The great tit HapMap project: a continental-scale analysis of genomic variation in a songbird

Lewis G. Spurgin[1,2]*, Mirte Bosse[3,4,5], Frank Adriaensen[6], Tamer Albayrak[7], Christos Barboutis[8], Eduardo Belda[9], Andrey Bushuev[10], Jacopo G. Cecere[11], Anne Charmantier[12], Mariusz Cichon[13], Niels J. Dingemanse[14], Blandine Doligez[15,16], Tapio Eeva[17], Kjell Einar Erikstad[18], Vyacheslav Fedorov[10]**, Matteo Griggio[19]**, Dieter Heylen[6,20,21], Sabine Hille[22], Camilla A. Hinde[23], Elena Ivankina[24], Bart Kempenaers[25], Anvar Kerimov[10], Milos Krist[26], Laura Kvist[27], Veronika N. Laine[3,28], Raivo Mänd[29], Erik Matthysen[6], Ruedi Nager[30], Boris P. Nikolov[31], A. Claudia Norte[32], Markku Orell[27], Jenny Ouyang[25,33], Gergana Petrova-Dinkova[31], Heinz Richner[34], Diego Rubolini[35], Tore Slagsvold[36], Vallo Tilgar[29], János Török[37], Barbara Tschirren[38], Csongor I. Vágási[39,40], Teru Yuta[41], Martien A.M. Groenen[4], Marcel E. Visser[3,4,42], Kees van Oers[3], Ben C. Sheldon[2] and Jon Slate[43]*

1. School of Biological Sciences, University of East Anglia, Norwich Research Park, United Kingdom
2. Edward Grey Institute, University of Oxford, United Kingdom
3. Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, the Netherlands
4. Animal Breeding and Genomics Centre, Wageningen University, the Netherlands
5. Department of Ecological Science, Animal Ecology Group, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
6. Evolutionary Ecology Group, Department of Biology, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium
7. Department of Biology, Mehmet Akif Ersoy University, Science and Art Faculty, Ortulu, Burdur, Turkey

8. Hellenic Ornithological Society / BirdLife Greece, Themistokleous 80, GR-10681 Athens

9. Institut d'Investigació per a la Gestió Integrada de Zones Costaneres, Campus de Gandia, Universitat Politècnica de València, Carrer Paranimf 1, E-46730 Grau de Gandia (València), Spain

10. Faculty of Biology, Lomonosov Moscow State University, Moscow 119234, Russia

11. Area Avifauna Migratrice, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), via Ca' Fornacetta 9, I-40064, Ozzano Emilia, (BO), Italy

12. CEFE-CNRS, UMR 5175, 1919, route de Mende, F34293 Montpellier Cedex 5, France

13. Institute of Environmental Sciences, Jagiellonian University, Gronostajowa 7, 30-387 Kraków, Poland

14. Behavioural Ecology, Department of Biology, Ludwig Maximilians University of Munich, Planegg-Martinsried, Germany

15. UMR CNRS 5558—LBBE, Biométrie et Biologie Évolutive, UCB Lyon 1 - Bât. Grégor Mendel, 43 bd du 11 novembre 1918, 69622 VILLEURBANNE cedex, France

16. Department of Ecology and Evolution, Animal Ecology, Evolutionary Biology Centre, Uppsala University, Sweden

17. Department of Biology, University of Turku, Turku 20014, Finland

18. Norwegian Institute for Nature Research, FRAM-High North Research Centre for Climate and the Environment, 9296 Tromsø, Norway

19. Department of Biology, University of Padova, Via U. Bassi 58/B, I-35131, Padova, Italy

20. Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, United States of America

21. Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium

22. Institute of Wildlife Biology and Game Management, University of Natural Resources and Life Science, A-1180 Vienna, Austria

23. Behavioural Ecology Group, Department of Life Sciences, Anglia Ruskin University, East Road Cambridge, Cambridgeshire, CB1 1PT, UK.

24. Zvenigorod Biological Station of Lomonosov Moscow State University, P.O. Box Shikhovo, Odintsovo District, Moscow 143036, Russia

25. Max Planck Institute for Ornithology, Department of Behavioural Ecology & Evolutionary Genetics, Eberhard-Gwinner-Straße, House 5, 82319 Seewiesen (Starnberg), Germany

26. Department of Zoology and Laboratory of Ornithology, Faculty of Science, Palacký University, Olomouc 77147, Czech Republic

27. Department of Ecology and Genetics, P.O.Box 3000, 90014-University of Oulu, Finland

28. Finnish Museum of Natural History, University of Helsinki, FI-00014, Finland

29. Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, Tartu 51014, Estonia

30. Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow G12 8QQ, United Kingdom

31. Bulgarian Ornithological Centre, Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, 1 Tsar Osvoboditel Blvd, 1000 Sofia, Bulgaria

32. MARE - Marine and Environmental Sciences Centre, Department of Life Sciences, Faculty of Sciences and Technology, University of Coimbra, Portugal

33. University of Nevada, Reno, NV, United States of America

34. Evolutionary Ecology Lab, Institute of Ecology and Evolution, University of Bern, Bern 3012, Switzerland

35. Dipartimento di Scienze e Politiche Ambientali, Università degli Studi di Milano, via Celoria 26, I-20133, Milano, Italy

36. Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, P.O. Box 1066, Blindern, 0316 Oslo, Norway

37. Behavioural Ecology Group, Department of Systematic Zoology and Ecology, Eötvös Loránd University, Budapest H-1117, Hungary

38. University of Exeter, Centre for Ecology and Conservation, Treliever Road, Penryn, TR10 9FE, United Kingdom

39. Evolutionary Ecology Group, Hungarian Department of Biology and Ecology, Babeş-Bolyai University, Cluj-Napoca, Romania

40. Behavioural Ecology Research Group, Department of Evolutionary Zoology, University of Debrecen, Debrecen, Hungary

41. Graduate School of Environmental Science, Hokkaido University, N10 W5 Sapporo, Hokkaido 060-0810, Japan

42. Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, the Netherlands

43. Department of Animal and Plant Sciences, University of Sheffield, United Kingdom

* **Correspondence**: l.spurgin@uea.ac.uk; j.slate@sheffield.ac.uk

** Deceased

Abstract

A major aim of evolutionary biology is to understand why patterns of genomic diversity vary among populations and species. Large-scale genomic studies of widespread species are useful for studying how the environment and demographic history shape patterns of genomic divergence, and with the continually decreasing cost of sequencing and genotyping, such studies are now becoming feasible. Here, we carry out one of the most geographically comprehensive surveys of genomic variation in a wild vertebrate to date; the great tit (*Parus major*) HapMap project. We screened *ca* 500,000 SNP markers across 647 individuals from 29 populations, spanning almost the entire geographic range of the European great tit subspecies. We found that genome-wide variation was consistent with a recent colonisation across Europe from a single refugium in South-East Europe, with bottlenecks and reduced genetic diversity in island populations. Differentiation across the genome was highly heterogeneous, with clear "islands of differentiation" even among populations with very low levels of genome-wide differentiation. Low local recombination rate in the genome was a strong predictor of high local genomic differentiation (F_{ST}), especially in island and peripheral mainland populations, suggesting that the interplay between genetic drift and recombination is a key driver of highly heterogeneous differentiation landscapes. We also detected genomic outlier regions that were confined to one or more peripheral great tit populations, most likely as a result of recent directional selection at the range edges of this species. Haplotype-based measures of selection were also related to recombination rate, albeit less strongly, and highlighted population-specific sweeps that likely resulted from positive selection. These regions under positive selection contained candidate genes associated with morphology, thermal adaptation and

colouration, providing promising avenues for future investigation. Our study highlights how comprehensive screens of genomic variation in wild organisms can provide unique insights into evolution.

Introduction

Since the first studies of allozyme variation in humans [1] and *Drosophila* [2,3], there has been great interest in explaining how evolutionary and ecological processes shape the patterns of genetic variation observed within and among natural populations. One focus of research and debate in this area has been on quantifying the roles of adaptive and neutral processes in explaining observed levels of genetic variation [4]. However, adaptation does not occur in isolation, but acts on genetic variation that is also shaped by mutation, recombination, gene flow, and genetic drift [5]. More recently there has been increased effort in understanding how these fundamental evolutionary forces operate in concert to generate and maintain the levels of genetic diversity commonly observed in natural populations [6,7].

The increasing feasibility of high-throughput sequencing and genotyping, alongside subsequent characterisation of genome-wide variation across large numbers of individuals, has revealed that at the genomic level, patterns of variation and divergence among natural populations and species are highly heterogeneous [8]. A key feature of these "genomic landscapes" of divergence that has received particular attention is the presence of so-called "islands of differentiation": outlier regions of the genome with high levels of divergence estimated from statistics such as F_{ST} or d_{xy} that consistently emerge at the same loci in different comparisons between populations or related species [8–11]. Initially these

139 regions were termed “islands of speciation”, and were thought to arise as a result of
140 reduced gene flow in genomic regions associated with reproductive isolation [8,12].
141 Subsequent research has revealed that highly heterogeneous patterns of genomic
142 divergence can occur even in the complete absence of gene flow, as a result of
143 recombination rate variation and linked selection [13,14]. In genomic regions of low
144 recombination, selection for beneficial mutations (positive selection), or against
145 deleterious mutations (background selection), will impact relatively large genomic regions
146 as a result of high levels of linkage disequilibrium (LD) among sites. Selection within these
147 regions reduces diversity within populations, and increases levels of differentiation among
148 them, resulting in “islands” of increased differentiation that persist over evolutionary time
149 [14,15]. Another, less well explored reason for which islands of divergence can arise is due
150 to the differential effects of genetic drift in response to variation in effective population size
151 across different genomic regions; something that may be particularly important in recently
152 colonised populations [16,17]. These circumstances promote fixation of haplotypes and
153 therefore result in either reduced or inflated local differentiation.

154 Comparing patterns of genomic differentiation among sets of populations or species at
155 different stages of the divergence/speciation continuum is a powerful way of disentangling
156 the forces that shape variation among populations. Martin et al. [18] showed that, across
157 multiple *Heliconius* butterfly populations and species, patterns of genomic variation were
158 shaped by a combination of gene flow and selection, particularly in genomic regions
159 harbouring genes involved in wing patterning. In contrast, Renaut et al. [10] showed that in
160 *Helianthus* sunflowers, genomic architecture was the main driver of genomic
161 differentiation across sets of populations. Similarly, recent research in birds has revealed

162 that differentiation landscapes are conserved across populations, species and even across
163 avian families, with the same islands of differentiation arising among populations of
164 distantly related species [19–21]. This latter pattern appears to have arisen, at least in part,
165 as a result of a highly conserved synteny and recombination landscape in birds [22], with
166 background selection in regions of low recombination producing recurrent islands of
167 differentiation [23].

168 It is now clear that the recombination landscape and linked selection are key drivers of
169 genomic variation within and among populations. However, we are only just beginning to
170 understand how this linked selection interacts with other evolutionary forces to shape
171 patterns of differentiation across natural populations and species [23–28]. A recent, large-
172 scale analysis of threespine sticklebacks (*Gasterosteus aculeatus*) showed that islands of
173 differentiation were more likely to arise in low recombination regions when gene flow
174 occurred between populations [29]. There is also a significant impact of divergence time; in
175 recently separated populations the differentiation landscape is most likely to reflect
176 selective sweeps. Then, as divergence accumulates, genomic architecture is expected to
177 play an increasingly important role in generating these genomic islands [23].

178 Widespread continental species are excellent models for studying how demography and
179 the environment shape genetic and phenotypic variation among populations, due to their
180 large effective population sizes and ecologically varied ranges. Insight into the evolutionary
181 history of such species can be gained if genetic variation is characterized across much of its
182 geographical range. Cross-population comparisons of genetic variation can then be utilised
183 to make inferences about phylogeography, levels of gene flow between populations and

184 how adaptation to different environmental and ecological conditions occurs [30]. The first
185 large-scale studies were performed in humans - i.e. the HapMap Projects [31–33] which
186 characterized human genetic variation on different continents, with a view to determining
187 the feasibility of association mapping studies. Similar studies have been conducted in
188 domesticated species and their wild ancestors [34–36], and in model organisms [37,38].
189 More recently, there is a growing appreciation that HapMap-type studies are useful for
190 studying signatures of selection and adaptation in natural populations of species with large
191 effective population sizes and high levels of gene flow [39–42].

192 The European great tit (*Parus major major*) is an excellent model for ecological and
193 evolutionary studies [43]. As is the case with several avian species which are amenable to
194 long-term study [44], a wealth of ecological data exists across multiple great tit populations
195 [45–47], enabling informed hypotheses about selection to be tested in this system.
196 Phyleogeographic research using mitochondrial DNA suggests that this species has recently
197 expanded across its European range, possibly from a single refugium in South-East Europe
198 [48]. Contemporary populations are characterised by large effective population sizes and
199 low levels of genetic differentiation [49,50]. However, these previous cross-population
200 molecular studies have relied on a modest number of microsatellite loci and mitochondrial
201 DNA, making the detection of genomic regions under selection impossible. The genome of
202 the great tit has recently been sequenced [51], and a high density panel of *ca* 500,000 SNP
203 markers has been developed [52]. A recent study of two European populations using this
204 marker panel suggests that rapid adaptation has occurred at the genomic and phenotypic
205 levels, with pronounced selection on morphology [53].

206 Here, we perform a HapMap study of 647 unrelated individuals across 29 populations (Fig.
207 1), to examine how genomic architecture, natural selection and population history have
208 shaped patterns of genomic variation across recently colonised European great tit
209 populations. Using a large SNP panel typed across all individuals, we first characterise
210 genome-wide patterns of variation within and among populations, in order to infer
211 population history. We then examine how variation is partitioned across the genome, and
212 test the hypothesis that highly divergent genomic regions have arisen in genomic regions of
213 low recombination [13,14]. Finally, we examine how genomic divergence accumulates
214 along the colonisation route of this species, with the aim of inferring how recent natural
215 selection and demography drive variation across the genome in the wild.

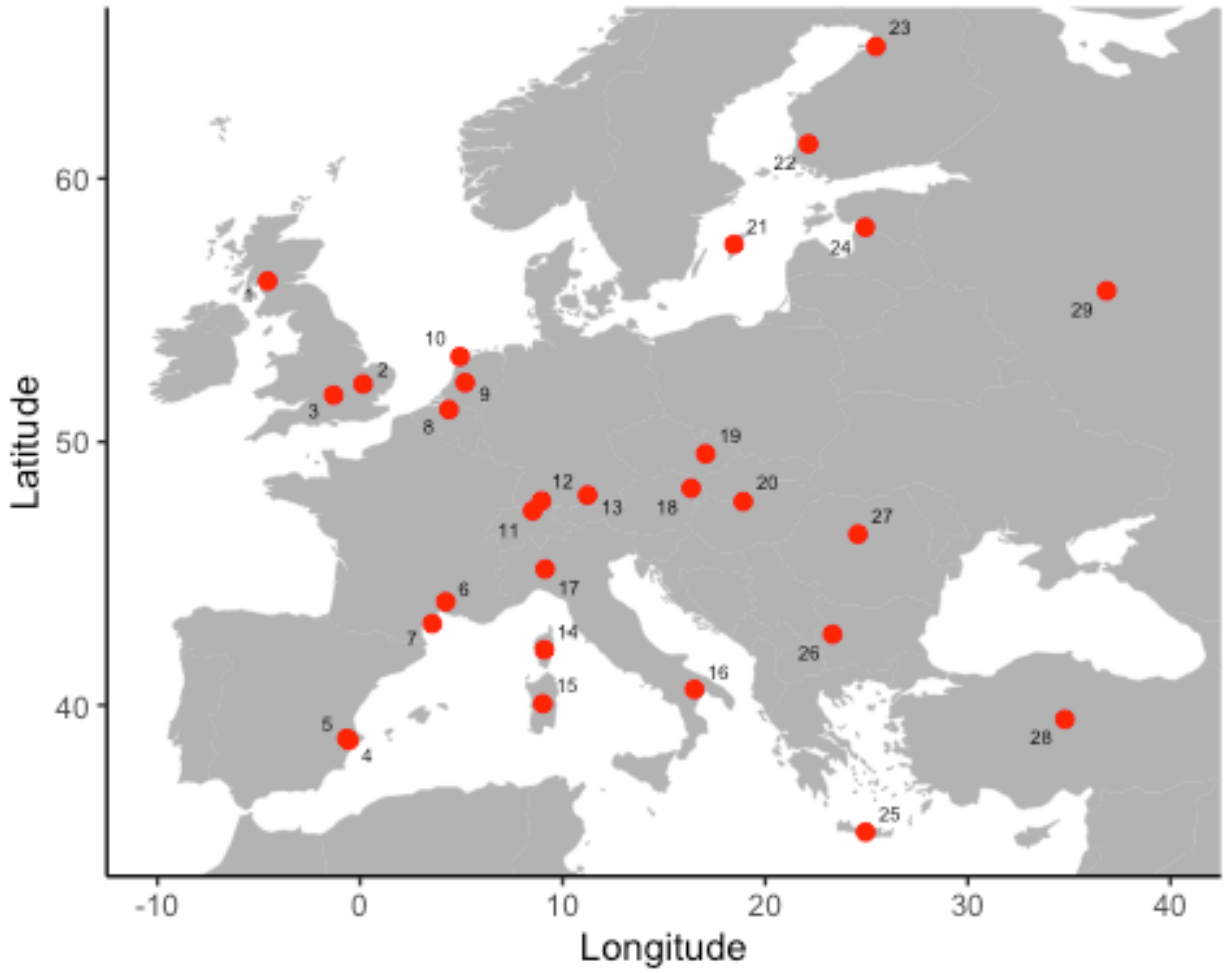


Figure 1 Sampling locations of great tit populations. Population names and sample sizes are given in Table S1, and numbers on the map correspond to the "code" column in Table S1.

Results and Discussion

Genetic diversity and population history

222 Sampling locations and sample sizes for each population are given in Table S1. Levels of
223 genetic diversity (π_{SNP}) were generally high, but we observed substantial differences
224 among populations (Fig. 2A). Similarly, LD declined rapidly with genomic distance in all
225 populations, reaching baseline levels within ~5kb in all populations, but also varied among
226 populations (Fig. S1). Highest levels of LD (and lowest levels of genetic diversity) were
227 observed in the Mediterranean island populations of Crete (Greece) and Sardinia (Italy),
228 with lowest levels of LD in central and western Europe (Fig. S1). This is consistent with
229 reduced effective population size in these island populations, either as result of the
230 colonisation process or more recent bottlenecks, along with low levels of subsequent gene
231 flow from the continent to the islands [54,55].

232 Mean genome-wide F_{ST} between all pairs of European great tit populations was 0.015, with
233 no significant pattern of isolation-by-distance (Mantel test; r = 0.13, p = 0.18; Fig. S2).
234 Instead, the highest levels of F_{ST} were found in comparisons involving the Mediterranean
235 island populations of Corsica (France), Sardinia and Crete (Fig. S2). Admixture analysis was
236 consistent with this pattern (Fig. S3); the K = 2 analysis assigned individuals in Sardinia
237 and Corsica to one genetic cluster, and the remaining populations to the second. Thus, it is
238 likely that much of the genetic structure between European great tit populations is a result
239 of genetic drift in these small island populations. Admixture analysis also revealed some
240 structure between (mainly peripheral) mainland and larger island populations. At K = 3
241 (the model that best fitted the great tit data; Fig. S4), Spain was separated from the rest of
242 mainland Europe. Increasing values of K resulted in the separation of populations in
243 Scotland (K = 4), Sardinia (from Corsica; K = 5), southern France (K = 6), Crete (K = 7) and
244 England (K = 8). The Admixture output at K = 8 is displayed in Fig. 2C as this gives the most

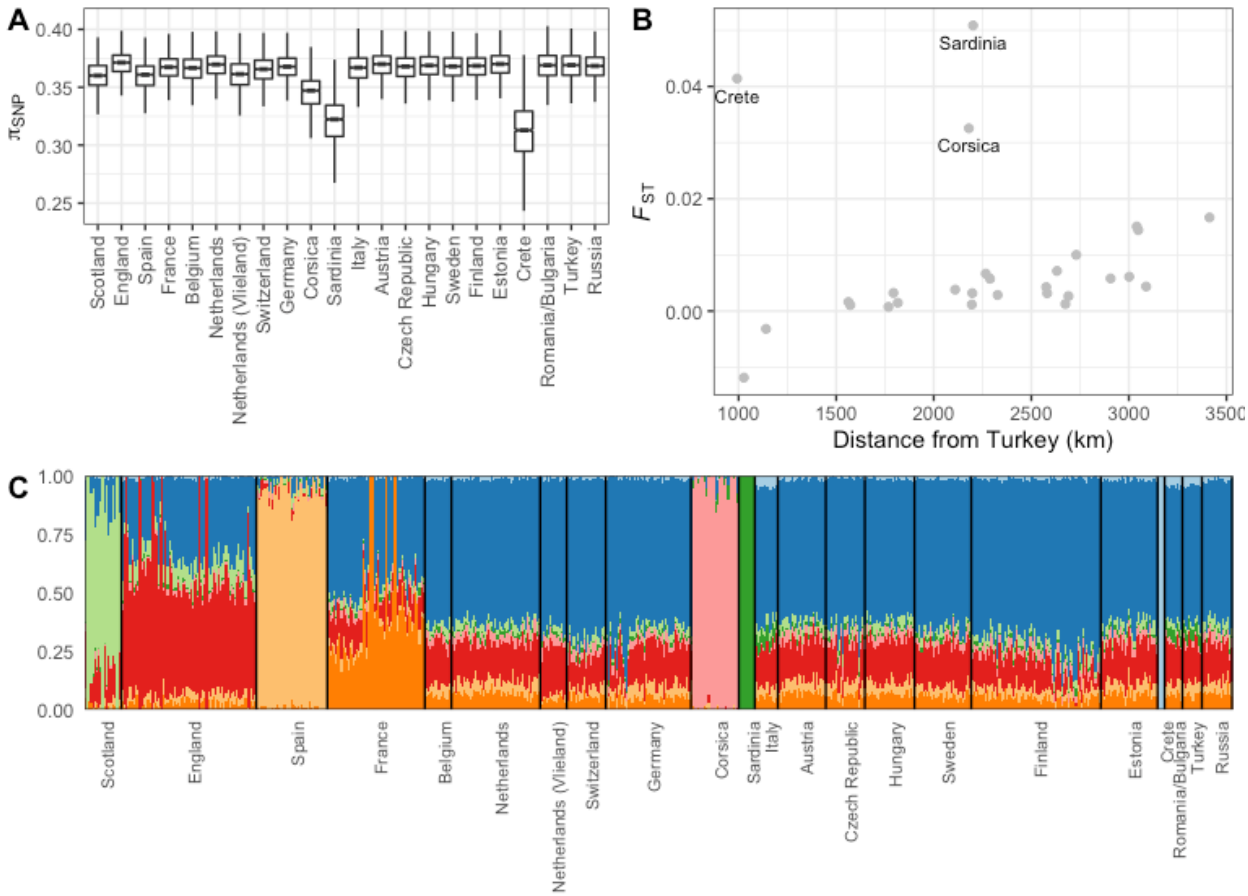
245 detailed picture of genetic structure among European great tit populations. Further
246 increases in K did not generate patterns of structure that corresponded to geographical
247 variation (Fig. S3), and were increasingly less well supported (Fig. S4). Thus, even with
248 hundreds of thousands of markers Admixture was unable to separate many of the
249 European populations, confirming that levels of divergence are extremely low [51]. PCA
250 largely corroborated the Admixture results, with PC1 separating Corsica and Sardinia from
251 the remaining populations, PC2 separating Spain, while PC3 and PC4 separated Scotland,
252 England, Corsica, Sardinia and Crete (Fig. S5).

253 Maximum likelihood analyses implemented in TreeMix showed that a model with no
254 migration explained 97.8% of variance in relatedness between populations [56]; increasing
255 the number of migration events substantially improved the percentage of relatedness
256 explained, up to 99.7% when 10 migration events were fitted (Fig. S6). In Figure 3 we
257 display the maximum likelihood tree with three migration events, after which the variance
258 in relatedness explained plateaued when more migration events were added (Fig. S6). The
259 tree was generally characterised by short branch lengths, with the exception of the island
260 populations of Sardinia and Crete, which were grouped with the population from mainland
261 Italy (Fig. 3). Thus, the TreeMix analysis is consistent with large populations and low
262 overall genomic divergence, with the exception of the Mediterranean island populations.
263 However, much (though not all) of the grouping that did occur among continental
264 populations made geographical sense, with populations from Finland and Estonia grouped
265 together, as were some populations from South-East Europe, and populations from
266 England and Scotland (Fig. 3). Interestingly, TreeMix grouped the Spanish and Corsican
267 populations, which is consistent with previous subspecies descriptions of European great

268 tits [57]. The three fitted migration edges all involved Sardinia, with migration from
269 eastern Europe to Sardinia, and from Sardinia to Corsica and Spain (Fig. 3).

270 We next tested the hypothesis that great tits colonised Europe from a single refugium in
271 South-East Europe. This scenario has been suggested before [48], but due to the low
272 number of genetic markers available there has been limited power with which to test this
273 hypothesis. Using our genome-wide panel of SNP markers, we compared genetic and
274 geographic distance between each population and the proposed refugial populations.
275 Because of the elevated structure in Corsica, Sardinia and Crete (Fig. S2), we excluded
276 comparisons involving these populations. We found that F_{ST} was significantly related to
277 distance from Turkey ($r = 0.81$, $p < 0.001$; Fig. 2B) and Romania/Bulgaria ($r = 0.44$, $p =$
278 0.001). The same relationship was not found for alternative potential refugial populations
279 [58] in Spain ($r = -0.09$, $p = 0.55$), or southern Italy ($r = 0.04$, $p = 0.77$). Our results
280 therefore lend empirical support to the hypothesis [48] that great tits colonised Europe
281 primarily from a single refugium in the south-east. Clearly, although our sampling was
282 extensive, it is not exhaustive, and more fine-scaled sampling in eastern Europe would be
283 required to determine the extact location and extent of refugial great tit populations.
284 Sampling in North Africa would also be useful to determine whether further refugia exist,
285 and to quantify the extent of admixture between European and African great tit
286 populations.

287



288

289 **Figure 2** Genetic diversity and structure in European Great tit populations. **A** Nucleotide

290 diversity within each population. **B** Pairwise F_{ST} in relation to geographic distance from the

291 Turkey, only including comparisons involving Turkey. **C** Output from Admixture analysis at

292 K = 8. Population details can be found in Table S1.

293

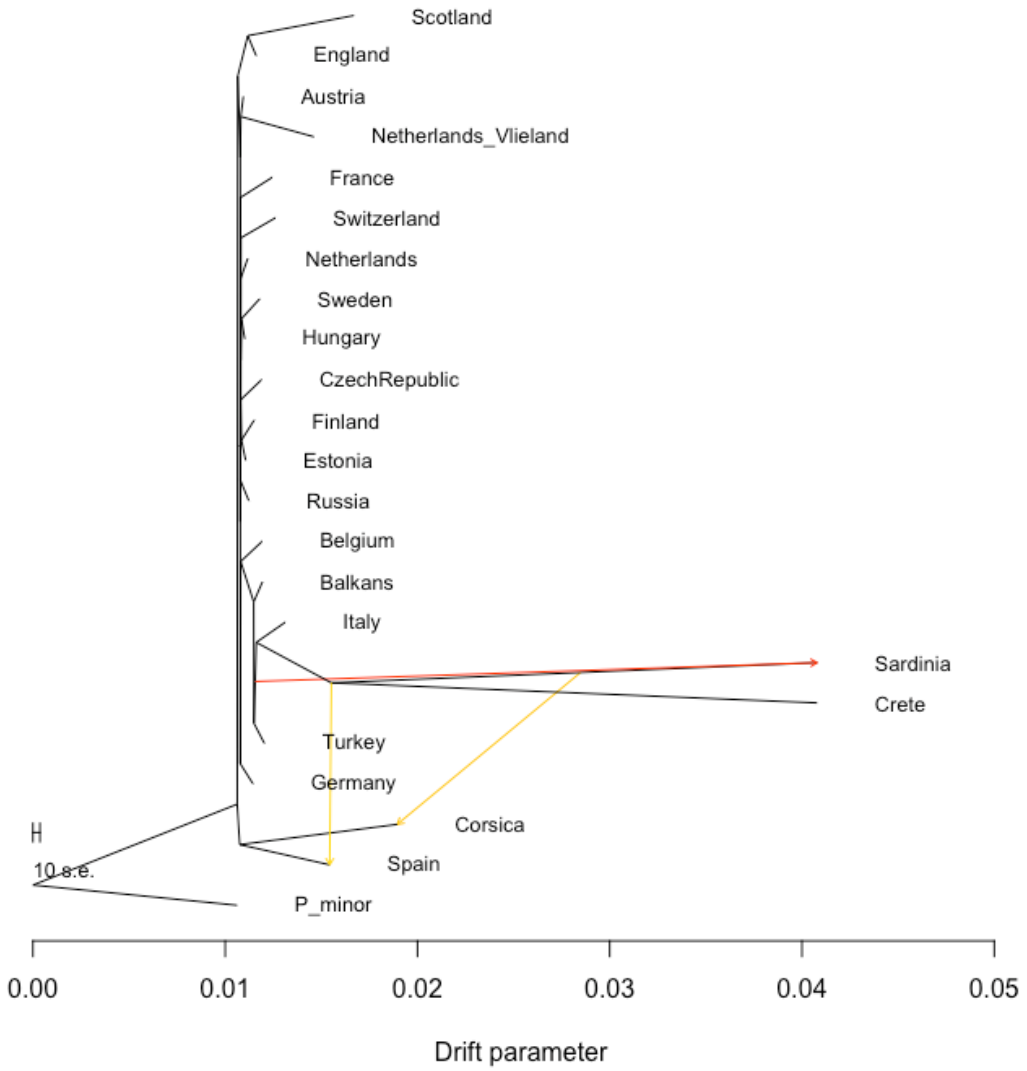


Figure 3 Maximum likelihood tree inferred by TreeMix, allowing three migration events. The two migration events (arrows) are coloured according to their weight (red = higher migration), and horizontal branch lengths are proportional to the amount of genetic drift that has occurred along the branch. A population of the great tit's sister species, *Parus minor*, was used as an outgroup. Population details are given in Table S1.

300

301 *Genomic landscapes of differentiation*

302 It is likely that many, and perhaps the majority, of wild populations are characterised by
303 highly heterogeneous patterns of differentiation across the genome [27]. This is the case in
304 European great tits - despite the extremely low average F_{ST}, we found genomic regions
305 with very high levels of genomic structure (maximum F_{ST} for 10kb and 500kb windows
306 was 0.98 and 0.07, respectively). To examine how landscapes of genomic divergence have
307 formed along the colonisation route of European great tits, we calculated windowed F_{ST} in
308 500kb bins between each population and the proposed refugial population in Turkey. We
309 found that F_{ST} varied markedly across the genome in all comparisons (Fig. 4; Fig. S7).
310 Outlier regions (windows with standardized F_{ST},hereafter zF_{ST}, > 10) were found in all
311 comparisons apart from Crete and Sardinia, in which overall levels of divergence were
312 highest, with some outlier regions found across multiple comparisons (Fig. 4). Our results
313 suggest, therefore, that genomic islands of differentiation can and do arise even among
314 recently separated populations.

315 Considering all populations, F_{ST} calculated in 500kb windows was strongly negatively
316 correlated with local variation in recombination rate (Spearman correlation, r = -0.50, p <
317 0.001). Although recombination rate and gene density were positively correlated, (r = 0.16,
318 p < 0.001), F_{ST} was only moderately correlated with gene density in 500kb windows (r = -
319 0.09, p < 0.001), and this correlation became weak when calculated in 10kb windows (r = -
320 0.01, p < 0.001). Thus it appears that the negative relationship between F_{ST} and
321 recombination rate is not driven by gene density. Examining how the relationship between
322 genomic differentiation and recombination varied among populations revealed that F_{ST}

323 was negatively related to recombination rate in almost all comparisons with Turkey (Fig.
324 5). The relationship between F_{ST} and recombination rate was generally weak, but in a
325 handful of populations this relationship was substantially stronger - most notably in the
326 island populations of Corsica, Sardinia and Crete, and in the peripheral mainland
327 populations of England, Scotland, Spain, Finland and France, which are the populations that
328 are likely most susceptible to drift (Fig. 5).

329 As a further examination of how allele frequency variation and haplotype structure may be
330 shaped by neutral and adaptive processes, we compared F_{ST} and recombination rate with
331 Rsb - a measure which aims to detect regions of the genome under positive selection by
332 comparing extended haplotype homozygosity profiles between populations, and is
333 expected to be less sensitive to variation in recombination rate because extended
334 haplotypes should be present in both populations when recombination frequency is lower
335 [59]. Specifically, we compared the distributions of these statistics in 500kb windows using
336 three populations in the Netherlands, Finland and Spain, with the aim of exploring how Rsb
337 varies with recombination rate and with F_{ST}. We found that, across the genome,
338 correlations between absolute Rsb and recombination rate were only slightly weaker
339 (Spearman rank r between -0.25 and -0.20) compared to those between F_{ST} and
340 recombination for the same pairwise comparisons (r between -0.36 and -0.20), and that the
341 regions of high F_{ST} in these comparisons tended to be in these same regions of low
342 recombination and high Rsb (Fig. S8). Notably, though, there were several cases where
343 windows with extreme Rsb values occurred in regions of high recombination, and most
344 high F_{ST} regions in high recombination regions were also Rsb outliers (Fig. S8). Below we

345 explore the possibility that such regions are strong candidates for recent adaptive
346 evolution.

347 Background selection is a key driver of genomic differentiation in birds [20,21], and other
348 organisms [60], but is not expected to play a major role in driving islands of differentiation
349 over the timescales (tens of thousands of years) relevant to this study [23,29]. On the other
350 hand, differential effects of drift operating across the genome could contribute
351 substantially to the highly heterogeneous patterns of differentiation observed in European
352 great tits. Our finding that the relationship between recombination rate and F_{ST} is strongest
353 in populations experiencing the greatest levels of drift (e.g. islands; Fig. 5) is consistent
354 with this - low recombining regions have reduced effective population size and increased
355 rates of lineage sorting due to drift. Drift has been suggested to be a driver of
356 heterogeneous genomic landscapes in other systems [16,17], and our study suggests that it
357 may play a key role in shaping genomic structure in widespread, continental species.

358 Outlier regions of very high differentiation (zF_{ST} > 10) mostly occurred in areas of the
359 genome with low recombination rates (Fig. 5) and the recombination rate of outlier regions
360 was (just) significantly lower than the genome-wide average (Wilcoxon test, P = 0.048). Of
361 the 11 outlier regions, nine were found in only one or two comparisons, while the other
362 two were found in 12 and 10 comparisons, respectively (Table S2). We hereafter refer to
363 outlier regions found in one or two comparisons as "population-specific" outlier regions,
364 and to those found in more than two comparisons as "shared" regions.

365 Regions of high differentiation that are not shared among populations could potentially
366 arise as a result of recent drift, but are also candidate regions for recent positive selection

367 [23]. Several lines of evidence suggest that this may be the case in European great tits.
368 Firstly, population-specific regions tended to occur in regions of higher recombination than
369 shared outliers, although low sample size prevents a formal statistical analysis of this.
370 Second, population-specific outliers tended to be found in the most peripheral European
371 great tit populations, with three found in Scotland, and two in England, Spain and Finland;
372 the remaining outlier regions were found in comparisons involving the Czech republic,
373 Russia, Vlieland (Netherlands) and Belgium (Table S2). Notably, these populations are not
374 the lowest in genetic diversity and therefore drift effects are not necessarily stronger in
375 these peripheral populations. Third, the Rsb tests, which indicate which population has
376 undergone a selective sweep, also tend to find evidence of positive selection in the more
377 peripheral population, as expected under an adaptive evolution scenario (Fig. S8). This
378 directionality cannot be assessed with F_{ST}, because it is a combined measure of allele
379 frequency differences. Of course, across the genome, both positive selection and linked
380 background selection are likely to be operating. Observational and experimental research
381 shows that adaptation at range edges is a key feature shaping divergence among recently
382 colonised and expanding populations [61–63]. Population-specific regions appeared less
383 likely to be in regions of low recombination than shared regions (Fig. 5), although the small
384 number of regions precluded testing this hypothesis formally. Thus, it is likely that genomic
385 architecture plays a key role in determining how both selection and drift have shaped
386 genomic variation across the recent evolutionary history of European great tits.

387 To further explore how selection may have shaped variation in F_{ST} outlier regions, we
388 estimated levels of nucleotide and haplotype diversity within these regions. Nucleotide
389 diversity (π_{SNP}) in outlier regions varied from 0.21 to 0.48, and diversity in these regions

390 was significantly lower than the genome-wide average (Wilcoxon test, P = 0.018; Fig. S9A).
391 However, there appeared to be little difference in nucleotide diversity between shared and
392 population-specific regions (Fig. S9A). Haplotype diversity varied substantially among
393 regions, with haplotype richness ranging from 72 to 1033. Both haplotype richness and
394 marker density in shared regions tended to be lower than those in population-specific
395 regions (Fig. S9B), consistent with lower recombination frequency. A detailed examination
396 of haplotype structure in one shared and one population-specific region is displayed in
397 Figure S10. The population-specific outlier region (to Finland, situated on chromosome 1A)
398 was characterised by a complex structure, with a single haplotype at high frequency in
399 Finland compared to other populations, indicating a population-specific selective sweep
400 and high background diversity (Fig. S10). In contrast, the shared region on chromosome 2
401 was much less complex, demonstrating high frequency haplotypes that are found across a
402 range of populations. Our data therefore suggest that examining patterns of haplotype
403 diversity in outlier regions may help to separate recent episodes of positive selection from
404 drift and background selection (Figs S9, S10).

405 Potential candidate genes found within shared and population-specific outlier regions are
406 displayed in Table S2. Perhaps most notable among these is *COL4A5*, a gene found to be
407 associated with bill length, and under selection between populations in England and the
408 Netherlands, in a recent great tit study [53]. Here we found that the *COL4A5* region is an F_{ST}
409 outlier in England and Scotland, but not in any other European populations (Table S2). UK
410 great tits have been described as a separate subspecies based on beak shape [64], and our
411 results here, combined with previous results, suggest that this divergence is the result of
412 recent natural selection in the UK [53]. Another notable candidate gene potentially

413 involved in beak morphology, and previously found to be under selection in UK great tits is
414 *BMPR1A*, which plays a key role in palate development [65] and in this study was found in
415 an outlier region in Scotland. Other candidate morphology and obesity genes in the
416 population-specific outlier regions in the UK included *PPP1CB*, which may play a role in
417 adipogenesis [66] and *GHITM*, which appears to have been subject to natural selection in
418 human pygmy populations [67]. Thus, morphological traits may frequently be involved in
419 adaptation in great tits.

420 In addition to morphological candidates in the UK, we found outlier regions specific to cold
421 populations in Scotland, Finland and Russia (Table S2), containing at least one candidate
422 gene for thermal stress (*CDKN1B*) [68]. Other genomic outlier regions contained potential
423 candidate genes associated with malaria infection (*MRPL33*) [69] and colour variation
424 (*SOX10*) [70]. This is thus far an exploratory analysis, and we are therefore reluctant to
425 speculate whether these candidate genes are genuine targets for natural selection, and
426 more reluctant still to speculate as to how selection might be driving variation at these
427 regions. Regardless, these candidates will provide useful starting points for future genomic
428 and ecological investigation.

429 HapMap style projects have been hugely informative in shaping our understanding of how
430 natural selection operates in humans and other model species [31,38]. This study is one of
431 the largest to date of genomic variation in a wild vertebrate, which has helped to reveal the
432 evolutionary history of great tits, and to identify candidate genes and traits that may have
433 been involved in adaptation during and/or after postglacial recolonisation. Further, this
434 work will form the foundation of many future analyses. Clearly, we have only touched on

435 haplotype-based methods to infer adaptation here, and this will be the subject of future
436 work. Environmental association approaches are also highly suited to detecting adaptation
437 in widespread continental species [71,72], and further work will test how variation in the
438 environment has shaped patterns of genomic variation in great tits [73]. This combination
439 of environmental and genomic data in species such as great tits, in which a wealth of
440 ecological and genomic resources are available, is likely to generate interesting insights
441 into the the genetic and phenotypic basis of natural selection.

442

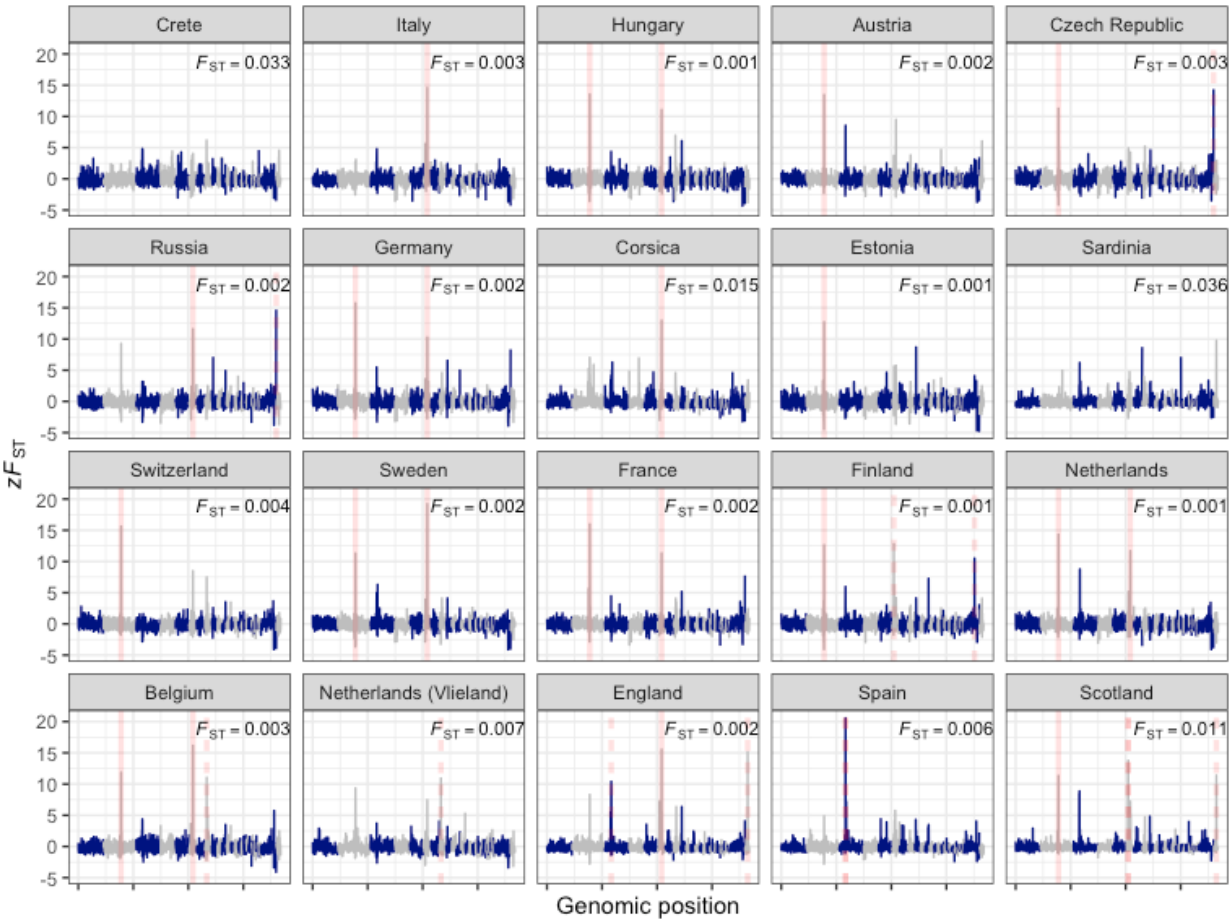


Figure 4 Landscapes of relative genomic differentiation in European great tit populations. zF_{ST} across the genome is averaged in 500kb windows, with each panel displaying a pairwise comparison with the proposed refugial population in Turkey. Red lines represent F_{ST} outliers (windows with mean F_{ST} values at least 10 standard deviations greater than the global mean for that comparison) shared across more than two comparisons (solid red lines), or specific to one or two comparisons (dashed red lines). Mean, untransformed F_{ST} values are given in the top-right of each panel, and are fully displayed in Fig. S7.

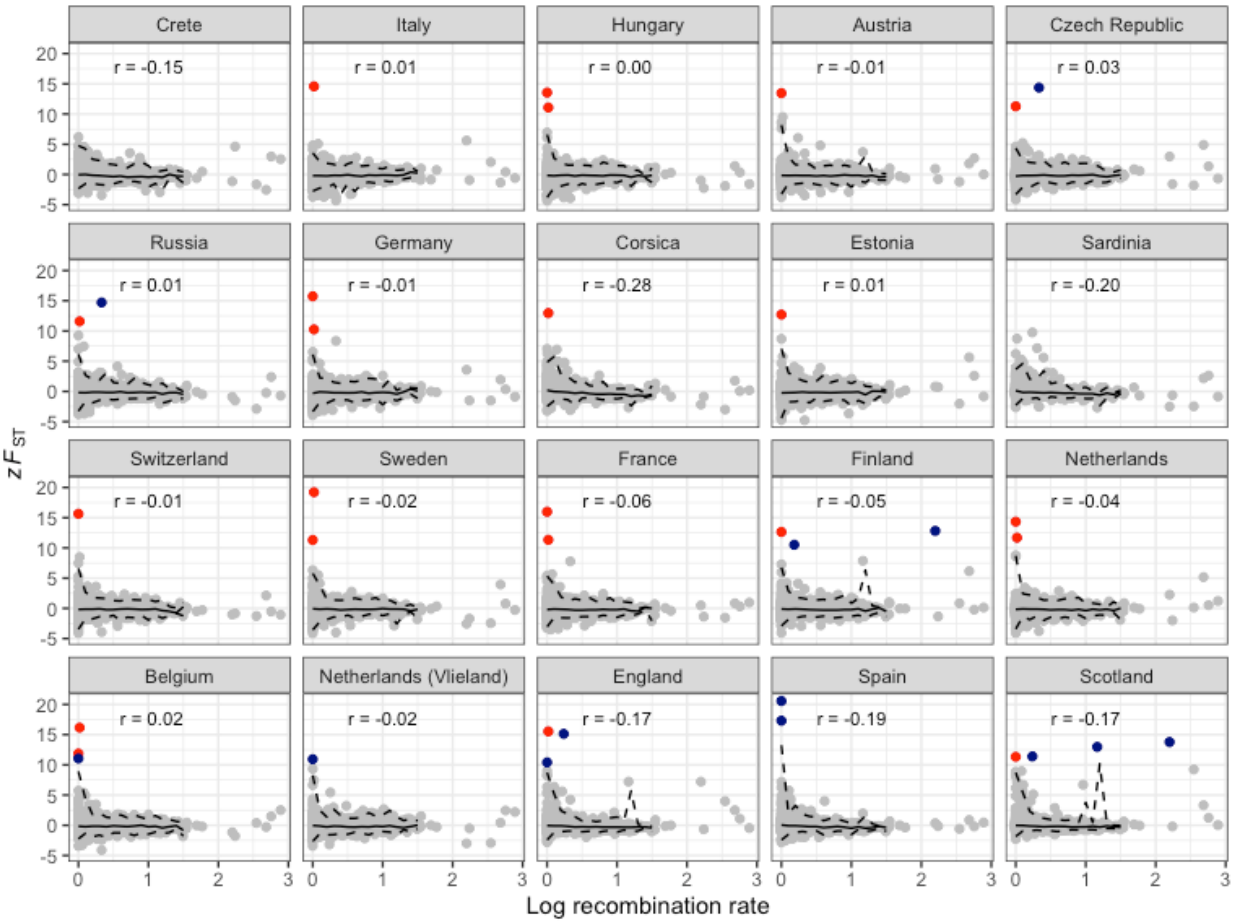


Figure 5 Genomic differentiation and recombination rate variation in European great tit populations. Each point is the mean of a 500kb window, with each panel displaying a pairwise comparison with the proposed refugial population in Turkey. Coloured points represent F_{ST} outliers (mean standardized F_{ST} values of zF_{ST} > 10) shared across more than two comparisons (red), or specific to one or two comparisons (dark blue). Solid and dotted lines represent median and 99% quantiles of F_{ST} windows from bins of 0.1 log cM/Mbp.

Materials and Methods

Sampling and molecular methods

Samples were collected from 29 populations from 22 regions across Europe (Fig. 1; Table S1). Samples were pooled into regions either based on geographical proximity (e.g. Cambridge and Wytham woods), or based on sample size (e.g. Romania and Bulgaria). An exploratory analysis considering all sampled populations separately yielded virtually identical results to those shown here, and in no cases did we observe substructure within pooled populations in our Admixture analyses (Fig. S3).

Birds were trapped from nest boxes, or using mist nets, and ringed with a uniquely numbered aluminium ring. Blood was taken via brachial or tarsal venipuncture, and stored in either 1 ml Cell Lysis Solution (Gentra Puregene Kit, Qiagen, USA), Queen's buffer, or absolute ethanol. All samples were genotyped using a custom-made Affymetrix® great tit 650K SNP chip at Edinburgh Genomics (Edinburgh, United Kingdom), following the approaches outlined in [52]

Analyses

Unless stated otherwise, all population genetic statistics were calculated in PLINK version 1.9 [74], and downstream analysis and plotting was carried out in R version 3.3 [75]. For population genetic analyses, we used the filtering approaches outlined in [53]. Briefly, we randomly removed individuals from pairs with relatedness values > 0.4, and for demographic analyses we used a LD pruned dataset (based on VIF > 0.2), with SNPs

480 associated with an inversion on chromosome 1A [**???**] removed. After filtering, a total of
481 647 samples typed at 483888 SNPs were retained for analysis.

482 In each population, we estimated LD (R^2) for each pair of markers within 50kb on the same
483 chromosome, and compared this to physical distance between marker pairs. We calculated
484 observed heterozygosity for each SNP and population using a reduced SNP dataset, which
485 was pruned based on LD to remove all markers with $R^2 > 0.1$, then thinned with a
486 probability of retaining each variant of 0.25. We calculated genome-wide (mean) F_{ST}
487 between each pair of populations using the pruned and thinned dataset described above.
488 Pairwise F_{ST} was compared to geographic distance between populations using Mantel tests,
489 implemented in the Ecodist package in R [76]. We tested whether genetic structure was
490 related to distance from candidate refugial populations (in Romania/Bulgaria, Turkey,
491 Spain and Italy), using Pearson correlations. We also estimated population structure using
492 Admixture version 1.3, with default settings [77]. We varied values of K from one to ten; by
493 which point increasing values of K provided no informative information about population
494 structure (see results). Model support for each value of K was estimated by calculating 5-
495 fold cross-validation error. Finally, we visualised the evolutionary history among European
496 great tit populations by generating a maximum likelihood tree in TreeMix version 1.13
497 [56]. We rooted the tree using a sample of *P. minor* individuals sampled from Amur, Russia
498 [52]. We fitted models allowing for range of migration events (0-10), and used a window
499 size of 500 SNPs [56]. To assess model fit, we calculated the proportion of variance in
500 relatedness between populations explained by each model [56].

501 Recombination rates at each locus were estimated by comparing the location of SNPs on
502 the genome assembly (v1.1) with their location on the great tit linkage map [78]. Previous
503 linkage mapping, using a lower density SNP chip, was independently carried out in UK and
504 Netherlands great tit populations and the two maps were almost identical [78]. For the
505 purposes of this analysis, we used SNPs and marker intervals from the UK comprehensive
506 map. A total of 2706 SNPs were located on both the genome assembly and the linkage map.
507 Thus a positon in Mb and cM of each of these SNPs is known. All other SNPs on the HD chip
508 have a physical (Mb) position but no known linkage map position. The great tit genome
509 v1.1 is 1.02 Gb long, so the average physical interval between mapped SNPs is ~376Kb. The
510 linkage map position of each unmapped SNP was estimated by interpolation; by taking the
511 closest mapped SNP in either direction, and assuming a constant recombination rate in the
512 interval between those SNPs. For example, an unmapped SNP with physical position 1.4Mb,
513 flanked by mapped SNPs at 1.0Mb/0.5cM and 2.0Mb/1.0cM, would be estimated to be
514 located at 0.5 + (1.4-1.0)/(2.0-1.0) * (1.0-0.5) = 0.7cM. Having interpolated cM position of
515 every SNP, the local recombination rate was calculated as the cM interval spanned by the
516 nearest neighbouring SNPs, divided by the physical distance (bp) spanned by those same
517 neighbouring SNPs. In other words, for the ith SNP, the recombination rate is estimated as
518 the linkage distance between the i-1th and i+1th SNP, divided by the physical distance
519 between the i-1th and i+1th SNP. For downstream analyses, local recombination rates were
520 estimated by averaging across all SNPs in each 500kb window. We calculated gene density
521 in 10kb and 500kb windows using custom R scripts and the annotated great tit genome (v
522 1.1) [51].

523 We examined the genomic landscape of differentiation across European great tit
524 populations by calculating F_{ST} in 10kb and 500kb bins, using python scripts obtained from
525 Github (https://github.com/simonhmartin/genomics_general). We did not estimate d_{xy}, as
526 this parameter is difficult to estimate accurately from single SNP loci [14]. We also
527 calculated standardised F_{ST} (zF_{ST}) by mean-centring windowed values and dividing them
528 by the standard deviation among windows. We defined outlier regions as 500kb bins with
529 zF_{ST} values greater than ten. Finally, Rsb [59] was calculated for three pairwise
530 comparisons (in Spain, Finland and the Netherlands), using the R package Rehh [79], and
531 this averaged into 500kb windows. Unlike F_{ST}, the Rsb statistic gives an indication of which
532 population an adaptively important haplotype is under positive selection in, and is thought
533 to be less sensitive to local recombination rates [59]. Rsb estimation requires phased
534 genotype data, so phasing was performed using shapeIT2 v 2.r837 [80]. The –duohmm
535 argument was used to ensure that family information, where available, was used to
536 improve the accuracy of phasing. The –effective-size parameter was set at 500,000
537 reflecting the large effective population size of European great tit populations [51]. Local
538 recombination rates (measured in cM/Mbp) were used in the map files. Comparisons
539 between F_{ST}, Rsb, recombination rate and gene density were carried out using Spearman
540 rank correlation (as all statistics were highly skewed).

541 *Ethics Statement*

542 Great tit DNA samples were provided by researchers studying hole-nesting great tit
543 populations with permissions and licenses to study and sample birds granted by the
544 following organisations: Bundesministerium für Wissenschaft und Forschung (Ref. II/3b

Gentechnik und Tierversuche), Czech Republic Ministry of Education Youth and Sports (MSMT-56147/2012-310), Dutch legal entity Dier Experimenten Comissie (NIOO 14.11, NIOO 07.04), Dutch Central Authority for Scientific Procedures on Animals (AVD-801002017831), Estonian Ministry of Agriculture - Animal Procedures Committee (License 108), Ethical Committee of the Canton of Bern, Ethics Committee for animal experimentation of Languedoc Roussillon CEEA-LR (APAFIS#8608-2017012011062214), Ethics Committee of Lomonosov Moscow State University (Approval numbers 26ch, 37zh, 49zh, 59zh, 88zh), Federal Office for the Environment - Switzerland (License 234), Flemish Ministry - Agency for Nature and Forest, Generalitat Valenciana - Conselleria de Medi Ambient, Hellenic Ministry of Environment and Energy, Istituto Superiore per la Protezione e la Ricerca Ambientale - ISPRA (Law 157/1992 Art. 4 (1) and Art. 7 (5)), Mehmet Akif Ersoy University - Local Ethical Committee on Animal Experiments (93773921-9), Middle Danube Valley Inspectorate for Environmental Protection, Nature Conservation and Water Management (43355-1/2008), Regierung Oberbayern (55.2-1-54-2532-140-11), Regierungspräsidium Freiburg - Departments for Animal Welfare and Nature Conservation, Regional State Administrative Agency of Southern Finland (ESAVI/1650/04.10.03/2012, ESAVI-1804/04.10.03/2011), Romanian Academy of Sciences (Permit number 2257), Swedish Committee for Experiments on Animals (Licence number C 108/7), UK Home Office (License PPL 30/2409), University of Oxford - Department of Zoology Local Ethical Review Committee, University of Glasgow AWERB committee, Veterinary Office of the Canton of Zurich (License 195/2010)

Samples were imported to the UK under DEFRA general import license IMP/GEN/2011/01, according to regulations outlined by the UK Animal Health and Veterinary Laboratories Agency.

Acknowledgements

This work was supported by grants from the Natural Environment Research Council (grant NE/J012599/1 to J.S. and B.C.S) and the European Research Council (grant 202487 to J.S. and grant 339092 – E-Response to MEV) L.G.S was supported by fellowships from the Edward Grey Institute for Ornithology and the BBSRC (BB/N011759/1). We thank Claire Bloor, Geoff Scopes and Alessandro Davassi of Affymetrix for their help during the chip design and genotyping calling processes. Richard Talbot and Alison Downing of Edinburgh Genomics provided the genotyping service.

Data Availability

All data and code to reproduce the results are available on Github:
https://github.com/lgs85/SpurginBosse_Hapmap

References

1. Harris H. Enzyme polymorphisms in man. Proceedings of the Royal Society B: Biological Sciences. 1966;164: 298–310. doi:10.1098/rspb.1966.0032

2. Lewontin RC, Hubby JL. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. Genetics. 1966;54: 595–609. doi:10.1111/j.1601-5223.1973.tb01163.x

3. Hubby JL, Lewontin RC. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in Drosophila pseudoobscura. Genetics. 1966;54: 577–594.

4. Nei M. Selectionism and Neutralism in Molecular Evolution. Molecular Biology and Evolution. 2005;22: 2318–2342. doi:10.1093/molbev/msi242

5. Hartl DL, Clark AG. Principles of population genetics. Sinauer associates Sunderland, MA; 1997.

6. Feder JL, Egan SP, Nosil P. The genomics of speciation-with-gene-flow. Trends in Genetics. 2012;28: 342–350. doi:10.1016/J.TIG.2012.03.009

7. Lanfear R, Kokko H, Eyre-Walker A. Population size and the rate of evolution. Trends in Ecology & Evolution. 2014;29: 33–41. doi:10.1016/J.TREE.2013.09.009

8. Turner TL, Hahn MW, Nuzhdin SV, Mitchison G, Spencer F. Genomic Islands of Speciation in Anopheles gambiae. PLoS Biology. 2005;3: e285. doi:10.1371/journal.pbio.0030285

602 9. Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, et al. Genomic
603 islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted
604 sequencing. Philosophical Transactions of the Royal Society of London B: Biological
605 Sciences. 2011;367. Available: doi: 10.1098/rstb.2011.0198

606 10. Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, et al. Genomic islands of
607 divergence are not affected by geography of speciation in sunflowers. Nature
608 Communications. 2013;4: 1827. doi:10.1038/ncomms2833

609 11. Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, et al. The genomic landscape
610 underlying phenotypic integrity in the face of gene flow in crows. Science. 2014;344: 1410–
611 1414. doi:10.1126/science.1253226

612 12. Wolf JBW, Ellegren H. Making sense of genomic islands of differentiation in light of
613 speciation. Nature Reviews Genetics. 2017;18: 87–100. doi:10.1038/nrg.2016.133

614 13. Noor MAF, Bennett SM. Islands of speciation or mirages in the desert? Examining the
615 role of restricted recombination in maintaining species. Heredity. 2009;103: 439–444.
616 doi:10.1038/hdy.2009.151

617 14. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are
618 due to reduced diversity, not reduced gene flow. Molecular Ecology. 2014;23: 3133–3157.
619 doi:10.1111/mec.12796

620 15. Turner TL, Hahn MW. Genomic islands of speciation or genomic islands and speciation?
621 Molecular Ecology. 2010;19: 848–850. doi:10.1111/j.1365-294X.2010.04532.x

622 16. Ma T, Wang K, Hu Q, Xi Z, Wan D, Wang Q, et al. Ancient polymorphisms and divergence
623 hitchhiking contribute to genomic islands of divergence within a poplar species complex.
624 Proceedings of the National Academy of Sciences of the United States of America. National
625 Academy of Sciences; 2018;115: E236–E243. doi:10.1073/pnas.1713288114

626 17. Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ. Distinguishing noise from signal
627 in patterns of genomic divergence in a highly polymorphic avian radiation. Molecular
628 Ecology. 2015;24: 4238–51. doi:10.1111/mec.13314

629 18. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al.
630 Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome
631 Research. 2013;23: 1817–1828. doi:10.1101/gr.159426.113

632 19. Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. Linked selection and
633 recombination rate variation drive the evolution of the genomic landscape of
634 differentiation across the speciation continuum of Ficedula flycatchers. Genome research.
635 2015;25: 1656–1665. doi:10.1101/gr.196485.115

636 20. Vijay N, Weissensteiner M, Burri R, Kawakami T, Ellegren H, Wolf JBW. Genomewide
637 patterns of variation in genetic diversity are shared among populations, species and
638 higher-order taxa. Molecular Ecology. 2017;26: 4284–4295. doi:10.1111/mec.14195

639 21. Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M. Correlated
640 patterns of genetic diversity and differentiation across an avian family. Molecular Ecology.
641 2017;26: 3982–3997. doi:10.1111/mec.14083

642 22. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights
643 into avian genome evolution and adaptation. Science. 2014;346: 1311–1320.

644 23. Burri R. Dissecting differentiation landscapes: a linked selection's perspective. Journal
645 of Evolutionary Biology. 2017;30: 1501–1505. doi:10.1111/jeb.13108

646 24. Ellegren H, Wolf JBW. Parallelism in genomic landscapes of differentiation, conserved
647 genomic features and the role of linked selection. Journal of Evolutionary Biology. 2017;30:
648 1516–1518. doi:10.1111/jeb.13113

649 25. Jiggins CD, Martin SH. Glittering gold and the quest for Isla de Muerta. Journal of
650 Evolutionary Biology. 2017;30: 1509–1511. doi:10.1111/jeb.13110

651 26. Lohse K. Come on feel the noise - from metaphors to null models. Journal of
652 Evolutionary Biology. 2017;30: 1506–1508. doi:10.1111/jeb.13109

653 27. Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, et al. Interpreting the
654 genomic landscape of speciation: a road map for finding barriers to gene flow. Journal of
655 Evolutionary Biology. 2017;30: 1450–1477. doi:10.1111/jeb.13047

656 28. Perrier C, Charmantier A. On the importance of time scales when studying adaptive
657 evolution. Evolution Letters. John Wiley & Sons, Ltd; 2019;3: 240–247.

658 29. Samuk K, Owens GL, Delmore KE, Miller SE, Rennison DJ, Schluter D. Gene flow and
659 selection interact to promote adaptive divergence in regions of low recombination.
660 Molecular Ecology. 2017;26: 4378–4390. doi:10.1111/mec.14226

661 30. Perrier C, Rougemont Q, Charmantier A. Demographic history and genomics of local
662 adaptation in blue tit populations. Evol Appl. John Wiley & Sons, Ltd; 2020;13: 1145–1165.

663 31. Altshuler D, Donnelly P. A haplotype map of the human genome. Nature. 2005;437:
664 1299–1320. doi:10.1038/nature04226

665 32. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second
666 generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449: 851–861.
667 doi:10.1038/nature06258

668 33. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, et al.
669 Integrating common and rare genetic variation in diverse human populations. Nature.
670 2010;467: 52–8. doi:10.1038/nature09298

671 34. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al. Genome-
672 wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong
673 recent selection. PLoS Biology. 2012;10: e1001258. doi:10.1371/journal.pbio.1001258

674 35. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, et al. Genome-
675 wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science.
676 2009;324: 528–532. doi:10.1126/science.1167936

677 36. Chia J-M, Song C, Bradbury PJ, Costich D, Leon N de, Doebley J, et al. Maize HapMap2
678 identifies extant variation from a genome in flux. Nature Genetics. 2012;44: 803–807.
679 doi:10.1038/ng.2313

680 37. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome
681 sequence, comparative analysis and haplotype structure of the domestic dog. Nature.
682 2005;438: 803–819. doi:10.1038/nature04338

683 38. Kirby A, Kang HM, Wade CM, Cotsapas C, Kostem E, Han B, et al. Fine mapping in 94
684 inbred mouse strains using a high-density haplotype resource. Genetics. 2010;185: 1081–
685 95. doi:10.1534/genetics.110.115014

686 39. Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, et al. Genome-wide
687 patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the
688 RegMap panel. Nature Genetics. 2012;44: 212–216. doi:10.1038/ng.1042

689 40. Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KM, et al. 1,135
690 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell.
691 2016;166: 481–491. doi:10.1016/j.cell.2016.05.063

692 41. Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, et al.
693 Evidence for widespread positive and negative selection in coding and conserved
694 noncoding regions of Capsella grandiflora. PLoS Genetics. 2014;10: e1004622.
695 doi:10.1371/journal.pgen.1004622

696 42. Mueller JC, Kuhl H, Boerno S, Tella JL, Carrete M, Kempenaers B. Evolution of genomic
697 variation in the burrowing owl in response to recent colonization of urban areas.
698 Proceedings of the Royal Society B: Biological Sciences. The Royal Society; 2018;285:
699 20180206. doi:10.1098/rspb.2018.0206

700 43. Gosler A. The great tit. Hamlyn Species Guides; 1993.

701 44. Culina A, Adriaensen F, Bailey LD, Burgess MD, Charmantier A, Cole EF, et al. Connecting
702 the data landscape of long-term ecological studies: The SPI-Birds data hub. J Anim Ecol.
703 John Wiley & Sons, Ltd; 2020;n/a.

704 45. Visser ME, Noordwijk AJv, Tinbergen JM, Lessells CM. Warmer springs lead to mistimed
705 reproduction in great tits (Parus major). Proceedings of the Royal Society B: Biological
706 Sciences. 1998;265: 1867–1870. doi:10.1098/rspb.1998.0514

707 46. Charmantier A, McCleery RH, Cole LR, Perrins C, Kruuk LEB, Sheldon BC. Adaptive
708 phenotypic plasticity in response to climate change in a wild bird population. Science.
709 2008;320: 800–803. doi:10.1126/science.1157174

710 47. Dingemanse NJ, Bouwman KM, Pol M van de, Overveld T van, Patrick SC, Matthysen E, et
711 al. Variation in personality and behavioural plasticity across four populations of the great
712 tit Parus major. Journal of Animal Ecology. 2012;81: 116–126. doi:10.1111/j.1365-
713 2656.2011.01877.x

714 48. Kvist L, Ruokonen M, Lumme J, Orell M. The colonization history and present-day
715 population structure of the European great tit (Parus major major). Heredity. 1999;82:
716 495–502. doi:10.1038/sj.hdy.6885130

717 49. Kvist L, Martens J, Higuchi H, Nazarenko AA, Valchuk OP, Orell M. Evolution and genetic
718 structure of the great tit (Parus major) complex. Proceedings of the Royal Society B:
719 Biological Sciences. 2003;270: 1447–1454. doi:10.1098/rspb.2002.2321

720 50. Lemoine M, Lucek K, Perrier C, Saladin V, Adriaensen F, Barba E, et al. Low but
721 contrasting neutral genetic differentiation shaped by winter temperature in European

722 great tits. Biological Journal of the Linnean Society. 2016;118: 668–685.
723 doi:10.1111/bij.12745

724 51. Laine V, Gossmann T, Schachtschneider K, Garroway C, Madsen O, Verhoeven K, et al.
725 Evolutionary signals of selection on cognition from the great tit genome and methylome.
726 Nature Communications. 2016;7: 10474. doi:10.1038/ncomms10474

727 52. Kim J-M, Santure AW, Barton HJ, Quinn JL, Cole EF, Great Tit HapMap Consortium ME, et
728 al. A high-density SNP chip for genotyping great tit (Parus major) populations and its
729 application to studying the genetic architecture of exploration behaviour. Molecular
730 Ecology Resources. 2018;18: 877–891. doi:10.1111/1755-0998.12778

731 53. Bosse M, Spurgin L, Laine V, Cole E, Firth J, Gienapp P, et al. Recent natural selection
732 causes adaptive evolution of an avian polygenic trait. Science. 2017;358: 365–368.
733 doi:10.1126/science.aal3298

734 54. Postma E, Noordwijk AJ van. Gene flow maintains a large genetic difference in clutch
735 size at a small spatial scale. Nature. 2005;433: 65–68. doi:10.1038/nature03083

736 55. James JE, Lanfear R, Eyre-Walker A. Molecular Evolutionary Consequences of Island
737 Colonization. Genome Biology and Evolution. 2016;8: 1876–1888.
738 doi:10.1093/gbe/evw120

739 56. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-
740 Wide Allele Frequency Data. PLoS Genetics. 2012;8: e1002967.
741 doi:10.1371/journal.pgen.1002967

742 57. Clements J. The Clements checklist of the birds of the world. New York: Cornell
743 University Press; 2007.

744 58. Hewitt GM. Postglacial re-colonisation of European biota. Biological Journal of the
745 Linnean Society. 1999;68: 87–112.

746 59. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect
747 recent positive selection in the human genome. PLoS Biology. 2007;5: 1587–1602.
748 doi:10.1371/journal.pbio.0050171

749 60. Comeron JM. Background selection as null hypothesis in population genomics: insights
750 and challenges from Drosophila studies. Philosophical Transactions of the Royal Society B:
751 Biological sciences. The Royal Society; 2017;372. doi:10.1098/rstb.2016.0471

752 61. Sexton JP, McIntyre PJ, Angert AL, Rice KJ. Evolution and Ecology of Species Range
753 Limits. Annual Review of Ecology, Evolution, and Systematics. 2009;40: 415–436.
754 doi:10.1146/annurev.ecolsys.110308.120317

755 62. Hill JK, Griffiths HM, Thomas CD. Climate Change and Evolutionary Adaptations at
756 Species' Range Margins. Annual Review of Entomology. 2011;56: 143–159.
757 doi:10.1146/annurev-ento-120709-144746

758 63. Weiss-Lehman C, Hufbauer RA, Melbourne BA. Rapid trait evolution drives increased
759 speed and variance in experimental range expansions. Nature Communications. 2017;8:
760 14303. doi:10.1038/ncomms14303

761 64. Gosler AG. A comment on the validity of the British Great Tit Parus major newtoni.
762 Bulletin of The British Ornithologists' Club. 1999;119: 47–55. Available:
763 http://biostor.org/reference/111964

764 65. Baek J-A, Lan Y, Liu H, Maltby KM, Mishina Y, Jiang R. Bmpr1a signaling plays critical
765 roles in palatal shelf growth and palatal bone formation. Developmental Biology. 2011;350:
766 520–531. doi:10.1016/J.YDBIO.2010.12.028

767 66. Cho Y-L, Min J-K, Roh KM, Kim WK, Han BS, Bae K-H, et al. Phosphoprotein phosphatase
768 1CB (PPP1CB), a novel adipogenic activator, promotes 3T3-L1 adipogenesis. Biochemical
769 and Biophysical Research Communications. 2015;467: 211–217.
770 doi:10.1016/J.BBRC.2015.10.004

771 67. Migliano AB, Romero IG, Metspalu M, Leavesley M, Pagani L, Antao T, et al. Evolution of
772 the Pygmy Phenotype: Evidence of Positive Selection from Genome-wide Scans in African,
773 Asian, and Melanesian Pygmies. Human Biology. 2013;85: 251–284.
774 doi:10.3378/027.085.0313

775 68. Logan CA, Somero GN. Effects of thermal acclimation on transcriptional responses to
776 acute heat stress in the eurythermal fish Gillichthys mirabilis (Cooper). American Journal of
777 Physiology-Regulatory, Integrative and Comparative Physiology. 2011;300: R1373–R1383.
778 doi:10.1152/ajpregu.00689.2010

779 69. Videvall E, Cornwallis CK, Palinauskas V, Valkiūnas G, Hellgren O. The Avian
780 Transcriptome Response to Malaria Infection. Molecular Biology and Evolution. 2015;32:
781 1255–1267. doi:10.1093/molbev/msv016

782 70. Gunnarsson U, Kerje S, Bed'hom B, Sahlqvist A-S, Ekwall O, Tixier-Boichard M, et al. The
783 Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10.
784 Pigment Cell & Melanoma Research. 2011;24: 268–274. doi:10.1111/j.1755-
785 148X.2011.00825.x

786 71. Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to
787 identify loci underlying local adaptation. Genetics. 2010;185: 1411–23.
788 doi:10.1534/genetics.110.114819

789 72. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci
790 and environmental gradients using latent factor mixed models. Molecular Biology and
791 Evolution. 2013;30: 1687–1699. doi:10.1093/molbev/mst063

792 73. Salmón P, Jacobs A, Ahrén D, Biard C, Dingemanse NJ, Dominoni DM, et al. Repeated
793 genomic signatures of adaptation to urbanisation in a songbird across europe. bioRxiv.
794 2020. p. 2020.05.05.078568.

795 74. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: A Tool
796 Set for Whole-Genome Association and Population-Based Linkage Analyses. The American
797 Journal of Human Genetics. 2007;81: 559–575. doi:10.1086/519795

798 75. R Development Core Team. R: A Language and Environment for Statistical Computing
799 [Internet]. Team RDC, editor. 2011. doi:10.1007/978-3-540-74686-7

800 76. Goslee SC, Urban DL. The ecodist Package for Dissimilarity-based Analysis of Ecological
801 Data. Journal of Statistical Software. 2007;22: 1–19. doi:10.18637/jss.v022.i07

802 77. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in
803 unrelated individuals. Genome Research. 2009;19: 1655–1664. doi:10.1101/gr.094052.109

804 78. Oers K van, Santure AW, De Cauwer I, Bers NE van, Crooijmans RP, Sheldon BC, et al.
805 Replicated high-density genetic maps of two great tit populations reveal fine-scale genomic
806 departures from sex-equal recombination rates. Heredity. 2014;112: 307–316.
807 doi:10.1038/hdy.2013.107

808 79. Gautier M, Vitalis R. Rehh: An R package to detect footprints of selection in genome-
809 wide SNP data from haplotype structure. Bioinformatics. 2012;28: 1176–1177.

810 80. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, et al. A general
811 approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet.
812 2014;10: e1004234.