$, $t = 3.33$, $P < 0.001$). The significant marker appears to have an additive effect on bill depth, where heterozygotes on average show an intermediate phenotype relative to the two homozygous genotypes and the minor allele (T) is associated with deeper bills (Appendix III; Fig. IIIC). Linear regression of MAF on hatch year for this marker, with sex as fixed factor is significant ($R^2 = 0.01$, $F(2, 1780) = 11.04$, $P < 0.001$), and shows increase an in the T allele frequency of 0.86% per year ($\beta = 0.0086 \pm 0.002$, $t = 4.603$, $P < 0.001$) (Appendix III; Fig. IIID). Bill size, which is contributed to by both bill length and depth, has previously been linked to increased fitness in this Helgeland metapopulation of house sparrows (Ringsby *et al.*, 2009, Jensen *et al.*, 2008). Using data from a sub-set of the populations used here, from the period

1993-2002 (which is mostly before the period in the present study), Jensen *et al.* (2008) found that observed response to selection on bill depth was negative for both sexes, which is not in agreement with predictions based on the strength and direction of selection acting on bill depth and correlated morphological traits. Results from Steinsland and Jensen (2010) suggested a weak positive relationship between bill depth breeding values and cohort, but this result was not replicated in Holand *et al.* (2013). In Holand *et al.* (2011) bill depth was shown to differ between different populations along the Norwegian coast, although the relative influence of drift was likely to be higher than that of selection for this phenotype. More broadly, bill morphology may directly affect survival in several avian species through interaction with diverse ecological and environmental factors, which give different bill phenotypes an advantage in different environments. Structural correspondence between bill morphology and food plants, as in the Hawaiian hummingbird, can affect foraging efficiency and indicate co-adaptation; evolution of bill morphology in hummingbirds may be driven by differing subsets of flowers in different environments (Temeles *et al.*, 2009). In Darwin's finches, large birds with big bills better survived a drought on Daphne Major Island, as they were more able to crack the large seeds which were more abundant in these conditions (Boag and Grant, 1981). Conversely, Lamichhaney *et al.* (2016) found that bill size was under negative selection in the medium ground finch during a drought, in part due to niche competition with the large ground finch. In our Helgeland metapopulation of house sparrows, bill size is correlated with rate of feeding of offspring in female house sparrows and parental investment represented by this feeding rate may increase successful recruitment (Ringsby *et al.*, 2009). In insular avian populations in general, larger bill and body size may improve survival through expanding dietary niche (Grant, 1965, Scott *et al.*, 2003).

Taken together, these studies indicate that directional selection on bill phenotype, particularly in insular populations, is possible and even commonplace. In the study metapopulation, high heritability and significant positive association between bill depth and hatch year, coupled with additive genetic effect of marker SNPa77348 with its plausible link to CRIM1, suggests that evolution of deeper bills has occurred in this system between 1997 and 2012. This may have been a result of selection for deeper bills, but a significant association between hatch year and bill depth may also be a function of fluctuating environmental conditions covarying with genetic change. Between 1995-2010 in Troms, Northern Norway, average temperature in May increased by around 2.5 °C, representing an increase of 0.19 °C per year (Barrett, 2011). Insect prey species and abundance fluctuates in line with temperature and season (Schwagmeyer and Mock, 2003), it is possible that insect abundance increased and frequency of different insect species changed in line with increasing temperatures in this time period. Better diet, particularly during

development, may lead to increased average body mass (Ringsby *et al.*, 2009) and increase in bill dimensions because these traits appear to be both phenotypically and genetically correlated (Jensen *et al.*, 2003, Jensen *et al.*, 2008). Alternatively, greater adult bill depth values may be due to proportional increase in insect prey relative to seeds, which reduces wear on the bill (Greenberg *et al.*, 2013). Here, body mass was not included as a covariate in the association analyses, despite positive correlation with bill dimensions, as any mechanism through which bill morphology may evolve is of interest. Inclusion of body mass as covariate in future analyses will help determine whether the implicated region affects bill development directly or whether the association is mediated by body size and may help unravel the complex mechanisms through which genes and environment interact to influence bill morphology changes in this study population

No significant association was found between any QTC and bill morphology traits using QTCAT analysis. This is unexpected, as the multi-marker association method should increase power to detect associations by removing the need for the GRM to correct for relatedness and population structure, hence permitting testing of the unrestricted hypothesis. This implies that the significant association between the SNPa77348 gene region and bill depth detected in RepeatABEL analysis should have also been detected in QTCAT. However, QTCAT does not permit use of repeated measures in the usual manner, by including ID as a random effect. Ability to account for within-individual variation may explain why association was only detected by RepeatABEL analysis, as this form of variation may be an important contributor to bill morphology variation in this population and therefore should be properly accounted for. It is possible to include repeated individual measures with the QTCAT method by numbering repeated measures for each individual and including measurement number as a covariate (Klasen *et al.*, 2016). This is something that could be explored in further analysis, and may improve detection power in QTCAT where within-individual variation is important.

ALX1 and FGF8 candidate gene SNPs were tested for association with bill morphology in a second set of island populations, as were the SNPs 55 Kbp either side of the significant marker, SNPa77348, from the RepeatABEL GWAS. No significant association was observed for either candidate gene for any bill morphology trait in this second set of populations, regardless of whether island was included as a fixed factor in analysis. This reduces confidence in the weak associations for these candidate genes detected in the Helgeland metapopulation. It is possible that effect of these genes may be population dependent in house sparrows, or that smaller sample size in the Southern populations could have reduced power to detect weak effects. Significant association was detected between SNP a77370 and bill shape only where island was not included as a fixed factor in the analysis (Appendix; Table IX, Fig. IV). Migration between the islands is all but non-existent

in the Southern populations (Skjelseth *et al.*, 2007) which may explain the apparent greater influence of island in this dataset (see Appendix I; Population Structure for further details). The association was not significant if α_{loc} was computed using all 258 markers used in result verification, therefore evidence for association with bill shape in the Southern dataset is weak and should be interpreted with caution.

Somewhat surprisingly, none of the candidate loci previously related to bill morphology in genome-wide scans or in functional expression studies were strongly associated with variation in bill morphology here (Appendix III; Fig. II). This illustrates the problems often associated with replicating candidate gene-trait associations in subsequent studies; detecting such relationships can be difficult, even with a large, high-quality dataset like in this study. This may simply be due to species or population-specific differences in the genetic architecture of bill morphology traits. While candidate genes examined in functional studies may govern development of diverse bill shapes between species (as evidenced by differential expression between *Geospiza* type; Abzhanov *et al.*, 2006, Mallarino *et al.*, 2011), or when artificially expressed (Abzhanov *et al.*, 2006, Abzhanov and Tabin, 2004), natural variation in developmental expression of these genes in the free-living metapopulation of house sparrows may not be so substantial. Alternatively, expression may be governed by polymorphisms in genes which regulate the differentially expressed gene(s) (Mundy, 2016), it is then these upstream polymorphisms that are detectable by GWAS (but not by the candidate gene approach). Detection of the significant marker in the CRIM1 gene region here gives some support to this hypothesis, based on function of CRIM1 as an upstream regulator of BMP4 (Wilkinson *et al.*, 2003). Difference in approaches used may also influence results: Lamichhaney *et al.* (2015, 2016) performed pairwise ZF_{st} scans in 15 Kbp windows across the whole genome, taking advantage of speciation to contrast large, medium, and small-billed finches, resulting in increased power to detect association compared to single marker GWAS. Haplotype association analysis of the region with highest ZF_{st}, which overlaps ALX1, was then performed, again increasing power substantially compared to single marker association (Lamichhaney *et al.*, 2015). Analysis in the house sparrow did not permit separation by species into bill size categories for F_{st} scans and accurate haplotype association analysis requires phased data. The BSLMM approach utilised by Chaves *et al.* (2016) purportedly yields less conservative *P* values than LMM methods like GRAMMAR-gamma and RepeatABEL when individuals are closely related or where strongly associated markers contribute to a significant proportion of variation in the phenotype (Zhou and Stephens, 2012). Preliminary analysis of the Helgeland dataset by postdoc Jostein Gohli, using the BSLMM method implemented in GEMMA, yielded a large proportion of SNPs in the collected MCMC samples with non-zero slopes, both for bill depth and bill length. Markers with

non-zero β have a measurable effect on their associated phenotype. Using a conservative posterior inclusion probability (PIP) cut-off of 0.1, as suggested in Chaves *et al.* (2016), Comeault *et al.* (2014) and Riesch *et al.* (2017), three markers were significantly associated with bill depth. The marker with the highest PIP for this trait is SNPa77348 on chromosome 3 ($\gamma = 0.36$, $\beta = 0.05$), increasing confidence in the association between the same marker and trait detected in RepeatABEL analysis here. Several significant markers were also detected in Gohli's analysis on bill length, whereas none were detected using LMM methods for this trait. BSLMM may represent a workable method for detection of significant associations with quantitative traits where marker effect size is low.

Lack of large effects on bill morphology in candidate gene regions may be explained by non-genetic control mechanisms for expression of candidate genes, low effective population size or rapid LD decay. Developmental expression of previously identified candidate genes for bill morphology may be under epigenetic control, as has been suggested for BMP4 by species-specific epimutations in the BMP pathway in Darwin's finches (Skinner *et al.*, 2014). This form of regulation is, however, not detectable by any genetic association method. Alternatively, rapid LD decay in our house sparrow population may affect the probability of detecting extant effects, as high marker density on the 200k SNP chip coupled with marker enrichment in candidate gene regions increases probability of detection (Hagen *et al.*, in prep. -b). Population structure may also play a role, if different markers are in LD with a causative variant in different populations then this will reduce power to detect the association. Low N_e in subpopulations could also 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). These factors bring us full circle to the relative importance of environment and genes to bill morphology in our free-living house sparrow population, and more broadly to relative contributions to phenotypic diversity in nature. Despite the unique challenges presented by quantitative genetic studies in such populations, these types of study allow examination of the genetic basis of ecologically important traits in their natural context and may aid in unravelling complex, environment-phenotype-genotype interactions and shed light on the mechanisms driving selection and adaptive evolution. Studies which examine these links in keystone species, such as the Atlantic salmon (Barson *et al.*, 2015, Johnston *et al.*, 2014) or Yellowstone wolf (Coulson *et al.*, 2011) may help predict spatiotemporal variation in selection and population fluctuations beyond the study system, aiding management decisions. Such work is important and conclusions invaluable even if detecting associations in natural populations may be more problematic than anticipated.

5: Conclusion

Results presented here support previous characterisation of bill morphology as highly heritable and governed by polygenic genetic architecture. A significant locus of small effect size on bill depth was identified, which may explain a small proportion of the variation in this trait. The locus is approximately 1 Mbp away from CRIM1, a gene which negatively regulates BMP4. One can hypothesise that the significant marker may be in LD with an enhancer element or lincRNA which is an upstream regulator of CRIM1, and may influence BMP4 expression via this pathway. Furthermore, increase in bill depth with hatch year was observed in the Helgeland metapopulation, as was increase in the frequency of the allele conferring deeper bills at the significant marker. Bill depth may be under positive selection in the Helgeland metapopulation or, alternatively, fluctuating environmental conditions covarying with genetic change may be driving variation. Weak associations were also detected for candidate genes ALX1 and FGF8 in the Helgeland dataset for bill shape and size respectively. All associations detected here warrant exploration in additional populations.

This work illustrates the difficulties involved in discovering loci for quantitative traits in natural populations. Careful selection of highly variable traits, with high heritability that are ideally under strong selection or are associated with niche adaptation and speciation may increase probability of detecting large effect variants. Where within-individual variation contributes highly to phenotype, association methods which allow repeated measures should be considered. Alternatively, recently developed multi-marker and BSLMM methods may improve chances of identifying genes for highly polygenic traits. Future work should seek to determine whether variants identified here affect bill morphology alone or if they act via alteration of body mass. Despite difficulties encountered in uncovering genes for ecologically important traits in natural populations, candidate gene and GWA studies can, on occasion, provide valuable insights into eco-evo dynamics in such populations. Consequently, association studies are likely to remain a valuable investigative tool in eco-evo research for the foreseeable future.

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**,** 1462-5.

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

ÅKESSON, M., BENSCH, S., HASSELQUIST, D., TARKA, M. & HANSSON, B. 2008. Estimating Heritabilities and Genetic Correlations: Comparing the 'Animal Model' with Parent-Offspring Regression Using Data from a Natural Population. *PLoS ONE,* 3**,** 1739.

AULCHENKO, Y. S. & KARSSEN, L. C. 2015. The GenABEL project developers. The GenABEL Tutorial. [Online] Zenodo. Available at; https://zenodo.org/record/19738#.WPSBzIVOL4g [Accessed 17 April 2017].

ALMASY, L. & BLANGERO, J. 2009. Human QTL Linkage Mapping. *Genetica,* 136**,** 333-340.

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

ASCHARD, H. 2016. A perspective on interaction effects in genetic association studies. *Genetic Epidemiology,* 40**,** 678-688.

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

BACKSTRÖM, N., QVARNSTRÖM, A., GUSTAFSSON, L. & ELLEGREN, H. 2006. Levels of linkage disequilibrium in a wild bird population. *Biology Letters,* 2**,** 435-438.

BARRETT, R. D. H. & HOEKSTRA, H. E. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics,* 12**,** 767-780.

BARRETT, R. T. 2011. Recent response to climate change among migrant birds in northern Norway. *Ringing & Migration,* 26**,** 83-93.

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.

BATES, D., MÄCHLER, M., BOLKER, B. & WALKER, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,* 67**,** 48.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological),* 57**,** 289-300.

BILLING, A. M., LEE, A. M., SKJELSETH, S., BORG, Å. A., HALE, M. C., SLATE, J. O. N., 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**,** 82-85.

BOCK, W. J. 1966. An Approach to the Functional Analysis of Bill Shape. *The Auk,* 83**,** 10-51.

BONFERRONI, C. E. ``Teoria statistica delle classi e calcolo delle probabilità." *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **8**, 3-62, 1936.

BRADSHAW, H. D., JR., OTTO, K. G., FREWEN, B. E., MCKAY, J. K. & SCHEMSKE, D. W. 1998. Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (Mimulus). *Genetics,* 149**,** 367-382.

BRODIE, A., AZARIA, J. R. & OFRAN, Y. 2016. How far from the SNP may the causative genes be? *Nucleic Acids Research,* 44**,** 6046-6054.

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.

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.

CHARLESWORTH, B. & CHARLESWORTH, D. 2010. *Elements of Evolutionary Genetics,* Greenwood Village, Colorado, USA, Roberts and Co.

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, 5282-5295.

CHEN, Z. & LIU, Q. 2011. A new approach to account for the correlations among single nucleotide polymorphisms in genome: wide association studies. *Human Heredity,* 72**,** 1-9.

CHEVERUD, J. M. 2001. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity (Edinburgh),* 87**,** 52-58.

CHO, K. W. Y. 2012. Enhancers: WIRE Review. *Wiley interdisciplinary reviews. Developmental biology,* 1**,** 469-478.

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.

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.

COULSON, T., MACNULTY, D. R., STAHLER, D. R., VONHOLDT, B., WAYNE, R. K. & SMITH, D. W. 2011. Modeling Effects of Environmental Change on Wolf Population Dynamics, Trait Evolution, and Life History. *Science,* 334**,** 1275-1278.

DAVIS, J. 1954. Seasonal Changes in Bill Length of Certain Passerine Birds. *The Condor,* 56**,** 142-149.

DELONG, J. P., FORBES, V. E., GALIC, N., GIBERT, J. P., LAPORT, R. G., PHILLIPS, J. S. & VAVRA, J. M. 2016. How fast is fast? Eco-evolutionary dynamics and rates of change in populations and phenotypes. *Ecology and Evolution,* 6**,** 573-581.

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. in press. The genomic mosaicism of hybrid speciation. *Science Advances*, xx, xx-xx.

ELLEGREN, H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution,* 29**,** 51-63.

ELLEGREN, H., SMEDS, L., BURRI, R., OLASON, P. I., BACKSTRÖM, N., KAWAKAMI, T., KÜNSTNER, A., MÄKINEN, H., NADACHOWSKA-BRZYSKA, K., QVARNSTRÖM, A., UEBBING, S. & WOLF, J. B. W. 2012. The genomic landscape of species divergence in Ficedula flycatchers. *Nature,* 491**,** 756-760.

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.

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.

GALWEY, N. W. 2009. A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. *Genetic Epidemiology,* 33**,** 559-568.

GAO, X., STARMER, J. & MARTIN, E. R. 2008. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology,* 32**,** 361-369.

GENZ, A. 1992. Numerical Computation of Multivariate Normal Probabilities. *Journal of Computational and Graphical Statistics,* 1**,** 141-149.

GILMOUR, A. R., THOMPSON, R. & CULLIS, B. R. 1995. Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics,* 51**,** 1440-1450.

GLICKMAN, M. E., RAO, S. R. & SCHULTZ, M. R. 2014. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical Epidemiology,* 67**,** 850-857.

GODDARD, M. E., KEMPER, K. E., MACLEOD, I. M., CHAMBERLAIN, A. J. & HAYES, B. J. 2016. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proceedings of the Royal Society B: Biological Sciences,* 283**,** 20160569.

GOWER, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika,* 53**,** 325-338.

GRANT, B. R. & GRANT, P. R. 2014. *40 years of evolution: Darwin's finches on Daphne Major Island.* Princeton (New Jersey): Princeton Univeristy Press.

GRANT, P. & GRANT, B. 1994. Phenotypic and genetic-effects of hybridization in darwins finches. *Evolution,* 48**,** 297-316.
GRANT, P. R. 1965. The Adaptive Significance of Some Size Trends in Island Birds. *Evolution,* 19**,** 355-367.
GRANT, P. R. & GRANT, B. R. 2002. Unpredictable Evolution in a 30-Year Study of Darwin's Finches. *Science,* 296**,** 707-711.
GREENBERG, R., ETTERSON, M. & DANNER, R. M. 2013. Seasonal dimorphism in the horny bills of sparrows. *Ecology and Evolution,* 3**,** 389-398.
HA, N.-T., FREYTAG, S. & BICKEBOELLER, H. 2014. Coverage and efficiency in current SNP chips. *Euroupean Journal of Human Genetics,* 22**,** 1124-1130.
HAAG, C. R., SAASTAMOINEN, M., MARDEN, J. H. & HANSKI, I. 2005. A candidate locus for variation in dispersal rate in a butterfly metapopulation. *Proceedings of the Royal Society B: Biological Sciences,* 272**,** 2449-2456.
HAGEN, I. J., BILLING, A. M., LIEN, S., ELGVIN, T. O., TRIER, C., HUSBY, A. & JENSEN, H. in prep.-a. Effects of population demography on linkage disequilibrium in the house sparrow (Passer domesticus) genome. *Molecular Ecology Resources.*
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.
HAGEN, I. J., LIEN, S., BILLING, A. M., TARKA, M., ELGVIN, T. O., TRIER, C. N., SÆTRE, G. & HENRIK, J. in prep.-b. A genome wide, medium density linkage map for the house sparrow, Passer domesticus.
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.*
HANSKI, I. & SACCHERI, I. 2006. Molecular-level variation affects population growth in a butterfly metapopulation. *PLoS Biology,* 4**,** 129.
HARTIGAN, J. A. & WONG, M. A. 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 28**,** 100-108.
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.
HENDRY, A. 2012. Eco-evolutionary dynamics. *Genes and Genetic Systems,* 87**,** 374-374.
HENDRY, A. P. 2016. *Eco-Evolutionary Dynamics,* Oxfordshire, Princeton University Press.
HINRICHS, A. L., LARKIN, E. K. & SUAREZ, B. K. 2009. Population Stratification and Patterns of Linkage Disequilibrium. *Genetic epidemiology,* 33**,** 88-92.
HOFFMANN, R. & VALENCIA, A. 2004. A gene network for navigating the literature. *Nature Genetics,* 36**,** 664-664.
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**,** 367-379.
HOLAND, A. M., STEINSLAND, I., MARTINO, S. & JENSEN, H. 2013. Animal models and integrated nested Laplace approximations. *G3 (Bethesda),* 3**,** 1241-1251.
HRDLICKOVA, B., DE ALMEIDA, R. C., BOREK, Z. & WITHOFF, S. 2014. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* 1842**,** 1910-1922.
HUNT, R., SAUNA, Z. E., AMBUDKAR, S. V., GOTTESMAN, M. M. & KIMCHI-SARFATY, C. 2009. Silent (Synonymous) SNPs: Should We Care About Them? *In:* KOMAR, A. A. (ed.) *Single Nucleotide Polymorphisms: Methods and Protocols.* Totowa, NJ: Humana Press.
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, 20150156.
HUSBY, A., SAETHER, 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**,** 1128-1139.
JAMES, N. M. S. & ZACH, R. 1979. Heritability of Some Morphological Characters in a Song Sparrow Population. *Evolution,* 33**,** 460-467.
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**,** 1792-1805.
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., SAETHER, 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**,** 1296-1307.

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.

JENSEN, H., SZULKIN, M. & SLATE, J. 2014. Molecular quantitative genetics. *Quantitative Genetics in the Wild.* Oxford: Oxford University Press.

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.

JONES, F. C., GRABHERR, M. G., CHAN, Y. F., RUSSELL, P., MAUCELI, E., JOHNSON, J., SWOFFORD, R., PIRUN, M., ZODY, M. C., WHITE, S., BIRNEY, E., SEARLE, S., SCHMUTZ, J., GRIMWOOD, J., DICKSON, M. C., MYERS, R. M., MILLER, C. T., SUMMERS, B. R., KNECHT, A. K., BRADY, S. D., ZHANG, H., POLLEN, A. A., HOWES, T., AMEMIYA, C., LANDER, E. S., DI PALMA, F., LINDBLAD-TOH, K. & KINGSLEY, D. M. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature,* 484**,** 55-61.

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.

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.

KINNISON, M. T., UNWIN, M. J. & QUINN, T. P. 2008. Eco-evolutionary vs. habitat contributions to invasion in salmon: experimental evaluation in the wild. *Molecular Ecology,* 17**,** 405-414.

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.

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.

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., PARN, H., HOLAND, H., ENGEN, S. & SÆTHER, B. E. in press. Reversal of response to artificial selection on body size in a wild passerine bird. *Evolution***,** xx, xx-xx.

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.

LEIPS, J. & MACKAY, T. F. 2000. Quantitative trait loci for life span in Drosophila melanogaster: interactions with genetic background and larval density. *Genetics,* 155**,** 1773-1788.

LI, J. & JI, L. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinburgh),* 95**,** 221-227.

LI, M. H. & MERILA, J. 2010. Extensive linkage disequilibrium in a wild bird population. *Heredity,* 104**,** 600-610.

LIN, J. Z. & RITLAND, K. 1997. Quantitative Trait Loci Differentiating the Outbreeding Mimulus Guttatus from the Inbreeding M. Platycalyx. *Genetics,* 146**,** 1115-1121.

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.

LOWE, W. H., KOVACH, R. P. & ALLENDORF, F. W. 2017. Population Genetics and Demography Unite Ecology and Evolution. *Trends in Ecology & Evolution,* 32**,** 141-152.

MACKAY, T. F. C. 2004. The genetic architecture of quantitative traits: lessons from Drosophila. *Current Opinion in Genetics & Development,* 14**,** 253-257.

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.

MARDIA, K. V. 1978. Some properties of classical multidimensional scaling. . *Communications on Statistics - Theory and Methods,* A7**,** 1233 - 1241.

MAROULI, E., GRAFF, M. & MEDINA-GOMEZ, C. 2017. Rare and low-frequency coding variants alter human adult height. *Nature,* 542**,** 186-190.

MCCARROLL, S. A. 2008. Extending genome-wide association studies to copy-number variation. *Human Molecular Genetics,* 17**,** 135-142.

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.

MIWA, T., HAYTER, A. J. & KURIKI, S. 2003. The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 65**,** 223-234.

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**,** 2383-2390.

MOSKVINA, V. & SCHMIDT, K. M. 2008. On multiple-testing correction in genome-wide association studies. *Genetic Epidemiology,* 32**,** 567-573.

MUNDY, N. I. 2016. Population genomics fits the bill: genetics of adaptive beak variation in Darwin's finches. *Molecular Ecology,* 25**,** 5265-5266.

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.

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

NYHOLT, D. R. 2004. A Simple Correction for Multiple Testing for Single-Nucleotide Polymorphisms in Linkage Disequilibrium with Each Other. *American Journal of Human Genetics,* 74**,** 765-769.

PARDO-DIAZ, C., SALAZAR, C. & JIGGINS, C. D. 2015. Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution,* 6**,** 445-464.

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.

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**,** 144-152.

PELLETIER, F., CLUTTON-BROCK, T., PEMBERTON, J., TULJAPURKAR, S. & COULSON, T. 2007. The Evolutionary Demography of Ecological Change: Linking Trait Variation and Population Growth. *Science,* 315**,** 1571-1574.

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, 559-575.

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/.

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**,** 82.

RINGSBY, T., BERGE, T., SAETHER, B.-E. & JENSEN, H. 2009. Reproductive success and individual variation in feeding frequency of House Sparrows (Passer domesticus). *Journal of Ornithology,* 150**,** 469-481.

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.

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

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.

SÆTHER, B.-E., ALTWEGG, R. & RINGSBY, T. H. 1999. *Temporal and spatial variation in survival rates of a house sparrow, Passer domesticus, metapopulation,* Copenhagen, OIKOS.

SÆTHER, B.-E. & ENGEN, S. 2015. The concept of fitness in fluctuating environments. *Trends in Ecology & Evolution,* 30**,** 273-281.

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.

SCHWAGMEYER, P. L. & MOCK, D. W. 2003. How Consistently are Good Parents Good Parents? Repeatability of Parental Care in the House Sparrow, Passer domesticus. *Ethology,* 109**,** 303-313.

SCOTT, S. N., CLEGG, S. M., BLOMBERG, S. P., KIKKAWA, J. & OWENS, I. P. 2003. Morphological shifts in island-dwelling birds: the roles of generalist foraging and niche expansion. *Evolution,* 57**,** 2147-2156.

SHAFFER, L. G., BEAUDET, A. L., BROTHMAN, A. R., HIRSCH, B., LEVY, B., MARTIN, C. L., MASCARELLO, J. T. & RAO, K. W. 2007. Microarray analysis for constitutional cytogenetic abnormalities. *Genetics in Medicine,* 9**,** 654-662.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research,* 13**,** 2498-2504.

SILVA, C. N. S., MCFARLANE, E. S., HAGEN, I. J., RÖNNEGÅRD, L., BILLING, A. M., KVALNES, T., KEMPPAINEN, P., RØNNING, B., RINGSBY, T. H., SÆTHER, B. E., QVARNSTRÖM, A., ELLEGREN, H., JENSEN, H. & HUSBY, A. in press. Insights into the genetic architecture of natural and sexually selected morphological traits in two passerine bird species. *Heredity,* xx**,** xx-xx.

SKINNER, M. K., GURERRERO-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**,** 1972-1989.

SKJELSETH, S., RINGSBY, T. H., TUFTO, J., JENSEN, H. & SÆTHER, B.-E. 2007. Dispersal of introduced house sparrows Passer domesticus: an experiment. *Proceedings of the Royal Society B: Biological Sciences,* 274**,** 1763-1771.

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.

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

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**,** e0129479.

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.

STEINSLAND, I. & JENSEN, H. 2010. Utilizing Gaussian Markov Random Field Properties of Bayesian Animal Models. *Biometrics,* 66**,** 763-771.

STEPHAN, W., SONG, Y. S. & LANGLEY, C. H. 2006. The Hitchhiking Effect on Linkage Disequilibrium Between Linked Neutral Loci. *Genetics,* 172**,** 2647.

SYMONDS, M. R. & TATTERSALL, G. J. 2010. Geographical variation in bill size across bird species provides evidence for Allen's rule. *American Naturalist,* 176**,** 188-197.
SZKLARCZYK, D., FRANCESCHINI, A., WYDER, S., FORSLUND, K., HELLER, D., HUERTA-CEPAS, J., SIMONOVIC, M., ROTH, A., SANTOS, A., TSAFOU, K. P., KUHN, M., BORK, P., JENSEN, L. J. & VON MERING, C. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research,* 43**,** 447-452.
TABOR, H. K., RISCH, N. J. & MYERS, R. M. 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Naure 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.
TOKITA, M., YANO, W., JAMES, H. F. & ABZHANOV, A. 2017. Cranial shape evolution in adaptive radiations of birds: comparative morphometrics of Darwin's finches and Hawaiian honeycreepers. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 372.
UNIPROT 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Research,* 45**,** 158-169.
VAN DER MEIJ, M. A. A. & BOUT, R. G. 2004. Scaling of jaw muscle size and maximal bite force in finches. *Journal of Experimental Biology,* 207**,** 2745-2753.
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.
WILSON, A. J. 2008. Why h2 does not always equal VA/VP? *Journal of Evolutionary Biology,* 21**,** 647-650.
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.
WRAY, N. R., YANG, J., HAYES, B. J., PRICE, A. L., GODDARD, M. E. & VISSCHER, P. M. 2013. Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics,* 14**,** 507-515.
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.
YANG, J., ZAITLEN, N. A., GODDARD, M. E., VISSCHER, P. M. & PRICE, A. L. 2014. Advantages and pitfalls in the application of mixed model association methods. *Nature genetics,* 46**,** 100-106.
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**,** 710-716.
ZHOU, X. & STEPHENS, M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics,* 44**,** 821-824.

Appendix

I: Addendum

Measures Data Adjustment

Adult individuals were measured to the nearest 0.01 mm for bill depth and length, these measurements were performed by several different fieldworkers. After an initial training period, each fieldworker measured approximately 30 adult individuals together with Thor Harald Ringsby or another experienced fieldworker (Kvalnes *et al.*, in press.). Subsequently, all linear measurements were adjusted according to T. H. R. by adding mean differences when found significant ($P < 0.05$) using paired T-tests. Further, many morphology measures, including those for bill morphology, show seasonal and age-related variation. Therefore, only summer (May – August) measures were used and phenotypic data was adjusted to a predicted value for age one in May to produce single measures data for use with the GCTA, GRAMMAR-gamma and QTCAT methods. Linear mixed effects models with age and age^2 as covariates were used to age standardise, month was also included as a fixed factor. Year, cohort and ring number were fitted as random intercepts with ring number as a random slope to partition within-individual variation. Likelihood ratio tests of nested models were used to determine significance of age and month effects for each trait. If effects were significant, predicted values from the model were used to adjust measurements to age one in May before calculation of individual means (Kvalnes *et al.*, in press.).

Population Structure

Genetic variation in the small, stochastic subpopulations in the Helgeland system is likely to be structured due to genetic drift and limited gene flow. House sparrow microenvironment is dissimilar between agricultural and residential islands, and annual conditions vary between islands (Pärn *et al.*, 2012). The adaptive landscape may also vary between islands for this reason. Difference in bill morphology is observed between subpopulations, as evidenced by presence of island as a fixed factor in all best models determined via AICc model selection (Appendix II; Table III). Controlling for island as a fixed factor could, however, control away the variation which encompasses association between bill morphology phenotype and genotype. Inclusion of hatch year as a fixed factor may pose a similar problem if bill morphology is under directional selection. To evaluate the extent of this effect on power to detect extant associations, GWA analyses were run both with and without island and hatch year as fixed factors. As expected, *P* value estimates

for GWAS run without including these factors were marginally lower (results not shown), but this did not yield any additional significant associations under either the GRAMMAR-gamma or RepeatABEL method. Inclusion of island and hatch year as fixed factors in main analysis cannot explain why more large effect loci were not detected in the Helgeland study system.

In the Southern study system, inclusion of island as a fixed factor in analysis did alter results. SNPa77370 showed significant association with bill shape at the single marker set significance level if island was not included as a fixed factor, but the link disappeared if island was included in the analysis. Population structure may be more important in the Southern population as a consequence of low migration between islands. This could be particularly important if the CRIM1 gene region is under differential selection on different islands due to differences in environment, as including island as a fixed factor may then control away the effect we are interested in. The association between the CRIM1 region and the interrelated bill depth and bill shape may be preserved across populations in *P. domesticus*, but strength of its influence may be controlled by environmental factors. If the association is genuine it is several orders of magnitude weaker in the Southern dataset than in the Helgeland metapopulation.

Aside from QTCAT, methods used control for relatedness between individuals by including the GRM as a random effect. Stratified relatedness as a consequence of differing environment between islands poses a problem for this approach, as modelling of the covariance between individuals in this manner tests the restricted hypothesis of locus effect on phenotype independent of population structure or genetic background. QTCAT deals effectively with this issue by simultaneously associating all markers to the phenotype (Klasen *et al.*, 2016). That no significant QTC was detected using this method may be because repeated measures cannot be included in the model as a random effect. Significant association between bill depth and marker SNPa77340 in this study was only revealed using RepeatABEL, suggesting within-individual variation is an important contributor to bill morphology variation in this population and should be properly accounted for.

Multiple Testing Correction

Multiple testing within a gene region is a problem in candidate gene studies due to LD, which may result in false positives and type-1 error. Two commonly used adjustments are Bonferroni (1936) and Benjamini Hochburg (FDR) correction (Benjamini and Hochberg, 1995). The first is a family-wise error rate (FWER) correction which rejects null hypotheses with *P* value less than the desired

significance level (eg. 0.05) divided by the number of tests (here number of SNPs tested), ensuring that the probability of at least one false positive is less than 0.05. This type of correction is however conservative, increases the chance of type-2 error and substantially decreases power for detection of true significant results (Glickman *et al.*, 2014). False discovery rate (FDR) is a less conservative alternative, which ranks obtained *P* values and calculates significance based on comparison of each p_i with a fraction of 0.05 (k/n of eg. 0.05 for k= 1,2…n). This produces a result more conservative than simply comparing all *P* values to a chosen significance level, but less conservative than Bonferroni which compares all *P* values to 0.05/n. FDR gives the assurance that, at an FDR of 0.05, at most 5% of significant results were false positives (Glickman *et al.*, 2014).

Although FDR increases power compared to Bonferroni adjustment, both methods in fact result in very conservative *P* values and reduced power. An alternative can be derived if correlation between markers, which reduces statistical independence, is taken into account. Based on early effective number of tests (M_{eff}) calculation approaches (Cheverud, 2001, Nyholt, 2004) and using various methods to determine extent of correlation between markers, (Chen and Liu, 2011, Galwey, 2009, Gao *et al.*, 2008, Li and Ji, 2005) an estimate for M_{eff} can be calculated and used in place of n in the Bonferroni or Šídák correction. There is, however, no mathematical justification that FWER, the probability of making at least one type-I error, is controlled using this approach. Alternatively, a method based on allelic tests can be used, where main output is estimation of a local alpha level (α_{loc}) and FWER is controlled if distribution of test statistics is monotonically sub-Markovian of order k (it is reasonable to assume this condition is satisfied in GWA data) (Halle *et al.*, 2016, Moskvina and Schmidt, 2008). Custom methods (pers. comm. Halle, K., Langaas, M. Bakke, Ø), based on the order k FWER approximation method outlined in Halle *et al.* (2016) and adapted for traits with Gaussian distribution, are used here. This approximation method conditions on the previous k-1 neighbouring SNPs across the selected region. Firstly, one of each pair of SNPs which are in perfect LD ($r > 0.99$) is removed, as these pairs represent duplicated tests and lower the α_{loc} unnecessarily. The mean is then imputed for genetic covariates and k^{th} order approximation for the score test is used to produce a vector of score statistics (T). Our collection of m null hypotheses (no association between phenotype and genotype at marker k) are then tested with corresponding score statistics, using a closed form equation to be solved for α_{loc} (Halle *et al.*, 2016). For candidate genes, a full order method was used to control for FWER, this utilises the Genz Bretz algorithm (Genz, 1992; possible for up to 1000 SNPs) and conditions on all SNPs in the candidate gene. For GWAS, an order 3 approximation, which utilises the Miwa algorithm (Miwa *et al.*, 2003), was used to control for FWER. As the value for α_{loc} is dependent on the response variable and covariates defined in the score test model, different values were computed

for all traits (see Appendix II; Table III for best models used). This was done individually for each candidate gene (as different markers are included). Finally, α_{loc} for all candidate genes combined was defined for each bill morphology trait, as candidate gene SNPs should surpass this significance level for the association to be truly valid.

Pathway Determination

Pathway analysis was carried out for the gene CRIM1 as it was plausibly linked with bill depth in RepeatABEL analysis in the Helgeland dataset. Online methods iHop (Hoffmann and Valencia, 2004) and STRING (Szklarczyk *et al.*, 2015) were utilised to discover genes which may be functionally linked to CRIM1, first in *Gallus gallus*, then in all species. Subsequently, evidence for these links was manually curated to confirm that links discovered via text mining were valid. Genes whose products were involved in protein-protein interactions, positive or negative regulation (either via direct binding or through upstream mechanisms) or co-expression with CRIM1 were retained and a pathway diagram containing these genes was produced in Cytoscape 3.2.1 (Shannon *et al.*, 2003). Candidate genes which were examined in this study were added to the diagram, where links could be drawn between them and CRIM1 or its associated genes. Subsequently, available literature on genes in the pathway diagram was examined to uncover additional links between nodes. Genes were categorised into those which have been previously related to craniofacial development, those which have been previously been implicated in skeletal development but not craniofacial development specifically, and genes with no previous link to skeletal or craniofacial development. Genes in the first and second categories may especially be worth examining for association with bill morphology in subsequent studies. See Appendix III; Fig. V.

II: Supplementary Tables

Table I: Descriptive statistics for bill morphology traits in the Helgeland metapopulation of house sparrows (N = 1857 individuals, N = 2439 measurements).

Trait	Mean	SD	Within Individual Variance	Between Individual Variance	Total Variance
Bill Depth	8.146	0.277	0.04	0.15	019
Bill Shape	-0.002	0.260	0.04	0.13	0.17
Bill Length	13.702	0.539	0.12	0.64	0.76
Bill Size	-0.004	0.547	0.12	0.66	0.78

Table II: Descriptive statistics for bill morphology traits in the Southern populations of house sparrows (N = 710 individuals, N = 1343 measurements).

Trait	Mean	SD	Within Individual Variance	Between Individual Variance	Total Variance
Bill Depth	8.096	0.272	0.04	0.12	0.16
Bill Shape	0.015	0.247	0.04	0.10	0.14
Bill Length	13.557	0.503	0.13	0.43	0.56
Bill Size	-0.010	0.516	0.13	0.46	0.59

Table III: AICc best models for each of the four bill morphology traits for both single and repeated measures data. Models within Δ2 AICc of the best model for each trait are shown in grey. id is individual ID as determined by ring number.

Method	Trait	Model	AICc	Δ AICc
GRAMMAR-gamma	Bill Depth	BillD ~ sex + island + hatchyear	437.59	0
	Bill Shape	PC2 ~ sex + island + hatchyear	132.84	0
	Bill Length	BillL ~ island + hatchyear	3235.46	0
	Bill Size	PC1 ~ island + hatchyear	3218.35	0
		PC1 ~ sex + island + hatchyear	3220.34	1.99
RepeatABEL	Bill Depth	BillD ~ sex + age + island +month + (1|id)	247.63	0
	Bill Shape	PC2 ~ sex + age + island + month + (1|id)	-83.92	0
	Bill Length	BillL ~ sex + island + month + (1|id)	6035.43	0
		BillL ~ sex + age + island +month + (1|id)	6037.13	1.70
	Bill Size	PC1 ~ sex + age + island + month +(1|id)	6092.31	0
		PC1 ~ age + island + month + (1|id)	6093.69	1.38
		PC1 ~ sex + island + month + (1|id)	6093.86	1.55

Table IV: Genomic inflation factor (λ) for each bill morphology trait, GenABEL and RepeatABEL estimates. Local alpha level (α_{loc}) for each trait estimated using single-measures data and the full order method for all candidate gene SNPs, or the order 3 method to derive genome-wide α_{loc}.

Trait	λ (GenABEL)	λ (RepeatABEL)	Candidate Gene α_{loc}	Genome-Wide α_{loc}
Bill Depth	1.01 ± 0.0001	1.05 ±3.60e^{-5}	0.00009	3.14e^{-7}
Bill Shape	0.97 ±0.1500	1.01 ± 5.25e^{-5}	0.00013	3.14e^{-7}
Bill Length	1.01 ± 0.0001	1.03 ±5.74e^{-5}	0.00010	3.15e^{-7}
Bill Size	0.99 ±0.2590	1.01 ± 5.02e^{-5}	0.00009	3.14e^{-7}

Table V: Heritability estimates from different packages for the four studied bill morphology traits.

Trait	h^2 (GRAMMAR-gamma)	h^2 (RepeatABEL)	h^2 (GCTA)
Bill Depth	0.531	0.35	0.52
Bill Shape	0.545	0.38	0.53
Bill Length	0.497	0.38	0.52
Bill Size	0.496	0.40	0.53

Table VI: Summary statistics for the top 5 SNPs associated with bill shape for ALX1 in the GenABEL analysis. GenABEL lambda estimate for bill shape = 0.97 so inflation was not corrected for here. Top SNP is significant at the single gene significance level, α_{loc} = 0.001. For each SNP the table shows its name, chromosome, position (bp), the reference allele A1, effect allele A2, minor allele frequency, estimated effect size of A2 with standard error, adjusted *P* value, marker call rate, and Hardy-Weinburg *P* value.

SNP	Chromosome	Position	A1	A2	MAF	Effect (SE)	*P* value	Call Rate	HWE *P* value
SNPa174071	1A	38525018	C	A	0.11	0.049 ±0.015	0.0009	0.993	0.727
SNPa174122	1A	38496160	C	T	0.14	0.044 ±0.015	0.0032	1.000	0.843
SNPa174126	1A	38492884	C	T	0.13	0.044 ±0.015	0.0032	0.990	0.548
SNPa174084	1A	38517324	G	A	0.11	0.046 ±0.016	0.0037	0.996	0.646
SNPa174087	1A	38515244	T	C	0.11	0.046 ±0.016	0.0038	0.999	0.646

Table VII: Summary statistics for the top 5 SNPs associated with bill size for FGF8 in the RepeatABEL analysis. Effect size and P value corrected for λ = 1.01. Top SNP is significant at the single gene significance level, α_{loc} = 0.0009. For each SNP the table shows its name, chromosome, position (bp), the reference allele A1, effect allele A2, minor allele frequency, estimated effect size of A2 with standard error, adjusted *P* value, marker call rate, and Hardy-Weinburg *P* value.

SNP	Chromosome	Position	A1	A2	MAF	Effect (SE)	*P* value	Call Rate	HWE *P* value
SNPa276021	6	13375629	C	T	0.44	-0.067 ±0.020	0.0007	0.992	0.962
SNPa276015	6	13382614	G	T	0.33	-0.041 ±0.020	0.0447	0.993	0.091
SNPa275990	6	13396985	C	T	0.17	0.049 ±0.025	0.0498	0.998	0.568
SNPa275995	6	13395062	A	G	0.19	0.045 ±0.025	0.0628	0.997	0.877
SNPa275994	6	13396178	T	G	0.20	0.043±0.024	0.0718	0.997	1.000

Table VIII: Summary statistics for the top 5 SNPs associated with bill depth in the RepeatABEL GWAS. Effect size and *P* value corrected for λ=1.05. Top SNP is significant at Bonferroni genome-wide significance level, α_{loc} = 3.14^{e-7}. For each SNP the table shows its name, chromosome, position (bp), the reference allele A1, effect allele A2, minor allele frequency, estimated effect size of A2 with standard error, adjusted *P* value, marker call rate, and Hardy-Weinburg *P* value.

SNP	Chromosome	Position	A1	A2	MAF	Effect (SE)	*P* value	Call Rate	HWE *P* value
SNPa77348	3	21541818	C	T	0.39	0.052 ±0.010	1.208e-7	0.996	0.408
SNPa77303	3	21474218	T	G	0.36	0.049 ±0.011	1.655e-6	0.995	0.013
SNPa500795	4	1474181	A	G	0.12	-0.067 ±0.015	3.825e-6	0.995	0.541
SNPa196812	7	21744692	G	A	0.15	0.063 ±0.015	5.665e-6	0.993	0.073
SNPa101419	3	94810097	A	G	0.35	-0.046 ±0.011	1.035e-5	0.999	0.913

Table IX: Summary statistics for the top 5 SNPs associated with bill shape in the Southern population. λ = 1 for this trait. Results where island was not included as a fixed factor are presented in black and results where island was included as fixed factor are presented below in grey for comparison. When island is not included as a fixed factor, SNPa77370 is significant at the single marker set significance level (α_{loc}, = 0.0032) but not at the level for all SNPs tested in this dataset (α_{loc} = 0.0004).

SNP	Chromosome	Position	A1	A2	MAF	Effect (SE)	*P* value	Call Rate	HWE *P* value
SNPa77370	3	21566606	T	C	0.16	-0.053 ±0.018	0.0031	0.999	0.067
						-0.049 ±0.018	0.0054		
SNPa77351	3	21544510	G	T	0.41	0.040 ±0.014	0.0048	0.993	0.816
						0.037 ±0.014	0.0077		
SNPa77332	3	21520307	T	C	0.32	0.043 ±0.015	0.0052	0.997	0.904
						0.037 ±0.015	0.0139		
SNPa77349	3	21542748	C	T	0.45	0.034 ±0.014	0.0173	1.000	0.541
						0.031 ±0.014	0.0266		
SNPa77347	3	21540984	G	A	0.34	0.025 ±0.014	0.1012	0.999	0.947
						0.024 ±0.015	0.1196		

Table X: Genes within 1 Mbp of SNPa77348, which was significantly associated with bill depth in RepeatABEL GWAS. SNPa77348 position = 21541818 bp, closest upstream flanking gene with known function is CEBPZ and downstream is CRIM1. The closest upstream gene codes for a protein of unknown function.

Gene	Position (bp)	Analogous Gene
IV00_00011376	20469672	Cysteine-rich motor neuron 1 protein (CRIM1)
IV00_00011398	21691531	Protein of unknown function
IV00_00011399	21691544	CCAAT/enhancer-binding protein zeta (CEBPZ)
IV00_00011400	21706132	NADH dehydrogenase [ubiquinone] assembly factor 7 (NDUFAF7)
IV00_00011401	21715543	Serine/threonine-protein kinase D3 (PRKD3)
IV00_00011404	21765455	Glutaminyl-peptide cyclotransferase (QPCT)
IV00_00011405	21778947	Zinc transporter 6 (SLC30A6)
IV00_00011407	21802297	Spastin SPAST)
IV00_00011408	21839922	Protein dpy-30 homolog (Dpy30)
IV00_00011409	21863121	Protein Memo1 (Memo1)
IV00_00011414	21941050	3-oxo-5-alpha-steroid 4-dehydrogenase 2 (SRD5A2)
IV00_00011415	21982413	Protein ELYS (AHCTF1)
IV00_00011418	22042867	Saccharopine dehydrogenase-like oxidoreductase (SCCPDH)
IV00_00011421	22083275	Consortin (CNST)
IV00_00011422	22128457	Dimethyladenosine transferase 2%2C mitochondrial (TFB2M)
IV00_00011431	22538506	Kinesin-like protein 26B (KIF26B)

III: Supplementary Figures

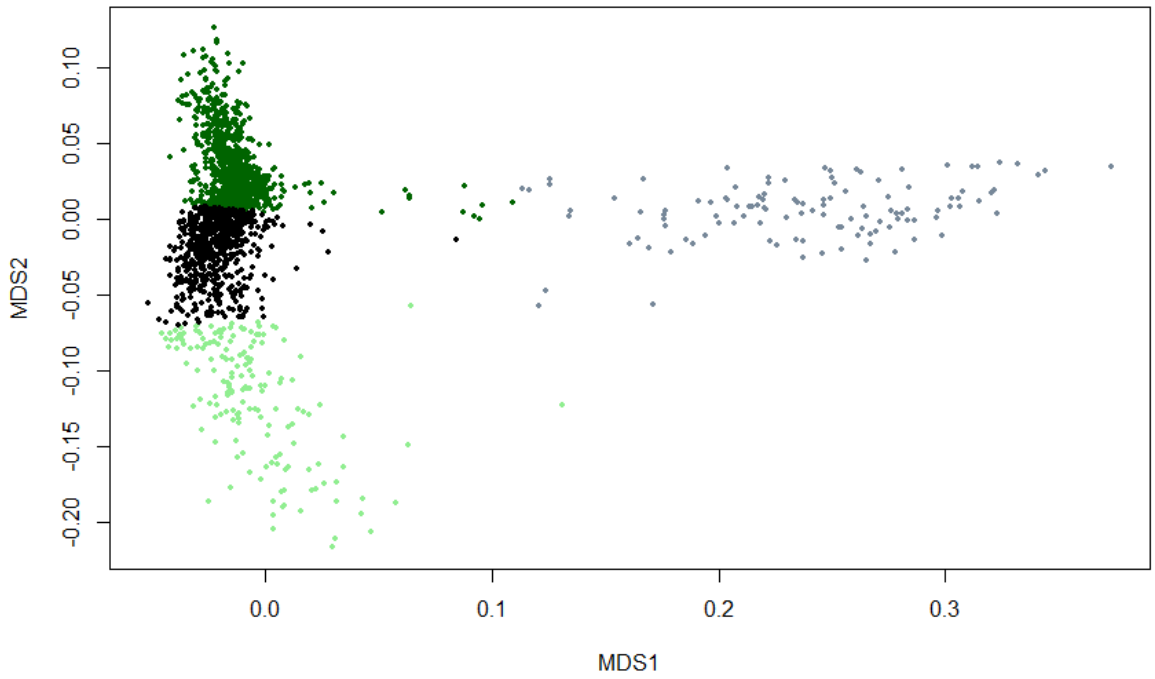

Figure I: Multidimensional scaling (k = 4) of the genomic relatedness matrix with each point representing one of 1857 individuals typed on 183109 SNPs. k = 4 was chosen as increasing number of dimensions beyond this point did not result in greater differentiation between islands by cluster. Genetic differences are partially sub-population related; grey points represent only individuals from Aldra and light green points only individuals from Gjerøy and Hestmannøy. Black points represent individuals from Nesøy, Træna, Selvær, Gjerøy, Hestmannøy, Indre Kvarøy and Aldra. Dark green points represent individuals from all the previous islands, plus Myken. The Aldra population is inbred and most genetically distinct.

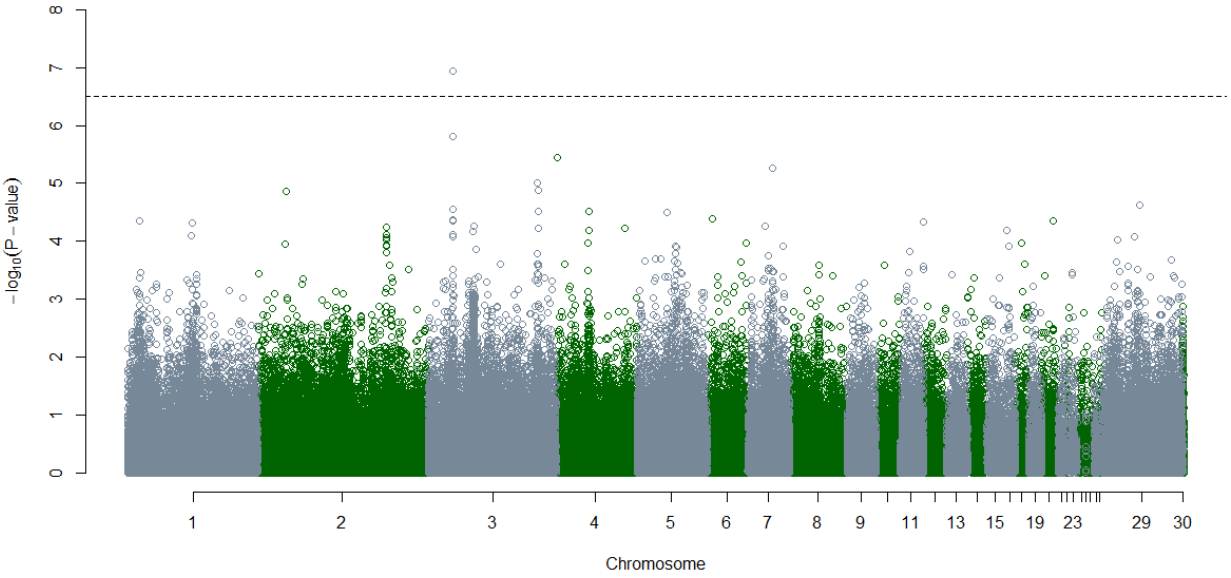

Figure II: Manhattan plot of repeated measures GWAS scan for bill depth (N= 4239 measurements) on 183109 SNPs. No SNPs on chromosome 16 or Z are included and neither are markers without a position (those with a zero value for chromosome). 'Chromosome' 30 is a linkage group with no chromosome name. Position of markers on the X axis corresponds to their bp position on their chromosome. Local alpha level when FWER of 0.05 is controlled for is $3.14e^{-7}$ (dotted line).

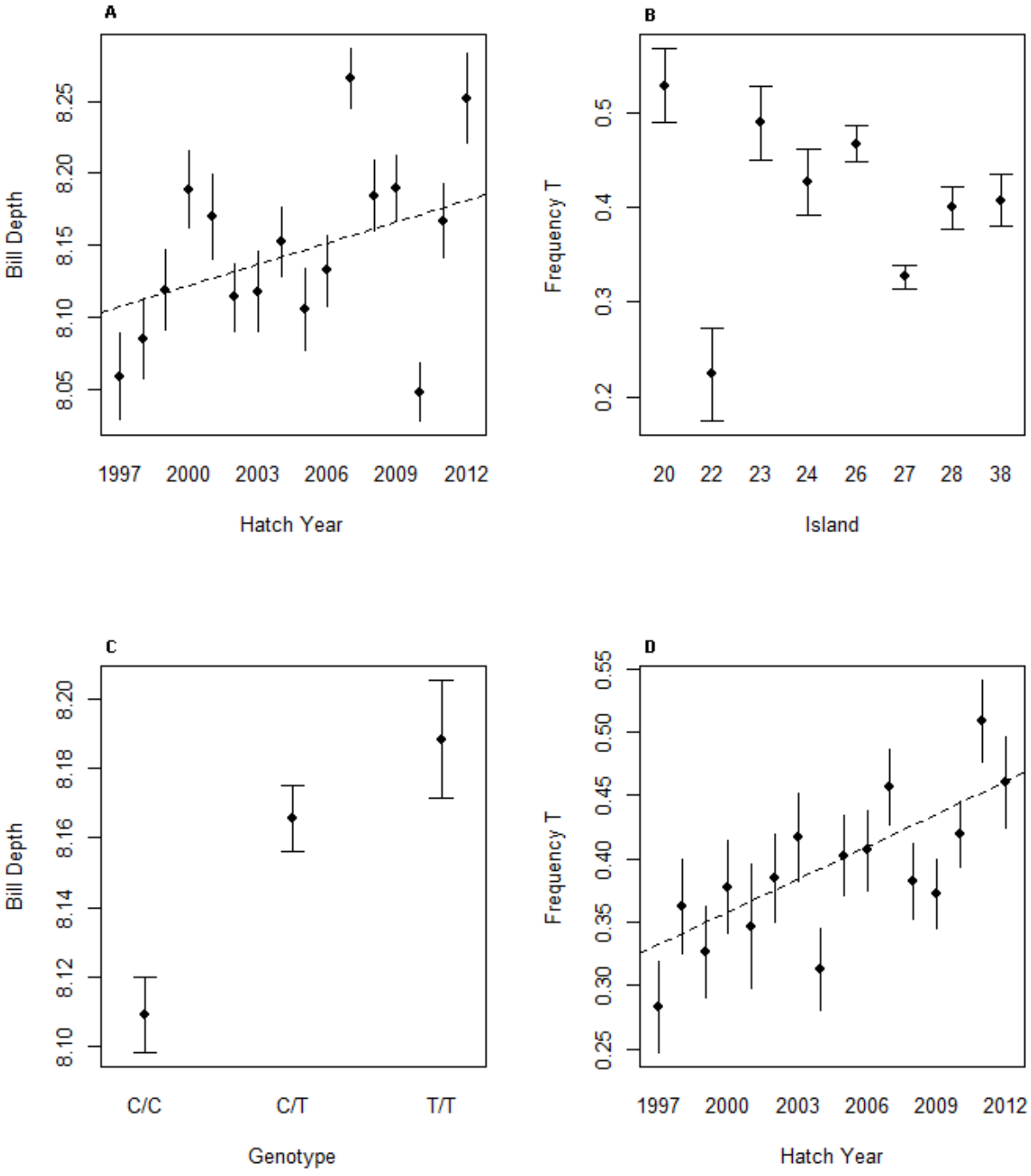

Figure III: Phenotypic and marker data exploration for bill depth and the marker significantly associated with this trait by RepeatABEL GWAS in the Helgeland metapopulation, SNPa77348. **A:** Mean adult bill depth (± SE) for hatch years 1997-2012. **B:** Frequency of the minor allele, T, (± SE) for each island. **C:** Genotype phenotype relationship and **D:** mean frequency (± SE) of the minor allele, T, by hatchyear. Regression of hatch year on bill depth is signifcant (β = 0.004 ± 0.001, t = 3.33, P <0.001). Regression of hatch year on MAF is also signifcant (β = 0.0086 ± 0.002, t = 4.603, P < 0.001).

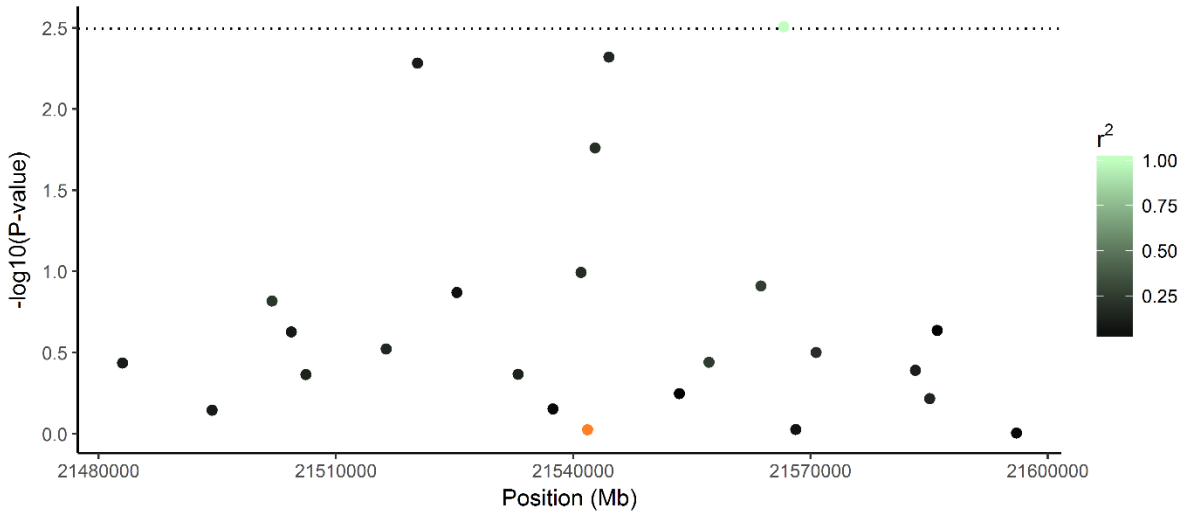

Figure IV: Manhattan-linkage disequilibrium plot for the 55 Kbp flanking regions either side of SNPa77348 in the Southern dataset. SNPa77348 is highlighted in orange. Result is derived from the RepeatABEL association test for bill shape where island was not included as a fixed factor, $\lambda = 1$ for this trait. SNPs are coloured after degree of LD with reference marker, SNPa77370 ($P = 0.0031$). The dashed line represents the single-marker set significance threshold ($\alpha l_{oc} = 0.00032$) where FWER is controlled to 0.05.

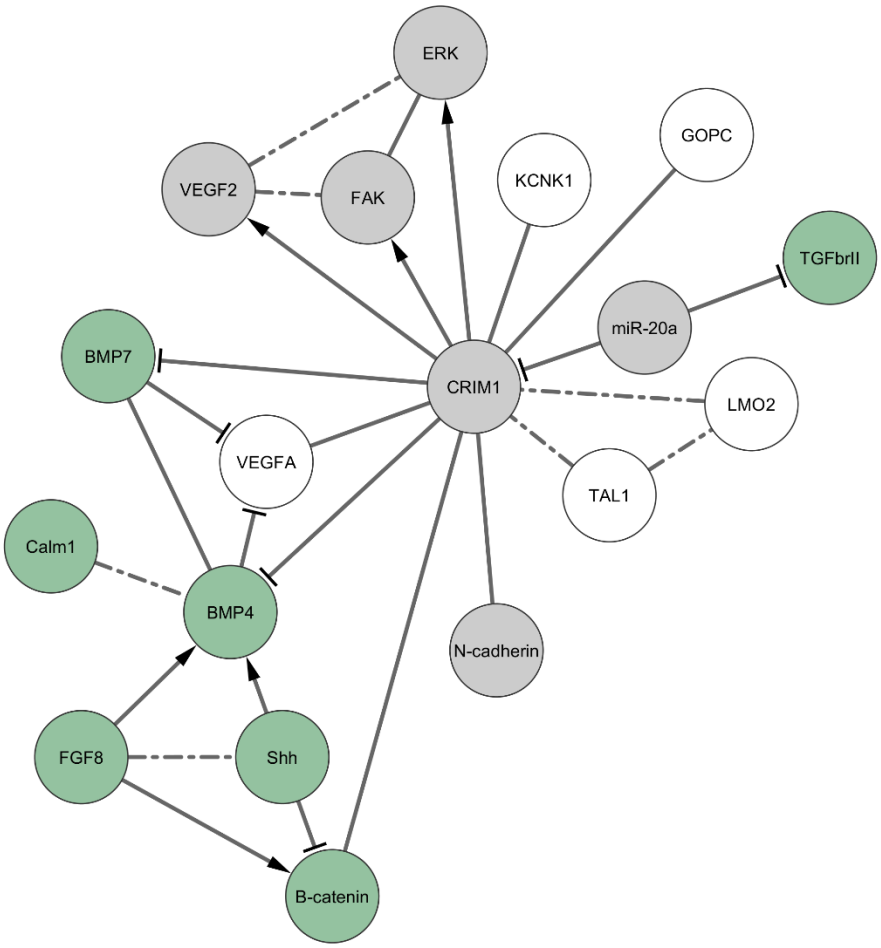

Figure V: Pathway diagram displaying links between CRIM1, plausibly identified as affecting bill depth by RepeatABEL GWAS in the Helgeland study system, and associated genes. Full edges indicate direct protein-protein interaction, dashed edges indicate co-expression, and bars and arrows indicate negative and positive regulation respectively. Nodes were categorised according to their potential for involvement in governing bill morphology. Green nodes represent genes which have previously been linked to craniofacial development in the literature, grey nodes indicate genes which have been linked to skeletal development but not to craniofacial development specifically, and white nodes indicate genes with no previous link to either craniofacial or skeletal development. Especially, genes represented by nodes coloured green or grey may be worth investigation for association with bill morphology in future candidate gene studies.