Genomic approaches to the evolutionary dynamics of malaria drug resistance

Caroline Buckee, Caroline Colijn, Christophe Fraser and Dominic Kwiatkowski

With MalariaGEN partners around the world (www.malariagen.net) we have generated a large open dataset of genome variation data on the malaria parasite *Plasmodium falciparum*. Over the next five years we will generate longitudinal genomic data on tens of thousands of samples collected across a range of spatial scales and a variety of epidemiological settings, along with data on malaria control interventions and key epidemiological variables at each sampling location. Our data analysis pipeline produces high quality genotype calls on approximately 3 million polymorphisms including SNPs, indels and copy number variations, thus enabling detailed analyses of population structure, drug resistance, gene flow, migration and natural selection at both the local and the global level.

This postdoctoral project will develop analytical approaches to utilise longitudinal genomic data to gain novel insights into the evolutionary, epidemiological and demographic processes that determine the emergence and spread of antimalarial drug resistance. This could include the development of new methods for integrating genetic data into epidemiological models, and for combining genetic data with mobile phone data to estimate patterns of human migration (Wesolowski et al., 2018). It could also include new methods for inference of evolutionary processes from longitudinal genomic data, such as phylodynamic analysis of specific drug resistance loci as well as computational approaches for constructing genealogies at all loci across the genome (Speidel et al., 2019, Kelleher et al., 2019). There are many different ways to approach this problem and it offers a great opportunity to explore this very rich biological dataset while also building new analytical and computational tools that will drive the field forward.

A striking feature of antimalarial drug resistance is that it evolves in several stages, typically starting with a soft selective sweep of multiple independent origins of resistance. Some resistant lineages rise to a much higher frequency and spread over a much wider geographic range than others and, as they expand, they can acquire additional mutations and recombinant forms that increase their biological fitness. This large longitudinal dataset will provide the opportunity to address fundamental evolutionary hypotheses, such as that drug resistance mutations often carry a fitness cost that is mitigated by compensatory mutations, and that major increases in fitness advantage can result from epistatic interactions with other loci. Other important questions include the role of within host competition versus transmission fitness in selecting for dominant lineages, and the role of local dispersal versus long-range human migration in spreading drug resistance? By combining human migration data with longitudinal genomic data, collected at high spatiotemporal resolution, we aim to map transmission hotspots in areas of heterogeneous transmission intensity, and to identify routes of human migration that are critical for controlling malaria and halting the spread of drug resistance.