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], Veronika N. Laine[3,25], Bart Kempenaers[26], Anvar Kerimov[10], Milos Krist[27], Laura Kvist[28], Raivo Mänd[29], Erik Matthysen[6], Ruedi Nager[30], Boris P. Nikolov[31], A. Claudia Norte[32], Markku Orell[28], Jenny Ouyang[26,33], 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 Animal Sciences, Wageningen University, Wageningen 6708 PB, The Netherlands

24. Zvenigorod Biological Station of Lomonosov Moscow State University, P.O. Box Shikhovo, Odintsovo District, Moscow 143092, Russia

25. Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, United States of America

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

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

28. Department of Ecology and Genetics, P.O.Box 3000, 90014-University of Oulu, 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, Babeș-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

Abstract

A major aim of evolutionary biology is to understand why patterns of genomic variation 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, such studies are now becoming feasible. Here, we carry out one of the most 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 the Balkans and/or Turkey, with bottlenecks and reduced genetic diversity in island populations. Differentiation across the genome was highly heterogeneous, with clear "islands of differentiation" even among populations which are ostensibly panmictic. These islands of differentiation were consistently found in regions of low recombination, suggesting that background selection can rapidly promote population differentiation among even the most recently colonised populations. We also detected genomic outlier regions that were unique to peripheral great tit populations, most likely as a result of recent directional selection at the range edges of this species. These "unique" outlier regions contained candidate genes for morphology, thermal adaptation and colouration, supporting previous research in this species, and providing avenues for future investigation. Our study suggests that comprehensive screens of genomic variation in wild organisms can provide unique insights into evolution.

Author summary

Studying patterns of genetic variation is a useful way of determining why populations and species differ in nature. Genetic variation is shaped by natural selection, but also by the present and past size of populations, the amount of migration, and by features of the genome, such as variation in recombination rate, of the organism being studied. Teasing apart the effects of these different processes on genomic diversity is difficult, but one way that this can be achieved is by studying genomic variation across the entire range of a species. We performed a continental-scale analysis of genetic variation in the great tit - a widespread songbird that has been the focus of extensive ecological research. We first used genomic data to reconstruct the historical colonisation of great tits across Europe, and showed that during the last ice age, this species was likely restricted to a single region in Eastern Europe, from which they spread across the continent. We then studied how patterns of variation differ along the genome, and show that recombination rate is a key driver of variation among all populations. Importantly, by comparing many populations we were able to identify

genes that have been subject to natural selection in specific geographical regions. We found that natural selection appeared to be strongest in populations on the edges of the great tit's range acting on traits such as morphology, stress response and colouration. Large-scale genetic analyses such as ours are therefore useful approaches for understanding how evolution operates in the wild.

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. 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 [5,6].

The increasing feasibility of high-throughput sequencing and subsequent characterisation of genome-wide variation across large numbers of individuals has revealed that at the genetic level, patterns of variation and divergence among natural populations and species are highly heterogeneous [7]. 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} [7–10]. Initially these regions were termed "islands of speciation", and were thought to arise as a result of reduced gene flow in genomic regions associated with reproductive isolation [7,11]. Subsequent research has revealed that highly heterogeneous patterns of genomic divergence can occur even in the complete absence of gene flow, as a result of recombination rate variation and linked selection [12,13]. In genomic regions of low recombination, selection for beneficial mutations (positive selection), or against deleterious mutations (background selection), will impact relatively large genomic regions as a result of high levels of linkage disequilibrium (LD) among sites. Selection within these regions reduces diversity within populations, and increases differentiation among them, resulting in "islands" of increased differentiation that persist over evolutionary time [13,14].

Comparing patterns of genomic differentiation among sets of populations or species at different stages of the divergence/speciation continuum is a powerful way of disentangling the forces that shape variation among populations. Martin et al. [15] showed that, across multiple *Heliconius* butterfly populations and species,

patterns of genomic variation were shaped by a combination of gene flow and selection, particularly in genomic regions harbouring genes involved in wing patterning. In contrast, Renaut et al. [9] showed that in *Helianthus* sunflowers, genomic architecture was the main driver of genomic differentiation across sets of populations. Similarly, recent research in birds has revealed that differentiation landscapes are conserved across populations, species and even across avian families, with the same islands of differentiation arising among populations of distantly related species [16–18]. This latter pattern appears to have arisen, at least in part, as a result of a highly conserved recombination landscape in birds, with background selection in regions of low recombination producing recurrent islands of differentiation [19].

It is now clear that the recombination landscape and linked selection are key drivers of genomic variation within and among populations. However, we are only just beginning to understand how this linked selection interacts with other evolutionary forces to shape patterns of differentiation across natural populations and species [19–23]. A recent, large-scale analysis of threespine sticklebacks (*Gasterosteus aculeatus*) showed that islands of differentiation were more likely to arise in low recombination regions when gene flow occurred between populations [24]. There is also a significant impact of divergence time; in recently separated populations the differentiation landscape is most likely to reflect selective sweeps. Then, as divergence accumulates, genomic architecture is expected to play an increasingly important role in generating these genomic islands [19].

Widespread continental species are excellent models for studying how demography and the environment shape genetic and phenotypic variation among populations, due to their large effective population sizes and ecologically varied ranges. Insight into the evolutionary history of such species can be gained if genetic variation is characterized across much of its geographical range. Cross-population comparisons of genetic variation can then be used to make inferences about phylogeography, levels of gene flow between populations and how adaptation to different environmental and ecological conditions occurs. Whole-genome resequencing and customized SNP genotyping arrays have made studies that characterize all, or much, of the genome tractable. The first large-scale studies were performed in humans - i.e. the HapMap Projects [25–27] which characterized human genetic variation on different continents, with a view to determining the feasibility of association mapping studies. Similar studies have been conducted in domesticated species and their wild ancestors [28–30], and in model organisms [31,32]. More recently, there is a growing appreciation that HapMap-type studies are useful for studying signatures of selection and adaptation in natural populations of species with large effective population sizes and high levels of gene flow [33–36].

The European great tit (*Parus major major*) is an excellent model for ecological and evolutionary studies [37]. A wealth of ecological data exists across multiple great tit populations [38–40], enabling informed hypotheses about selection to be tested in this system. Phyleogeographic research using mitochondrial DNA suggests that

this species has recently expanded across its European range, possibly from a single refugium in the Balkans [41]. Contemporary populations are characterised by large effective population sizes and high levels of gene flow among populations, resulting in low levels of genetic differentiation [42,43]. However, these previous cross-population molecular studies have relied on a modest number of microsatellite loci and mitochondrial DNA, making the detection of genomic regions under selection impossible. The genome of the great tit has recently been sequenced [44], and a high density panel of SNP markers has been developed [45]. A recent study of two European populations using this marker panel suggests that rapid adaptation has occurred at the genomic and phenotypic levels, with pronounced selection on morphology [46].

Here, we perform a HapMap study of 647 unrelated individuals across 29 populations (Fig. 1), to examine how genomic architecture, natural selection and population history have shaped patterns of genomic variation across recently colonised European great tit populations. Using a large SNP panel typed across all individuals, we first characterise genome-wide patterns of variation within and among populations, in order to infer population history. We then examine how variation is partitioned across the genome, and test the hypothesis that highly divergent genomic regions have arisen in genomic regions of low recombination [12,13]. Finally, we examine how genomic divergence accumulates along the colonisation route of this species, with the aim of inferring how recent natural selection and background selection drive variation across the genome in the wild.

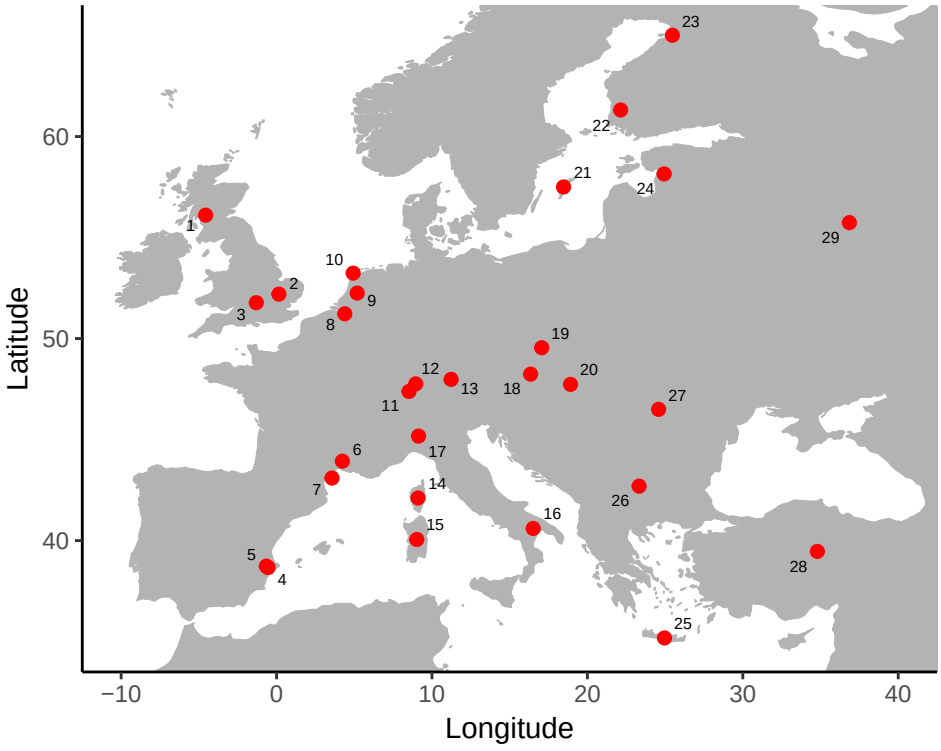

Figure 1 Sampling locations of great tit populations. Population names and sample sizes are given in Table S1, and numbers on the map correspond to the "code" column in Table S1.

Results and Discussion

Genetic diversity and population history

Sampling locations and sample sizes for each population are given in Table S1. Levels of genetic diversity (π_{SNP}) were generally high, but we observed substantial differences among populations (Fig. 2A). Similarly, LD declined rapidly with genomic distance in all populations, reaching baseline levels within ~5kb in all populations, but also varying among populations (Fig. S1). Highest levels of LD (and lowest levels of genetic diversity) were observed in the Mediterranean island populations of Crete (Greece) and Sardinia (Italy), with lowest levels of LD in central and western Europe (Fig. S1). This is consistent with reduced effective population size in these island populations, either as result of the colonisation process or more recent bottlenecks, along with low levels of subsequent gene flow from the continent to the islands [47,48].

Genome-wide F_{ST} between European great tit populations was 0.007, with no significant pattern of isolation-by-distance (Mantel test; r = 0.13, p = 0.18; Fig. S2). Instead, the highest levels of F_{ST} were found in comparisons involving the Mediterranean island populations of Corsica (France), Sardinia and Crete (Fig. S2). Admixture analysis was consistent with this pattern (Fig. S3); the $K = 2$ analysis assigned individuals in Sardinia and Corsica to one genetic cluster, and the remaining populations to the second. Thus, it is likely that much of the genetic structure between European great tit populations is a result of genetic drift in these small island populations. Admixture analysis also revealed some structure between (mainly peripheral) mainland and larger island populations. At $K = 3$ (the model that best fitted the great tit data; Fig. S4), Spain was separated from the rest of mainland Europe. Increasing values of K resulted in the separation of populations in Scotland ($K = 4$), Sardinia (from Corsica; $K = 5$), southern France ($K = 6$), Crete ($K = 7$) and England ($K = 8$). The Admixture output at $K = 8$ is displayed in Fig. 2C as this gives the most detailed picture of genetic structure among European great tit populations. Further increases in K did not generate patterns of structure that corresponded to geographical variation (Fig. S3), and were increasingly less well supported (Fig. S4). Thus, even with hundreds of thousands of markers Admixture was unable to separate many of the European populations, confirming that levels of divergence are extremely low. PCA largely corroborated the Admixture results, with PC1 separating Corsica and Sardinia from the remaining populations, PC2 separating Spain, while PC3 and PC4 separated Scotland, England, Corsica, Sardinia and Crete (Fig. S5).

Maximum likelihood analyses implemented in TreeMix showed that a model with no migration explained 97.8% of variance in relatedness between populations [49]; increasing the number of migration events substantially improved the percentage of relatedness explained, up to 99.7% when 10 migration events were fitted (Fig.

S6). In Figure 3 we display the maximum likelihood tree with two migration events, after which the variance in relatedness explained plateaued when more migration events were added (Fig. S6). The tree was generally characterised by short branch lengths, with the exception of the island populations of Sardinia and Crete, which were grouped with the population from mainland Italy (Fig. 3). Thus, the TreeMix analysis is consistent with the pattern of low overall genomic divergence, with the exception of the Mediterranean island populations. However, much (though not all) of the grouping that did occur among continental populations made geographical sense, with populations from Finland and Estonia grouped together, as were populations from Turkey and the Balkans, and populations from England and Scotland (Fig. 3). Interestingly, TreeMix grouped the Spanish and Corsican populations, which is consistent with previous subspecies descriptions of European great tits [50]. The two fitted migration edges both involved Sardinia, with migration from eastern Europe to Sardinia, and from Sardinia to Corsica (Fig. 3).

We next tested the hypothesis that great tits colonised Europe from a single refugium in Turkey and the Balkans. This scenario has been suggested before [41], but due to the low number of genetic markers available there has been limited power with which to test this hypothesis. Using our genome-wide panel of SNP markers, we compared genetic and geographic distance between each population and the proposed refugial populations. Because of the elevated structure in Corsica, Sardinia and Crete (Fig. S2), we excluded comparisons involving these populations. We found that F_{ST} was significantly related to distance from Turkey ($r = 0.81$, $p < 0.001$; Fig. 2B) and the Balkans ($r = 0.44$, $p = 0.001$). The same relationship was not found for alternative potential refugial populations in Spain ($r = -0.09$, $p = 0.55$), or southern Italy ($r = 0.04$, $p = 0.77$). Our results therefore lend empirical support to the hypothesis [41] that great tits colonised Europe from a single refugium in the south-east. Clearly, although our sampling was extensive, it is not exhaustive, and more fine-scaled sampling in eastern Europe would be required to determine the extact location and extent of refugial great tit populations. Sampling in North Africa would also be useful to determine whether further refugia exist, and to quantify the extent of admixture between European and African great tit populations.

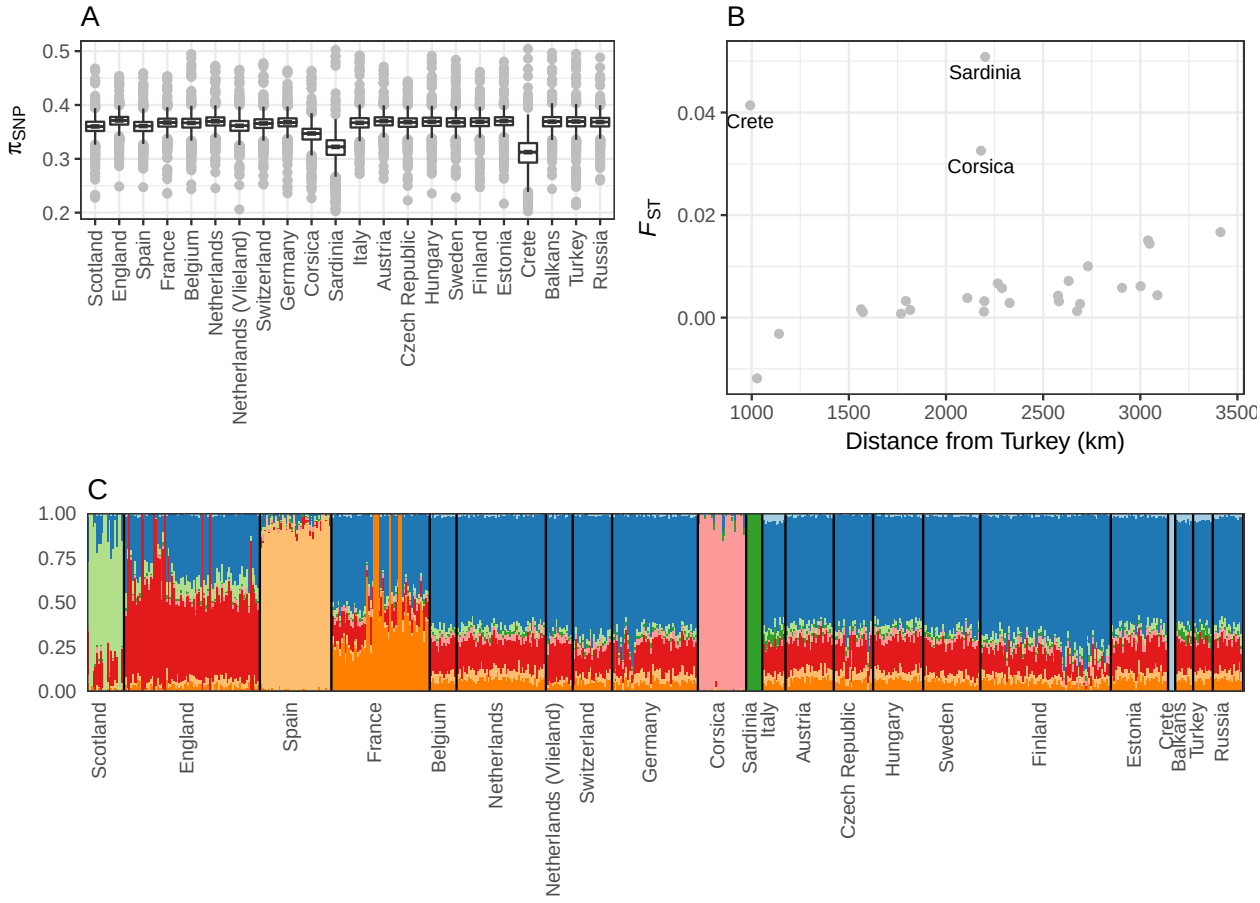

Figure 2 Genetic diversity and structure in European Great tit populations. **A** Nucleotide diversity within each population. **B** Pairwise F_{ST} in relation to geographic distance from the Turkey, only including comparisons involving Turkey. **C** Output from Admixture analysis at $K = 8$. Population details can be found in Table S1.

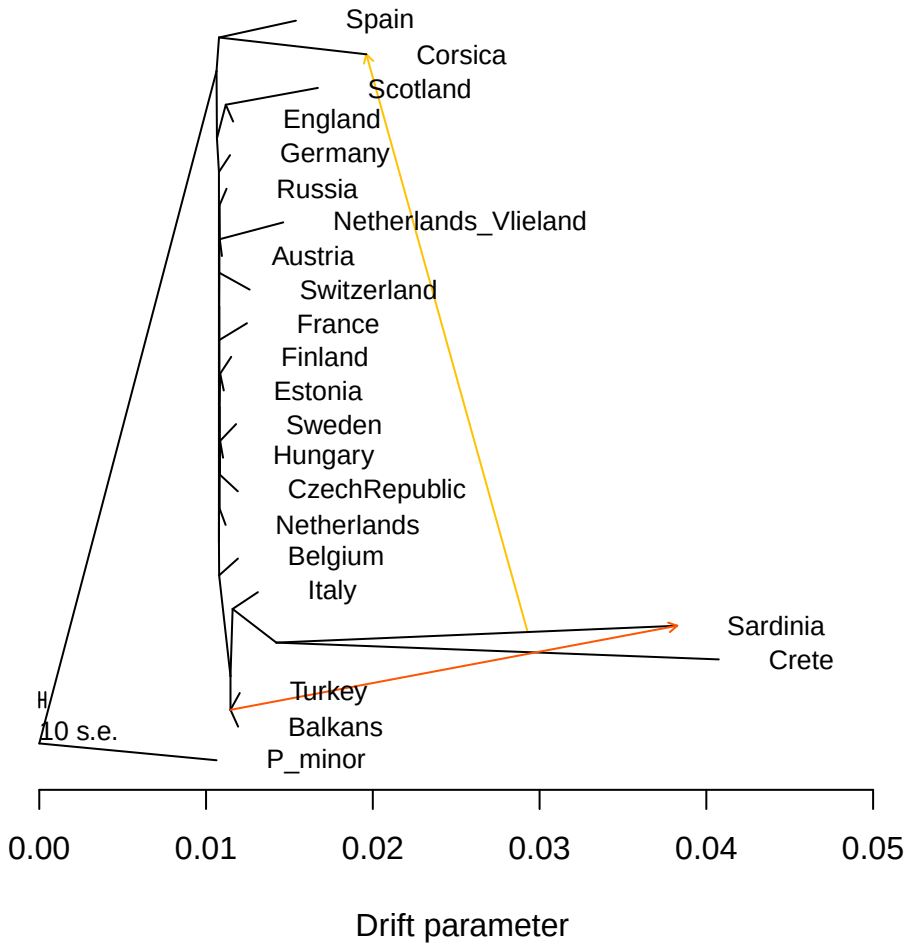

Figure 3 Maximum likelihood tree inferred by TreeMix, allowing two migration events. The two migration events (arrows) are coloured according to their weight (red = higher migration), and horizontal branch lengths are proportional to the amount of genetic drift that has occurred along the branch. A population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

Genomic landscapes of differentiation

It is likely that many, and perhaps the majority, of wild populations are characterised by highly heterogeneous patterns of differentiation across the genome [23]. To examine how landscapes of genomic divergence have formed along the colonisation route of European great tits, we calculated windowed F_{ST} in 500kb bins between each population and the proposed refugial population in Turkey. We found that F_{ST} varied markedly across the genome in all comparisons (Fig. 4; Fig. S7). Outlier regions (windows with standardized F_{ST},hereafter zF_{ST}, > 10) were found in all comparisons apart from Crete and Sardinia, in which overall levels of divergence were highest (Fig. 4). Our results suggest, therefore, that genomic islands of differentiation can and do arise even among populations that could be considered panmictic.

Genomic differentiation was negatively related to recombination rate in almost all comparisons with Turkey (Fig. 5). The relationship between F_{ST} and recombination rate was generally weak, with correlation coefficients < 0.1. In a handful of populations this relationship was substantially stronger - most notably in the island populations of Corsica, Sardinia and Crete, and in England, Scotland, Spain, Finland and France (Fig. 5). Outlier regions of very high differentiation ($zF_{ST} > 10$) almost exclusively occurred in relatively low recombining regions, and accordingly, the recombination rate of outlier regions was lower than the genome-wide average, albeit not significantly (Wilcoxon test, $P = 0.079$). Nonetheless, there is a suggestion that linked selection in regions of low recombination may play a key role in driving patterns of genomic differentiation even in the very earliest stages of population separation.

Of the 11 outlier regions, nine were found in only one or two comparisons, while the other two were found in 12 and 10 comparisons, respectively (Table S2). We hereafter refer to outlier regions found in one or two comparisons as "unique" outlier regions, and to those found in more than two comparisons as "shared" regions. It appears that European great tits have colonised from a single refugial population, and as such truly independent comparisons are not available. Therefore an outlier region shared among multiple populations could represent either selection in the ancestral population, or background selection. Given the pervasiveness of background selection in birds [17,18], and other organisms [51], and the overall negative relationship between F_{ST} and recombination rate in this study (Fig. 5), it is likely that this background selection is the driver of the shared outlier regions in great tits. Thus, it appears that background selection can generate islands of differentiation in the very earliest stages of population separation. This is not necessarily what we would expect - it is often assumed that peaks of high differentiation in recently separated populations are the product of selective sweeps, and that only after time do correlated patterns of genomic divergence arise [19,24]. Further research into genomic landscapes of differentiation among widespread continental species will help us better understand on the role of background selection in shaping genomic divergence over short

evolutionary timescales.

Regions of high differentiation that are not shared among populations are more likely to be the result of recent positive selection [19]. We found that unique outlier regions tended to be found in the most peripheral European great tit populations, with three found in Scotland, two in England, Spain and Finland; the remaining outlier regions were found in comparisons involving the Czech republic, Russia, Vlieland (Netherlands) and Belgium (Table S2). Observational and experimental research shows that adaptation at range edges is a key feature shaping divergence among recently colonised and expanding populations [52–54]. There appeared to be no difference in the recombination rate between shared and unique outlier regions (Fig. 5), although the small number of regions precluded testing this hypothesis formally. Thus, it is likely that genomic architecture plays a key role in determining how both positive and background selection have shaped genomic variation across the recent evolutionary history of European great tits.

Genes found within shared and unique outlier regions are displayed in Table S2. Perhaps most notable among these is *COL4A5*, a gene found to be associated with bill length, and under selection between populations in England and the Netherlands, in a recent great tit study [46]. Here we found that the *COL4A5* region is an F_{ST} outlier in England and Scotland, but not in any other European populations (Table S2). UK great tits have been described as a separate subspecies based on beak shape [55], and our results here, combined with previous results, suggest that this divergence is the result of recent natural selection in the UK [46]. Another notable candidate gene potentially involved in beak morphology, and previously found to be under selection in UK great tits is *BMPR1A*, which plays a key role in palate development [56] and in this study was found in an outlier region in Scotland. Other candidate morphology and obesity genes in the unique outlier regions in the UK included *PPP1CB*, which may play a role in adipogenesis [57] and *GHITM*, which appears to have been subject to natural selection in human pygmy populations [58]. Thus, morphological traits may frequently be involved in adaptation in great tits.

In addition to morphological candidates in the UK, we found outlier regions unique to cold populations in Scotland, Finland and Russia (Table S2), containing at least one candidate gene for thermal stress (*CDKN1B*) [59]. Other genomic outlier regions contained potential candidate genes for malaria infection (*MRPL33*) [60] and colour variation (*SOX10*) [61]. This is thus far an exploratory analysis, and we are therefore reluctant to speculate whether these candidate genes are genuine targets for natural selection, and more reluctant still to speculate as to how selection might be driving variation at these regions. Regardless, these candidates will provide useful starting points for future genomic and ecological investigation.

To further explore how selection may have shaped variation in F_{ST} outlier regions, we estimated levels

of nucleotide and haplotype diversity within these regions. Nucleotide diversity (π_{SNP}) in outlier regions varied from 0.21 to 0.48, and diversity in these regions was significantly lower than the genome-wide average (Wilcoxon test, $P = 0.018$; Fig. S8A). However, there appeared to be no difference in nucleotide diversity between shared and unique regions (Fig. S8A). Haplotype diversity varied substantially among regions, with haplotype richness ranging from 72 to 1033. Both haplotype richness and marker density in shared regions tended to be lower than those in unique regions (Fig. S8B). A detailed examination of haplotype structure in one shared and one unique region is displayed in Figure S9. The unique outlier region (to Finland, situated on chromosome 1A) was characterised by a complex structure, with a single haplotype at high frequency in Finland compared to other populations, indicating a population-specific selective sweep (Fig. S9A,C). In contrast, the shared region on chromosome 2 was much less complex, demonstrating higher haplotype frequencies across a range of populations. Our data therefore suggest that examining patterns of haplotype diversity in outlier regions may help to separate recent episodes of positive selection from the gradual process of background selection (Figs S8, S9).

HapMap style projects have been hugely informative in shaping our understanding of how natural selection operates in humans and other model species [25,32]. This study is one of the largest to date of genomic variation in a wild vertebrate, which has helped to reveal the evolutionary history of great tits, and to identify candidate genes and traits that may have been involved in adaptation during and/or after postglacial recolonisation. Further, this work will form the foundation of many future analyses. Clearly, we have only touched on haplotype-based methods to infer adaptation here, and this will be the subject of future work. Environmental association approaches are also highly suited to detecting adaptation in widespread continental species [62,63], and further work will test how variation in the abiotic environment has shaped patterns of genomic variation in great tits. This combination of environmental and genomic data in species such as great tits, in which a wealth of ecological and genomic resources are available, is likely to generate interesting insights into the the genetic and phenotypic basis of natural selection.

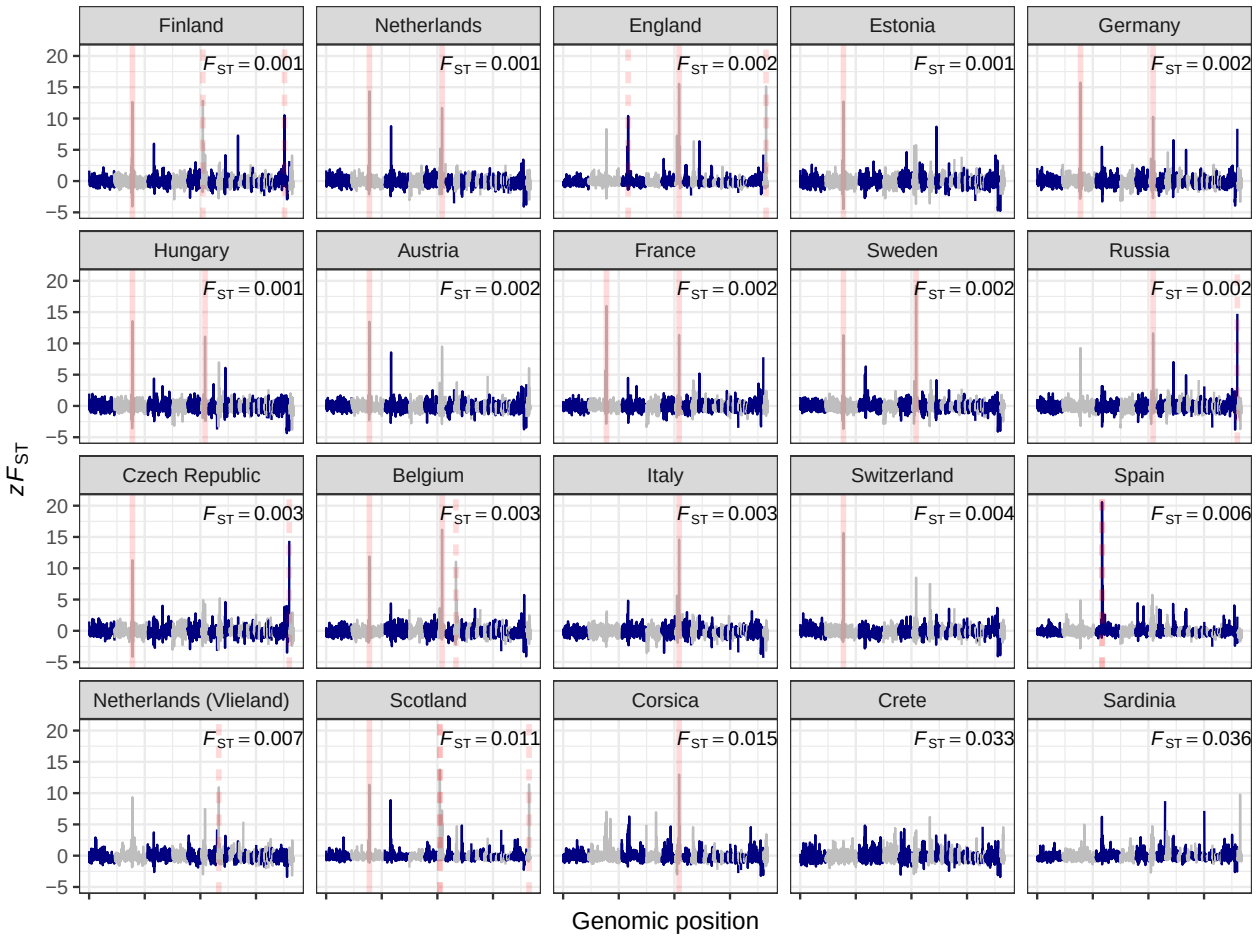

Figure 4 Landscapes of relative genomic differentiation in European great tit populations. zF_{ST} across the genome is averaged in 500kb windows, with each panel displaying a pairwise comparison with the proposed refugial population in Turkey. Red lines represent F_{ST} outliers (windows with mean F_{ST} values at least 10 standard deviations greater than the global mean for that comparison) shared across more than two comparisons (solid red lines), or unique 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.

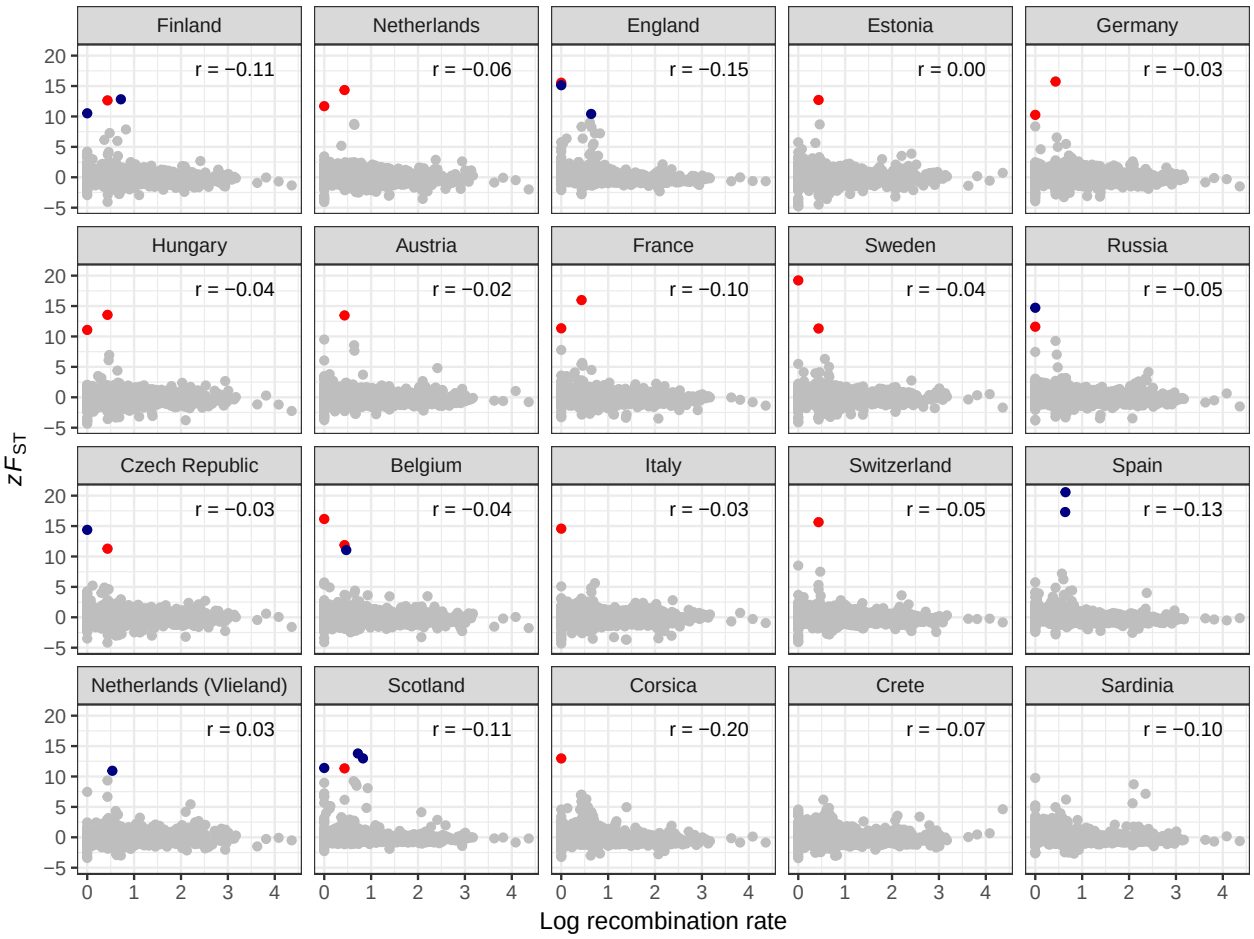

Figure 5 Genomic differentiation and recombination rate variation in European great tit populations. Each point is the mean of a 500kb window, with each panel displaying a pairwise comparison with the proposed refugial population in Turkey. Coloured points represent F_{ST} outliers (mean standardized F_{ST} values of $zF_{ST} > 10$) shared across more than two comparisons (red), or unique to one or two comparisons (dark blue).

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 prximity (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 [45], and the filtering approaches outlined in [46]. After filtering, a total of 647 samples typed at 483888 SNPs were retained for analysis.

Analyses

Unless stated otherwise, all population genetic statistics were calculated in PLINK version 1.9 [64], and downstream analysis and plotting was carried out in R version 3.3 [65]. In each population, we estimated LD (R^2) for each pair of markers within 50kb on the same chromosome, and compared this to physical distance between marker pairs. We calculated observed heterozygosity for each SNP and population using a reduced SNP dataset, which was pruned based on LD to remove all markers with $R^2 > 0.1$, then thinned with a probability of retaining each variant of 0.25.

We calculated genome-wide (mean) F_{ST} between each pair of populations using the pruned and thinned dataset described above. Pairwise F_{ST} was compared to geographic distance between populations using Mantel tests, implemented in the Ecodist package in R [66]. We tested whether genetic structure was related to distance from candidate refugial populations (in the Balkans, Turkey, Spain and Italy), using Pearson correlations. We also estimated population structure using Admixture version 1.3, with default settings [67]. We varied values of K from one to ten; by which point increasing values of K provided no informative information about population structure (see results). Model support for each value of K was estimated by calculating 5-fold cross-validation error. Finally, we visualised the evolutionary history among European great tit populations by generating a maximum likelihood tree in TreeMix version 1.13 [49]. We rooted the tree using a sample of *P. minor* individuals sampled from Amur, Russia. We fitted models allowing for range

of migration events (0-10), and used a window size of 500 SNPs [49]. To assess model fit, we calculated the proportion of variance in relatedness between populations explained by each model [49].

We examined the genomic landscape of differentiation across European great tit populations by calculating F_{ST} in 500kb bins, using python scripts obtained from Github (https://github.com/simonhmartin/genomics_general). We did not estimate d_{xy}, as this parameter is difficult to estimate accurately from single SNP loci [13]. We also calculated standardised F_{ST} (zF_{ST}) by mean-centring windowed values and dividing them by the standard deviation among windows. We defined outlier regions as 500kb bins with zF_{ST} values greater than ten. We tested whether the landscape of genomic differentiation was related to recombination rate variation using a recombination map previously developed for the great tit using a 10K SNP chip [68], from which we estimated recombination rates using third-order polynomials [69].

Acknowledgements

This work was supported by grants from the European Research Council (grant 202487 to J.S. and grant 339092 – E-Response to MEV) and Natural Environment Research Council (grant NE/J012599/1 to J.S. and B.C.S). 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.

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. 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

6. 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

7. 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

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

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

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

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

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

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

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

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

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

17. Vijay N, Weissensteiner M, Burri R, Kawakami T, Ellegren H, Wolf JBW. Genomewide patterns of

447 variation in genetic diversity are shared among populations, species and higher-order taxa. Molecular Ecology.
448 2017;26: 4284–4295. doi:10.1111/mec.14195

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

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

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

456 21. Jiggins CD, Martin SH. Glittering gold and the quest for Isla de Muerta. Journal of Evolutionary Biology.
457 2017;30: 1509–1511. doi:10.1111/jeb.13110

458 22. Lohse K. Come on feel the noise - from metaphors to null models. Journal of Evolutionary Biology.
459 2017;30: 1506–1508. doi:10.1111/jeb.13109

460 23. Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, et al. Interpreting the genomic
461 landscape of speciation: a road map for finding barriers to gene flow. Journal of Evolutionary Biology.
462 2017;30: 1450–1477. doi:10.1111/jeb.13047

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

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

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

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

472 28. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al. Genome-wide analysis
473 of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biology.
474 2012;10: e1001258. doi:10.1371/journal.pbio.1001258

475 29. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, et al. Genome-wide

476 survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 2009;324: 528–532.
477 doi:10.1126/science.1167936

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

480 31. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome se-
481 quence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005;438: 803–819.
482 doi:10.1038/nature04338

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

485 33. Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, et al. Genome-wide patterns of
486 genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nature Genetics.
487 2012;44: 212–216. doi:10.1038/ng.1042

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

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

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

497 37. Gosler A. The great tit. Hamlyn Species Guides; 1993.

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

501 39. Charmantier A, McCleery RH, Cole LR, Perrins C, Kruuk LEB, Sheldon BC. Adaptive pheno-
502 typic plasticity in response to climate change in a wild bird population. Science. 2008;320: 800–803.
503 doi:10.1126/science.1157174

504 40. Dingemanse NJ, Bouwman KM, Pol M van de, Overveld T van, Patrick SC, Matthysen E, et al. Variation

505 in personality and behavioural plasticity across four populations of the great tit Parus major. Journal of
506 Animal Ecology. 2012;81: 116–126. doi:10.1111/j.1365-2656.2011.01877.x

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

509 42. Kvist L, Martens J, Higuchi H, Nazarenko AA, Valchuk OP, Orell M. Evolution and genetic structure
510 of the great tit (Parus major) complex. Proceedings of the Royal Society B: Biological Sciences. 2003;270:
511 1447–1454. doi:10.1098/rspb.2002.2321

512 43. Lemoine M, Lucek K, Perrier C, Saladin V, Adriaensen F, Barba E, et al. Low but contrasting neutral
513 genetic differentiation shaped by winter temperature in European great tits. Biological Journal of the Linnean
514 Society. 2016;118: 668–685. doi:10.1111/bij.12745

515 44. Laine V, Gossmann T, Schachtschneider K, Garroway C, Madsen O, Verhoeven K, et al. Evolutionary
516 signals of selection on cognition from the great tit genome and methylome. Nature Communications. 2016;7:
517 10474. doi:10.1038/ncomms10474

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

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

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

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

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

530 50. Clements J. The Clements checklist of the birds of the world. New York: Cornell University Press; 2007.

531 51. Comeron JM. Background selection as null hypothesis in population genomics: insights and challenges
532 from Drosophila studies. Philosophical Transactions of the Royal Society B: Biological sciences. The Royal

533 Society; 2017;372. doi:10.1098/rstb.2016.0471

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

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

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

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

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

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

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

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

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

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

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

561 63. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental

562 gradients using latent factor mixed models. Molecular Biology and Evolution. Oxford University Press;
563 2013;30: 1687–1699. doi:10.1093/molbev/mst063

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

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

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

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

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

576 69. Corcoran P, Gossmann TI, Barton HJ, Great Tit HapMap Consortium, Slate J, Zeng K. Determinants of
577 the Efficacy of Natural Selection on Coding and Noncoding Variability in Two Passerine Species. Genome
578 Biology and Evolution. 2017;9: 2987–3007. doi:10.1093/gbe/evx213