PROJECT TITLE: Divide and conquer strategies for population genomics
DTP Research Theme: Living World
Lead Institution: University of Bristol
Lead Supervisor: Prof Mark Beaumont, University of Bristol, School of Biological Sciences
Co-Supervisor: Dr Kari Heine, University of Bath, Dept of Mathematical Sciences
Co-Supervisor: Dr Anthony Lee, University of Bristol, School of Mathematics
Co-Supervisor: Prof Christophe Andrieu, University of Bristol, School of Mathematics
Project Enquiries: m.beaumont@bristol.ac.uk
Project keywords: Population genomics, demographic history, Monte Carlo, Machine Learning

An example genome alignment of a sample of individuals based on 100kb from Scaffold 0 of the cichlid *Astatotilapia calliptera*. Reference/alternate alleles at SNPs coloured red/white.

We have access to whole-genome data from many of the subspecies of wildcat, including domestic cats. Hypotheses to test include the timing of domestication, nature of hybridisation between wildcats and domestics, and, more generally, hypotheses concerning the global demographic history of *F. sylvestris* and natural selection associated with domestication.

Project Background
DNA sequences contain a wealth of information about the history of a population. Typical genomes contain about $3 \times 10^9$ bases, of which around $10^7$ are variable in typical samples, reflecting 10 million mutations in the ancestry of the sample. The patterns of these mutations – the number of copies in a sample, their geographic distribution, and their relationships to each other – can tell us a great deal about changes in past population size, the structure of populations and migration rates between subpopulations, as well as the effects of mutation, recombination, and natural selection. Although many complex models of the evolution of DNA sequences in populations have been proposed, it is still very challenging to efficiently estimate the parameters of these models, and to compare models and test hypotheses.

Project Aims and Methods
The project aims to adapt for population genomics, and to validate, recent methods that have been developed for inferring parameters in complex models (so-called intractable likelihood problems). A particular focus is to apply ‘divide and conquer’ strategies to population genomic data. In this case the genome is divided into manageable subsets, for which Bayesian inference is applied in parallel, and then the resulting inferences are combined together. In MB’s research group a methodological approach that appears to work well with genomic data is the use of expectation propagation (EP), which can be broadly applied for a variety of methods that are used to solve intractable likelihood problems. One method for intractable likelihoods that is often used in genomics, recently in conjunction with machine-learning techniques, is approximate Bayesian computation (ABC). Additionally, full-likelihood methods have been developed for genome data (Heine et al, 2018). The student will have full access to scripts and programs that have been developed for these problems, and will adapt them and validate them for population genomic analysis. However, depending on the interests
of the student, we would be happy for additional methodological development to be explored also. There are a number of genomic data sets that are being obtained in MB’s group in Bristol – in particular whole-genome sequences for wildcats (*Felis sylvestris*), and also, through collaborations, whole-genome data for a number of Lepidoptera species. These data are the subject of a number of hypotheses of interest - for example times of divergence between populations, changes in past population size, and evolutionary responses to environmental change.

**Candidate Requirements**

Candidates should have an interest in computational biology, and be able to demonstrate experience in previous projects that require some level of programming and/or scripting. Suitable candidates from all STEM backgrounds will be considered. It is important to have an interest in biological problems, but no formal training in biology or genomics is required.

**Training**

The candidate will obtain training in population genomics and evolutionary biology, and advanced computational statistics – exact-approximate methods for intractable likelihoods, ABC, machine learning methodology. Training will also be provided in use of High Performance Computing (HPC) techniques, via Bristol’s BlueCrystal supercomputer. There will be opportunities to be involved in Biology undergraduate teaching via practical demonstration and tutorials. Students would be encouraged to participate in overseas training courses in bioinformatics and genomics.

**References / Background reading list**


**Links:**

School URL: [http://www.bristol.ac.uk/biology/courses/postgraduate/](http://www.bristol.ac.uk/biology/courses/postgraduate/)

NERC GW4+ DTP Website: [http://nercgw4plus.ac.uk/](http://nercgw4plus.ac.uk/)

Bristol NERC GW4+ DTP Prospectus: [http://www.bristol.ac.uk/study/postgraduate/2020/doctoral/phd-great-western-four-dtp/](http://www.bristol.ac.uk/study/postgraduate/2020/doctoral/phd-great-western-four-dtp/)

How to apply to the University of Bristol: [http://www.bristol.ac.uk/study/postgraduate/apply/](http://www.bristol.ac.uk/study/postgraduate/apply/)

The application deadline is 1600 hours GMT Monday 6 January 2020 and interviews will take place between 10 and 21 February 2020

**General Enquiries:**

Bristol NERC GW4+ DTP Administrator

Email: bristol-nercgw4plusdtp-admin@bristol.ac.uk