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

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

**Affiliations:**

1 Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, the Netherlands

2 Animal Breeding and Genomics Centre, Wageningen University, the Netherlands

3 Edward Grey Institute, Department of Zoology, University of Oxford, United Kingdom

4 School of Biological Sciences, University of East Anglia, Norwich Research Park, United Kingdom

5 Department of Animal and Plant Sciences, University of Sheffield, United Kingdom

6 Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn, United Kingdom

*Correspondence to: j.slate@sheffield.ac.uk

† These authors contributed equally to this manuscript

**One Sentence Summary:** We identify genomic regions that have evolved under selection, and that explain variation in bill length and fitness in great tits.

**Abstract**: We use extensive data from a long-term study of great tits (*Parus major*) in the UK 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 UK. Finally, bill length and *COL4A5* variation were associated with usage of feeders, suggesting that longer bills may have evolved in the UK as a response to supplementary feeding.

## Main Text:

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 by natural selection usually involve strong selection (‘hard sweeps’) on genetic variants, that may be recently derived, with a major effect on variation in preselected phenotypes (*1*–*3*). However, most quantitative phenotypes are polygenic (*4*) and for these traits selection is likely to act on many pre-existing 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 (e.g  (*7*)). Furthermore, even if population genomics 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 combine 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 UK (Wytham, n = 949) and the Netherlands (Oosterhout, n = 254 and Veluwe, n = 1812; Fig. 1A).

After filtering (see methods), our dataset comprised 2322 great tits typed at 485,122 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, B and C). 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 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, S2), which were supported by $F_{\mathrm{ST}}$ analyses (fig. S3). The majority of these outlier regions contained candidate genes (e.g. *COL4A5, SIX2, TRPS1, NELL1*) involved in skeletal development and morphogenesis (Fig. 2, A to C, table S1 and external database S1). Genes associated with the ontology term "palate development" (GO:0060021; genes *ALX4, BMPR1A, SATB2, INHBA, GLI3*) were more significantly overrepresented than any other GO term (Fig. 2C; Bonferroni-corrected $p = 2.9 \times 10^{-5}$; external database S1). The strongest single-marker signal was found at the *LRRIQ1* gene (table S1, external database S1), where there was evidence of selection in Wytham, but not Veluwe (fig. S4). LRRIQ1 is one of four genes located in the 240kb 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 *VPS13B*, a gene also associated with bill 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 to the rest of mainland Europe based on 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 512-7163), 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 (*4*) (fig. S5). Second, and consistent 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 $F_{\mathrm{ST}}$ analyses (Fig. 2, A and B, 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 expected 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 inter-population 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* 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*), that has also been identified as a candidate for craniofacial disorders (*22*). The ~400kb 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 *COL4A5*; Table S2), and belongs to the top three regions under strongest divergent selection between birds from the UK and 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 focus on the *COL4A5* locus hereafter. Patterns of genetic variation at *COL4A5* reveal a clear signature of recent selection for longer bills in the UK. 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 UK (0.54, bootstrap 95% confidence intervals = 0.52-0.56) compared to the two Dutch populations (Veluwe: 0.28, CI = 0.27-0.29; Oosterhout: 0.26, CI = 0.23-0.29). 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 to Veluwe, and LD at this locus is much higher in Wytham, suggesting selection is UK-specific (fig. S8). Third, SNP data from 15 European populations, including 3 UK populations, shows that the *COL4A5-C* allele is at a higher frequency across the UK than across Europe (LGS *et al.* In Prep), 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 GLMM, interaction between genotype and population: *n* = 3076 breeding attempts from 1790 birds, estimate = -0.40 ± 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 UK, 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 ± 0.11, p = 0.046, Fig. 3E; 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 ± 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 ± 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 ± 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 UK and mainland Europe, the UK individuals had considerably longer bills (*n* = 291, estimate = 0.40 ± 0.06 mm, p = $5.2 \times 10^{-12}$, $R^2 = 0.16$, Fig. 4A), in accordance with a previous study (*17*). Using a 26-year dataset from live birds in Wytham, we found that bill length has increased significantly over recent years (1982-2007; *n* = 2489, estimate = 0.004 ± 0.001 mm per year, p = 0.0038, $R^2$ of year effect = 0.004, Fig. 4B, table S3; with tarsus length fitted as a covariate, the significant temporal increase in bill length remained significant - *n* = 2485, estimate = 0.005 ± 0.001 mm per year, p = 0.0001, $R^2$ of year effect = 0.003). This effect, though 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 UK 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 it occurs in over 50% of gardens (*25*) and that the UK'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 utilization 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 to heterozygotes or short-billed homozygotes ($n$ = 444, estimate = -0.17 ± 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 using a more recent dataset gathered from high-resolution RFID feeders (but on un-genotyped birds) showed a positive relationship between feeding propensity and bill length (n = 1806 observations of 183 birds, estimate = 0.15 ± 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 significantly enhance our understanding of both the mechanistic basis and evolutionary consequences of natural selection.

## References and Notes:

1. C. R. Linnen, E. P. Kingsley, J. D. Jensen, H. E. Hoekstra, On the origin and spread of an adaptive allele in deer mice. *Sci. (New York, NY)*. **325**, 1095–1098 (2009).
2. S. Rost *et al.*, Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature*. **427**, 537–41 (2004).
3. S. Lamichhaney *et al.*, A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* . **352** (2016).
4. J. Yang *et al.*, Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
5. M. C. Turchin *et al.*, Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* **44**, 1015–9 (2012).
6. J. K. Pritchard, J. K. Pickrell, G. Coop, The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Curr. Biol.* **20** (2010), , doi:10.1016/j.cub.2009.11.055.
7. J. J. Berg, G. Coop, A Population Genetic Signal of Polygenic Adaptation. *PLoS Genet.* **10**, e1004412 (2014).
8. R. D. H. Barrett, H. E. Hoekstra, Molecular spandrels: tests of adaptation at the genetic level. *Nat. Rev. Genet.* **12**, 767–780 (2011).
9. C. Pardo-Diaz, C. Salazar, C. D. Jiggins, Towards the identification of the loci of adaptive evolution. *Methods Ecol. Evol.* **6**, 445–464 (2015).
10. J. R. Stinchcombe, H. E. Hoekstra, Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*. **100**, 158–170 (2007).
11. A. Gosler, *The great tit* (Hamlyn Species Guides, 1993).
12. V. Laine *et al.*, Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* (2016), doi:10.1038/ncomms10474.

13. N. E. M. Van Bers *et al.*, The design and cross-population application of a genome-wide SNP chip for the great tit Parus major. *Mol. Ecol. Resour.* **12**, 753–770 (2012).

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

15. S. Lamichhaney *et al.*, Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*. **518**, 371–375 (2015).

16. I. Balikova *et al.*, Deletions in the *VPS13B* ( *COH1* ) gene as a cause of Cohen syndrome. *Hum. Mutat.* **30**, E845–E854 (2009).

17. A. G. Gosler, A comment on the validity of the British Great Tit Parus major newtoni. *Bull. Br. Ornithol. Club*. **119**, 47–55 (1999).

18. G. Moser *et al.*, Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. *PLOS Genet.* **11**, e1004969 (2015).

19. N. Hagiwara, Sox6, jack of all trades: A versatile regulatory protein in vertebrate development. *Dev. Dyn.* **240**, 1311–1321 (2011).

20. H. M. Kronenberg, PTHrP and Skeletal Development. *Ann. N. Y. Acad. Sci.* **1068**, 1–13 (2006).

21. D. F. Barker *et al.*, Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science*. **248**, 1224–1227 (1990).

22. J. J. Jonsson *et al.*, Alport syndrome, mental retardation, midface hypoplasia, and elliptocytosis: a new X linked contiguous gene deletion syndrome? *J Med Genet*. **35**, 273–278 (1998).

23. A. P. Hendry, T. J. Farrugia, M. T. Kinnison, Human influences on rates of phenotypic change in wild animal populations. *Mol. Ecol.* **17**, 20–29 (2008).

24. P. R. Grant, B. R. Grant, Unpredictable evolution in a 30-year study of Darwin's finches. *Science*. **296**, 707–711 (2002).

25. M. E. Orros, M. D. E. Fellowes, Wild Bird Feeding in an Urban Area: Intensity, Economics and

Numbers of Individuals Supported. *Acta Ornithol.* **50**, 43–58 (2015).

26. D. N. Jones, S. James Reynolds, Feeding birds in our towns and cities: a global research opportunity. *J. Avian Biol.* **39**, 265–271 (2008).

27. P. Tryjanowski *et al.*, Who started first? Bird species visiting novel birdfeeders. *Sci. Rep.* **5**, 11858 (2015).

28. G. Rolshausen, G. Segelbacher, K. A. Hobson, H. M. Schaefer, Contemporary Evolution of Reproductive Isolation and Phenotypic Divergence in Sympatry along a Migratory Divide. *Curr. Biol.* **19**, 2097–2101 (2009).

29. R. A. Crates *et al.*, Individual variation in winter supplementary food consumption and its consequences for reproduction in wild birds. *J. Avian Biol.* (2016), doi:10.1111/jav.00936.

30. S. Purcell *et al.*, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

31. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

32. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

33. R Development Core Team, R. D. C. Team, Ed., R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* (2011), , doi:10.1007/978-3-540-74686-7.

34. P. C. Sabeti *et al.*, Detecting recent positive selection in the human genome from haplotype structure. *Nature*. **419**, 832–837 (2002).

35. O. Delaneau, J. Marchini, J.-F. Zagury, A linear complexity phasing method for thousands of genomes. *Nat. Methods*. **9**, 179–181 (2011).

36. M. Gautier, R. Vitalis, rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*. **28**, 1176–1177 (2012).

37. K. Tang, K. R. Thornton, M. Stoneking, A new approach for using genome scans to detect

recent positive selection in the human genome. *PLoS Biol.* **5**, 1587–1602 (2007).

38. G. Bindea *et al.*, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. **25**, 1091–1093 (2009).

39. A. G. Gosler, Pattern and process in the bill morphology of the Great Tit Parus major. *Ibis (Lond. 1859).* **129**, 451–476 (2008).

40. J. D. Hadfield, MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Softw.* **33**, 1–22 (2010).

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

42. Y. S. Aulchenko, D.-J. de Koning, C. Haley, Genomewide Rapid Association Using Mixed Model and Regression: A Fast and Simple Method For Genomewide Pedigree-Based Quantitative Trait Loci Association Analysis. *Genetics*. **177** (2007).

43. S. Bouwhuis *et al.*, Great tits growing old: selective disappearance and the partitioning of senescence to stages within the breeding cycle. *Proc. R. Soc. B Biol. Sci.* **276**, 2769–77 (2009).

44. M. Erbe *et al.*, Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* **95**, 4114–4129 (2012).

**Acknowledgements:** We thank the many researchers who collected material and data for the Wytham study, Louis Vernooij, Piet de Goede and Henri Bouwmeester for fieldwork on the Dutch populations and Christa Mateman for lab 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 ERC (339092 to MEV, 250164 to BCS and 202487 to JS), BBSRC (BB/N011759/1 to LGS) and NERC (NE/J012599/1 to JS). Author contributions: 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.vO., B.C.S and M.E.V. coordinated and collected ecological data and DNA

samples. M.A.M.G., K vO., M.E.V. and J.S. coordinated collection of SNP data. M.B., L.G.S, V.N.L. and J.S. cleaned and QC 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 XXX.

## Supplementary Materials

Materials and Methods
Supplementary Text
Tables S1 – S3
Fig S1 – S9
Caption for database S1
References (*30*–*44*)

**Fig. 1. Population structure of Western European great tits.** (**A**) Worldwide distribution of *P. major* and sampling locations in Wytham (▲) Oosterhout (■) and Veluwe (●). (**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).

**Fig. 2. Differentiation and regions under selection across two great tit populations.** (**A**) Upper panel: EigenGWAS on PC1 across all autosomes, averaged over 200kb sliding windows. Genes surrounding or covering peaks are indicated. Gene names highlighted in bold green belong to the most significant GO-term 'palate development'. Lower panel: GWAS for bill length in the UK population, averaged over 200kb 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 200kb windows. Color-highlighted points correspond with the highlighted regions in (**A**). (**C**) Gene Ontology 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**) 2Mb zoom of EigenGWAS (green triangles) and GWAS (black circles) p-values at the *COL4A5* region (highlighted blue in Fig. 2A). 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. Note the extended haplotype at the *COL4A5-C*-allele in (**C**), relative to the shorter haplotypes at the *COL4A5-T* allele in (**B**), consistent with a recent selective sweep around the *COL4A5-C* allele in the UK. (**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-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 UK 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 ($R^2 = 0.004$); note different scales on axes in **A** and **B**.