The evolution 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 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. Other variants are in regions of the protein that are internal to the cell and thus cannot directly interact with insecticide molecules. However, functional studies have demonstrated that substitutions within internal linker domains can substantially enhance the level of resistance when present in combination with channel mutations (@@REFs). 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

To identify single nucleotide polymorphisms (SNPs) with a potentially functional role in pyrethroid resistance, we extracted SNPs from the Ag1000G phase 1 data resource 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 confer resistance are expected to increase in frequency under selective pressure, and we refined the list of functional SNPs to retain only those at an appreciable frequency (>5%) in one or more populations (Table 1). The resulting list comprises 20 SNPs, including the known L995F, L995S and N1570Y variants, and a further 17 SNPs not previously described in these species. We reported 15 of these novel SNPs in our initial analysis of the Ag1000G phase 1 data (@@REF Ag1000G), and we extend the analysis here to incorporate two tri-allelic SNPs affecting codons 402 and 410.

The two 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 (Table 1; @@REF Ag1000G). Both alleles were present in populations sampled from Cameroon and Gabon, including some individuals with a hybrid L995S/F genotype. Within these populations, the L995S and L995F alleles were (@@TODO were not?) in Hardy-Weinberg equilibrium (P=@@), thus there does not appear to be any selection against hybrids.

The I1527T allele is present in *An. coluzzii* from Burkina Faso at 14% frequency, and there is evidence that haplotypes carrying this allele have been positively selected (@@REF Ag1000G). Codon 1527 occurs within a trans-membrane domain, immediately adjacent to a second predicted binding pocket for pyrethroid molecules, thus it is plausible that I527T could alter insecticide binding (@@REF). We also found that the two variant alleles affecting codon 402, both of which induce a V402L substitution, were in strong linkage with I1527T (D’>@@N), and almost all haplotypes carrying I1527T also carried a V402L substitution. The most parsimonious explanation for this pattern of linkage is that the I527T mutation occurred first, and mutations in codon 402 subsequently arose on this genetic background. Codon 402 also occurs within the channel, and the V402L substitution by itself has been shown experimentally to increase pyrethroid resistance in @@species and in Xenopus oocytes (@@REFs). However, because the V402L alleles appear secondary to I527T in our cohort, we classify I1527T as a putative resistance driver and V402L as a putative enhancer. We hypothesize that the I527T/V402L combination represents a pyrethroid resistance allele that arose natively within West African *An. coluzzii* populations; however, the L995F allele is at higher frequency (85%) in our Burkina Faso *An. coluzzii* population and is known to be increasing in frequency (@@REFs), therefore L995F may provide a stronger resistance phenotype and is replacing I527T/V402L in these populations.

Of the 16 other SNPs, 13 occur almost exclusively in combination with L995F (Figure 1; @@REF Ag1000G). These include the N1570Y substitution known to enhance pyrethroid resistance in *An. gambiae* in combination with L995F (@@REF). These also include two substitutions in codon 1874 (P1874S, P1874L). P1874S has previously been found in a colony of the crop pest *Plutoblah blah* with a pyrethroid resistance phenotype (@@REF), but has not been shown to confer pyrethroid resistance experimentally. 10 of these variants, including N1570Y and P1874S/L, occur within internal linker domains of the protein, and so fit the model of variants that may enhance the resistance phenotype by modulating channel gating behaviour (Table 1; @@REFs). The remaining 3 variants are found within trans-membrane domains, and so may enhance resistance by @@TODO how. Because of the tight linkage between these 13 SNPs and the L995F allele, we classify all as putative L995F enhancers, although experimental work is required to confirm a resistance phenotype.

The remaining 3 variants (M490I, A1125V and V1254I) do not occur in combination with any known resistance allele, and do not appear to be associated with haplotypes under selection. A possible exception is the M490I allele found at 18% frequency in the Kenyan population, although the fact that the Kenyan population has experienced a recent population crash makes it difficult to test for evidence of selection (@@REF Ag1000G). All 3 variants occur in internal linker domains, and so do not fit the model of a resistance driver, although experimental work is required to rule out a resistance phenotype.

# Sandbox

, 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)