**PROJECT DESCRIPTION** (for Pilot projects and non NERC-grant funded projects)

Maximum 750 words excluding references. Submit as a PDF.

Please explain adequately the following aspects:

* background
* objectives of the research and the relevance of the proposed molecular genetic analyses to these objectives
* your contribution to the work, including, for example, any bioinformatic analyses
* any wider justification for the project.

Only applications from NERC eligible PIs with all documents submitted will be considered.

Background:

Comparing the genomes of museum specimens with those of contemporary samples provides an opportunity for studying adaptive evolution in response to temporally-changing environments (e.g. climate change). We propose to take this approach to address whether great tits (*Parus major*), a widely-studied bird with an assembled and annotated genome, show signatures of selection to warmer climates over the last hundred or so years. The work will form a central part of the PhD of Joanne Stonehouse, who is funded by the NERC ACCE DTP.

Jo has recently completed a large analysis of genomic and climate data in a set of 20 great tit populations, sampled across the species’ range as part of the Great Tit HapMap Project coordinated by J Slate. The populations were typed at >400,000 SNPs (15-48 birds per population, 535 samples in total). Using data from the WorldClim database, multivariate climate data were reduced to four principal components which between them explain more than 90% of the variation in 22 climatic variables. These 4 main PCs were then used as environmental covariates in a genome-environment association scan, using BayPass 2.1 [1], to identify SNPs that are associated with adaptation to climate. Upto 250 SNPs associated with PC1 and 430 SNPs associated with PC2, were identified at the threshold of Bayesian Factor >10 (‘strong evidence’). PCs 1 and 2 largely describe year-round temperature and precipitation levels in warmer months respectively. Briefly, the data suggest that there has been adaptation to the wide range of climates across Europe, and that this adaptation involves a large set of genes with different gene ontologies, although some ontologies are over-represented among adaptation loci. Adaptation to climate appears to be complex and polygenic.

The next stage of the work is to formally test whether signatures of selection can be identified in the same genes, in an independent temporal dataset of great tits that have been sampled in the UK since the late 1800s. We have visited natural history museums around the UK and extracted DNA from toe pads of 119 birds, that were caught between 1868 and 2000. Most samples (around 2/3 of the total) have over 100ng of extracted DNA. We go into this work fully aware that DNA sequencing from low yield samples may not be ideal and are prepared to take that risk. The aim is to sequence the genomes of 100 birds at approximately 12x coverage. Twenty of the birds will be from our contemporary UK sample archive, as they have been typed on a high density SNP chip but not whole genome sequenced. The sequenced genomes will be assembled using the latest great tit genome, v1.1 (<https://www.ensembl.org/Parus_major/Info/Annotation>) as a reference.

We will test the following predictions:

1. Signatures of selection in the contemporary samples will show that genes identified in the previous Genome-Environment analyses are more likely to show signatures of selection than other parts of the genome.
2. SNPs in these genes will show greater changes in allele frequency over the last 150 years, compared to the rest of the genome i.e. exceed expectations from genetic drift. Note that haplotypes showing signatures of positive selection in Obj 1, are expected to be increasing in frequency in the time series.

It is worth noting that when the museum samples were collected, we measured bill lengths as well. This builds on our previous research showing the microevolution of longer bills in the UK population of great tits relative to European populations [2]. Thus, the dataset we will collect has added value beyond the immediate project.

Following data QC, transfer and storage by NEOF-Liverpool, we will conduct the downstream analyses. We will use appropriate sequence assembly tools (e.g. Bowtie), and then perform variant calling using GATK. Signatures of selection in contemporary samples will be performed using tests that examine the site frequency spectra (e.g. Tajima’s D and related statistics). Temporal trends in allele frequencies will be calculated by linear regression of allele count (binned by decade and weighted by sample size per bin) on time. The distribution of these regression coefficients will be compared between candidate regions and the rest of the genome; the rest of the genome can be considered a null distribution of what might be expected by chance under a scenario of drift with no adaptive evolution.

1. Gautier, M., *Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates.* Genetics, 2015. **201**(4): p. 1555-1579.

2. Bosse, M., et al., *Recent natural selection causes adaptive evolution of an avian polygenic trait.* Science, 2017. **358**(6361): p. 365--368.