

# Genomic epidemiology of malaria vectors in the *Anopheles gambiae* species complex

Alistair Miles

25th November 2020

## Abstract

In this thesis I ask, how can the study of genome variation within malaria vector populations contribute to the control of malaria in sub-Saharan Africa.

In the **first chapter** I provide an introduction to the current situation in malaria control in sub-Saharan Africa, and the role played by large scale mosquito control programmes using insecticide-based interventions. I also introduce high-throughput whole-genome sequencing and its potential applications to the study and surveillance of malaria vectors.

In the **second chapter** I introduce the *Anopheles gambiae* species complex, and provide historical context by describing how the species complex was discovered, which marked the introduction of genetic methods into the study and surveillance of African malaria vectors. I conclude that there are important parallels between past and present efforts towards malaria elimination, but also new opportunities afforded by genomic epidemiology.

In the **third chapter** I describe the production of a genome variation data resource derived from whole-genome sequencing of *Anopheles gambiae* and *Anopheles coluzzii* mosquitoes from 8 African countries, carried out as part of the first phase of the *Anopheles gambiae* 1000 Genomes (Ag1000G) Project. This chapter establishes and validates methods for robust discovery of nucleotide variation from Illumina deep whole-genome sequencing of individual mosquitoes, and confirms that *Anopheles* mosquitoes are among the most genetically diverse organisms in the natural world. Subsequent chapters all perform analyses using this data resource.

In the **fourth chapter** I identify genetically distinct populations among the mosquitoes sampled in Ag1000G phase 1, and quantify genetic diversity within and differentiation between these population. I show that there is strong population structure and marked differences in diversity between populations, suggesting important heterogeneities in population size and rates of gene flow. These results are an essential foundation on which to build analyses of recent evolution in subsequent chapters.

In the **fifth chapter** I search for signals of recent positive selection among the populations sampled in Ag1000G phase 1, to identify which genes are most important

in generating an adaptive response to the use of insecticides in malaria vector control. I show that there are strong signals of recent selection both at known insecticide resistance genes and at previously unknown genes with a plausible link to insecticide resistance.

In the **sixth chapter** I perform a detailed analysis of the voltage-gated sodium channel gene, where genetic changes cause target-site resistance to pyrethroid insecticides, the main ingredient in insecticide-treated bednets. I identify previously unknown mutations within this gene, and use haplotype data to show that resistance mutations have spread over large geographical distances and between mosquito species.

In the **final chapter**, I discuss the potential applications of genome sequencing for operational malaria vector surveillance and insecticide resistance management. I do this by identifying key use cases for genetic data in routine entomological surveillance, and by exploring the analogy between insecticide resistance outbreaks and outbreaks of infectious diseases.