The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors *Anopheles gambiae* and *Anopheles coluzzii*

Chris S. Clarkson, Alistair Miles, …, and the *Anopheles gambiae* 1000 Genomes Consortium

**Abstract**

Resistance to pyrethroid insecticides is a major concern for malaria vector control, because these are the only compounds currently approved for use in insecticide treated bed-nets (ITNs) and are also widely used for indoor residual spraying (IRS). Pyrethroids target the voltage-gated sodium channel (VGSC), an essential component of the mosquito nervous system, but mutations in the *Vgsc* gene can disrupt the activity of these insecticides, inducing a “knock-down resistance” (*kdr*) phenotype. Here we use Illumina whole genome sequence data from phase 1 of the *Anopheles gambiae* 1000 Genomes Project (Ag1000G) to provide a comprehensive account of genetic variation at the *Vgsc* locus in mosquito populations from 8 African countries. In addition to three known variants that alter the protein-coding sequence of the gene, we describe 19 previously unknown non-synonymous variants at appreciable frequency in one or more populations. We report a highly significant enrichment for non-synonymous variation occurring in linkage with the known L995F resistance allele, indicating the evolution of multiple secondary variants that may either enhance or compensate for the L995F phenotype. We also describe a possible resistance variant I1527T, which occurs in linkage with two non-synonymous variants in codon 402. We use an analysis of haplotype sharing on the flanks of the gene to refine our understanding of the origins and geographical spread of resistance variants within the gene. We characterise 11 distinct lineages, each of which carries one or more resistance alleles and appears to be undergoing a rapid and recent expansion in one or more populations. We provide preliminary evidence that the most successful and widespread resistance lineage (F1) originates in West Africa and has subsequently spread to countries in Central and Southern Africa. We also reconstruct a putative ancestral haplotype for each lineage, and analyse patterns of recombination to show that lineages are unrelated and thus represent independent outbreaks of resistance. Our data demonstrate that the molecular basis of pyrethroid resistance in African malaria vectors is more complex than previously appreciated, and provide a foundation for the development of new genetic tools to predict resistance phenotype and track the further spread of resistance.

# Introduction

An estimated 663 million cases of malaria were averted in Africa between 2000 and 2015 due to public health interventions, of which 68% were prevented by insecticide-treated bed-nets (ITNs) and @@N% through indoor residual spraying (IRS) of insecticides @@REF. However, over this same period, insecticide resistance has become increasingly prevalent in malaria vector populations @@REF. Four chemical classes of insecticides – organophosphates, carbamates, pyrethroids and organochlorines – are licensed for use in public health, but only pyrethroids are approved by the World Health Organisation (WHO) for use in ITNs. Pyrethroids are also commonly used for IRS and in agriculture, and mosquito populations are under pressure to evolve molecular mechanisms of pyrethroid resistance. There is evidence that pyrethroid resistance has a direct impact on the effectiveness of ITNs (@@REF) and IRS (@@REF), although assessing the impact is difficult and hampered by the fact that pyrethroid resistance is now so pervasive that it is nearly impossible to find fully susceptible mosquito populations to use as controls (@@REF). Nevertheless, the position of the WHO remains that insecticide resistance poses a grave threat to the future of malaria control in Africa (@@REF GPIRM). Improvements are needed in our ability to monitor resistance, and gaps must be filled in our knowledge of the molecular mechanisms of resistance.

The voltage-gated sodium channel (VGSC) is the physiological target of pyrethroids, and of the organochlorine DDT. The VGSC protein is integral to the insect nervous system, involved in the transmission of nerve impulses. Both pyrethroids and DDT have a similar mode of action, binding to sites within the protein channel and preventing normal nerve function, causing paralysis (“knock-down”) and then death (@@REFs). However, amino acid substitutions at key positions within the channel can alter the interaction between the channel and the insecticide molecule, and thereby substantially increase the dosage of insecticide required for knock-down (@@REFs). If this tolerance exceeds the dosage present in ITNs or on indoor surfaces following IRS, these interventions may be rendered ineffective. In the African malaria vectors *Anopheles gambiae* and *Anopheles coluzzii*, three substitutions have been found in natural populations and shown to cause pyrethroid and DDT resistance. Two of these substitutions occur in codon 995[[1]](#footnote-1), with the Leucine→Phenylalanine (L995F) substitution prevalent in West and Central Africa, and the Leucine→Serine (L995S) substitution found in Central and East Africa (@@REFs). A third variant N1570Y has been found in association with L995F in West and Central Africa and shown to increase resistance above L995F alone (@@REF).

Target-site resistance to pyrethroids and DDT has also been studied in other insect species that experience insecticide pressure, including disease vectors as well as domestic and crop pests (@@REF). Because of its essential function, the VGSC protein is highly conserved across these species, and knowledge gained from one species is relevant to another. Many resistance-associated variants have been described in these other species, and thus there are many possible amino acid substitutions that could induce a resistance phenotype in malaria vectors other than the known variants in codons 995 and 1570. Some of these variants are within the channel and thus may directly interfere with insecticide binding, however others are in regions of the protein that are internal to the cell and thus cannot directly interact with insecticide molecules. In the absence of any insecticide pressure, it has been shown that substitutions within the channel cause a reduction in fitness, and therefore substitutions within internal domains may provide compensatory changes to the gating dynamics of the channel (@@REF). Most previous studies of *An. gambiae* and/or *An. coluzzii* have performed targeted sequencing of small regions within the gene, and there has been no comprehensive survey of variation across the entire gene in multiple populations.

Insecticide resistance monitoring in malaria vector populations now regularly incorporates some form of genetic assay to detect the allele present at *Vgsc* codon 995. Both alleles are present at high frequency in multiple geographical locations, and the L995F allele is also present in both *An. gambiae* and *An. coluzzii*. Although the extent of mosquito migration remains an open question, mosquitoes do travel between different locations and have the potential to spread resistance alleles from one population to another (adaptive gene flow). Hybridization between mosquito species also occurs and has the potential to transfer resistance alleles between species (adaptive introgression). Studies in West Africa have shown that the L995F allele has been introduced from *An. gambiae* into *An. coluzzii* populations (@@REFs). A resistance allele may also arise independently in multiple populations, either because of multiple mutational events occurring after insecticides are introduced (selection on new mutations), or because resistance alleles were already present at low frequency in mosquito populations prior to the introduction of insecticides (selection on standing variation). Previous studies have found evidence that the L995F allele occurs on multiple genetic backgrounds, suggesting multiple origins of resistance (@@REF). However, these studies have used information from only a small region of the gene, and have limited resolution to make inferences about geographical origins or history of spread of resistance alleles. Better information about the origins and spread of resistance could improve resistance monitoring and inform strategies for insecticide resistance management.

Here we elaborate on analyses of genetic variation within the *Vgsc* gene that were carried out as part of phase 1 of the *Anopheles gambiae* 1000 Genomes Project (@@REF). We draw on genotype and haplotype data derived from whole genome Illumina sequencing of 765 individual mosquitoes from 8 African countries to survey molecular diversity and study the evolutionary and demographic history of insecticide resistance at the *Vgsc* locus. These results provide a new foundation for functional studies to unravel the molecular basis of resistance, and a resource for the development of improved molecular diagnostics for monitoring insecticide resistance. @@TODO something needed here to round off.

# Results

## Functional variation

We used the Ag1000G phase 1 data resource to identify single nucleotide polymorphisms (SNPs) that alter the amino acid sequence of the VGSC protein, and computed their allele frequencies among 9 populations defined by species and country of origin. SNPs that play a functional role in insecticide resistance are expected to increase in frequency under selective pressure, and we refined the list of functional SNPs to retain those at an appreciable frequency (>5%) in one or more populations (Table 1). As described in @@REF, the two resistance alleles in codon 995 are clearly the main drivers of resistance at this locus, with the L995F allele at high frequency in populations of both species from West, Central and Southern Africa, and the L995S allele at high frequency among *An. gambiae* populations from Central and East Africa. Both alleles were present in populations sampled from Cameroon and Gabon, including some individuals with a hybrid L995S/F genotype. Within these populations, the two codon 995 alleles were (@@TODO were not?) in Hardy-Weinberg equilibrium (P>@@), thus there does not appear to be any selection against hybrids.

In addition to the known polymorphisms in codons 995 and 1570, we discovered a further 17 non-synonymous SNPs. We reported 15 of those SNPs in @@REF, and we extend the analyses here to incorporate data on two multiallelic SNPs affecting codons 402 and 410. There is a clear pattern of linkage disequilibrium among these SNPs (Figure 1), and we use this information in combination with previously reported evidence from this and other datasets to provide a preliminary classification for the functional role each SNP plays in pyrethroid resistance. We classify 13 of these SNPs as putative secondary resistance variants enhancing the L995F phenotype, because they occur in high LD (D’>0.92) with the L995F allele. This set includes the N1570Y allele which is known to enhance resistance in combination with L995F, and two SNPs affecting codon 1874 which has been associated with pyrethroid resistance in the diamond-back moth (@@REF). We classify I1527T as a putative primary resistance allele, because it does not occur in linkage with either L995F or L995S, but there is evidence it has been positively selected in the Burkina Faso *An. coluzzii* population (@@REF). We then classify both alleles of the multiallelic SNP affecting codon 402 as putative secondary resistance alleles enhancing the I1527T phenotype, because these alleles occur in strong linkage with I1527T. We classify the two remaining SNPs affecting codons 1125 and 1254 as unknown, because there is no evidence for selection or linkage with primary resistance alleles, but we cannot rule out a resistance phenotype without further sampling or functional work.

, including two multiallelic SNPs affecting codons 402 and 490, and a pair of adjacent SNPs both affecting codon 1874

The I1527T allele appears to be an alternative primary resistance allele. was present at 14% frequency in *An. coluzzii* from Burkina Faso.

To investigate the evolutionary relationships between non-synonymous polymorphisms, we used phased haplotypes to compute the linkage disequilibrium coefficient (D’) between all pairs of SNPs (Figure 1). We also used haplotypes including synonymous and intronic SNPs within the *Vgsc* gene to construct median-joining haplotype networks (Figure 2).

These analyses reveal a clear pattern for how these SNPs are organised.

To investigate the role of the remaining 16 non-synonymous SNPs

@@TODO

# Discussion

@@TODO

# Methods

@@TODO

1. Codon numbering is given relative to transcript @@TODO as defined in the AgamP4.@@N gene annotations. A mapping of codon numbering from *An. gambiae* @@TODO transcript to *Musca domestica* @@TODO transcript is given in Table 1. [↑](#footnote-ref-1)