

## EVOLUTION

# Recent natural selection causes adaptive evolution of an avian polygenic trait

Mirte Bosse,<sup>1,2\*</sup> Lewis G. Spurgin,<sup>3,4\*</sup> Veronika N. Laine,<sup>1</sup> Ella F. Cole,<sup>3</sup> Josh A. Firth,<sup>3</sup> Phillip Gienapp,<sup>1</sup> Andrew G. Gosler,<sup>3</sup> Keith McMahon,<sup>3</sup> Jocelyn Poissant,<sup>5,6</sup> Irene Verhagen,<sup>1</sup> Martien A. M. Groenen,<sup>2</sup> Kees van Oers,<sup>1</sup> Ben C. Sheldon,<sup>3</sup> Marcel E. Visser,<sup>1,2</sup> Jon Slate<sup>5†</sup>

We used extensive data from a long-term study of great tits (*Parus major*) in the United Kingdom and Netherlands to better understand how genetic signatures of selection translate into variation in fitness and phenotypes. We found that genomic regions under differential selection contained candidate genes for bill morphology and used genetic architecture analyses to confirm that these genes, especially the collagen gene *COL4A5*, explained variation in bill length. *COL4A5* variation was associated with reproductive success, which, combined with spatiotemporal patterns of bill length, suggested ongoing selection for longer bills in the United Kingdom. Last, bill length and *COL4A5* variation were associated with usage of feeders, suggesting that longer bills may have evolved in the United Kingdom as a response to supplementary feeding.

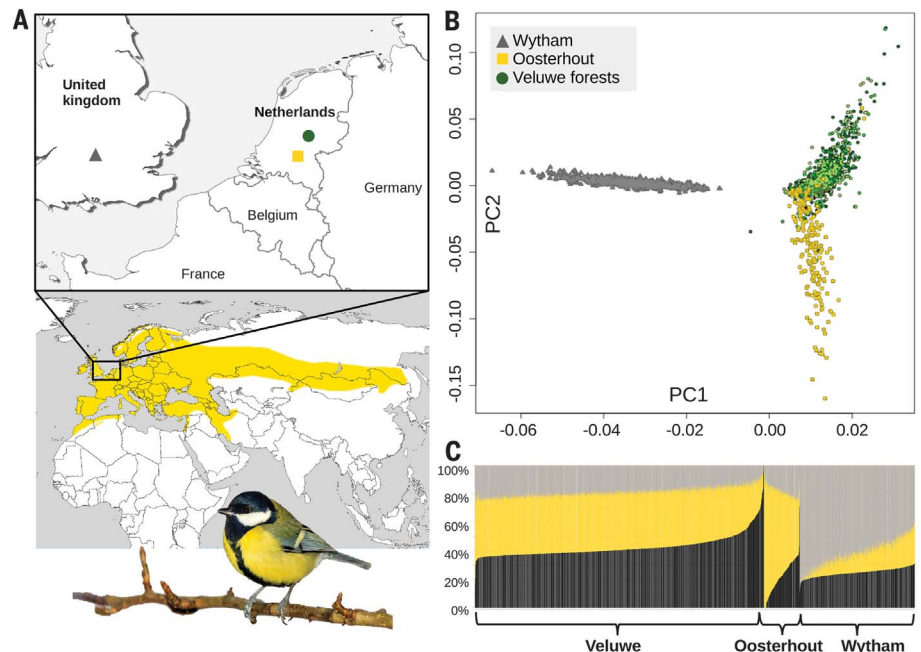
To demonstrate evolutionary adaptation in wild populations, we must identify phenotypes under selection, understand the genetic basis of those phenotypes along with effects on fitness, and identify potential drivers of selection. The best-known demonstrations of genes underlying evolution through natural selection usually involve strong selection (“hard sweeps”) on genetic variants that may be recently derived, with a major effect on variation in pre-selected phenotypes (1–3). However, most quantitative phenotypes are polygenic (4), and for these traits, selection is likely to act on many preexisting genetic variants of small effect (5). Detecting so-called polygenic selection is challenging because selection acts on multiple loci simultaneously, and selection coefficients are likely to be small (6). Most attempts to detect polygenic selection have focused on gene sets rather than individual loci (7). Furthermore, even if population genomic analyses identify genes under selection, these analyses are rarely combined with detailed ecological and behavioral data (8–10), and as a result, linking all three components of the genotype-phenotype-fitness continuum remains a challenge. In this study, we combined fine-scale ecological and genomic data to study adaptive evolution in the great tit (*Parus major*), a widespread and abundant passerine bird and well-known ecological model system (11) with excellent genomic resources (12).

To do so, we analyzed genomic variation within and among three long-term study populations from the United Kingdom (Wytham,  $n = 949$  birds) and the Netherlands (Oosterhout,  $n = 254$  birds and Veluwe,  $n = 1812$  birds) (Fig. 1A).

After filtering (supplementary materials, materials and methods), our data set comprised 2322 great tits typed at 485,122 single-nucleotide polymorphisms (SNPs). Levels of genetic diversity

were high, and linkage disequilibrium (LD) decayed rapidly within all three sample sites (fig. S1). Admixture and principal component analyses (PCA) both suggest that genetic structure is low (Fig. 1). These findings demonstrate a large effective population size and confirm high levels of gene flow in the species (12, 13), making the long-term study populations well suited to studying evolutionary adaptation.

To identify loci under divergent selection between the UK and Dutch populations, we ran a genome-wide association study (GWAS) using the first eigenvector from the PCA as a “phenotype” (EigenGWAS) (14). We identified highly significant outlier regions of the genome likely to be under divergent selection (Fig. 2A and fig. S2), which were supported by fixation index ( $F_{ST}$ ) analyses (fig. S3). The majority of these outlier regions contained candidate genes (such as *COL4A5*, *SIX2*, *TRPS1*, and *NELL1*) involved in skeletal development and morphogenesis (Fig. 2, A to C; table S1; and database S1). Genes associated with the ontology term “palate development” [Gene Ontology (GO) 0060021; genes *ALX4*, *BMPRIA*, *SATB2*, *INHBA*, and *GLI3*] were more significantly over-represented than any other GO term (Bonferroni-corrected  $P = 2.9 \times 10^{-5}$ ) (Fig. 2C and database S1). The strongest single-marker signal was found at the *LRRIQ1* gene (table S1 and database S1), where there was evidence of selection in Wytham, but not Veluwe (fig. S4). *LRRIQ1* is one of four genes located in the 240-kb region associated with beak shape in Darwin’s finches—arguably the best-known example of a trait undergoing adaptive evolution in the wild (15). Another EigenGWAS peak contained *VPSI3B*, a gene also associated with bill



**Fig. 1. Population structure of Western European great tits.** (A) Worldwide distribution of *P. major* and sampling locations in Wytham (gray triangle), Oosterhout (yellow square), and Veluwe (green circle). (B) Principal component analysis of genotype data. (C) ADMIXTURE plot with  $K = 3$ , which is both the most likely number of clusters and the number of geographically distinct sampling sites. Levels of genetic structure are low ( $F_{ST}$  Veluwe-Wytham = 0.006, and  $F_{ST}$  Veluwe-Oosterhout = 0.003).

<sup>1</sup>Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands. <sup>2</sup>Wageningen University and Research—Animal Breeding and Genomics, Netherlands. <sup>3</sup>Edward Grey Institute, Department of Zoology, University of Oxford, UK. <sup>4</sup>School of Biological Sciences, University of East Anglia, Norwich Research Park, UK. <sup>5</sup>Department of Animal and Plant Sciences, University of Sheffield, UK. <sup>6</sup>Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn, UK.

\*These authors contributed equally to this work. †Corresponding author. Email: j.slate@sheffield.ac.uk

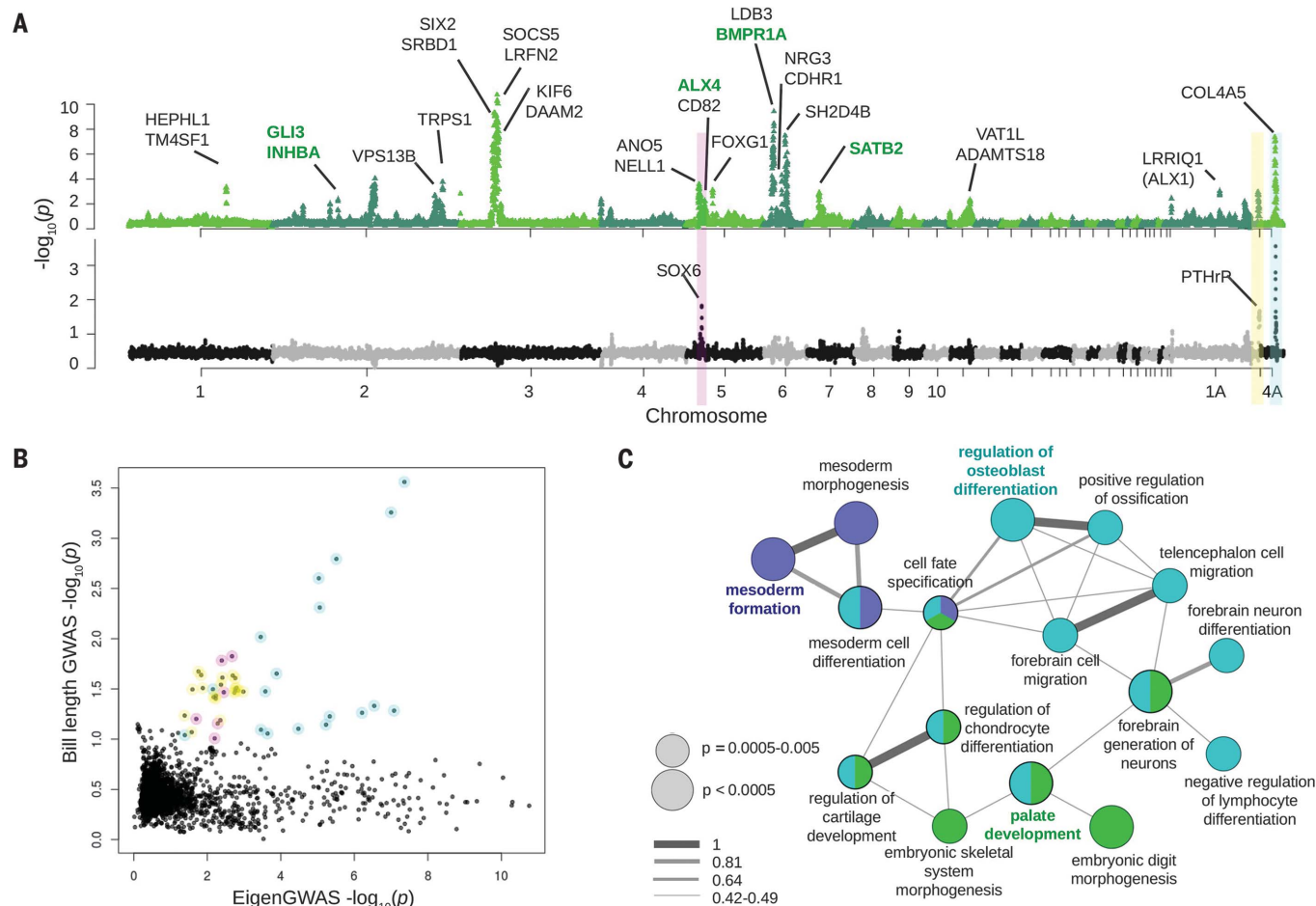
morphology in the Darwin's finch study and with facial dysmorphism in humans (16).

Our genetic analyses therefore suggested bill morphology as a key trait involved in differentiation between UK and Dutch great tit populations. Previously, UK great tit populations have been characterized as a different subspecies (*P. major newtoni*) compared with the rest of mainland Europe on the basis of bill length, but this classification is disputed (17), and it is unknown whether any bill length differences are adaptive in this species. We examined the genetic architecture of bill length in the UK population using two complementary approaches. First, we fitted all SNPs simultaneously in a mixture model analysis (18) and estimated that 3009 (95% credible interval 512 to 7163), or 0.8%, of the SNPs contributed to bill length variation, suggesting that bill length is highly polygenic. Collectively, these SNPs explained ~31% of the phenotypic variation. The proportion of variance in bill length explained by each chromosome scaled with its size, which is also consistent with a polygenic architecture (fig. S5) (4). Second, and consist-

ent with the mixture model analysis, we found multiple nominally significant SNPs in a GWAS on bill length in Wytham, but even the most significant ( $P = 1.6 \times 10^{-6}$ ) was not genome-wide significant after accounting for multiple testing, perhaps as a consequence of small effect size and modest sample size. Nonetheless, the SNPs were associated with bill length variation independently of overall body size (table S2). Using a sliding window approach, we found that the most significant GWAS regions largely overlapped with the most significant regions in the EigenGWAS and  $F_{ST}$  analyses (Fig. 2, A and B, and fig. S3), suggesting that genes involved in bill length have been under divergent selection between populations. We extracted SNPs from the most significant EigenGWAS peaks, calculated the summed effect of those SNPs on bill length, and compared this against a null distribution generated by randomly resampling the same number of SNPs and regions from across the genome. The regions under selection explained a small amount of variation (0.54%) in bill length in the UK population, but this is more than ex-

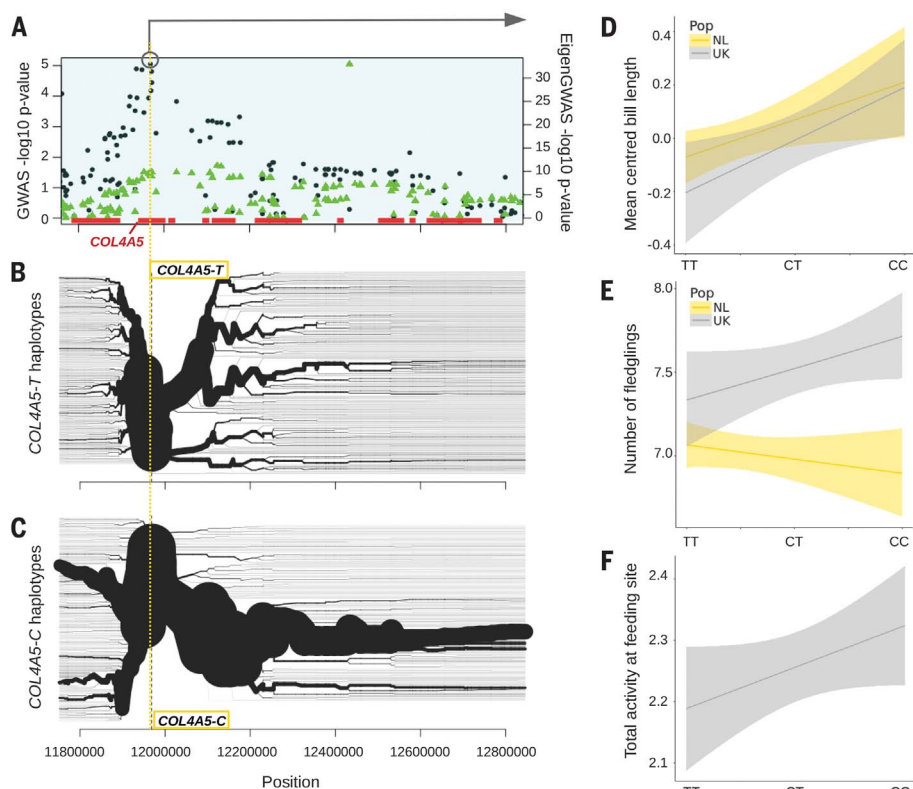
pected by chance ( $P = 0.004$ ) (fig. S6). Moreover, genomic prediction analysis using just the SNPs from the EigenGWAS peaks showed that UK birds had breeding values for longer bills than birds from the Netherlands (fig. S7), confirming that interpopulation differences in bill length is at least partially attributable to the loci that have been under recent selection.

The three genomic regions most notably associated with bill length variation and under likely divergent selection (Fig. 2, A and B) all contained genes with annotations that make them candidates for involvement in bill length. *SOX6* is a transcription factor, and *PTHrP* is a member of the parathyroid hormone family; both are essential for bone development (19, 20). *COL4A5* is a type IV collagen gene best known for its association with Alport's syndrome in humans (21) and has also been identified as a candidate for craniofacial disorders (22). The ~400-kb region of chromosome 4A containing the *COL4A5* gene was the region most notably associated with bill length (4 of the 24 most significant SNPs in the GWAS were in

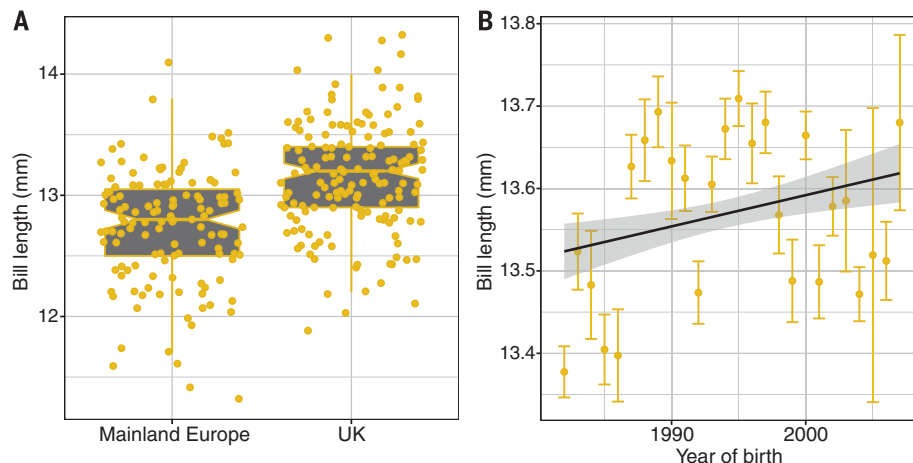


**Fig. 2. Differentiation and regions under selection across two great tit populations.** (A) (Top) EigenGWAS on PC1 across all autosomes, averaged over 200-kb sliding windows. Genes surrounding or covering peaks are indicated. Gene names highlighted in bold green belong to the most significant GO-term "palate development." (Bottom) GWAS for bill length in the UK population, averaged over 200-kb sliding windows. Color-

highlighted regions indicate peaks found in both the GWAS and EigenGWAS analyses. (B) EigenGWAS  $P$  values in relation to bill length GWAS  $P$  values averaged over 200-kb windows. Color-highlighted points correspond with the highlighted regions in (A). (C) GO network of genes in or surrounding the EigenGWAS peaks. Size of circles indicates significance, and line thickness indicates proportion of shared genes.



**Fig. 3. COL4A5 locus on chromosome 4A.** (A) Two-Mb zoom of EigenGWAS (green triangles) and GWAS (black circles)  $P$  values at the COL4A5 region (Fig. 2A, highlighted in blue). Red horizontal bars indicate gene locations. (B and C) Bifurcation diagram for haplotypes in Wytham, starting from the two alleles at the most significant GWAS SNP. Shown is the extended haplotype at the COL4A5-C-allele in (C), relative to the shorter haplotypes at the COL4A5-T allele in (B), which is consistent with a recent selective sweep around the COL4A5-C allele in the United Kingdom. (D) Bill length and COL4A5 genotype. The C allele is associated with longer bills ( $R^2 = 0.035$ ). (E) The COL4A5-C allele is associated with greater annual fledgling production in the UK population ( $R^2 = 0.015$ ). (F) COL4A5-C allele birds display greater winter feeding site activity. The y axis is  $\log_{10}$  transformed cumulative activity records ( $R^2 = 0.01$ ). Lines and shaded areas in (D) to (F) are fitted values and 95% confidence limits from general(ized) linear models (full data are plotted in figs. S8 and S9).



**Fig. 4. Spatiotemporal variation in bill length.** (A) Bill lengths of museum samples from the United Kingdom and mainland Europe. (B) Temporal variation in bill length in the Wytham population plotting annual means with standard error from 1982–2007. Line and (narrow) shaded area in (B) are fitted values and 95% confidence limits from a linear regression ( $R^2 = 0.004$ ); there are different scales on axes in (A) and (B).

COL4A5) (table S2) and belongs to the top three regions under strongest divergent selection between birds from the United Kingdom and the Netherlands (Fig. 2, A and B). A closer inspection of the individual SNPs within *SOX6* and *PTHrP* reveals numerous SNPs that are nominally significantly associated with bill length, but none as strongly as the COL4A5 SNPs; thus, we focused on the COL4A5 locus. Patterns of genetic variation at COL4A5 reveal a clear signature of recent selection for longer bills in the United Kingdom. First, the allele at the SNP that is most significantly associated with increased bill length (hereafter “COL4A5-C”) (Fig. 3D) is at higher frequency in the United Kingdom [0.54, bootstrap 95% confidence intervals (CIs) = 0.52 to 0.56] as compared with the two Dutch populations (Veluwe, 0.28; CI = 0.27 to 0.29; Oosterhout, 0.26; CI = 0.23 to 0.29). Second, extended haplotype homozygosity tests confirm that the haplotype carrying the COL4A5-C allele extends further than alternative haplotypes within Wytham (Fig. 3, A to C). The COL4A5-C haplotype is longer and more abundant in Wytham compared with Veluwe, and LD at this locus is much higher in Wytham, suggesting that selection is United Kingdom-specific (fig. S8). Third, SNP data from 15 European populations, including three UK populations, show that the COL4A5-C allele is at a higher frequency across the United Kingdom than across Europe, which is consistent with selection on this gene in the UK.

To further elucidate how natural selection has shaped variation in bill length across the two populations, we tested how variation at the COL4A5 locus was related to annual reproductive success. We found differences in the relationship between COL4A5 genotype and the number of chicks fledged between the two populations [zero-inflated Poisson generalized linear mixed model (GLMM), interaction between genotype and population,  $n = 3076$  breeding attempts from 1790 birds; estimate =  $-0.40 \pm 0.17$ ;  $P = 0.016$ ] (Fig. 3E). The interaction was significant because the associations between genotype and bill length in the two populations were in opposite directions; in the United Kingdom, the number of copies of the “long-billed” COL4A5-C allele was positively associated with fledgling production ( $n = 868$  breeding attempts from 516 birds; estimate =  $0.23 \pm 0.11$ ;  $P = 0.046$ ) (Fig. 3E and fig. S9), whereas in the Dutch birds, COL4A5-C was negatively, but not significantly, associated with fewer fledglings ( $n = 2208$  breeding attempts from 1274 birds; estimate =  $-0.16 \pm 0.10$ ;  $P = 0.093$ ). The relationship between fledgling production and COL4A5 genotype did not arise because long-billed genotype birds were more likely to produce offspring (binomial GLMM,  $n = 3076$  breeding attempts from 1790 birds; estimate =  $-0.20 \pm 0.17$ ;  $P = 0.91$ ); rather, when we only considered “successful” breeding attempts in which at least one fledgling was produced, long-billed genotype birds produced more fledglings (Poisson GLMM,  $n = 2690$  breeding attempts from 1612 birds; estimate =  $0.058 \pm 0.024$ ;  $P = 0.018$ ). Thus, we suggest that the COL4A5 allele associated with longer bills confers a fitness advantage in the UK population.



To better understand the evolutionary consequences of selection for longer bills in the UK population, we examined spatiotemporal variation in bill length. In museum samples from the United Kingdom and mainland Europe, the UK individuals had considerably longer bills [ $n = 291$  museum specimens; estimate =  $0.40 \pm 0.06$  mm;  $P = 5.2 \times 10^{-12}$ ; coefficient of determination ( $R^2$ ) = 0.16] (Fig. 4A), in accordance with a previous study (17). Using a 26-year data set from live birds in Wytham, we found that bill length has increased significantly over recent years (1982–2007;  $n = 2489$  birds; estimate =  $0.004 \pm 0.001$  mm per year;  $P = 0.0038$ ;  $R^2$  of year effect = 0.004; with tarsus length fitted as a covariate, the significant temporal increase in bill length remained significant,  $n = 2485$  birds; estimate =  $0.005 \pm 0.001$  mm per year;  $P = 0.0001$ ;  $R^2$  of year effect = 0.003) (Fig. 4B and table S3). This effect, although weak in terms of the variance explained, is not due to stochastic variation among years (randomization test,  $P = 0.02$ ) (supplementary materials) and is equivalent to an evolutionary rate of change of 0.0154 Haldanes; in a large review of phenotypic change in wild-animal populations, this rate was exceeded in just 641 of 2420 estimates (23).

Selection on bill-length has been documented multiple times in birds and is typically associated with variation in food availability (24). No differences in the natural diet of great tits between the United Kingdom and mainland Europe are known. In contrast, bird feeding by the public has been widespread in the UK since the 19th century; it is estimated that it occurs in over 50% of gardens (25) and that the United Kingdom's expenditure on bird seed is twice that spent in the whole of mainland Europe (26). Great tits are particularly good at exploiting bird feeders (27), and therefore, we investigated whether supplementary feeding could have been a driver of selection on bill length in UK great tits, similar to that proposed in UK blackcap (*Sylvia atricapilla*) populations (28). Radio frequency identification (RFID) bird feeders throughout Wytham recorded RFID-tagged great tit use of supplementary food

over the course of three winters (29). We found that *COL4A5-C* homozygotes displayed a higher propensity to use the feeders compared with that of heterozygotes or short-billed homozygotes ( $n = 444$ ; estimate =  $-0.17 \pm 0.08$ ;  $P = 0.03$ ) (Fig. 3F). There was some variation in the extent of this effect across winter seasons (fig. S10), and the strength and consistency of this effect, along with the mechanisms behind it, requires further investigation. Encouragingly, however, a follow-up analysis that used a more recent data set gathered from high-resolution RFID feeders (but on ungenotyped birds) showed a positive relationship between feeding propensity and bill length ( $n = 1806$  observations of 183 birds; estimate =  $0.15 \pm 0.05$ ;  $P = 0.004$ ) (fig. S11).

Together, our results provide a detailed example of natural selection in a wild animal. Starting with a bottom-up analysis of genomic data, and no preselected phenotypes, we have demonstrated polygenic adaptation by providing associations between loci that have responded to selection, fitness variation, phenotypic variation, microevolutionary change, and a possible driver of selection. Combining large-scale genomic and ecological data in natural populations will substantially enhance our understanding of both the mechanistic basis and evolutionary consequences of natural selection.

## REFERENCES AND NOTES

- C. R. Linnen, E. P. Kingsley, J. D. Jensen, H. E. Hoekstra, *Science* **325**, 1095–1098 (2009).
- S. Rost et al., *Nature* **427**, 537–541 (2004).
- S. Lamichhaney et al., *Science* **352**, 470–474 (2016).
- J. Yang et al., *Nat. Genet.* **43**, 519–525 (2011).
- M. C. Turchin et al., *Nat. Genet.* **44**, 1015–1019 (2012).
- J. K. Pritchard, J. K. Pickrell, G. Coop, *Curr. Biol.* **20**, R208–R215 (2010).
- J. J. Berg, G. Coop, *PLOS Genet.* **10**, e1004412 (2014).
- R. D. H. Barrett, H. E. Hoekstra, *Nat. Rev. Genet.* **12**, 767–780 (2011).
- C. Pardo-Diaz, C. Salazar, C. D. Jiggins, *Methods Ecol. Evol.* **6**, 445–464 (2015).
- J. R. Stinchcombe, H. E. Hoekstra, *Heredity* **100**, 158–170 (2008).
- A. Gosler, *The Great Tit* (Hamlyn Species Guides, 1993).
- V. N. Laine et al., *Nat. Commun.* **7**, 10474 (2016).
- N. E. M. Van Bers et al., *Mol. Ecol. Resour.* **12**, 753–770 (2012).
- G.-B. Chen, S. H. Lee, Z.-X. Zhu, B. Benyamin, M. R. Robinson, *Heredity* **117**, 51–61 (2016).
- S. Lamichhaney et al., *Nature* **518**, 371–375 (2015).
- I. Balikova et al., *Hum. Mutat.* **30**, E845–E854 (2009).
- A. G. Gosler, *Bull. Br. Ornithol. Club* **119**, 47–55 (1999).
- G. Moser et al., *PLOS Genet.* **11**, e1004969 (2015).
- N. Hagiwara, *Dev. Dyn.* **240**, 1311–1321 (2011).
- H. M. Kronenberg, *Ann. N. Y. Acad. Sci.* **1068**, 1–13 (2006).
- D. F. Barker et al., *Science* **248**, 1224–1227 (1990).
- J. J. Jonsson et al., *J. Med. Genet.* **35**, 273–278 (1998).
- A. P. Hendry, T. J. Farrugia, M. T. Kinnison, *Mol. Ecol.* **17**, 20–29 (2008).
- P. R. Grant, B. R. Grant, *Science* **296**, 707–711 (2002).
- M. E. Orros, M. D. E. Fellowes, *Acta Ornithol.* **50**, 43–58 (2015).
- D. N. Jones, S. James Reynolds, *J. Avian Biol.* **39**, 265–271 (2008).
- P. Tryjanowski et al., *Sci. Rep.* **5**, 11858 (2015).
- G. Rolshausen, G. Segelbacher, K. A. Hobson, H. M. Schaefer, *Curr. Biol.* **19**, 2097–2101 (2009).
- R. A. Crates et al., *J. Avian Biol.* **47**, 678–689 (2016).

## ACKNOWLEDGMENTS

We thank the many researchers who collected material and data for the Wytham study: L. Vernooij, P. de Goede, and H. Bouwmeester for fieldwork on the Dutch populations; and C. Mateman for laboratory assistance. The Natural History Museums in London (NHM) and Oxford (OUMNH) kindly granted us access to great tit specimens. R. Butlin, N. Nadeau, and P. Nosil provided helpful comments on the manuscript. This work was supported by grants from the European Research Council (grants 339092 to M.E.V., 250164 to B.C.S., and 202487 to J.S.), Biotechnology and Biological Sciences Research Council (grant BB/N011759/1 to L.G.S.), and Natural Environment Research Council (grant NE/J012599/1 to J.S.). M.B., L.G.S., M.A.M.G., B.C.S., M.E.V., and J.S. designed the study. M.B., L.G.S., and J.S. analyzed the genomic data for signatures of selection. V.N.L., P.G., and J.S. analyzed estimated trait genetic architectures. V.N.L. performed gene ontology analyses. A.G.G., K.M., J.P., and I.V. measured and analyzed bills. L.G.S., E.F.C., and J.A.F. collected and analyzed bird feeding station data. E.F.C., J.A.F., A.G.G., K.v.O., B.C.S., and M.E.V. coordinated and collected ecological data and DNA samples. M.A.M.G., K.v.O., M.E.V., and J.S. coordinated collection of SNP data. M.B., L.G.S., V.N.L., and J.S. cleaned and quality-control checked SNP data. M.B., L.G.S., and J.S. wrote the manuscript with input from all other authors. The data described in the paper are archived on Dryad with accession number doi:10.5061/dryad.p03j0.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/358/6361/365/suppl/DC1  
Materials and Methods  
Supplementary Text  
Tables S1 to S3  
Figs. S1 to S9  
References (30–44)  
Database S1

2 November 2016; resubmitted 19 May 2017  
Accepted 12 September 2017  
10.1126/science.aal3298

## Recent natural selection causes adaptive evolution of an avian polygenic trait

Mirte Bosse, Lewis G. Spurgin, Veronika N. Laine, Ella F. Cole, Josh A. Firth, Phillip Gienapp, Andrew G. Gosler, Keith McMahon, Jocelyn Poissant, Irene Verhagen, Martien A. M. Groenen, Kees van Oers, Ben C. Sheldon, Marcel E. Visser and Jon Slate

*Science* **358** (6361), 365-368.  
DOI: 10.1126/science.aal3298

### Recent natural selection in a wild bird

Many studies have found evidence of rapid evolution in response to environmental change. In most cases, there has been some suggestion of which traits might be most responsive ahead of time. Bosse *et al.* turn this approach on its head by using genomic regions with a signature of selection to identify traits that are changing. In great tits (*Parus major*) in the United Kingdom, genomic regions showing selection invariably contained genes associated with bill growth. Indeed, U.K. birds not only have longer bills, but these longer bills are associated with increased fitness. These changes likely reflect an increase in domestic garden bird feeders over the past several decades.

*Science*, this issue p. 365

#### ARTICLE TOOLS

<http://science.sciencemag.org/content/358/6361/365>

#### SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2017/10/19/358.6361.365.DC1>

#### REFERENCES

This article cites 43 articles, 7 of which you can access for free  
<http://science.sciencemag.org/content/358/6361/365#BIBL>

#### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)