**Genomic insights into continental-scale population structure and adaptation in a model passerine**

Lewis G. Spurgin,1 Mirte Bosse2, Martien A. M. Groenen3, Veronika Laine2, Kees van Oers2, The Great Tit Hap Map Consortium, Ben C. Sheldon1, Marcel E. Visser2,3, Jon Slate4

1 Edward Grey Institute, University of Oxford, United Kingdom  
2 Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, the Netherlands  
3 Animal Breeding and Genomics Center, Wageningen University, the Netherlands  
4 Department of Animal and Plant Sciences, University of Sheffield, United Kingdom

**Abstract**

Understanding how demography and selection shape genetic and phenotypic variation among natural populations is a major aim of biology. The European great tit (Parus major major) is a widespread passerine bird that inhabits a wide range of ecological environments. Moreover, the great tit is a model species for ecological and behavioural research, with excellent genomic resources. Here we use a panel of ~500,000 SNPs and samples from 23 European populations to disentangle the the roles of demographic and adaptive processes in shaping genomic variation across this species' range. Patterns of genomic diversity are consistent with recent spread and admixture from multiple southern refugia. While some structure exists between southern and northern populations, we find that the majority of northern and central European great tits exist as a single, panmictc population. Selection analyses reveal clear evidence for genomic regions under selection between southern and northern European populations, as well as within the northern panmictic cluster. We identify candidate genes for adaptation to temperature and latitude. Our results provide novel insights into ecological adaptation in widespread species.

**Introduction**

In this study we analyse genomic variation across 23 great tit populations from across Europe. Our aims were twofold. First, we quantify levels of fenetic variation and population structure among European great tit populations, and relate this to hypotheses of colonisaiton history. Second, we identify regions of the genome involved in ecological adaptation, with the aim of identifying biological processes involved in adaptation across this species' range.

**Materials and methods**

*Sampling and molecular methods*

Samples were obtained from etc.

DNA extraction details

SNP chip details

*Statistical analyses*

Summary statistics were calculated using PLINK v 1.9. Levels of genetic diversity, inbreeding, linkage etc.

We calculated pairwise FST between each air of populations, at every marker, and in 200kb sliding windows, using VCFtools.

To better understand populations structure, we used multidimensional scaling (MDS) plots, implemented in PLINK, and the software ADMIXTURE. For the latter, we ran etc etc

We used two approaches to identify regions of the genome under selection across European great tit populations. First, we identified outlier regions of high differentiation among genetic clusters (identified using ADMIXTURE), using EigenGWAS.

Second, using individuals from a single, widespread genetic cluster (see results) we used BAYESCENV - an FST-based approach to identify regions of the genome significantly associated with ecological variation. For this analysis we used DETAILS and latitude as an environmental variable. Outlier regions from BAYESCENV were confirmed using two less conservative approaches - BAYENV and latent factor mixed models.

**Results**

In total, XX birds were genetyped at between YY and ZZ SNPs. Details of population-level sample sizes and SNP success rates are provided in Table S1.

*Genetic diversity and population structure*

Differences in levels of genetic diversity were predominantly between southern and northern populations, with lowest levels of genetic diversity in Corsican and Spanish populaitons (Fig. 1). Levels of genetic diversity among central and northern European populations were almost identical (Fig. 1).

Linkage disequilibrium declined very rapidly, to background levels of XX-YY, in all populations (Fig. 2). However, there were significant differences in the initial rate of LD decay (), as well as in levels of long-range LD (). At very short inter-SNP distances, highest levels of LD were found in Corsica, Spain and Scotland (Fig. 2), and LD in Corsica decayed less rapidly than in all other populations.

Multi-dimensional scaling analysis revealed three genetically distinct clusters - one for Corsica, one for the two Spanish populations and one for the rest of Europe (Fig. 3). Admixture confirmed that three genetic clusters best explained the data

*Selection*

**Discussion**

**Figures**

**References**