



HAL
open science

Genetic variation at the Cyp6m2 putative insecticide resistance locus in *Anopheles gambiae* and *Anopheles coluzzii*

Martin G Wagah, Petra Korlević, Christopher Clarkson, Alistair Miles, Mara K N Lawniczak, Alex Makunin, Michael C. Fontaine

► **To cite this version:**

Martin G Wagah, Petra Korlević, Christopher Clarkson, Alistair Miles, Mara K N Lawniczak, et al.. Genetic variation at the Cyp6m2 putative insecticide resistance locus in *Anopheles gambiae* and *Anopheles coluzzii*. *Malaria Journal*, 2021, 20 (1), pp.234. 10.1186/s12936-021-03757-4 . hal-03861207

HAL Id: hal-03861207

<https://hal.science/hal-03861207>

Submitted on 19 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH

Open Access



Genetic variation at the *Cyp6m2* putative insecticide resistance locus in *Anopheles gambiae* and *Anopheles coluzzii*

Martin G. Wagah^{1*} , Petra Korlević^{1,2}, Christopher Clarkson¹, Alistair Miles³, The *Anopheles gambiae* 1000 Genomes Consortium¹, Mara K. N. Lawniczak¹ and Alex Makunin¹

Abstract

Background: The emergence of insecticide resistance is a major threat to malaria control programmes in Africa, with many different factors contributing to insecticide resistance in its vectors, *Anopheles* mosquitoes. *CYP6M2* has previously been recognized as an important candidate in cytochrome P450-mediated detoxification in *Anopheles*. As it has been implicated in resistance against pyrethroids, organochlorines and carbamates, its broad metabolic activity makes it a potential agent in insecticide cross-resistance. Currently, allelic variation within the *Cyp6m2* gene remains unknown.

Methods: Here, Illumina whole-genome sequence data from Phase 2 of the *Anopheles gambiae* 1000 Genomes Project (Ag1000G) was used to examine genetic variation in the *Cyp6m2* gene across 16 populations in 13 countries comprising *Anopheles gambiae* and *Anopheles coluzzii* mosquitoes. To identify whether these alleles show evidence of selection either through potentially modified enzymatic function or by being linked to variants that change the transcriptional profile of the gene, hierarchical clustering of haplotypes, linkage disequilibrium, median joining networks and extended haplotype homozygosity analyses were performed.

Results: Fifteen missense biallelic substitutions at high frequency (defined as > 5% frequency in one or more populations) are found, which fall into five distinct haplotype groups that carry the main high frequency variants: A13T, D65A, E328Q, Y347F, I359V and A468S. Despite consistent reports of *Cyp6m2* upregulation and metabolic activity in insecticide resistant Anophelines, no evidence of directional selection is found occurring on these variants or on the haplotype clusters in which they are found.

Conclusion: These results imply that emerging resistance associated with *Cyp6m2* is potentially driven by distant regulatory loci such as transcriptional factors rather than by its missense variants, or that other genes are playing a more significant role in conferring metabolic resistance.

Keywords: Mosquito, *Cyp6m2*, Metabolic resistance, Allelic variants, Selection

Background

Malaria remains a pernicious public health problem that plagues the African region, which has over 90% of the world's malaria cases and deaths [1]. Although concerted vector control interventions, such as long-lasting insecticidal nets (LLINs) and indoor residual spraying (IRS), have led to the attainment of key milestones, global

*Correspondence: mw21@sanger.ac.uk

¹ Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SD, UK
Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

progress has stagnated and case numbers are stable or on the rise in many countries in Africa [1–3]. This is due to multiple factors, including the emergence of insecticide resistance, which threaten the effectiveness of vector control interventions [4].

The mechanisms of insecticide resistance are classified into four main functional categories: target-site insensitivity, penetration resistance, behavioural resistance and metabolic sequestration and detoxification; the latter being the primary focus. Mutations that diminish the binding capacity of insecticides to their targets confer target-site resistance [5], those that lead to a reduction in the insecticide absorption rate confer penetration resistance [6], those that lead to behavioural modification to circumvent the lethal effects of insecticides confer behavioural resistance [7], and those that increase insecticide detoxification and elimination rates confer metabolic resistance [8]. These mechanisms of insecticide resistance may occur concurrently within a single population or even within a single mosquito [9, 10]. They have led to increasing resistance to all four common insecticide classes—pyrethroids, organochlorines, carbamates and organophosphates—in all major malaria vectors across Africa [11, 12].

Metabolic resistance occurs mainly through the elevated activity of large and functionally diverse multigene enzyme families: glutathione S-transferases (GSTs), carboxylesterases (COEs) and cytochrome P450 monooxygenases (P450s) [13, 14]. Metabolic resistance is considered as one of the most serious threats to mosquito control [13], because the only widely accepted occurrence of malaria vector control failure was attributed to the elevated expression of resistance-associated P450s in *Anopheles funestus* [15–17]. A comprehensive understanding of metabolic resistance is therefore vital, and it must involve disambiguating the roles that individual enzymes play and the genetic backgrounds that underlie their significance in vector populations.

The CYP6M2 enzyme exhibits complex insecticide metabolism associated with multiple binding modes for insecticides [18]. Its gene is located within a cluster of 14 Cyp6 P450 genes on chromosome 3R of *Anopheles gambiae* [19], and is among the 111 known P450 genes across the *An. gambiae* genome [20, 21]. In this genomic region, *Cyp6m2* is nested within a sub-cluster of P450s containing *Cyp6m3* and *Cyp6m4* which have also been associated with xenobiotic detoxification [22].

Cyp6m2 is notably one of the few specific P450s that have shown a consistent association with metabolic resistance [9]. Metabolic resistance is mainly assessed through transcriptional profiling of genes involved in xenobiotic detoxification. Transcriptomic experiments such as quantitative PCR and microarray assays have

established a link between *Cyp6m2* overexpression and the resistance phenotype in field populations of *An. gambiae*, *Anopheles coluzzii*, *Anopheles arabiensis* and *Anopheles sinensis*, irrespective of the presence of knock-down resistance (*kdr*) mutations such as L995F or L995S (corresponding to L1014F or L1014S in *Musca domestica*) in the voltage gated sodium channel (VGSC) [9, 23–25]. In DDT resistant *An. gambiae* in Ghana, *Cyp6m2* has been found to be overexpressed 3.2 to 5.2-fold in combination with the upregulation of additional P450s like *Cyp6z2* [22]. In DDT resistant *An. coluzzii* collected in Benin, *Cyp6m2* was also found to be overexpressed 1.2 to 4.6-fold in combination with *Gste2* from the *GST* gene family and in the presence of fixed *kdr* alleles in the *Vgsc* gene [26]. In Nigeria, the 2.4 to 2.7-fold upregulation of *Cyp6m2* was found to be associated with high levels of permethrin resistance [9] and *An. gambiae* that exhibited a strong resistance to bendiocarb in Côte d'Ivoire also had an elevated (up to eightfold) expression of the *Cyp6m2* gene [24]. In the same study, transgenic expression of *Cyp6m2* in *Drosophila melanogaster* was shown to produce resistance to both DDT and bendiocarb. In vivo functional analysis of multi-tissue overexpression induced by genetic modification has also shown *Cyp6m2* to be sufficient in conferring resistance to permethrin and deltamethrin [27]. However, this overexpression also increased the susceptibility of mosquitoes to the organophosphate malathion (a negative cross-resistance mechanism). Collectively, these studies indicate that *Cyp6m2* can confer metabolic resistance against insecticides in 3 of the 4 known classes: both type I and type II pyrethroids [18, 22, 27], organochlorines [28], and carbamates [24]. It has, therefore, a high potential for cross-resistance, which may make the problem of malaria vector control even more intractable by limiting the options available to malaria control programmes for insecticide rotation or combination. The negative cross-resistance associated with malathion hereby points to potential mitigating strategies [27].

The frequent association of *Cyp6m2* with insecticide resistance described above warrants further investigation into whether there is evidence of copy number variation (CNV) or missense mutations at the locus. CNVs have been implicated in augmenting gene dosage leading to increased transcription of metabolic enzymes [29, 30]. A genome-wide CNV analysis conducted on the Ag1000G dataset and described in detail elsewhere [29] found CNVs to be significantly enriched in metabolic resistance-associated gene families and to be undergoing positive selection. These CNVs were identified across P450s (such as *Cyp9k1* and at both the *Cyp6z3–Cyp6z1* and the *Cyp6aa1–Cyp6p2* gene clusters) and GSTs (at the *Gstu4–Gste3* cluster). However, CNVs across the

Cyp6m2 locus were found to be rare, even in populations that are known to exhibit *Cyp6m2*-mediated resistance [29]. This indicates that CNVs alone are not sufficient to explain the widespread occurrence of the *Cyp6m2*-associated resistance phenotype: additional factors such as allelic variation might contribute to resistance associated with *Cyp6m2* activity.

Allelic variation can play an additional role in P450-mediated resistance by modifying either enzyme catalytic activity or gene expression levels [31]. Allelic variation has been shown to be key in inducing high metabolic efficiency of *Cyp6P9b* and in conferring metabolic resistance to *An. funestus* [32]. Allelic variants in metabolic genes have also been identified to reliably and reproducibly associate with resistance, such as in *Cyp4J5* and *Coeae1d* in *An. gambiae*, and can serve as diagnostic markers of phenotypic resistance [33]. However, there is still a paucity of information about allelic variation associated with metabolic resistance when compared to the well-characterized target-site mutations [33]. Mutations that may modulate metabolic resistance by either altering function or modifying expression in *Cyp6m2* are yet to be described.

Following the consistent association of *Cyp6m2* with insecticide resistance in many populations, whole-genome Illumina sequence data was examined from phase 2 of the *Anopheles gambiae* 1000 Genomes Project (Ag1000G) [34], which consists of 1142 wild-caught mosquitoes sequenced to a mean depth above 14×, and report a comprehensive analysis of genetic variation within the *Cyp6m2* gene. The wider haplotypes around *Cyp6m2* spanning across the *Cyp6m* sub-cluster and the larger *Cyp6* supercluster are also examined for signatures of selection.

Methods

Data collection and analysis

This study followed the species nomenclature of Coetzee et al. [35], where *An. gambiae* refers to *An. gambiae* sensu stricto (*s.s.*) (S form) and *An. coluzzii* refers to *An. gambiae s.s.* (M form). A detailed description of the Ag1000G sample collection, DNA extraction, sequencing, variant calling, quality control and phasing can be found here [36]. Briefly, Anopheline samples were collected from 33 sampling sites across 16 populations in 13 countries in sub-Saharan Africa (Table 1 and Additional file 1). The sampling procedure covered different ecosystems and aimed at collecting a minimum of 30 specimens per country. The specimens consisted of *An. gambiae* and *An. coluzzii*: only *An. coluzzii* were sampled from Angola, both *An. gambiae* and *An. coluzzii* were sampled from Burkina Faso, while all other populations consisted of *An.*

gambiae, except Kenya, The Gambia and Guinea Bissau where the species identity was indeterminate.

Whole genome sequencing of all mosquitoes was performed on the Illumina HiSeq 2000 platform. The generated 100 base paired-end reads were aligned to the *An. gambiae* AgamP3 reference genome assembly [37] and variants were called using GATK UnifiedGenotyper. Samples with mean coverage $\leq 14\times$ and variants with attributes that correlated with Mendelian error in genetic crosses were removed during quality control.

The SnpEff v4.1b software was used for the functional annotation of Ag1000G variant data [38] using locations from geneset AgamP4.12. All variants in transcript AGAP008212-RA with a SnpEff annotation of “missense” were regarded as nonsynonymous variants. The *Cyp6m2* gene has not been shown to exhibit alternative splicing, and no alternative transcripts have been reported.

Haplotype clustering, linkage disequilibrium and mapping of haplotype clusters

To reveal the haplotype structure at *Cyp6m2*, *Cyp6m* sub-cluster, *Cyp6* supercluster, *HAM*, *ODR-2* and *SH2*, the Hamming distance is computed between all haplotype pairs and hierarchical clustering of haplotypes is performed. Arbitrary clustering threshold values were worked through to cut the dendrograms at genetic distances that would best highlight the most relevant clusters. Lewontin’s D' [39] was used to compute the linkage disequilibrium (LD) between all pairs of missense *Cyp6m2* mutations. Image rendering for the haplotype clustering, linkage disequilibrium and haplotype cluster frequencies map was performed using the matplotlib Python package version 3.1.2 [40]. Geography handling for the haplotype cluster frequencies map was done using cartopy version 0.17.0 [41].

Haplotype networks

Haplotype networks are constructed using the median-joining algorithm [42] implemented in Python [43]. Haplotypes carrying the main high frequency mutations were analysed with a maximum edge distance of two SNPs. The Graphviz library was used to render the networks and the composite figure was constructed in Inkscape [44].

Extended haplotype homozygosity

The core haplotype is defined on a 1689 base region spanning the *Cyp6m2*, from chromosome arm 3R, starting at position 6928858 and ending at position 6930547. This region was selected to ensure a 1:1 haplotype correspondence with that used in the hierarchical clustering analysis. Extended haplotype homozygosity (EHH) was computed across all core haplotypes in all populations

Table 1 Allele frequencies of non-synonymous variants in *Cyp6m2* coding region

Variants	Population allele frequency (%)																	
	Ag ^b	All	AOcol ^c	GHcol	BFcol	Clcol	GNcol	GW	GM	CMgam	GHgam	BFgam	GNgam	GAgam	UGgam	GQgam	FRgam	KE
6928945 G>A A13T	2.8	39.7	0	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6929047 G>C G47R	0.6	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6929050 T>A S48T	2.8	9	0.9	4	2.1	0	3.3	2.3	3.2	4.2	2.7	2.5	0	2.2	0	0	0	0
6929102 A>C D65A	6.0	0	0	0	0	0	0.5	0	1.9	4.2	2.7	3.8	42.8	2.7	0	0	0	52.1
6929375 A>T K156I	0.4	0	0	0	0	0	0	0	0	0	0	0	1.4	0	44.4	0	0	0
6929506 A>G N200D	0.4	0	0	0	0	0	0	0	0	0	0	0	1.4	0	44.4	0	0	0
6929770 T>A S288T	1.1	0	5.5	0.7	0.7	0	0	0.8	4.2	2.7	1.3	0.7	1.3	0	0	0	0	0
6929881 G>A E325K	1.2	0	7.3	3.3	2.1	0	3.8	2.3	0	0.5	0	0	0	0	0	0	0	0
6929890 G>C E328Q	2.9	0.6	13.6	7.3	7	12.5	9.9	6.2	0.3	0	0	0	0	0	0	0	0	0
6929948 A>T Y347F	2.4	0	0	0	0	0	0	0.2	0	0	0	0	0	1.8	0	0	0	52.1
6929983 A>G I359V	16.0	0	4.5	0	7	37.5	16.5	20	19.4	25	17.4	16.3	48.6	19.6	0	0	0	0
6930206 C>T P407L	0.7	0	2.7	0.7	7	0	0	0.2	0	0	0	0	0	0	0	0	0	0
6930242 A>T E419V	0.4	5.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6930269 A>G K428R	2.0	0	0	0.7	0	0	0.5	0.8	3.5	0	2.7	2.5	0.7	6.3	0	0	0	0
6930388 G>T A468S	2.6	0	0	0	0	0	0	0.3	0	42	0	0	0	0	0	0	0	0

^a Position relative to the AgamP4 reference sequence, chromosome 3R

^b Codon numbering according to *Anopheles gambiae* AGAP008212-RA transcript in geneset AgamP4.12

^c AOcol, Angola An. coluzzii; GHcol, Ghana An. coluzzii; BFcol, Burkina Faso coluzzii; Clcol, Côte d'Ivoire An. coluzzii; GNcol, Guinea An. coluzzii; GW, Guinea Bissau; GM, The Gambia; CMgam, Cameroon; GHgam, Ghana An. gambiae; BFgam, Burkina Faso An. gambiae; GAgam, Gabon An. gambiae; UGgam, Uganda An. gambiae; GOgam, Equatorial Guinea An. gambiae; FRgam, Mayotte An. gambiae; KE, Kenya

as described in Sabeti et al. [45] using scikit-allel version 1.1.9 [46]. EHH composite plots were made using the matplotlib Python package version 3.1.2 [40].

Results

Cyp6m2 non-synonymous nucleotide variation

Short-read whole-genome sequence data from the Ag1000G phase 2 data resource [47] were used to investigate genetic variation at the *Cyp6m2* locus across 16 populations of *An. gambiae* and *An. coluzzii* (n=1142 total individuals) collected between 2000 and 2012 (Table 1, Additional file 1). The single nucleotide polymorphisms (SNPs) studied here were discovered and QC'd using methods described elsewhere [36]. Focus is placed on SNPs that change the amino acid sequence of the *CYP6M2* enzyme as they have a potential to alter gene function in *Cyp6m2*-associated insecticide resistance (n=193) (Additional file 2). None of these reported mutations are synonymous or found within *cis* regulatory elements. As putative resistance variants under selection pressure from insecticides are expected to increase in frequency over time, allele frequencies were subsequently computed for every non-synonymous SNP in each population with reference to species and country of origin. The list is filtered to focus only on those variants that were at high frequency within populations or across populations (defined as >5% frequency in one or more populations). In total, this resulted in 15 non-synonymous variants that are explored further (Table 1).

Analysis of the patterns of polymorphism of *Cyp6m2* from different populations showed both relative homogeneity within some geographical regions and distinct variants across different regions. The variants with the highest overall frequency were **I359V** (16%) and **D65A** (6%) (Table 1). The most widespread variant was **I359V**, which was present in West, Central and East African populations of both *An. gambiae* and *An. coluzzii*. Populations with the highest frequency of **I359V** were Gabon (49%) and Ghana (25%) for *An. gambiae*, and Guinea (37.5%) for *An. coluzzii*. Another mutation, **E328Q**, was found across West Africa's *An. coluzzii* populations in Burkina Faso, Côte d'Ivoire, Ghana, Guinea and The Gambia and ranged in frequency from 6.2 to 13.6%. Several variants were found to exceed the 5% threshold only in one or two populations: **A13T** and **Y347F**, in Angola's *An. coluzzii* (39.7%) and in Kenya (52.1%) respectively and **D65A** only in Gabon's *An. gambiae* and in Kenya's populations at 42.8% and 52.1%, respectively (Table 1).

Haplotypic backgrounds of non-synonymous alleles

The Ag1000G data resource contains data that not only spans across exonic regions of any given gene, but also intronic and intergenic regions. This enables a

comprehensive analysis of haplotypes that contain putative insecticide resistance alleles, but is constrained by the fact that this resource does not contain samples whose resistance status or *Cyp6m2* expression levels are known.

Selection pressure acting upon missense variants or linked *cis* regulatory variants is likely to affect the haplotype structure of the gene. To study haplotype structure at *Cyp6m2*, biallelic SNPs across the entire 1689 bp *Cyp6m2* gene were extracted to calculate the number of SNP differences between all pairs of 2,284 haplotypes derived from the mosquitoes. A clustering threshold of seven SNPs was manually identified, where the haplotype clusters corresponded to the haplotypes carrying the high frequency alleles (Table 1, Fig. 1). These haplotypes could mostly be grouped into five distinct clusters (labelled C1–C5): C1 contained haplotypes carrying **A13T**; C2 contained most haplotypes carrying **D65A**, **A468S**, and some haplotypes carrying **I359V**; C3 contained most haplotypes carrying both **D65A** and **Y347F**, and C5 contained haplotypes carrying **E328Q**. C4 contained haplotypes with no signature missense mutation (Fig. 1).

Overall, haplotype cluster distribution resembled the whole genome groupings of individuals described elsewhere using the same dataset [34]: Cluster C5 contained haplotypes from West African *An. coluzzii*; C4 contained *An. gambiae* from West, Central and near-East Africa; and the rest of the clusters contained haplotypes from samples from a single country and species (Fig. 2). The variation across the haplotypes largely showed no strict or systematic difference between the two species or across broad geographic regions, which is in line with recent whole genome sequencing reports [36].

Patterns of association among these non-synonymous variants were investigated by computing the normalized coefficient of linkage disequilibrium (*D'*) using haplotypes from the Ag1000G phase 2 resource. Of the two highest frequency variants, **I359V** was found to be in perfect linkage with **A468S** but this was driven only by one population (Gabon) with most backgrounds carrying **I359V** not showing linkage with any other missense mutations (Fig. 1 and Additional file 3: Fig. S1). **D65A** was in perfect linkage with **A468S** and **Y347F**, showing that **D65A** was almost only ever found on haplotypes carrying either **A468S** or **Y347F**. **I359V** and **D65A**, the highest frequency mutations across all populations, were found to be only in moderate linkage disequilibrium (0.36) (Additional file 3: Fig. S1). Other variants were found to be in weak linkage disequilibrium with the six main high frequency alleles and segregated independently within their own populations. While some strong associations were observed through linkage disequilibrium analysis across

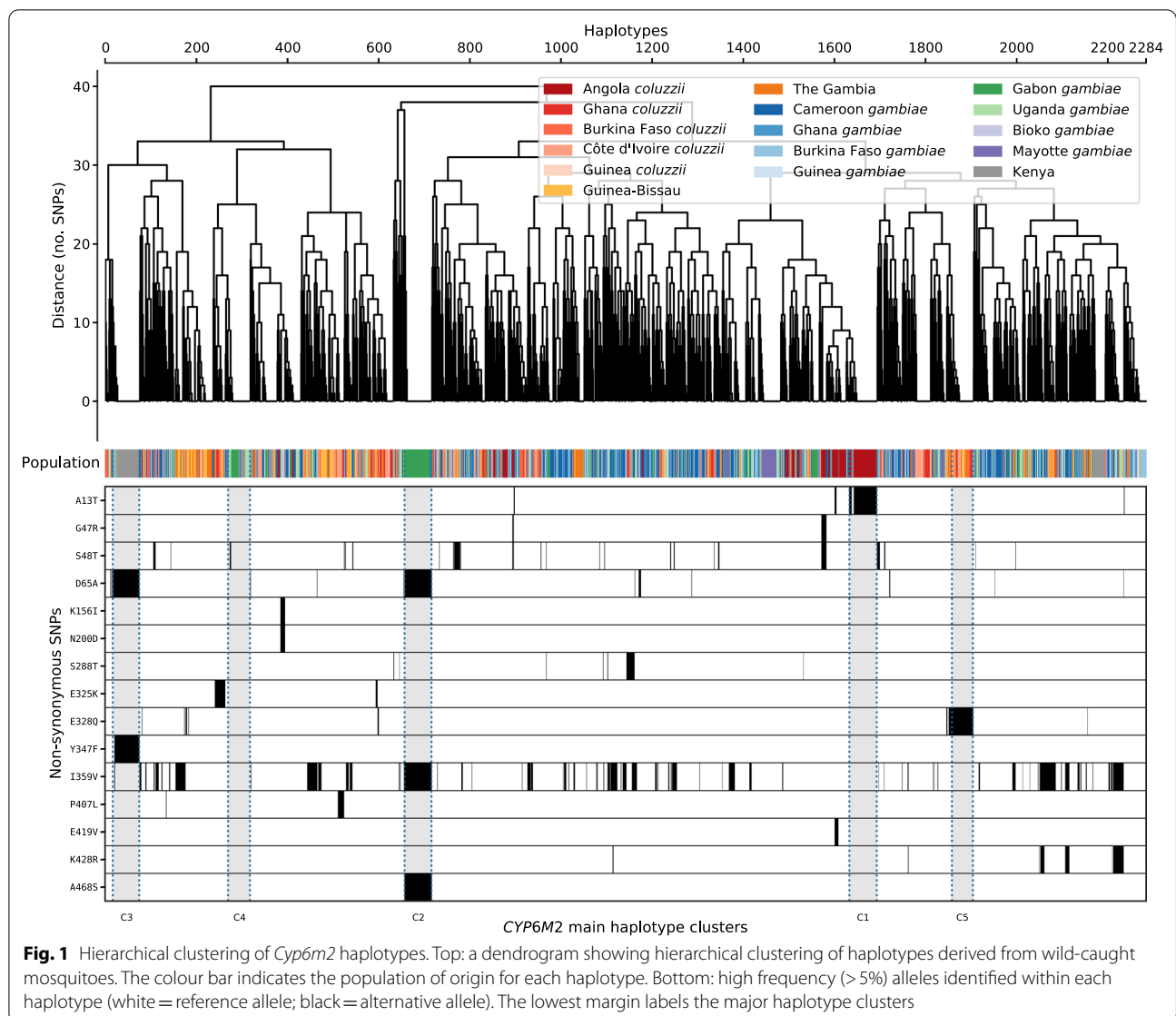
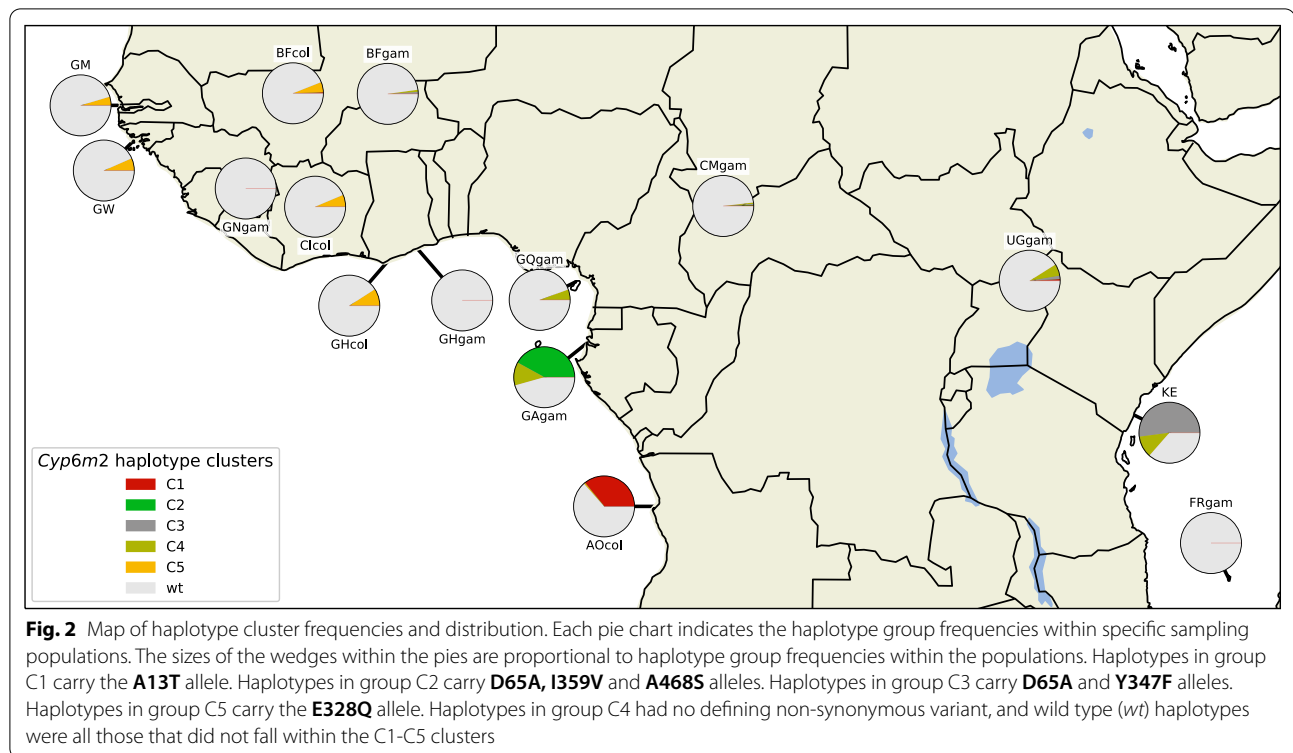


Fig. 1 Hierarchical clustering of *Cyp6m2* haplotypes. Top: a dendrogram showing hierarchical clustering of haplotypes derived from wild-caught mosquitoes. The colour bar indicates the population of origin for each haplotype. Bottom: high frequency (> 5%) alleles identified within each haplotype (white = reference allele; black = alternative allele). The lowest margin labels the major haplotype clusters

all populations, a deeper investigation revealed that these associations were driven by population specific dynamics in populations (such as Kenya) where bottlenecking has been an issue [36]. It is, therefore, unlikely that the identified variants are conferring some selective advantage against existing insecticide pressures.

The surrounding genomic region was explored to identify whether it showed a similar hierarchical clustering pattern to *Cyp6m2*, which might be indicative of either dominant demographic effects or selection acting at other linked loci that is having a major impact on variation within *Cyp6m2*. The downstream genes selected coded for proteins that were 1-to-1 orthologs with *Drosophila melanogaster* genes. Three genes were selected: *ODR2* [48], *HAM* [49] and *SH2* [50], which were 81280

bases, 457164 bases and 1198636 bases downstream of the *Cyp6m2* gene, respectively. The distinctive haplotype clustering pattern observed for *Cyp6m2* in the Kenya, Angola and Gabon populations persisted across these genes, indicating that in these populations, the diversity reduction in and downstream of *Cyp6m2* is more likely driven by demography rather than by a selective sweep (Additional files 4, 5, 6: Fig. S2–S4). Biallelic SNPs were also extracted across the *Cyp6m* sub cluster of 3 genes (*Cyp6m2*, *Cyp6m3* and *Cyp6m4*) and across the *Cyp6* supercluster of 14 genes within which the *Cyp6m* sub cluster is located (*Cyp6s2*, *Cyp6s1*, *Cyp6r1*, *Cyp6n2*, *Cyp6y2*, *Cyp6y1*, *Cyp6m1*, *Cyp6n1*, *Cyp6m2*, *Cyp6m3*, *Cyp6m4*, *Cyp6z3*, *Cyp6z2* and *Cyp6z1*), and performed hierarchical clustering across these regions as described



above. The typical geographical stratification of haplotypes persisted, suggesting the absence of a selective sweep across this region (Additional files 7, 8: Fig. S5 & S6).

The genetic backgrounds carrying these alleles were examined further by constructing median joining networks (MJNs) [42] using the Ag1000G Phase 2 haplotype data. This enabled the resolving of the radiation of DNA substitutions arising on haplotypes carrying the identified variants. It also allowed the reconstruction and positioning of intermediate haplotypes while revealing the non-hierarchical relationships between haplotypes that could not be resolved by hierarchical clustering alone. The MJNs were constructed with reference to a maximum edge distance of two SNPs. This ensured that the connected components captured only closely related haplotypes. The resulting MJNs had a close correspondence with the hierarchical clustering output in assignment of haplotypes to clusters (88% overall concordance across all clusters).

The median joining networks showed more clearly the distinctive demographic stratification of the high frequency variants that was highlighted by the hierarchical clustering networks (Fig. 3). Most nodes containing secondary variants arising from the main nodes were small, which is inconsistent with directional selection where larger nodes are expected. Only one of the **I359V** nodes

contained haplotypes from mosquitoes of both species, however the secondary nodes did not contain haplotypes from more than one species. This indicates that although **I359V** is shared by both *An. gambiae* and *An. coluzzii*, it is unlikely that this is because of an introgression event across the *Cyp6m2* gene.

Positive selection of non-synonymous alleles

Extended Haplotype Heterozygosity (EHH) decay [45] was calculated to explore evidence for directional selection on the haplotypes carrying high frequency non-synonymous variants. It is expected that the presence of ongoing or recent directional selection pressure would lead to the increase in frequency of haplotypes, which on average will have longer regions of haplotype homozygosity relative to haplotypes that are not under selection. This diversity reduction would produce signatures of selection that would be conspicuous across a large genomic region, including across nearby *cis*-regulatory elements. EHH analysis would therefore be able to detect diversity reduction caused by ongoing directional selection being driven either by amino acid substitutions identified within the gene or by mutations within *cis*-acting elements next to the gene that may be under selection.

To perform the EHH decay analysis, a core region of 1689 bases that spans across the entire gene was defined. This was identical to what was used to differentiate the

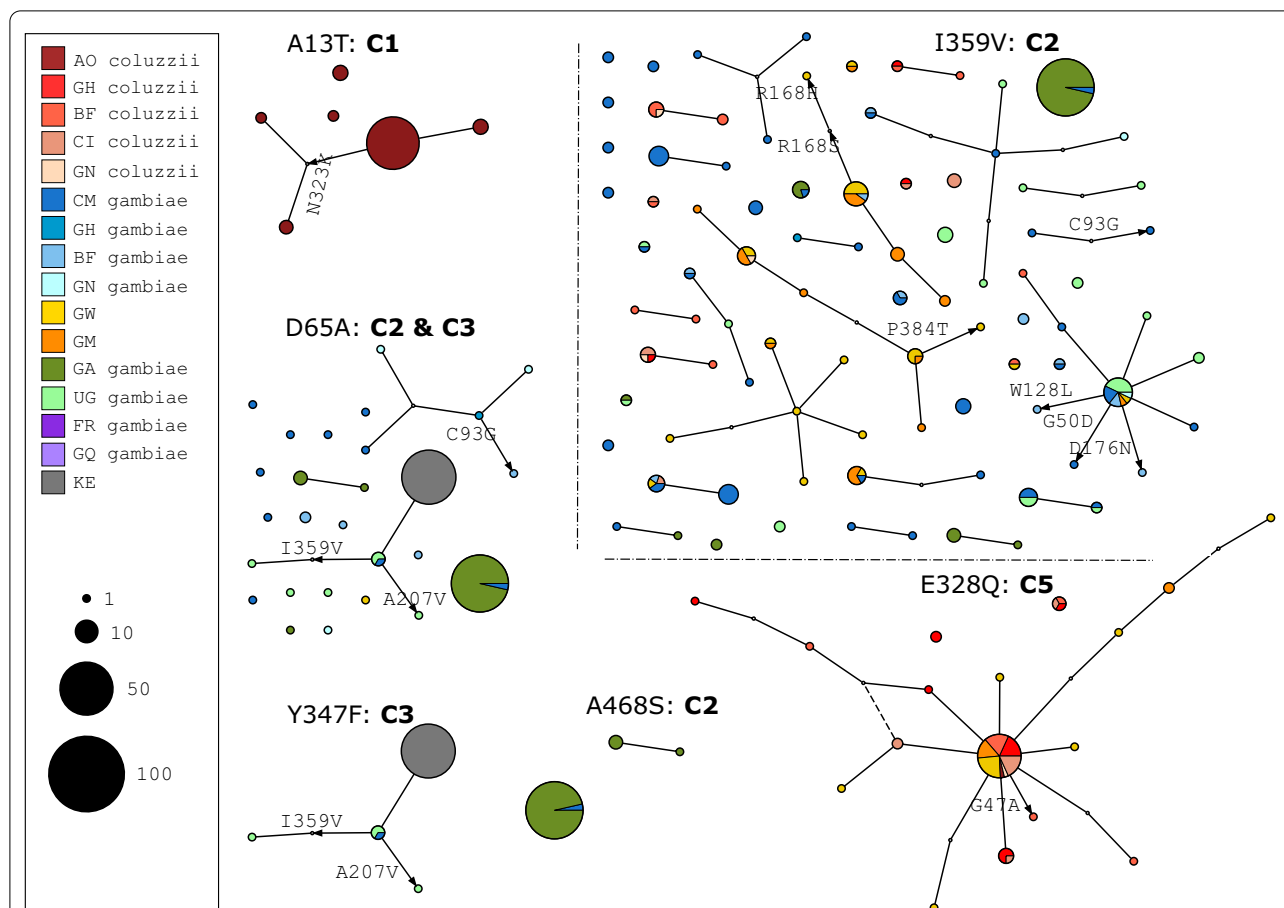


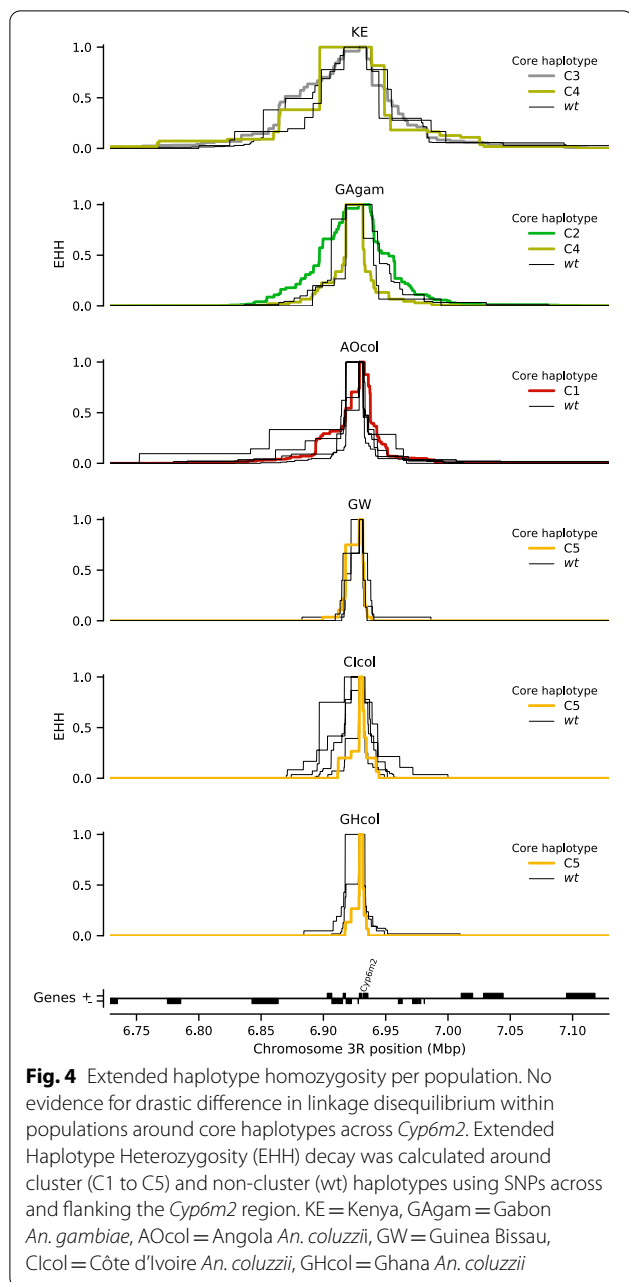
Fig. 3 Haplotype networks. Median joining network for haplotypes carrying **A13T**, **D65A**, **E328Q**, **Y347F**, **I359V** and **A468S**, with a maximum edge distance of two SNPs. Node size indicates haplotype counts and node colour indicates the population/species of haplotypes. AO = Angola; GH = ; BF = Burkina Faso; CI = Côte d'Ivoire; GN = Guinea; CM = Cameroon; GW = Guinea Bissau; GM = The Gambia; GA = Gabon; UG = Uganda; FR = Mayotte; GQ = Equatorial Guinea; KE = Kenya

identified haplotype groups though hierarchical clustering. This region contained multiple distinct haplotypes above 1% frequency within the cohort, including haplotypes corresponding to the C1–C5 haplotype clusters. All haplotypes that did not correspond to C1–C5 were considered to be wild type (wt). Although there were several different haplotypes in each population that fit this description, they are not distinguished in this paper and all these are called wild type, as *Cyp6m2* has no known resistance alleles and a true wild type remains to be discovered. EHH decay was then computed for each core haplotype up to 200 kilobases upstream and downstream (Additional file 9: Fig. S7): beyond 200 kb, the EHH had decayed to zero.

It is noted that haplotype clusters containing high frequency variants (C1–C5) did not exhibit a significantly slower EHH decay relative to the wild types, showing no evidence of positive selection. However, one Kenyan wild type haplotype group had a dramatically slower EHH

decay relative to wild type haplotypes from other populations. In order to account for this difference within wild type groups across multiple populations and to reveal potential signs of selection that would be obscured by a collective analysis across all populations, the haplotypes were separated by population and species and EHH decay was recomputed for each core haplotype as above.

Kenyan mosquito populations are known to have an extreme demographic history, as they have experienced a severe recent bottleneck, and the Angola and Gabon populations are known to be geographically unique populations which are strongly differentiated from all other populations [36]. Hence, their haplotypes exhibited a considerably slower decay than West African haplotypes (first three panels: Fig. 4). However, the putative resistance haplotypes C1–C5 did not experience a slower EHH decay relative to their wild type haplotypes, showing no evidence of positive selection acting upon those haplotypes in those populations.



As expected, the West African *An. coluzzii* haplotypes exhibited a much faster decay of EHH than specimens from Kenya, Angola, or Gabon, highlighting the demographic differences previously observed for these collections [36] (last three panels, Fig. 4). The C5 haplotype was a promising candidate for potential selection as it occurred within a more diverse population, and it was interesting to note that some wild type haplotypes in Côte d'Ivoire's *An. coluzzii* had a slightly slower decay than others within West Africa (fifth panel, Fig. 4). However, these haplotypes were not part of the C5 cluster,

and did not carry the widespread E328Q mutation. The C5 haplotype did not exhibit a dramatically slower decay of EHH than wild type haplotypes in the populations in which it was found, suggesting that it is not under positive selection.

Discussion

Cyp6m2 has been implicated in many *Anopheles* populations as a key P450 that contributes to the insecticide resistance phenotype [9, 18, 24, 28]. It has been reported that allelic variants across some P450s can affect enzyme conformational dynamics and substrate binding affinity [32], offering potential mechanisms that may modulate enzyme activity and efficiency, and thus account for additional *Cyp6m2* resistance where CNVs alone may not suffice. However, little is also known about *Cyp6m2* allelic variation across Africa.

In this study, a comprehensive account of the distribution of amino acid substitutions occurring within the *Cyp6m2* gene is reported. The haplotype structure of the gene is also examined to probe for selective sweeps by performing hierarchical clustering of haplotypes. The genetic background upon which the missense variants are found is also examined by plotting both median joining networks and decay of extended haplotype homozygosity, which are useful for revealing signatures of selection. It is noted that the distinct haplotype groups therein are stratified demographically and largely correspond to signature missense variants found in specific populations. This is in contrast to the strong signals of recent positive selection at other cytochrome P450 gene loci such as at *Cyp6p3* [36] which is often upregulated in tandem with *Cyp6m2* in multiple pyrethroid resistant populations [9, 51, 52].

It is still unclear how the identified non-synonymous variants may modulate *Cyp6m2* binding activity, in either the presence or absence of multiple competitive substrates and metabolites. The two aromatic residues (Phe 108 and Phe 121) that have been previously identified to be vital in deltamethrin orientation in the *Cyp6m2* active site [18] were not found to contain high frequency variants in our dataset. Candidate-gene based genome-wide association studies can be used to supplement this exploration of natural population variation in the identification of variants that reproducibly associate with resistance [33]. Computational techniques can also provide a quick and cheaper alternative to experimental approaches to explore and predict the effects of these identified variants on enzyme–substrate interactions [53]. However, subsequent in vivo functional analysis will be necessary to validate any causative links between variants and resistance phenotypes. This may involve transgenic expression of mutant enzymes, either in cell lines or in mosquito

lines, and the generation of enzyme activity profiles and mortality rates respectively, in the presence of various insecticides.

None of the haplotype groups identified that carried missense variants were found to be under directional selection. This is despite the existence of a widespread variant (E328Q) in both *An. gambiae* and *An. coluzzii* linked to a geographic region (West Africa) where *Cyp6m2* upregulation has been associated with emerging metabolic resistance [24, 51]. In *An. coluzzii* originating from both Côte d'Ivoire and Ghana, the C5 haplotype that carried E328Q was shown to have an even faster decay of EHH than the wild type haplotypes, further indicating an absence of directional selection. The stratification of other main haplotype clusters from Angola (C1), Gabon (C2) and Kenya (C3) was also consistent with the strong demographic differentiation and overall reduced heterozygosity of these populations described elsewhere [36].

While the genomic data quality across the *Cyp6m2* gene and its putative promoter region was satisfactory, there was a ~10,000 base region of inaccessibility upstream of *Cyp6m2* that cut across the intergenic region into *Cyp6n1* [54]. A similar inaccessible region was also present 1 kb downstream of the gene in the intergenic region between *Cyp6m2* and *Cyp6m3*, which is likely caused by the presence of repeats that inhibit read mapping. Although it is possible that the upstream region of inaccessibility could contain a regulatory variant that is susceptible to selection, it is unlikely to obscure signatures of selection.

It has been shown in multiple studies that target-site resistance (i.e. VGSC-*kdir*) provides a strong persistent baseline of resistance as it rises towards fixation within populations [55]. In the presence of insecticide selection pressure, target-site mutations and metabolic resistance have also been shown to act synergistically to confer a stronger resistance phenotype to pyrethroids [33, 56]. While signatures of selection have previously been identified in some metabolic gene clusters within populations that have a high *kdir* frequency [36], further studies need to examine whether directional selection occurring on one locus can obscure selection on another locus. To resolve this conundrum, genomic analysis must be performed on populations sampled across generations and whose transcriptomic and phenotypic characteristics are known, in order to tease out the individual contributions of specific sources of resistance.

Independent studies employing different experimental designs have also shown that metabolic resistance manifests as a cascade of multiple upregulated genes [57]. These genes, like *Cyp6m2*, are part of the normal cellular mechanism for xenobiotic detoxification that involves a

linked, coordinated response of large multi-gene enzyme families in complicated pathways. Therefore, it is likely that identifying signatures of selection due to insecticide pressure will involve thorough analysis across this vast network. The Cap 'n' Collar isoform-C (CncC) transcription factor sub-family has been shown to work in tandem with other transcription factors to regulate the transcription of phase I, II and III detoxification loci of multiple insects such as *Culex quinquefasciatus* and *D. melanogaster* [58, 59]. CncC knockdown or upregulation has been shown to directly affect phenotypic resistance in *An. gambiae* as well, modulating the expression of key P450s enzymes such as *Cyp6z2*, *Cyp6z3* and *Cyp6m2* that are located in the same genomic region [58, 59]. Given that no evidence of selection on amino acid variants were detected in the *Cyp6m2* gene, it is possible that the emergence of *Cyp6m2* associated resistance is being driven by selection pressures acting upon genes coding for distant regulatory proteins, such as transcription factors. These transcription factors can regulate downstream gene expression across large genomic distances. These transcription factors have also been implicated in the differential expression of other detoxification enzyme families also associated with insecticide resistance (GSTs, COEs, UDP-glucuronosyltransferases (UGTs) and ABC transporters). It is, therefore, likely that the centre of selection leading to the *Cyp6m2* associated resistance phenotype will be identified through whole genome selection scans of susceptible and resistant populations rather than by single loci analysis. Further research on Anopheline epigenomics, transcriptomics, proteomics and systems biology will also be game changers in mapping the complex regulatory network of insecticide resistance, aiding the identification of critical targets and the development of new strategies to control the spread of metabolic insecticide resistance.

Conclusion

The scale-up of insecticide-based interventions has caused increased selection pressure and higher levels of insecticide resistance across Africa. While the *CYP6M2* enzyme has been associated with emerging metabolic resistance in Africa, our data indicates that allelic variation within the *Cyp6m2* gene itself or across its *Cyp6* supercluster has not been subject to recent positive selection in any of the populations sampled. This is in contrast to other Cytochrome P450 genes where CNV alleles are clearly under strong selection. Our results do not rule out a role for *Cyp6m2* in insecticide resistance in natural populations, but highlight the need for a deeper understanding of the regulatory networks affecting Cytochrome P450 gene expression in malaria vectors. This will require large-scale, holistic experimental work

that collects genomic, transcriptomic and phenotypic datasets which when juxtaposed can resolve the complexities of metabolic resistance.

Abbreviations

CncC: Cap'n' Collar isoform-C; CNV: Copy Number Variation; COEs: Carboxylesterases; DDT: Dichlorodiphenyltrichloroethane; EHH: Extended Haplotype Heterozygosity; GSTs: Glutathione S-Transferases; HAM: Transcription Factor Hamlet; IRS: Indoor Residual Spraying; *Kdr*: Knock-Down Resistance; LLINs: Long Lasting Insecticidal Nets; MJNs: Median-Joining Networks; ODR2: Odd-Skipped Related 2; P450s: Cytochrome P450 Monooxygenases; SH2: SRC Homology 2; SNPs: Single Nucleotide Polymorphisms; UGTs: UDP-glucuronosyltransferases; VGSC: Voltage Gated Sodium Channel; wt: Wild type.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12936-021-03757-4>.

Additional file 1: List of *An. gambiae* and *An. coluzzii* genome samples and haplotypes from Ag1000G Phase 2-AR3. Table showing Ag1000G Phase 2-AR3 sample properties such as population, country, region, sex, species identity and haplotype cluster.

Additional file 2: List of synonymous and non-synonymous genetic variants in *Cyp6m2*. Table showing Ag1000G Phase 2-AR3 *Cyp6m2* variant calls and variant properties stratified by population and effect.

Additional file 3: Fig. S1. Linkage disequilibrium (D') between non-synonymous variants. A value of 1 shows perfect linkage between the alleles. A value of -1 shows that the alleles are never found conjointly. The bar plot indicates allele frequencies within the Ag1000G phase 2 cohort.

Additional file 4: Fig. S2. Hierarchical clustering and missense mutations for *ODR2*. Top: a dendrogram showing hierarchical clustering of haplotypes across the *ODR2* gene. The gene is located at position 7,059,422 to 7,119,244: 128,875 bases downstream of *Cyp6m2*. The colour bar indicates the population of origin for each haplotype. Bottom: high frequency (> 5%) alleles identified within each haplotype (white = reference allele; black = alternative allele).

Additional file 5: Fig. S3. Hierarchical clustering and missense mutations for *HAM*. Top: a dendrogram showing hierarchical clustering of haplotypes across the *HAM* gene. The gene is located at position 7,435,306 to 7,485,012: 504,759 bases downstream of *Cyp6m2*. The colour bar indicates the population of origin for each haplotype. Bottom: high frequency (> 5%) alleles identified within each haplotype (white = reference allele; black = alternative allele).

Additional file 6: Fig. S4. Hierarchical clustering and missense mutations for *SH2*. Top: a dendrogram showing hierarchical clustering of haplotypes across the *SH2* gene. The gene is located at position 8,176,778 to 8,183,084: 1,246,231 bases downstream of *Cyp6m2*. The colour bar indicates the population of origin for each haplotype. Bottom: high frequency (> 5%) alleles identified within each haplotype (white = reference allele; black = alternative allele).

Additional file 7: Fig. S5. Hierarchical clustering and missense mutations for *Cyp6m* sub cluster. Top: a dendrogram showing hierarchical clustering of haplotypes across the *Cyp6m* sub cluster of genes containing *Cyp6m2*, *Cyp6m3* and *Cyp6m4*. The genes are located at position 6928858 to 6935721. The colour bar indicates the population of origin for each haplotype. Bottom: high frequency (> 5%) alleles identified within each haplotype (white = reference allele; black = alternative allele).

Additional file 8: Fig. S6. Hierarchical clustering and missense mutations for *Cyp6* supercluster. Top: a dendrogram showing hierarchical clustering of haplotypes across the *Cyp6* supercluster of 14 P450 genes containing *Cyp6s2*, *Cyp6s1*, *Cyp6r1*, *Cyp6n2*, *Cyp6y2*, *Cyp6y1*, *Cyp6m1*, *Cyp6n1*, *Cyp6m2*, *Cyp6m3*, *Cyp6m4*, *Cyp6z3*, *Cyp6z2* and *Cyp6z1*. The genes are located at position 6903106 to 6978142. The colour bar indicates the population of

origin for each haplotype. Bottom: high frequency (> 70%) alleles identified within each haplotype (white = reference allele; black = alternative allele).

Additional file 9: Fig. S7. Extended haplotype homozygosity across all populations. A rapid decay of EHH in comparison to other haplotypes implies absence of positive selection.

Acknowledgements

The authors would like to thank the staff of the Wellcome Sanger Institute Sequencing and Informatics facilities for their contributions.

Authors' contributions

AM and ML designed the study. AM and CC developed the base code. MGW and AM performed all analyses. MGW drafted the manuscript. All authors read and approved the final manuscript.

Funding

The Wellcome Sanger Institute is funded by the Wellcome Trust (Grant 206194/Z/17/Z), which supports M.K.N.L. and part of the sequencing, analysis, informatics, and management of the *Anopheles gambiae* 1000 Genomes Project.

Availability of data and materials

Jupyter Notebooks and scripts containing all analyses, tables and figures can be found in the GitHub repository [43]. Variant calls and phased haplotype data from the Ag1000G Phase 2 AR3 data release were used, and can be found here [60].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SD, UK.

²European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridgeshire CB10 1SD, UK. ³University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK.

Received: 11 November 2020 Accepted: 8 May 2021

Published online: 25 May 2021

References

- WHO. World Malaria Report 2019. Geneva, World Health Organization; 2019. Accessed 02 May 2021.
- Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*. 2015;526:207–11.
- Cibulskis RE, Alonso P, Aponte J, Aregawi M, Barrette A, Bergeron L, et al. Malaria: global progress 2000–2015 and future challenges. *Infect Dis Poverty*. 2016;5:61.
- Ranson H, Lissenden N. Insecticide resistance in African *Anopheles* mosquitoes: a worsening situation that needs urgent action to maintain malaria control. *Trends Parasitol*. 2016;32:187–96.
- Ranson H, Jensen B, Vulule JM, Wang X, Hemingway J, Collins FH. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. *Insect Mol Biol*. 2000;9:491–7.
- Yahouédo GA, Chandre F, Rossignol M, Ginibre C, Balabanidou V, Mendez NGA, et al. Contributions of cuticle permeability and enzyme

- detoxification to pyrethroid resistance in the major malaria vector *Anopheles gambiae*. *Sci Rep*. 2017;7:11091.
7. Main BJ, Lee Y, Ferguson HM, Kreppel KS, Kihonda A, Govella NJ, et al. The genetic basis of host preference and resting behavior in the major African malaria vector, *Anopheles arabiensis*. *PLoS Genet*. 2016;12:e1006303.
 8. Hemingway J. The molecular basis of two contrasting metabolic mechanisms of insecticide resistance. *Insect Biochem Mol Biol*. 2000;30:1009–15.
 9. Djouaka RF, Bakare AA, Coulibaly ON, Akogbeto MC, Ranson H, Hemingway J, et al. Expression of the cytochrome P450s, CYP6P3 and CYP6M2 are significantly elevated in multiple pyrethroid resistant populations of *Anopheles gambiae* s.s. from Southern Benin and Nigeria. *BMC Genomics*. 2008;9:538.
 10. Djouaka R, Riveron JM, Yessoufou A, Tchigossou G, Akoton R, Irving H, et al. Multiple insecticide resistance in an infected population of the malaria vector *Anopheles funestus* in Benin. *Parasit Vectors*. 2016;9:453.
 11. WHO. Global report on insecticide resistance in malaria vectors: 2010–2016. Geneva, World Health Organization 2018. Report No. 9241514051. Accessed 02 May 2021.
 12. WHO. Malaria Threats Map. Geneva, World Health Organization, 2020 (Insecticide resistance status). <https://apps.who.int/malaria/maps/threats>. Accessed 09 Nov 2020.
 13. Ranson H, N'Guessan R, Lines J, Moiroux N, Nkuni Z, Corbel V. Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends Parasitol*. 2011;27:91–8.
 14. WHO. Global report on insecticide resistance in malaria vectors: 2010–2016. Geneva: World Health Organization; 2018.
 15. Hargreaves K, Koekemoer LL, Brooke BD, Hunt RH, Mthembu J, Coetzee M. *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. *Med Vet Entomol*. 2000;14:181–9.
 16. Wondji CS, Morgan J, Coetzee M, Hunt RH, Steen K, Black WC, et al. Mapping a quantitative trait locus (QTL) conferring pyrethroid resistance in the African malaria vector *Anopheles funestus*. *BMC Genomics*. 2007;8:34.
 17. Wondji CS, Irving H, Morgan J, Lobo NF, Collins FH, Hunt RH, et al. Two duplicated P450 genes are associated with pyrethroid resistance in *Anopheles funestus*, a major malaria vector. *Genome Res*. 2009;19:452–9.
 18. Stevenson BJ, Bibby J, Pignatelli P, Muangnoicharoen S, O'Neill PM, Lian L-Y, et al. Cytochrome P450 6M2 from the malaria vector *Anopheles gambiae* metabolizes pyrethroids: sequential metabolism of deltamethrin revealed. *Insect Biochem Mol Biol*. 2011;41:492–502.
 19. VEuPathDB. Chromosome 3R: 6,928,825–6,930,580 - Region in detail - *Anopheles gambiae* - VectorBase <https://vectorbase.org>: VEuPathDB Bioinformatics Resource Center 2020. https://www.vectorbase.org/Anopheles_gambiae/Location/View?db=core;g=AGAP008212;r=3R:6928825-6930580;t=AGAP008212-RA. Accessed 09 Nov 2020.
 20. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Shakhova MV, et al. Evolution of supergene families associated with insecticide resistance. *Science*. 2002;298:179–81.
 21. Ranson H, Paton MG, Jensen B, McCarroll L, Vaughan A, Hogan JR, et al. Genetic mapping of genes conferring permethrin resistance in the malaria vector, *Anopheles gambiae*. *Insect Mol Biol*. 2004;13:379–86.
 22. Müller P, Donnelly MJ, Ranson H. Transcription profiling of a recently colonised pyrethroid resistant *Anopheles gambiae* strain from Ghana. *BMC Genomics*. 2007;8:36.
 23. Nardini L, Christian RN, Coetzer N, Ranson H, Coetzee M, Koekemoer LL. Detoxification enzymes associated with insecticide resistance in laboratory strains of *Anopheles arabiensis* of different geographic origin. *Parasit Vectors*. 2012;5:113.
 24. Edi CV, Djogbénou L, Jenkins AM, Regna K, Muskavitch MAT, Poupardin R, et al. CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito *Anopheles gambiae*. *PLoS Genet*. 2014;10:e1004236.
 25. Yan Z-W, He Z-B, Yan Z-T, Si F-L, Zhou Y, Chen B. Genome-wide and expression-profiling analyses suggest the main cytochrome P450 genes related to pyrethroid resistance in the malaria vector, *Anopheles sinensis* (Diptera Culicidae). *Pest Manag Sci*. 2018;74:1810–20.
 26. Djègbè I, Agossa FR, Jones CM, Poupardin R, Cornelie S, Akogbèto M, et al. Molecular characterization of DDT resistance in *Anopheles gambiae* from Benin. *Parasit Vectors*. 2014;7:409.
 27. Adolli A, Poulton B, Anthousi A, MacIwee S, Ranson H, Lycett GJ. Functional genetic validation of key genes conferring insecticide resistance in the major African malaria vector, *Anopheles gambiae*. *Proc Natl Acad Sci USA*. 2019;116:25764–72.
 28. Mitchell SN, Stevenson BJ, Müller P, Wilding CS, Egyir-Yawson A, Field SG, et al. Identification and validation of a gene causing cross-resistance between insecticide classes in *Anopheles gambiae* from Ghana. *Proc Natl Acad Sci USA*. 2012;109:6147–52.
 29. Lucas ER, Miles A, Harding NJ, Clarkson CS, Lawniczak MKN, Kwiatkowski DP, et al. Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. *Genome Res*. 2019;29:1250–61.
 30. Weetman D, Djogbenou LS, Lucas E. Copy number variation (CNV) and insecticide resistance in mosquitoes: evolving knowledge or an evolving problem? *Curr Opin Insect Sci*. 2018;27:82–8.
 31. Schuler MA, Berenbaum MR. Structure and function of cytochrome P450s in insect adaptation to natural and synthetic toxins: insights gained from molecular modeling. *J Chem Ecol*. 2013;39:1232–45.
 32. Ibrahim SS, Riveron JM, Bibby J, Irving H, Yunta C, Paine MJ, et al. Allelic variation of cytochrome P450s drives resistance to bednet insecticides in a major malaria vector. *PLoS Genet*. 2015;11:e1005618.
 33. Weetman D, Wilding CS, Neafsey DE, Müller P, Ochomo E, Isaacs AT, et al. Candidate-gene based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing genetic variation in East African *Anopheles gambiae*. *Sci Rep*. 2018;8:2920.
 34. Clarkson CS, Miles A, Harding NJ, Lucas ER, Battey C, Amaya-Romero JE, et al. Genome variation and population structure among 1142 mosquitoes of the African malaria vector species *Anopheles gambiae* and *Anopheles coluzzii*. *Genome Res*. 2020;30:1533–46.
 35. Coetzee M, Hunt RH, Wilkerson R, Della Torre A, Coulibaly MB, Besansky NJ. *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa*. 2013;3619:246–74.
 36. Consortium TAGG. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*. 2017;552:96–100.
 37. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*. 2002;298:129–49.
 38. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012;6:80–92.
 39. Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*. 1964;49:49–67.
 40. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9:90–5.
 41. Cartopy. Using cartopy with matplotlib—cartopy 0.18.0 documentation <https://scitools.org.uk/2020> (cited Windows). 0.17.0. <https://scitools.org.uk/cartopy/docs/latest/matplotlib/intro.html>. Accessed 02 May 2021
 42. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16:37–48.
 43. Wagah MG. Ag1000G-phase2-cyp6m2 [Jupyter Notebooks]. <https://github.com/2020> (9/11/2020). <https://github.com/malariagen/ag1000g-phase2-cyp6m2>. Accessed 02 May 2021.
 44. Harrington B. Inkscape 2005 [1.0.1]. <http://www.inkscape.org>. Accessed 02 May 2021.
 45. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002;419:832–7.
 46. Miles A. scikit-allel - Explore and analyse genetic variation—scikit-allel 1.3.2 documentation. <https://github.com/2018>.
 47. The Anopheles gambiae Genomes Consortium, Clarkson CS, Miles A, Harding NJ, Lucas ER, Battey CJ, et al. Genome variation and population structure among 1,142 mosquitoes of the African malaria vector species *Anopheles gambiae* and *Anopheles coluzzii*. *Genome Res*. 2020;30:1533–46.
 48. VEuPathDB. Chromosome 3R: 7,059,422 - 7,119,244 - Region in detail - *Anopheles gambiae* - VectorBase <https://vectorbase.org>: VEuPathDB Bioinformatics Resource Center 2020. <https://vectorbase.org/vectorbase/app/record/gene/AGAP008222>. Accessed 09 Nov 2020.
 49. VEuPathDB. Chromosome 3R: 7,435,306 - 7,485,012 - *Anopheles gambiae* - VectorBase <https://vectorbase.org>: VEuPathDB Bioinformatics Resource Center 2020. : <https://vectorbase.org/vectorbase/app/record/gene/AGAP008232>. Accessed 09 Nov 2020.

50. VEuPathDB. Chromosome 3R: 8,176,778–8,183,084 - Region in detail - *Anopheles gambiae* – VectorBase. <https://vectorbase.org>. VEuPathDB Bioinformatics Resource Center 2020. <https://vectorbase.org/vectorbase/app/record/gene/AGAP008273>. Accessed 09 Nov 2020.
51. Müller P, Warr E, Stevenson BJ, Pignatelli PM, Morgan JC, Steven A, et al. Field-caught permethrin-resistant *Anopheles gambiae* over-express CYP6P3, a P450 that metabolises pyrethroids. *PLoS Genet*. 2008;4:e1000286.
52. Stica C, Jeffries CL, Irish SR, Barry Y, Camara D, Yansane I, et al. Characterizing the molecular and metabolic mechanisms of insecticide resistance in *Anopheles gambiae* in Faranah. *Guinea Malar J*. 2019;18:244.
53. Pandurangan AP, Blundell TL. Prediction of impacts of mutations on protein structure and interactions: SDM, a statistical approach, and mCSM, using machine learning. *Protein Sci*. 2020;29:247–57.
54. Consortium TAgG. Ag1000G - AR3 Panoptes genome browser Panoptes: *Anopheles gambiae* 1000 genomes project 2015. https://www.malariagen.net/apps/ag1000g/phase1-AR3/index.html?dataset=Ag1000G&workspace=workspace_1&view=f6c6c7c8-23c9-11eb-a4f3-22000a6287ed&state=genomebrowser. Accessed 02 May 2021.
55. Clarkson CS, Miles A, Harding NJ, Weetman D, Kwiatkowski D, Donnelly M, et al. The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors *Anopheles gambiae* and *Anopheles coluzzii*. *Mol Ecol*. 2021. <https://doi.org/10.1111/mec.15845> (online ahead of print).
56. Hemingway J. The role of vector control in stopping the transmission of malaria: threats and opportunities. *Philos Trans R Soc Lond B Biol Sci*. 2014;369:20130431.
57. Liu N. Insecticide resistance in mosquitoes: impact, mechanisms, and research directions. *Annu Rev Entomol*. 2015;60:537–59.
58. Ingham VA, Pignatelli P, Moore JD, Wagstaff S, Ranson H. The transcription factor Maf-S regulates metabolic resistance to insecticides in the malaria vector *Anopheles gambiae*. *BMC Genomics*. 2017;18:669.
59. Wilding CS. Regulating resistance: CncC:Maf, antioxidant response elements and the overexpression of detoxification genes in insecticide resistance. *Curr Opin Insect Sci*. 2018;27:89–96.
60. Consortium TAgG. Ag1000G phase 2 AR1 data release MalariaGen Genomic Epidemiology Network. 2017. <https://www.malariagen.net/data/ag1000g-phase-2-ar1>. Accessed 02 May 2021.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

