Paleo-population genomics of mosquitoes: using museum specimens to investigate the past century of evolutionary responses to vector control in *Anopheles* mosquitoes.

Supervisors: Mara Lawniczak (WTSI) & Paul Kersey (EBI)

In the past 15 years, insecticide treated bednets have prevented an estimated 450 million cases of malaria in sub-Saharan Africa, but insecticide resistance is a serious threat to this progress [1]. Insight into population structure and a deeper understanding of the genomic responses of vectors to our efforts to control them will help us navigate the inevitable evolution of resistance. As part of the Anopheles 1000 genomes project [2], we have sequenced the whole genomes of 765 contemporary wild mosquitoes from 8 countries spanning Africa. We discovered an astonishing reservoir of genetic diversity, with a polymorphism every other nucleotide on average. We also discovered extreme losses of genetic diversity in genomic regions containing both known and novel insecticide resistance loci. While species and geographical boundaries exist, they are permeable when selection pressure is strong enough. One population in the 1000 genomes project has evidence of a massive bottleneck, with long stretches of complete homozygosity suggesting considerable levels of inbreeding. Whether this population crash was driven by a scale-up in bednet distribution is unclear. However, if we were able to examine levels of diversity in historic populations from this area, this might inform us of the timing of the crash.

This project aims to understand how mosquitoes from hundred-to-thousands of generations ago fit into our current understanding of population structure, and whether genomic changes over time are correlated with the historic scale up in vector control programs.

Scientific Aims

1. To develop DNA extraction and library-prep methods for pinned museum specimens using minimal levels of sample destruction (e.g. a leg rather than a whole body), aiming to generate genome sequences of ancient mosquitoes, as well as any ancient malaria parasites internal to the mosquito specimens.
2. To specifically explore *An. coluzzii, An. gambiae*, and *An. arabiensis* sympatric populations and historic evidence of gene flow prior to widespread insecticide use.
3. To develop an analytical framework to understand how genomes change over time in several locations where we will have 10+ individuals sampled every 10-30 years for the past 100 years.
4. To incorporate the history of human-led interventions into this understanding, researching historical patterns of insecticide and land use in the regions from which specimens originate.

In 1939, Paul Hermann Müller discovered DDT was a potent insecticide, a finding that won him the Nobel Prize [3]. DDT came into wide use in Africa in the 1950s. Pyrethrum was first used in Kenya in the late 1930s, and now pyrethroid-treated bednets are widespread and key in the fight against malaria. These insecticides target the voltage gated sodium channel (VGSC) in many insects, including *Anopheles*. *An. coluzzii* and *An. gambiae*, formerly known as the M and

S forms, are closely related sister species [4]. It is clear that hybridization between these species has facilitated the spread of insecticide resistance mutations at the VGSC gene. In fact, we have discovered 9 independent selective sweeps at this locus, some of which have crossed species boundaries. It is of key interest to ask whether hybridization was evident in the genomes of these species prior to the pressure from insecticides, or whether our efforts to control mosquitoes can cause increased frequencies of previously rare occurrences. Data generated in this project will make a historical view of diversity at this single gene possible, but more exciting will be to extend questions of this nature to the entire genome. Looking into the past through sequencing museum specimens will enable us to understand what the genomes and population structure looked like before our vector control efforts were scaled up through the widespread use of insecticides. This project will also improve our understanding of contemporary reference and population genome sequences.

The fellow will benefit from spending half of their time in each institute. At the Sanger, sequencing support is key to this project, as well as access to other team members working on the analysis of modern-day mosquito genomes. At the EBI, the fellow will benefit the interactions with the VectorBase [6] group. VectorBase houses genome-centric databases that enable and empower mosquito genetic research, and it has a suite of pipelines for comparative analysis and functional annotation available that will be used in the project, including software for inferring the sequence of ancestral genomes. Both teams (Kersey and Lawniczak) will benefit from the bridge created by this project, facilitating further interactions between groups on campus working on these important insects.

1. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* (2015). doi:10.1038/nature15535
2. Ag1000G: Anopheles gambiae 1000 Genomes | www.malariagen.net. at <http://www.malariagen.net/projects/vector/ag1000g>
3. Wikipedia contributors. DDT. *Wikipedia, The Free Encyclopedia* (2016). at <https://en.wikipedia.org/w/index.php?title=DDT&oldid=709512797>
4. Lehmann, T. & Diabate, A. The molecular forms of Anopheles gambiae: a phenotypic perspective. *Infect. Genet. Evol.* **8,** 737–746 (2008).
5. Dao, A. *et al.* Signatures of aestivation and migration in Sahelian malaria mosquito populations. *Nature* (2014). doi:10.1038/nature13987
6. Giraldo-Calderón, G. I. *et al.* VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.* **43,** D707–13 (2015).