



OPEN

# Cryptic population structure and insecticide resistance in *Anopheles gambiae* from the southern Democratic Republic of Congo

Tristan P. W. Dennis<sup>1,11</sup>, Poppy Pescod<sup>1,11</sup>, Sonia Barasa<sup>2,11</sup>, Louise T. Cerdeira<sup>1</sup>, Eric R. Lucas<sup>1</sup>, Chris S. Clarkson<sup>3</sup>, Alistair Miles<sup>3</sup>, Alex Asidi<sup>4,5</sup>, Emile Z. Manzambi<sup>5</sup>, Emery Metelo<sup>5,6</sup>, Josue Zanga<sup>7</sup>, Steve Nsalambi<sup>5,8</sup>, Seth R. Irish<sup>9</sup>, Martin James Donnelly<sup>1</sup>, Fiacre Agossa<sup>5,7,10</sup>, David Weetman<sup>1,11</sup> & Francis Wat'senga Tezzo<sup>5,11</sup>

The Democratic Republic of Congo (DRC) suffers from one of the highest malaria burdens worldwide, but information on its *Anopheles* vector populations is relatively limited. Preventative malaria control in DRC is reliant on pyrethroid-treated nets, raising concerns over the potential impacts of insecticide resistance. We sampled *Anopheles gambiae* from three geographically distinct populations (Kimpese, Kapolowe and Mikalayi) in southern DRC, collecting from three sub-sites per population and characterising mosquito collections from each for resistance to pyrethroids using WHO tube bioassays. Resistance to each of three different pyrethroids was generally high in *An. gambiae* with <92% mortality in all tests, but varied between collections, with mosquitoes from Kimpese being the most resistant. Whole genome sequencing of 165 *An. gambiae* revealed evidence for genetic differentiation between Kimpese and Kapolowe/Mikalayi, but not between the latter two sample sites despite separation of approximately 800 km. Surprisingly, there was evidence of population structure at a small spatial scale between collection subsites in Kimpese, despite separation of just tens of kilometres. Intra-population (H12) and inter-population ( $F_{ST}$ ) genome scans identified multiple peaks corresponding to genes associated with insecticide resistance such as the voltage gated sodium channel (*Vgsc*) target site on chromosome 2L, a *Cyp6* cytochrome P450 cluster on chromosome arm 2R, and the *Cyp9k1* P450 gene on chromosome X. In addition, in the Kimpese subsites, the P450 redox partner gene *Cpr* showed evidence for contemporary selection (H12) and population differentiation ( $F_{ST}$ ) meriting further exploration as a potential resistance associated marker.

**Keywords** Insecticide resistance, WGS, Malaria, Population genomics, *Anopheles gambiae* complex

Malaria is a major burden in the Democratic Republic of the Congo (DRC), with 12% of all global malaria cases occurring in the country<sup>1</sup>. Several *Anopheles* species are vectors of malaria in DRC, including the primary vector

<sup>1</sup>Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK. <sup>2</sup>Pan-African Mosquito Control Association (PAMCA), Nairobi, Kenya. <sup>3</sup>Wellcome Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. <sup>4</sup>University Hospitals of Derby and Burton NHS Foundation Trust, Uttoxeter Road, Derby DE22 3NA, UK. <sup>5</sup>Unit of Entomology, Department of Parasitology, Institut National de Recherche Biomédicale, 5345 Avenue De La Démocratie, Gombe, Kinshasa, République Démocratique du Congo. <sup>6</sup>Faculty of Science and Technology, University of Kinshasa, Kinshasa, République Démocratique du Congo. <sup>7</sup>Faculty of Medicine, University of Kinshasa, B.P. 834 KIN XI, Kinshasa, République Démocratique du Congo. <sup>8</sup>Faculty of Veterinary Medicine, National University of Education, B.P 8815 Kinshasa, Kinshasa, République Démocratique du Congo. <sup>9</sup>Swiss Tropical and Public Health Institute (Swiss TPH), Kreuzstrasse 2, 4123 Allschwil, Switzerland. <sup>10</sup>U.S. President's Malaria Initiative (PMI) Evolve Project, Abt Associates, 6130 Executive Boulevard, Rockville, MD, USA. <sup>11</sup>These authors contributed equally: Tristan P. W. Dennis, Poppy Pescod, Sonia Barasa, David Weetman and Francis Wat'senga Tezzo. ✉email: tristanpwdennis@gmail.com; David.Weetman@lstmed.ac.uk

*Anopheles gambiae* and secondary vectors *An. funestus* sensu stricto (s.s.), *An. coluzzii*, *An. arabiensis* and several uncommon or suspected vector species (e.g. *An. mouchetti*, *An. paludis*, *An. demeilloni*, etc.)<sup>2</sup>. Resistance to pyrethroid insecticides, the main insecticide used on ITNs, the first line of defence against malaria, has been documented in *An. gambiae* in DRC; high frequencies of resistance-conferring mutations have been observed in the voltage-gated sodium channel<sup>3–5</sup> and a recent selective sweep in the Cyp6aa/Cyp6p cluster of Cytochrome P450 genes has been implicated in enhancing metabolic resistance<sup>6</sup>.

Malaria control programmes in DRC have depended heavily on the distribution of insecticide-treated bed-nets (ITNs) primarily using pyrethroids, with several nationwide distribution cycles over the last two decades contributing to a reduction in malaria incidence<sup>7,8</sup>. The scale of pyrethroid resistance across the African continent is prompting widespread deployment of synergists such as PBO and chlorfenapyr, which can help combat metabolic resistance<sup>9</sup>. However, new generation nets are considerably more expensive than standard nets, and a better understanding of the distribution of resistance mechanisms would aid a more targeted distribution of these second generation nets<sup>10,11</sup>.

Despite the high impact of malaria in DRC there is a paucity of studies on its malaria vectors. The first iteration of the *Anopheles gambiae* 1000 genomes project (Ag1000G) increased understanding of the genetic variation underlying insecticide resistance in sub-Saharan Africa, but with a notable gap in the map covering the logistically complicated DRC<sup>12</sup>. Subsequent additions to the Ag1000g dataset have facilitated analysis of genomic regions under recent selection in insecticide-resistant mosquito populations across Africa, demonstrating that the strongest selective sweeps in the genome include insecticide resistance genes<sup>12,13</sup>. Here we describe the genomes of *An. gambiae* from collections made in three geographically distinct locations in southern DRC, phenotypically characterised for resistance to pyrethroids. We investigate the level of genomic differentiation at both fine and more distant scales, and identify both known and novel signals of selection likely linked to insecticide resistance in the region.

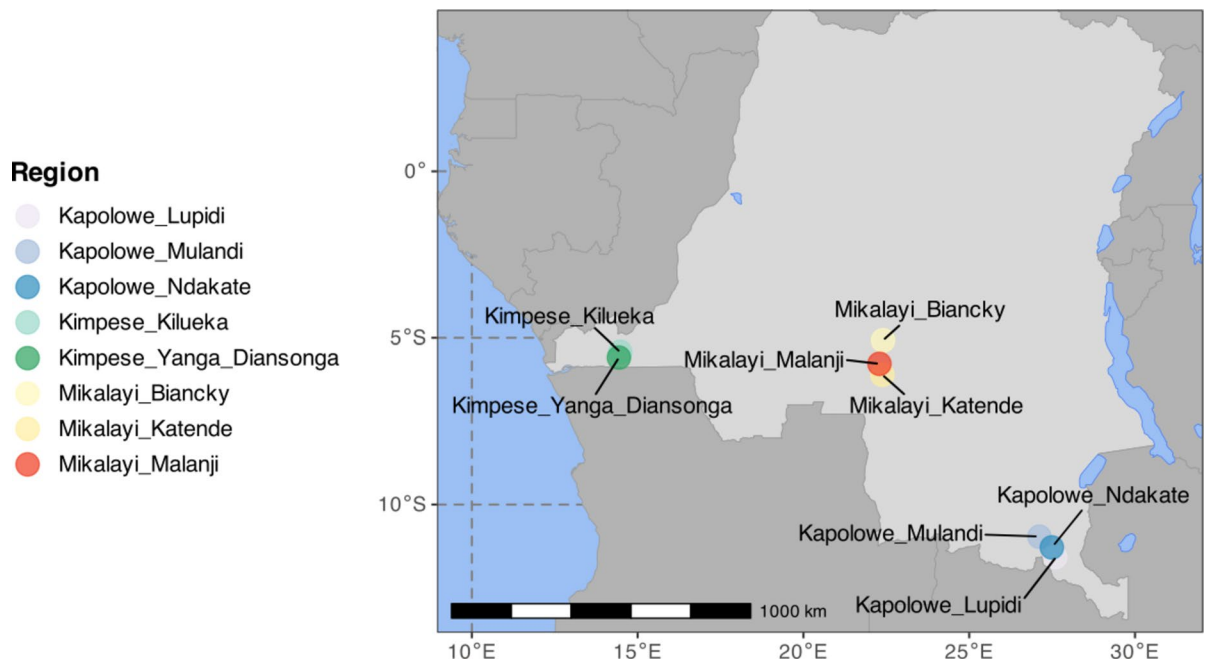
## Results

Adult mosquitoes were collected for sequencing, and larval stages collected and reared to adulthood for phenotypic assays, at the sites shown in Fig. 1.

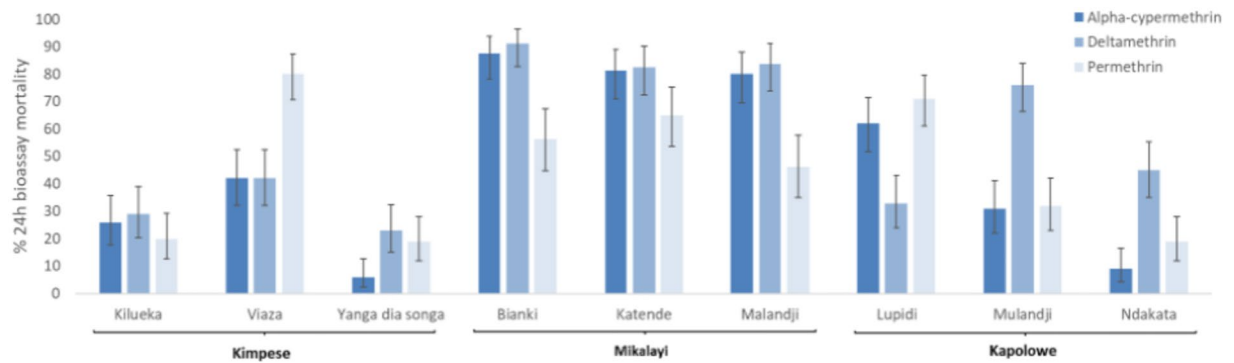
### Insecticide resistance bioassays

Larvae collected from each of the nine sites and reared to adulthood were tested against three insecticides at standard discriminating doses: alpha-cypermethrin (0.05%), deltamethrin (0.05%) and permethrin (0.75%). Bioassay results for *An. gambiae* s.l. are shown in Fig. 2 and given in full in Table S2. In all but one of the 27 site and insecticide combinations, mortality was less than 90%, indicating resistance according to WHO criteria<sup>14</sup>—mortality after deltamethrin exposure was 91% in Bianki, Mikalayi, indicating suspected resistance.

Although all sites showed resistance to all three pyrethroids according to WHO criteria, there were significant differences between sites (Table S3); with Mikalayi showing least resistance and Kimpese showing strongest resistance (Table S4). Within sites there were also statistically significant differences in mortality to all three insecticides between some sub-sites; these differences were strongest between sub-sites in Kimpese, with Yanga



**Fig. 1.** Map of the DRC showing the collection sites, where each circle represents a locality/village, coloured by the Province and location of each site.



**Fig. 2.** Pyrethroid insecticide resistance profiles for *Anopheles gambiae s.l.* from the nine collection sites (three sites, each containing three sub-sites, 80–100 mosquitoes tested per site) assessed using WHO diagnostic bioassays. Error bars are 95% confidence limits.

Diansonga showing the highest resistance (Wald  $\chi^2 = 7.42$ ,  $P = 0.006$ ) and Viaza the lowest (Wald  $\chi^2 = 53.54$ ,  $P < 0.001$ ). The insecticide used also had a significant impact on mortality (Wald  $\chi^2 = 24.35$ ,  $P < 0.001$ ), with deltamethrin overall producing lower mortality than alpha-cypermethrin and permethrin ( $P < 0.001$ ).

### Adult mosquito collection

Over the collection period, a total of 2153 adult *Anopheles* mosquitoes were collected in Mikalayi, Kimpese and Kapolowe by pyrethrum spray catch with *An. funestus sensu lato (s.l.)* being the predominant species (64%), followed by *An. gambiae s.l.* (36%) and *An. paludis* (0.2%). The mean indoor resting density of *An. gambiae s.l.* was 4.8 and of *An. funestus s.l.* was 7.7 per house across the three provinces, with *An. gambiae s.l.* mean densities of 6.1, 4.7 and 3.7 observed in Kapolowe, Kimpese and Mikalayi, respectively. The mean indoor resting density of *An. funestus s.l.* was 6.6, 14.2 and 2.3 in Kapolowe, Kimpese and Mikalayi, respectively. In all sites, the majority of *An. gambiae s.l.* (84.3%, 730/866) and *An. funestus s.l.* (82.5%, 1140/1382) collected indoors were blood-fed. One of the Kimpese sub-sites, Viaza, did not yield enough adult *An. gambiae* to be sequenced as part of the *Anopheles gambiae* 1000 genome project<sup>12</sup>. Of the 237 samples sequenced, 165 passed quality control (QC) filtering. The samples were processed as part of the *Anopheles gambiae* genomic surveillance MalariaGEN vector observatory (VObs) project under sample set ID 1264-VO-CD-WATSENGA-VMF00164.

### Population structure and diversity

Principal component analysis (PCA) of all sequenced individuals, (colour-labelled by collection site) for each chromosome is shown in Fig. S1. PCA showing their placement in the *An. gambiae s.s.* group with respect to other *An. gambiae* and *An. coluzzi* samples spanning the African continent, from the phase 3.0 release of the *An. gambiae* 1000 genome project is shown in Fig. S2. At least three differentiation patterns are evident in the plots. On chromosomes X and arms 3L/3R clear separation of the Kimpese (Kongo Central province) samples from those collected in Mikalayi (Kasai Central province) and Kapolowe (Haut Katanga province). Re-running the PCA as above, on chromosome 3L, on these two clusters (samples from Kimpese, and samples from Kapolowe/Mikalayi) (Fig. 3), show no major hidden structure within them, though there is some separation between Kimpese Kilueka and Kimpese Yanga Diansonga (Fig. 3A), and between Mikalayi and Kapolowe (Fig. 3B). Separation within Kimpese is despite the sample sites being only approximately 20 km apart, the smallest pairwise distance among any of the sites and far less than the distance between Mikalayi and Kapolowe (approximately 800 km). As samples from sites within Mikalayi and Kapolowe districts appear to represent similar populations, these were pooled by district in subsequent analyses.

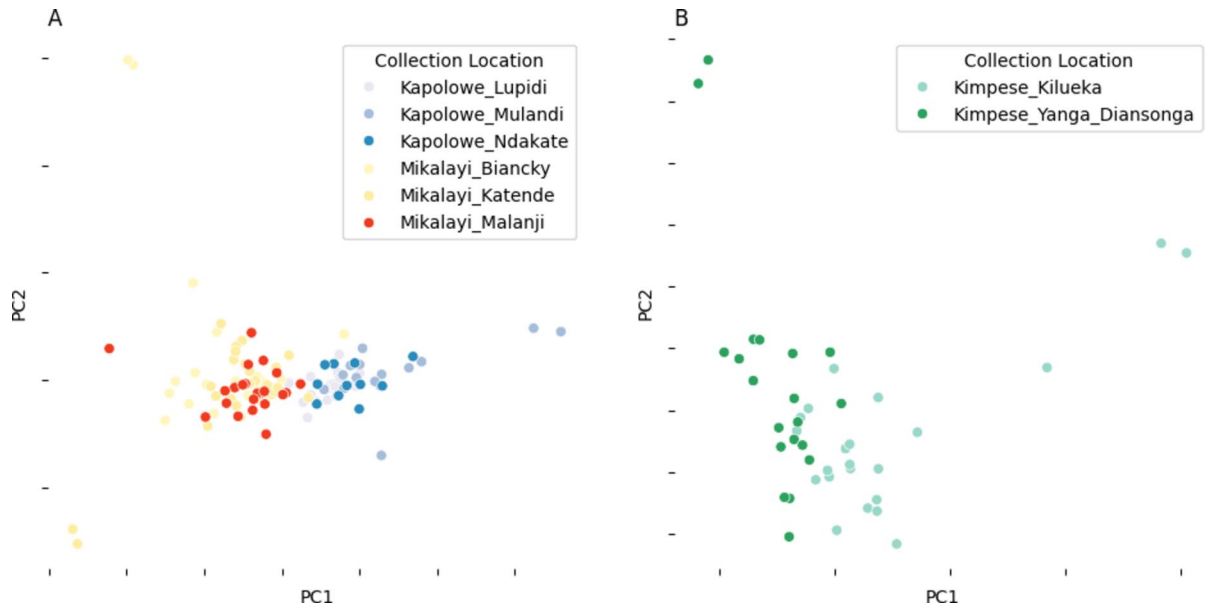
Further differentiation between sites is evident for chromosome 2, likely because of polymorphism in paracentric chromosomal inversions. This is most clearly evident for 2L, suggestive of 2La inversion karyotype variation. Separation is also evident on chromosome 2L PC2 and 2R PC2, also reflecting population structure among locations. In each case, more clearly for 2R, there is some evidence of separation of the Mikalayi and Kapolowe collection locations, likely suggesting differentiation in inversion frequencies. This contrasts with results for the other chromosomes (which lack polymorphic inversions in *An. gambiae*), indicating that differentiation between Mikalayi and Kapolowe is focal to inversions rather than genomewide.

### Genomic diversity

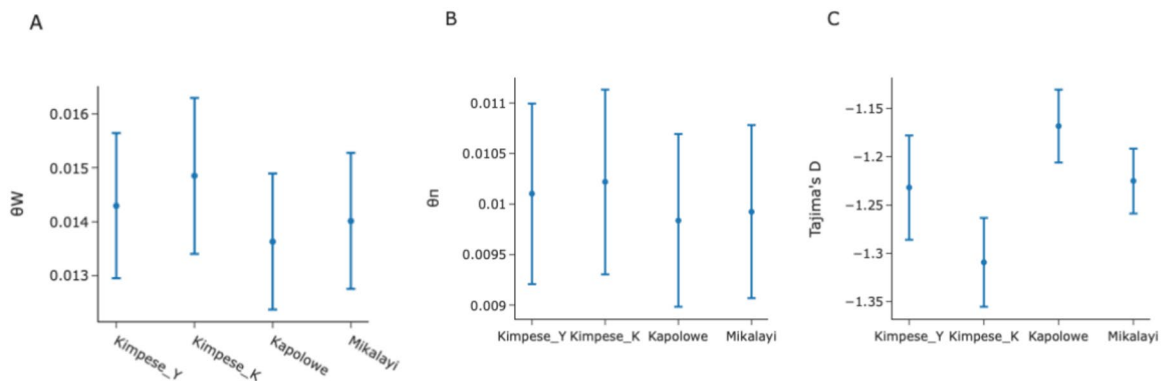
$\Pi$ , Watterson's  $\Theta$  ( $\Theta_W$ ), and Tajima's D were calculated and are shown in Fig. 4.  $\Theta_W$  was highest in Kimpese-Kilueka and Kimpese-Yanga Diansonga, and lowest in Kapolowe (Fig. 4B).  $\pi$  was highest in the two Kimpese cohorts compared to Mikalayi and Kapolowe. Tajima's D was negative in all cases, and highest in Kapolowe. Kimpese-Kilueka had the lowest Tajima's D. (Fig. 4C).

### Genomewide scans for signatures of selection

The statistic H12 was used to investigate signals of selection within each population (Fig. 5). On chromosome 2, a pronounced peak around 28.5 Mb is present in Kimpese-Kilueka, Mikalayi and Kapolowe (Fig. 5), and to



**Fig. 3.** Principal component analysis (PCA) of samples from Kapolowe/Mikalayi (A) and Kimpese (B). PCA performed with genotypes from chromosome arm 3L. Points are coloured by collection location.

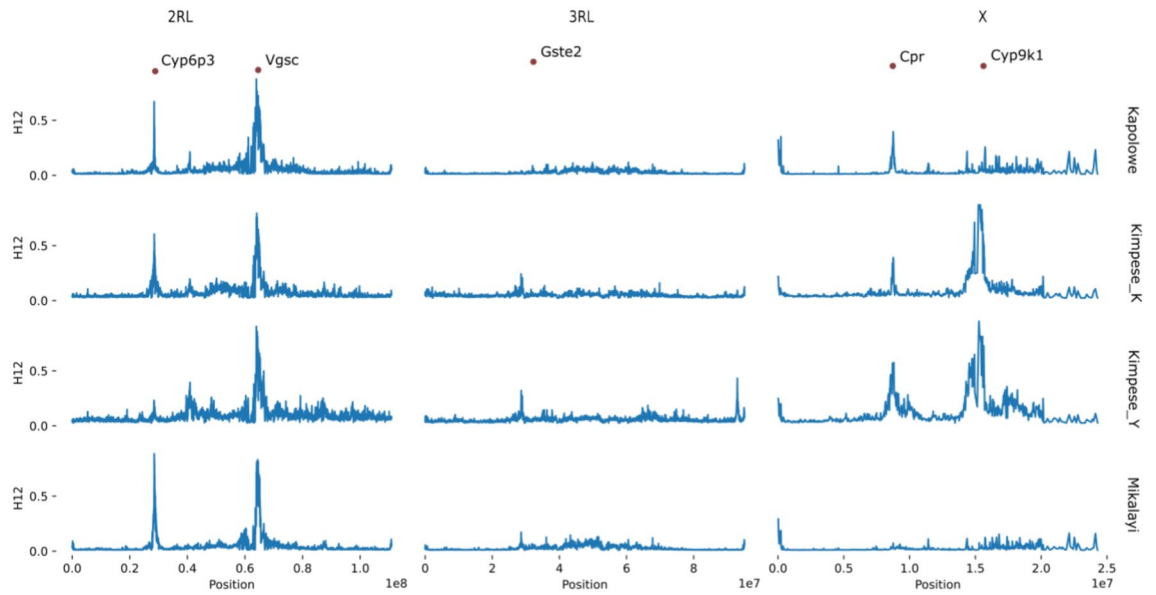


**Fig. 4.** Genomic diversity statistics from (Kimpese\_Y = Yanga Diansonga; Kimpese\_K = Kilueka) with 95% confidence intervals.

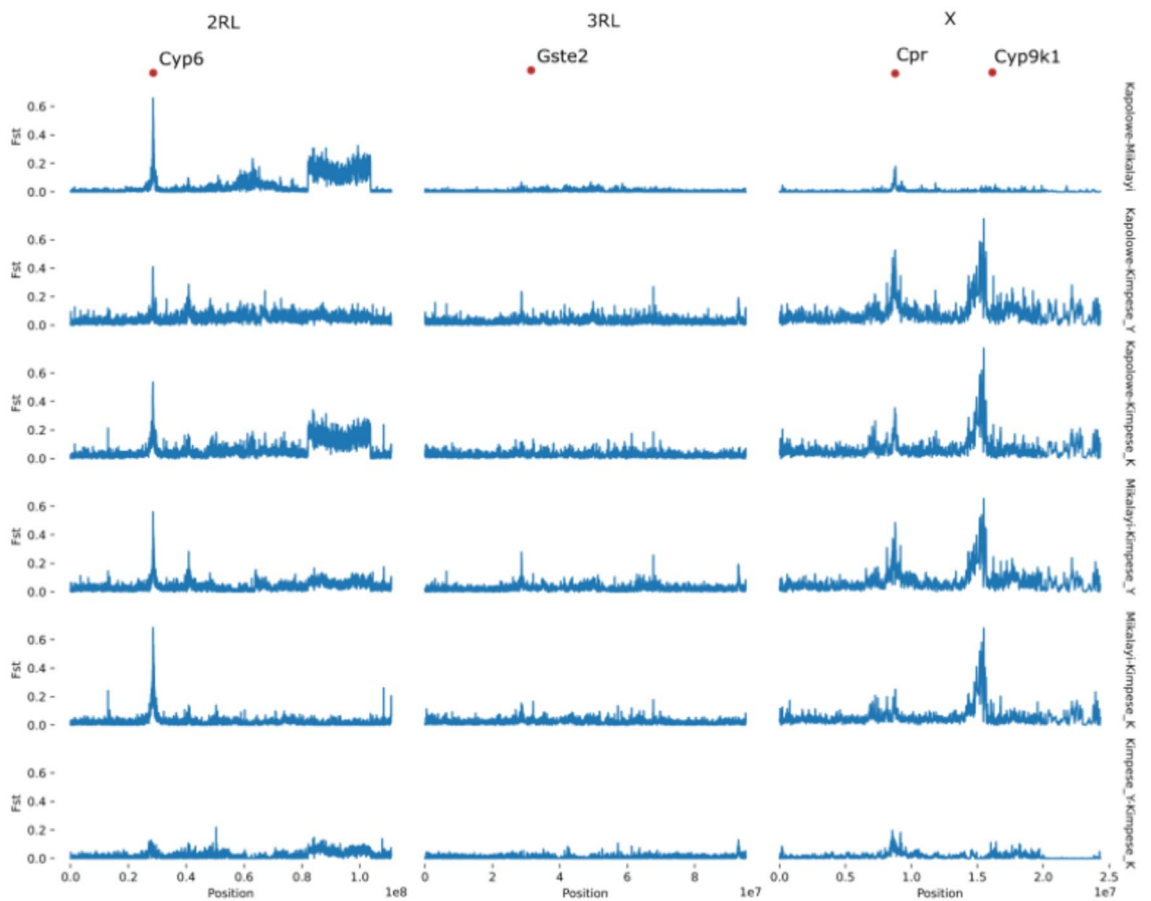
a much lesser extent in Kimpese-Yanga Diansonga, in which the largest peak is located at around 40.5 Mb. The 28.5 Mb peak centres on a cluster of P450 genes, including the proven pyrethroid-metabolising genes *Cyp6aa1*, *Cyp6p3* and *Cyp6p4*<sup>15,16</sup>. The 40.5 Mb peak centres on a gene *AGAP003623* (long-chain acyl-CoA synthetase), which has no previous association with insecticide resistance. On chromosome 2L, the only clear area highlighted by H12 is a protracted region evident in all populations centred on the voltage-gated sodium channel (*Vgsc*), which contains well known resistance-conferring *kdr* mutations (Fig. 5). The only obvious peak on chromosome 3RL is in Kimpese-Yanga Diansonga, and is a clear signal evident just before 40 Mb (Fig. 5). This peak centres on *AGAP012290*, an unnamed gene with transmembrane neurotransmitter transporter activity, and six *Cyp9* subfamily P450 genes (*CYP9L1-3*, *CYP9J3-5*), which are not currently known resistance candidate genes. Chromosome X shows considerable heterogeneity in signals between populations. In Mikalayi (Fig. 5), there are no obvious peaks, barring that proximate to the telomere (far left in Fig. 5), which is present in all populations and contains an uncharacterised gene *AGAP000002*. In both Kimpese populations, a very high and broad peak centres on the well-established pyrethroid metabolising gene *Cyp9k1*, with a second strong peak evident near 8.8 Mb in Kimpese-Yanga Diansonga and to a slightly lesser extent in Kapolowe and Kimpese-Kilueka. The closest gene to this peak is *AGAP000500* (NADPH cytochrome P450 reductase, *Cpr*), which was identified in a previous genome wide scan as a strong potential candidate for resistance because of its essential role in all P450-mediated metabolism reactions<sup>17</sup>.

#### Genomic comparison between populations using $F_{ST}$

Genetic differentiation between the Mikalayi and Kapolowe populations, which showed differences in insecticide resistance profiles (see Fig. 2) are shown in Fig. 6. Moderate differentiation was seen in the *Vgsc* gene (2RL at approx 60 Mb) between Mikalayi and Kapolowe, with stronger differentiation spanning the entire area of the 2La



**Fig. 5.** Genome-wide selection scan using the H12 statistic. X axis indicates position in a chromosome (plot columns), Y axis indicates H12 in a cohort (plot rows). Annotated points indicate IR genes of interest.



**Fig. 6.**  $F_{ST}$  comparison plots between population pairs. X axis indicates chromosomal position (chromosome denoted on the top of the plot), Y axis indicates  $F_{st}$ . Each row of plots denotes the labelled population pair. Genes of interest are highlighted on the top of the plot with red dots.

inversion visible, concordant with the differentiation between these sites on PC1 in Fig. 3. Differentiation at the *Vgsc* between populations combined with a clear signal of selection (see Fig. 5) may indicate population-specific haplotypes of *Vgsc*, or different frequencies of the same haplotype.

Differentiation on chromosome 2R between Mikalayi and Kapolowe is dominated by an extremely strong, sharp peak centred on the *Cyp6* P450 cluster (Fig. 6). Since H12 signals were similarly strong, this may indicate selection on different haplotypes in each population, or different frequencies of the same haplotype. A major peak area between 8.5 and 9 Mb dominates the  $F_{ST}$  profile between Mikalayi and Kapolowe and the Kimpese sites for chromosome X, in proximity to the *Cpr* gene (Fig. 6), and likely reflects a difference in the strength of H12 signal here between populations, which are present in Kapolowe but not in Mikalayi (Fig. 5).

Differentiation between the two sites within Kimpese again highlights the *Cyp6* cluster on 2R (Fig. 6), which was observed to show a much stronger H12 signal in Kilueka (Fig. 5) than Yanga Diansonga (Fig. 5). A second peak is also visible near 50 Mb. On chromosome X, differentiation around the *Cpr* gene is also evident, although far less pronounced than in comparisons with Mikalayi and Kapolowe (Fig. 6).

Profiles of differentiation between the Kimpese populations on chromosomes 2L, 3R and 3L are less clear for the other chromosomes owing to relatively high baselines and lower peak  $F_{ST}$ . Some differentiation of the 2La inversion region is present (Fig. 6), but other peaks on each chromosome are generally not supported by multiple contiguous windows.

### Frequencies of known and candidate resistance markers

SNP data from were screened for non-synonymous (protein altering) mutations in the resistance-associated genes *Vgsc*, *Cyp4j5*, *Cyp6p4*, *Rdl*, and in the *Cpr* gene newly implicated via the signals of selection analyses, and can be seen in Table 1 (more details in Table S4). The G280S (previously G119S) resistance mutation in the acetylcholinesterase (*Ace-1*) target site gene was screened but was absent in the samples and is not shown.

The most common mutation in *Vgsc* was 995F, which was present at high frequency in all populations, though notably lower in Mikalayi, which showed the highest frequency of the other well-known *Vgsc* mutant 995S; notably, the wild-type L995 was absent. Other resistance-associated non-synonymous SNPs in *Vgsc*, such as N1570Y, were absent; the only other present was R254K, which has previously been observed in close linkage with L995F but with an unknown effect on phenotype<sup>18</sup>. This variation between populations in the balance between 995F and S likely explains the *Vgsc* differentiation (Fig. 6) between Mikalayi and Kapolowe. Variants in the *Rdl* gene—a resistance-causing mutant (A296G) and linked, potentially compensatory variant (T345M)—were present at very low frequency, and the other known resistance-associated *Rdl* variant A296S was absent from the population<sup>19</sup>.

*Cyp4j5* is used as a predictive resistance marker in East African populations using the L43F mutation as a diagnostic SNP<sup>20</sup>. Twenty four substitutions were detected in *Cyp4j5*; L43F is present in all populations analysed here, with frequencies ranging from 0.19 to 0.64 (Table 1). Four non-synonymous variants with unknown phenotypic effects were detected in *Cyp6p4* (Table 1), which is within a P450 cluster containing multiple duplications implicated in increased insecticide resistance<sup>17</sup>. Most notably, the L161F mutation was absent in both Kapolowe and Mikalayi while being present in Kimpese at 0.35 and 0.61 frequencies in Yanga Diansonga and Kilueka respectively.

The three non-synonymous polymorphisms (K75N, T319N and A583S) at the candidate gene *Cpr* identified from genome scans show variation between populations. Mikalayi, which notably is the location with the lowest pyrethroid resistance in bioassays (Fig. 2), showed an extremely low frequency of each SNP (0.01 for each), whereas Kapolowe showed a relatively high frequency of the K75N mutant (0.51) but an absence of the others. In the two Kimpese populations, which showed higher overall resistance than Kapolowe and Mikalayi (Fig. 2), mutant T319N was at higher frequency in Kimpese Yanga Diansonga (which had the most resistant mosquitoes and the strongest H12 signal at *Cpr*) than Kimpese Kilueka (0.7 vs 0.3), whilst the A583S mutant showed the opposite pattern (0.18 vs 0.40) (Table 1). Noting also that differentiation in the  $F_{ST}$  scans was very high in this area of the genome between Kapolowe and Mikalayi and between the two Kimpese populations, these three mutants follow a pattern which could be compatible with conferring different levels of resistance.

Frequencies of known copy number variants (i.e. gene amplifications) are shown in Table 2 for resistance candidate genes (more details in Table S5). Copy number variation (CNV) frequencies at *Gste2* are similar for the two Kimpese populations (0.65 and 0.59), and at lower frequency in Kapolowe than Mikalayi (0.27 vs. 0.61). (Table S6). Although these CNVs have been linked to signals of selection in other populations previously<sup>17</sup>, no convincing evidence was found from selection scans for peaks at the eGST cluster situated at approximately 28 Mb on chromosome 3R in the populations surveyed here. Amplification of *Cpr* was only detected in Kapolowe, and at a similar frequency to the K75N mutant (0.57 and 0.51, respectively), suggesting possible association. Amplifications at the two adjacent *Cyp6aa* genes were absent or very rare in the Kimpese populations but at high frequency in Kapolowe and Mikalayi. Duplications of *Cyp6aa1* are associated with insecticide resistance, with one specific duplication (*Dup1*) co-occurring exclusively with the nearby *Cyp6p4* I236M mutation<sup>6</sup>. While *Cyp6aa1* CNVs were present at frequencies of 0.65 and 0.93 in Kapolowe and Mikalayi respectively, the I236M mutation was absent in all sites (Table 2), suggesting a different duplication to the *Dup1* identified in east Africa previously.

*Cyp9K1* amplification was fixed in the Kimpese populations but absent from the other locations. This is of interest because a substantial H12 signal at *Cyp9K1* was seen only in the two Kimpese populations (Fig. 5) and was absent from Kapolowe and Mikalayi (Fig. 5), suggesting the duplication may be the target of selection in Kimpese where mosquito populations show highest pyrethroid resistance.

Gene name	Substitution	Kimpese		Kapolowe	Mikalayi
		Yanga Diansonga	Kilueka		
<i>Vgsc</i>	R254K (G>A)	0.85	0.77	0.92	0.77
	L995S (T>C)	0.03	0.11	0.08	0.23
	L995F (A>T)	0.97	0.89	0.92	0.77
	F1920S (T>C)	0.13	0.09	0.00	0.00
<i>Cyp4j5</i>	Q530E (G>C)	0.73	0.43	0.82	0.40
	R509S (G>T)	0.30	0.02	0.09	0.01
	L416V (G>C)	0.13	0.00	0.00	0.00
	H415N (G>T)	0.10	0.05	0.18	0.05
	I414N (A>T)	0.10	0.05	0.18	0.05
	H377P (T>G)	0.28	0.02	0.02	0.01
	N371D (T>C)	0.10	0.05	0.04	0.12
	L354P (A>G)	1.00	0.95	0.96	0.99
	I243T (A>G)	1.00	1.00	0.99	1.00
	K216E (T>C)	0.13	0.11	0.09	0.18
	A211T (C>T)	0.90	0.70	0.93	0.76
	D177N (C>T)	0.10	0.05	0.03	0.12
	W176R (A>G)	0.08	0.11	0.21	0.05
	K175M (T>A)	0.10	0.07	0.17	0.05
	V128I (C>T)	0.45	0.23	0.71	0.19
	A69T (C>T)	0.10	0.07	0.21	0.10
	R63G (T>C)	1.00	1.00	1.00	1.00
	M58I (C>T)	0.10	0.07	0.08	0.19
	M58V (T>C)	0.13	0.02	0.03	0.05
	A47P (C>G)	0.50	0.27	0.75	0.19
	L43F (G>A)	0.50	0.27	0.74	0.19
	W30C (C>G)	0.10	0.09	0.21	0.08
	T17S (T>A)	0.00	0.00	0.11	0.01
	P12L (G>A)	0.13	0.02	0.00	0.00
	S11A (A>C)	0.13	0.02	0.00	0.00
<i>Cyp6p3</i>	D418E (G>C)	0.35	0.61	0.00	0.00
	L292V (A>C)	1.00	1.00	1.00	1.00
	S287T (C>G)	0.08	0.02	0.76	0.11
	I158L (T>G)	0.98	1.00	1.00	1.00
	D155N (C>T)	0.30	0.30	0.00	0.00
<i>Cyp6p4</i>	R473H (C>T)	0.10	0.16	0.00	0.00
	I333V (T>C)	0.10	0.05	0.75	0.10
	V177L (C>G)	0.15	0.25	0.22	0.17
	L161F (C>G)	0.35	0.61	0.00	0.00
<i>Gste2</i>	T154S (G>C)	0.90	0.80	0.71	0.75
	K92R (T>C)	0.15	0.14	0.28	0.16
	G26S (C>T)	0.15	0.20	0.35	0.25
	N3K (G>[C,T])	0.98	0.95	0.86	0.88
<i>Cpr</i>	K75N (G>T)	0.00	0.00	0.51	0.01
	T319N (C>A)	0.70	0.30	0.00	0.01
	A583S (G>T)	0.18	0.41	0.00	0.01
<i>Cyp9k1</i>	D350E (G>T)	0.00	0.00	0.23	0.06

**Table 1.** Frequencies of non-synonymous SNPs in insecticide resistance-implicated genes across the three study districts, with Kimpese district separated into its two sub-districts.

## Discussion

Despite the high burden of malaria in DRC, research on the major *Anopheles* vectors lags behind that from many other sub-Saharan African countries. The present study is the first population genomic investigations of insecticide resistance in *Anopheles gambiae* from the DRC. A striking finding was that of pronounced difference in population structure between the south-western sample site of Kimpese and the central and south-eastern locations of Mikalayi and Kapolowe, respectively. Despite comparable inter-sample geographic distance, *An.*

Gene name	Kimpese		Kapolowe	Mikalayi
	Yanga Diansonga	Kilueka		
<i>CYP9K1</i>	1.00	1.00	0	0.04
<i>CYP6aa1</i>	0.05	0	0.93	0.65
<i>CYP6aa2</i>	0	0	0.51	0.43
<i>Cpr</i>	0	0	0.58	0
<i>GSte2</i>	0.65	0.59	0.27	0.61
<i>CYP6P3</i>	0	0	0	0.01
<i>CYP6P4</i>	0	0	0.02	0.01

**Table 2.** Frequencies of known copy number variants in candidate insecticide resistance genes.

*gambiae* from Mikalayi and Kapolowe grouped together in PCAs of chromosome 3 and X, and differentiation present on chromosome 2 was mostly attributable to inversion frequency variation. Marked differentiation between locations is uncommon in *An. gambiae*, with broad scale homogeneity across vast spatial scales<sup>12,13,21</sup>. In this study, genomic differentiation appeared to be concentrated mostly in regions with genes associated with insecticide resistance and, in the case of *Cpr*, a gene with a mechanism for plausible resistance involvement. This difference is most striking between Kimpese: non-Kimpese sites, mirroring the trend of increased resistance in Kimpese compared to the other sites. Additionally, some differentiation within Kimpese is also concentrated in resistance-associated genomic locations, suggesting that barriers to gene flow may be raised by differential exposure to insecticides within Kimpese.

Resistance to pyrethroid insecticides was present in all sites, with strength of resistance varying depending on location, including between adjacent sub-sites. Several mechanisms which could underpin pyrethroid resistance were identified in the genome scans and candidate gene SNP analysis. Consistently strong H12 signals around *Vgsc* were detected in each location. High frequencies of the 995F mutation were detected with complementary presence of 995S mutants and complete absence of wild type leucine alleles, such that all individuals possessed a genotype expected to confer some level of knockdown resistance to pyrethroids<sup>22</sup>. Although, with the exception of the R254K mutation, which is present on the 995F haplotype background and has an unclear role in resistance<sup>18</sup>, other variants were absent.

Additional proven and potential mechanisms of pyrethroid resistance were identified in the genome scans and candidate SNP analysis but were variable among populations. Each site showed H12 peaks centred on the *Cyp6* cluster of P450 genes on chromosome arm 2R although this was less pronounced in the Kimpese subsites Kilueka and Yanga Diansonga than in Kapolowe and especially Mikalayi. A strong, sharp *F<sub>ST</sub>* peak between Kapolowe and Mikalayi centred on this gene cluster suggests that the causal variation(s) underlying the H12 signal may have been different between populations. While *Cyp6p4* frequencies were very low in each population, moderate-to-high CNV frequencies affecting the *Cyp6aa* genes were identified, which are also associated with enhanced pyrethroid resistance<sup>6,17</sup>. *Cyp9k1* has previously been associated with pyrethroid resistance in genome scans<sup>12,23</sup> and transcriptomic association studies, and has been shown to metabolise pyrethroids<sup>24</sup>. Here, the H12 signal at *Cyp9k1* was restricted to the two Kimpese sub-locations, which were also the most resistant to pyrethroids. Moreover, whilst non-synonymous variation was not found in *Cyp9k1*, a CNV was fixed in Kimpese, but absent or nearly so, in Mikalayi and Kapolowe. Whilst, as with many CNVs in *An. gambiae* detoxification-linked genes, conclusive evidence of causal relationships with resistance is lacking<sup>25</sup>, these results are consistent with a link between the *Cyp9k1* CNV and pyrethroid resistance.

A novel potential resistance mechanism was also highlighted in the H12 scans for chromosome X, with peaks of variable strength around the *Cpr* gene in three of the four locations. *Cpr*, also known as NADPH cytochrome P450 reductase (AGAP000500), is an essential electron donor in P450-mediated reactions, including insecticide metabolism. Increased *Cpr* expression has been implicated in insecticide resistance in *Culex*<sup>26</sup>, and knockdown of the gene has been demonstrated to elevate permethrin susceptibility in *An. gambiae*<sup>27</sup>. Differentiation in *F<sub>ST</sub>* plots and frequency differences in non-synonymous variants were also evident among samples at this locus, with near absence of each of the three SNPs identified in Mikalayi, in which no H12 signal is present, intermediate frequency of a K75N mutant in Kapolowe, and also of the T319N and A583S mutants in the Kimpese subsamples. Correspondence of these variants with patterns of pyrethroid resistance across the populations and H12 signal suggest further exploration of their role as possible resistance markers is warranted. The final major H12 signal, which was present only in the Kimpese Yanga Diansonga subsite on chromosome 3L, covered several genes, including a cluster of *Cyp9j* P450s. *Aedes aegypti* *Cyp9j* P450s have been shown to metabolise pyrethroids<sup>28</sup> and *Cyp9j5* is also a weak metaboliser of pyriproxyfen<sup>29</sup>.

Beyond P450s and associated genes (*Cpr*), we also detected the presence of CNVs covering epsilon glutathione-S-transferases. Variants in, or overexpression of, these genes, particularly *GSTe2*, are most frequently associated with DDT resistance<sup>16,30,31</sup> but have also been linked with resistance to pyrethroids<sup>31,32</sup> and organophosphates<sup>16</sup>. Though H12 signals that centre on the eGST cluster have been detected in multiple *An. gambiae*



populations<sup>12</sup>, we detected little evidence in our genome scans, suggesting that variation in these genes may be less important in southern DRC populations.

Overall our results highlight the value of newer generation nets, which either evade pyrethroid mechanisms by inclusion of an additional insecticide or bypass metabolic resistance mechanisms, such as *Cyp6 P450s*, *Cyp9k1* and *Cpr* by inclusion of the synergist PBO. The results also highlight that geographic proximity may be a poor predictor of patterns of pyrethroid resistance and associated mechanisms, with distance an unreliable proxy for genomic differentiation, and that specific investigation of resistance in different localities will be required.

## Conclusion

Our study gives a first insight into the genomic landscape of resistance in samples from multiple highly-resistant populations from the DRC. The selective signals detected appear to focus on genes and variants known or plausibly linked to pyrethroid resistance; strong resistance to three types of pyrethroids was described in all study sites, with the caveat that the samples sequenced were not the same samples used for genomic analyses. The level of resistance and strong selection signals seen are consistent with the heavy use of pyrethroids via country-wide ITN distribution programmes in DRC. In very large countries with logistically-difficult travel, broad scale sentinel site-based surveillance of insecticide resistance is challenging and molecular surveillance of resistance, which can be applied to samples collected and preserved for later analysis, has particular value. Several of the regions highlighted in our analysis already have candidate DNA markers, or we have detected novel variants which can be tested for applicability as part of a diagnostic panel. Implementation of such panels, or, where possible whole genome sequencing based surveillance to monitor spatial variation to help guide ITN-type distribution decisions, and to assess changes over time in areas where particular products are distributed, represents a key aim for operational decision making and resistance management.

## Methods

### Sample collection

Three study sites were chosen to cover a breadth of geographical locations across southern DRC: Kimpese, located in the Southwest, Mikalayi in the South-Central and Kapolowe in the South-East. Within each of these, three sub-sites (health areas) were identified for collections in 2019 and 2020. The locations and their geographical coordinates are provided in Table S1. Adult *Anopheles* were collected in the early morning (06:00–09:00) from houses at each site using pyrethrum spray catches (PSC), according to the standard PMI-Vectorlink protocol<sup>33</sup>. Female *Anopheles* were identified using keys<sup>34</sup> and stored for later analysis.

### Insecticide bioassays

Larvae were collected from various water sources at each site and reared locally to adults for insecticide resistance testing using standard WHO tube bioassay procedures<sup>35</sup> with pyrethroid papers each at the diagnostic concentrations: deltamethrin 0.05%; alpha-cypermethrin 0.05%; and permethrin 0.75%. Between 80 and 100 3–5 day old adult females were tested for each subsite and insecticide in four tubes alongside a control tube containing 25 females. After a one-hour exposure period, females were transferred back to their holding tubes and maintained with sugar water for 24 h when final mortality was assessed. 95% confidence intervals were calculated using the Newcombe method<sup>36</sup>. Results were compared among locations and insecticides using a binomial generalised linear model.

### Whole genome sequencing (WGS)

A total of 253 collected adult *An. gambiae s.l* samples were sequenced on an Illumina HiSeq (150 bp paired-end reads) as part of the MalariaGEN Vector Observatory release 3.5, which comprises samples from eight of the nine collection sites; Viza in Kimpese was dominated by *An. funestus*, providing too few *An. gambiae* for sequencing. Full details of library preparation, sequencing, alignment, SNP calling, CNV calling and phasing are detailed on the Ag1000G website (<https://malariagen.github.io/vector-data/ag3/methods.html>). Samples with coverage < 10X on an Illumina HiSeq (150 bp paired-end reads) were removed. Sex was checked using the modal coverage ratio between chromosomes X and 3R (ratio between 0.4 and 0.6 = male, the ratio between 0.8 and 1.2 = female, other ratios would lead to sample exclusion).

### Read alignment and variant calling

Reads were aligned to the AgamP4 PEST reference genome using BWA<sup>37</sup>, and indel realignment was performed using GATK version 3.7–0<sup>38</sup>. Genotypes were called for each sample independently using GATK version 3.7–0 UnifiedGenotyper in genotyping mode, given all possible alleles at all genomic sites where the reference base was not 'N'. Additional details regarding thresholds and detailed information regarding read alignment, SNP calling, CNV calling and haplotype phasing are detailed on the Ag1000G website (<https://malariagen.github.io/vector-data/ag3/methods.html>). This material includes detailed information on assignment of species within the *An. gambiae* species complex using ancestry informative markers and PCA. The CartoDB library was used to plot the map layer, in Python and R.

### Population genomics

Population genomic analysis was performed using the malariagen\_data Ag3 API (<https://malariagen.github.io/malariagen-data-python/latest/Ag3.html>). Genetic diversity was assessed using nucleotide diversity ( $\pi$ ), which assesses the average number of pairwise nucleotide differences between DNA sequences in the sample; Watterson's estimator ( $\theta$ ), which is a measure reflecting the number of segregating (i.e. polymorphic) sites in the sample;

and Tajima's  $D$ , which is calculated by comparing the two previous estimators, and for which negative or positive values indicate an excess or scarcity of rare alleles, respectively, relative to neutral model expectations. Genetic differentiation among individuals from each site was assessed using principal components analysis (PCA) run separately for each chromosome arm.

### SNP and CNV analysis

Frequencies of known mutants, either non-synonymous single nucleotide polymorphisms or copy number variants, were calculated for known or candidate resistance markers in a selection of genes, including the voltage-gated sodium channel (*Vgsc*) and gaba-gated chloride channel (*Rdl*) target sites, the P450 genes *Cyp4jJ5*, *Cyp6p4*, *Cyp6aa1*, *Cyp9k1*, and several Epsilon GSTs, as well as a relatively novel candidate, the P450 redox partner cytochrome P450 reductase, *Cpr*.

### Signals of selection analyses

Signatures of selection were assessed by genomewide scans using two statistical methods. The first,  $H_{12}$ <sup>39</sup>, was calculated using phased biallelic SNPs in 1000 SNP windows, using the `garuds_h` implementation in `scikit-allele`<sup>40</sup>.  $H_{12}$  measures expected haplotype homozygosity when the two most common haplotypes within a single population are merged. Strong selection, primarily acting on one or two haplotypes, will increase the frequency and positivity of the  $H_{12}$  statistic. When present in multiple contiguous windows, this produces a pronounced peak in regions undergoing a selective sweep. The second statistic,  $F_{ST}$  (estimated using the method of Patterson in Bhatia et al.<sup>41</sup> via the `moving_patterson_fst` function in `scikit-allele`), compares allele frequencies in each window between populations; pronounced peaks suggest differentiation in that area of the genome, which may indicate differential selection between the populations.

### Data availability

The authors confirm that all supporting data, code, and protocols are provided or referenced in the article or through supplementary data files. Code and data used in this study can be accessed at [https://colab.research.google.com/drive/12dsseyUnndHqaMOzc\\_R2-IN7OWInN3II?usp=sharing](https://colab.research.google.com/drive/12dsseyUnndHqaMOzc_R2-IN7OWInN3II?usp=sharing). All sequence data generated in this study have been deposited in the European Nucleotide Archive (ENA) under project accession number PRJEB2141. Individual accession numbers are listed in Table S1.

Received: 19 March 2024; Accepted: 22 August 2024

Published online: 18 September 2024

### References

- World Health Organization. *World Malaria Report 2023* (World Health Organization, 2023).
- Irish, S. R., Kyalo, D., Snow, R. W. & Coetzee, M. Updated list of *Anopheles* species (Diptera: Culicidae) by country in the Afrotropical region and associated islands. *Zootaxa* <https://doi.org/10.11646/zootaxa.4747.3.1> (2020).
- Watsenga, F. et al. Nationwide insecticide resistance status and biting behaviour of malaria vector species in the Democratic Republic of Congo. *Malar. J.* **17**, 129. <https://doi.org/10.1186/s12936-018-2285-6> (2018).
- The PMI VectorLink Project. The Democratic Republic of Congo annual entomological monitoring report, January–December 2020, USAID, Rockville, MD. <https://uat.pmi.gov/wp-content/uploads/2021/08/Entomological-Monitoring-Report-DRC-2020.pdf> Accessed on 01 Feb 2024 (2021).
- The PMI VectorLink Project. The Democratic Republic of Congo annual entomological monitoring report, January–December 2021, USAID, Rockville, MD. <https://www.pmi.gov/wp-content/uploads/2022/05/Entomological-Monitoring-Report-DR-Congo-2021.pdf> Accessed on 01 Feb 2024 (2022).
- Njoroge, H. et al. Identification of a rapidly-spreading triple mutant for high-level metabolic insecticide resistance in *Anopheles gambiae* provides a real-time molecular diagnostic for antimalarial intervention deployment. *Mol. Ecol.* **31**, 4307–4318. <https://doi.org/10.1111/mec.16591> (2022).
- Karemere, J. et al. Associating the scale-up of insecticide-treated nets and use with the decline in all-cause child mortality in the Democratic Republic of Congo from 2005 to 2014. *Malar. J.* **20**, 241. <https://doi.org/10.1186/s12936-021-03771-6> (2021).
- U.S. President's Malaria Initiative, Democratic Republic of the Congo Malaria Operational Plan FY 2024. <https://www.pmi.gov/wp-content/uploads/2024/04/FY-2024-DR-Congo-MOP-1.pdf> Accessed on 01 May 2024 (2024).
- Hien, A. S. et al. Evidence supporting deployment of next generation insecticide treated nets in Burkina Faso: Bioassays with either chlorfenapyr or piperonyl butoxide increase mortality of pyrethroid-resistant *Anopheles gambiae*. *Malar. J.* **20**, 406. <https://doi.org/10.1186/s12936-021-03936-3> (2021).
- Gleave, K., Lissenden, N., Richardson, M., Choi, L. & Ranson, H. Piperonyl butoxide (PBO) combined with pyrethroids in insecticide-treated nets to prevent malaria in Africa. *Cochrane Database Syst. Rev.* **2018**, CD012776 (2018).
- World Health Organization, WHO Guidelines for Malaria. <https://www.who.int/publications/i/item/guidelines-for-malaria> Accessed on 01 Feb 2024 (2023).
- Anopheles gambiae* 1000 Genomes Consortium. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature* **552**, 96–100. <https://doi.org/10.1038/nature24995> (2017).
- Anopheles gambiae* 1000 Genomes Consortium. Genome variation and population structure among 1142 mosquitoes of the African malaria vector species *Anopheles gambiae* and *Anopheles coluzzii*. *Genome Res.* **30**, 1533–1546. <https://doi.org/10.1101/gr.262790.120> (2020).
- World Health Organization. Standard operating procedure for testing insecticide susceptibility of adult mosquitoes in WHO tube tests. <https://www.who.int/publications-detail-redirect/9789240043831> Accessed on 01 Feb 2024. (2024).
- Ibrahim, S. S. et al. Pyrethroid resistance in the major malaria vector *Anopheles funestus* is exacerbated by overexpression and overactivity of the P450 CYP6AA1 across Africa. *Genes* <https://doi.org/10.3390/genes9030140> (2018).
- Adolfi, A. et al. Functional genetic validation of key genes conferring insecticide resistance in the major African malaria vector, *Anopheles gambiae*. *Proc. Natl. Acad. Sci.* **116**, 25764–25772. <https://doi.org/10.1073/pnas.1914633116> (2019).
- Lucas, E. R. et al. Genome-wide association studies reveal novel loci associated with pyrethroid and organophosphate resistance in *Anopheles gambiae* and *Anopheles coluzzii*. *Nat. Commun.* <https://doi.org/10.1038/s41467-023-40693-0> (2023).
- Clarkson, C. S. et al. The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors *Anopheles gambiae* and *Anopheles coluzzii*. *Mol. Ecol.* **30**, 5303–5317. <https://doi.org/10.1111/mec.15845> (2021).

19. Grau-Bové, X. *et al.* Evolution of the insecticide target *Rdl* in African *Anopheles* Is driven by interspecific and interkaryotypic introgression. *Mol. Biol. Evol.* **37**, 2900–2917. <https://doi.org/10.1093/molbev/msaa128> (2020).
20. Weetman, D. *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.* <https://doi.org/10.1038/s41598-018-21265-5> (2018).
21. Schmidt, H. *et al.* Transcontinental dispersal of *Anopheles gambiae* occurred from West African origin via serial founder events. *Commun. Biol.* **2**, 473. <https://doi.org/10.1038/s42003-019-0717-7> (2019).
22. Reimer, L. *et al.* Relationship between *kdr* mutation and resistance to pyrethroid and DDT insecticides in natural populations of *Anopheles gambiae*. *J. Med. Entomol.* **45**, 260–266 (2008).
23. Main, B. J. *et al.* Complex genome evolution in *Anopheles coluzzii* associated with increased insecticide usage in Mali. *Mol. Ecol.* **24**, 5145–5157. <https://doi.org/10.1111/mec.13382> (2015).
24. Vontas, J. *et al.* Rapid selection of a pyrethroid metabolic enzyme CYP9K1 by operational malaria control activities. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4619–4624. <https://doi.org/10.1073/pnas.1719663115> (2018).
25. Lucas, E. R. *et al.* A high throughput multi-locus insecticide-resistance marker panel for tracking resistance emergence and spread in *Anopheles gambiae*. *Sci. Rep.* <https://doi.org/10.1038/s41598-019-49892-6> (2019).
26. Gong, Y., Li, T., Li, Q., Liu, S. & Liu, N. The central role of multiple P450 genes and their co-factor CPR in the development of permethrin resistance in the mosquito *Culex quinquefasciatus*. *Front. Physiol.* <https://doi.org/10.3389/fphys.2021.802584> (2022).
27. Lycett, G. J. *et al.* *Anopheles gambiae* P450 reductase is highly expressed in oenocytes and in vivo knockdown increases permethrin susceptibility. *Insect Mol. Biol.* **15**, 321–327. <https://doi.org/10.1111/j.1365-2583.2006.00647.x> (2006).
28. Stevenson, B. J., Pignatelli, P., Nikou, D. & Paine, M. J. I. Pinpointing P450s associated with pyrethroid metabolism in the dengue vector, *Aedes aegypti*: Developing new tools to combat insecticide resistance. *PLoS Negl. Trop. Dis.* **6**, e1595. <https://doi.org/10.1371/journal.pntd.0001595> (2012).
29. Yunta, C. *et al.* Pyriproxyfen is metabolized by P450s associated with pyrethroid resistance in *An. gambiae*. *Insect Biochem. Mol. Biol.* **78**, 50–57. <https://doi.org/10.1016/j.ibmb.2016.09.001> (2016).
30. Mitchell, S. N. *et al.* Metabolic and target-site mechanisms combine to confer strong ddt resistance in *Anopheles gambiae*. *PLoS One* **9**, e92662. <https://doi.org/10.1371/journal.pone.0092662> (2014).
31. Riveron, J. M. *et al.* A single mutation in the *GSTe2* gene allows tracking of metabolically based insecticide resistance in a major malaria vector. *Genome Biol.* **15**, R27. <https://doi.org/10.1186/gb-2014-15-2-r27> (2014).
32. Lucas, E. R. *et al.* Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. *Genome Res.* **29**, 1250–1261. <https://doi.org/10.1101/gr.245795.118> (2019).
33. The PMI VectorLink Project. Tools and Innovations—The U.S. President’s Malaria Initiative VectorLink Project. <https://pmivectorlink.org/resources/tools-and-innovations/>. Accessed on 01 Feb 2024. (2024).
34. Coetzee, M. Key to the females of Afrotropical *Anopheles* mosquitoes (Diptera: Culicidae). *Mala. J.* **19**, 70. <https://doi.org/10.1186/s12936-020-3144-9> (2020).
35. World Health Organization, “Test procedures for insecticide resistance monitoring in malaria vector mosquitoes, 2nd ed. <https://fctc.who.int/publications/i/item/9789241511575>. Accessed on 29 Jan 2024. (2016).
36. Newcombe, R. G. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat. Med.* **17**, 857–872 (1998).
37. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595. <https://doi.org/10.1093/bioinformatics/btp698> (2010).
38. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. <https://doi.org/10.1101/gr.107524.110> (2010).
39. Garud, N. R. & Petrov, D. A. Elevated linkage disequilibrium and signatures of soft sweeps are common in *Drosophila melanogaster*. *Genetics* **203**, 863–880. <https://doi.org/10.1534/genetics.115.184002> (2016).
40. A. Miles *et al.*, *cggh/scikit-allele: v1.3.7*. <https://doi.org/10.5281/zenodo.10876220> (2023).
41. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: The impact of rare variants. *Genome Res.* **23**, 1514–1521. <https://doi.org/10.1101/gr.154831.113> (2013).

## Author contributions

T.P.W.D performed the analysis and wrote the main manuscript text. P.P. performed the analysis and wrote the main manuscript text. S.B. performed the analysis. L.T.C. performed the analysis and wrote the main manuscript text. E.R.L. provided conceptual and technical support. C.C. provided conceptual support and oversaw integration of data into MalariaGEN. A.M. provided conceptual support and oversaw integration of data into MalariaGEN. A.A. collected the samples and provided conceptual support. E.Z.M collected the samples. E.M. provided conceptual support. J.Z. provided conceptual support. S.N. collected the samples. S.R.I. provided conceptual support. M.J.D. provided conceptual support. F.A. supervised the study and provided conceptual support. D.W. supervised the study, performed the analysis and provided conceptual support. F.W.T. supervised the study, collected the samples and provided conceptual support. All authors reviewed the manuscript.

## Funding

This work was supported by the Bill and Melinda Gates Foundation (OPP1210331) and the National Institute of Allergy and Infectious Diseases (NIAID R01-AI116811).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-70885-7>.

**Correspondence** and requests for materials should be addressed to T.P.W.D. or D.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** 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/>.

© Crown 2024