SCIENTIFIC
REPORTS
natureresearch

OPEN

A high throughput multi-locus insecticide resistance marker panel for tracking resistance emergence and spread in *Anopheles gambiae*

Eric R. Lucas [1], Kirk A. Rockett [2], Amy Lynd[1], John Essandoh[1,3], Nelson Grisales[1,4], Brigid Kemei[5], Harun Njoroge[1], Christina Hubbart[2], Emily J. Rippon[1], John Morgan[1], Arjen E. Van't Hof [1], Eric O. Ochomo [5], Dominic P. Kwiatkowski[2,6], David Weetman[1] & Martin J. Donnelly[1,6]

The spread of resistance to insecticides in disease-carrying mosquitoes poses a threat to the effectiveness of control programmes, which rely largely on insecticide-based interventions. Monitoring mosquito populations is essential, but obtaining phenotypic measurements of resistance is laborious and error-prone. High-throughput genotyping offers the prospect of quick and repeatable estimates of resistance, while also allowing resistance markers to be tracked and studied. To demonstrate the potential of highly-mulitplexed genotypic screening for measuring resistance-association of mutations and tracking their spread, we developed a panel of 28 known or putative resistance markers in the major malaria vector *Anopheles gambiae*, which we used to screen mosquitoes from a wide swathe of Sub-Saharan Africa (Burkina Faso, Ghana, Democratic Republic of Congo (DRC) and Kenya). We found resistance association in four markers, including a novel mutation in the detoxification gene *Gste2* (*Gste2*-119V). We also identified a duplication in *Gste2* combining a resistance-associated mutation with its wild-type counterpart, potentially alleviating the costs of resistance. Finally, we describe the distribution of the multiple origins of *kdr* resistance, finding unprecedented diversity in the DRC. This panel represents the first step towards a quantitative genotypic model of insecticide resistance that can be used to predict resistance status in *An. gambiae*.

It is estimated that the use of insecticide-treated nets and indoor residual spraying of insecticide has been responsible for 78% of 663 million cases of malaria averted in the period 2000–2015[1]. The documented rise and spread of resistance to currently-used insecticides in the major malaria mosquito vectors therefore presents a worrying trend that may negatively impact malaria control programmes[2–4]. In particular, the major malaria vector in Sub-Saharan Africa (SSA), *Anopheles gambiae*, now has widespread resistance to pyrethroids, and is showing pockets of high resistance to carbamates and organophosphates[5]. To maintain effective malaria control, accurate assessment of the insecticide resistance profile of mosquito populations is essential. Furthermore, as new insecticides are brought to market, it will be crucial to identify and track resistance to these new products in order to react appropriately before the failure of malaria control.

Phenotypic assessment of insecticide resistance requires live mosquitoes of fixed age to be assayed in carefully controlled conditions. The effort and care required to perform these assays make them difficult to perform on a large scale and prone to inconsistency if conditions are not rigorously controlled[6]. Moreover, phenotyping approaches are only sensitive when resistance has reached appreciable frequency in the population. For these reasons, there is great interest in the prospect of screening mosquitoes for the genetic signatures of insecticide

[1]Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK. [2]Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK. [3]Department of Wildlife and Entomology, University of Cape Coast, Cape Coast, Ghana. [4]Abt Associates, 6130 Executive BLVD, Rockville, MD, United States. [5]Kenya Medical Research Institute (KEMRI), Centre for Global Health Research, Kisumu, Kenya. [6]Wellcome Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK. Correspondence and requests for materials should be addressed to E.R.L. (email: eric.lucas@lstmed.ac.uk)

resistance, which would provide a convenient and simple way of assessing resistance from any type of mosquito collection method, including dead mosquitoes routinely collected by monitoring and evaluation programmes. Genetic screening carries the additional benefit of improving our understanding of the origins of resistance mutations, which can inform resistance management policies. For example, by understanding where resistance originates, research can begin to elucidate the environmental and population-level contexts which favour its rise and spread. This has been seen in the evolution of drug resistance in the malaria parasite, which repeatedly appears and spreads from an area in Cambodia[7], driving research into understanding the causes of this pattern. Our aim is therefore to develop a highly-multiplexed genotypic panel that can be used in the first instance to quantify the insecticide-resistance association of mutations, and in the second instance to track these mutations across mosquito populations.

There are two main challenges involved in the development of genetic screening assays for insecticide resistance. First, genetic markers of resistance need to be identified and their effect size estimated. Several single nucleotide polymorphism (SNP) mutations associated with insecticide resistance have already been discovered in *Anopheles*[8–15], providing the first opportunities for genetic screening. However, a substantial proportion of the variance in resistance remains unexplained, making it crucial to identify new markers and to quantify their impact.

Second, new approaches are required for large-scale, high-throughput screening of large marker panels, to reduce the costs and time required to screen mosquito populations. The Agena Biosciences iPLEX MassARRAY technology is a platform capable of accurate genotyping combined with high level multiplexing capability. MassARRAY panels have already been developed for a range of infectious disease questions, such as studying resistance to malaria in humans[16], mapping genetic diversity in parasites[17] and detecting ancestry and inter-specific hybridisation in *Anopheles*[18].

We present the development and application of an iPLEX MassARRAY panel of 28 genotypic markers for screening *Anopheles gambiae* populations to quantify the effects of known and putative insecticide resistance markers, and to track the origins of spread of knockdown resistance variants (*kdr*) in the Voltage-gates sodium channel (*Vgsc*) gene. We applied our panel to samples from four different countries from West (Burkina Faso and Ghana), Central (Democratic Republic of Congo - DRC) and East Africa (Kenya). These samples were assayed for susceptibility to permethrin (all four countries), deltamethrin (Ghana, DRC and Kenya) and DDT (Ghana). This genetic screen:

1. revealed a heterologous duplication in *Gste2* that combines resistance-associated and wild-type alleles,
2. identified an unprecedented diversity of *kdr* genetic backgrounds in the DRC,
3. reported the first evidence of a role for the *Gste2*-119V mutation in insecticide resistance in *An. gambiae*.
4. found evidence that a mutation found alongside *kdr* reduces insecticide resistance, possibly indicating that this mutation serves to compensate for physiological costs of *kdr*.

Methods

Sample collection.   *Samples from Burkina Faso.*   *Anopheles* larvae were collected using hand dippers from semi-permanent and temporary water bodies between June and October 2013 at four sites (Bakaridjan: 10.407 N, −4.562 W; Bounouba: 10.357 N, −4.439 W; Naniagara: 10.536 N, −4.669 W and Tiefora: 10.632 N, −4.556 W). Samples were transported to the insectaries in Banfora where they were maintained at a temperature of 27 °C (±2 °C) and a relative humidity of 80% (±10%), and fed with TetraMin Baby**®**. Adult mosquitoes were tested at a range of permethrin concentrations (5–120 ppm) using an adaptation of the CDC bottle bioassay[19] where we exposed the mosquitoes for 60 minutes and recorded mortality after 24 hours. For this study, we selected mosquitoes exposed to 20 ug/ml because this was the concentration that came closest to killing 50% of the mosquitoes (bioassay results at 20 ppm shown in Supplementary Table S1). All mosquitoes were stored individually in perforated PCR tubes and placed into sealed bags with silica gel to avoid the decomposition and ensure DNA preservation. DNA was extracted using the LIVAK method[20] and samples were identified to species using SINE PCR[21].

Samples from DRC.   Samples from the DRC were the same as previously reported[22]. *An. gambiae* mosquitoes were collected in March and April 2016 from three rural collection sites (Pambwa, 3.937 N, 20.772 E; Fiwa, 4.318 N, 20.778 E; Bassa, 4.267 N, 21.283 E) in the area surrounding the major town of Gbadolite, near the border with the Central African Republic. Adult mosquito collections were carried out in all three villages using both manual and mechanical ('Prokopack'[23]) aspirators and were maintained in a field insectary until egg-laying. Larvae were collected from breeding sites in Fiwa and Pambwa. Larvae collected directly, and those raised from eggs, were reared until the adult stage. All mosquitoes were identified to species group using phenotypic keys[24] and insecticide resistance testing was carried out on 3–5 day-old adult *An. gambiae s.l.* to assess resistance to permethrin (0.75%) and deltamethrin (0.05%) using standard World Health Organization (WHO) protocols[25] (overall bioassay results for the population shown in Supplementary Table S2). All mosquitoes were stored on silica gel in 0.2 ml tubes for later DNA analyses. DNA was extracted from individual mosquitoes using Nexttec (Nexttec, Biotechnologie GmbH) extraction plates according to manufacturer's instructions, and mosquitoes were identified to species using SINE PCR[21].

Samples from Ghana.   Ghanaian samples were collected from Keta (5.917 N, 0.991 E), an urban community in the Volta Region of Ghana. Keta is located within the coastal savannah agroclimatic zone, with vegetation consisting of shrubs, grasses and a few scattered trees.

An. gambiae s.l. larvae were collected from September to November, 2016, using hand-held ladles. Collections were made from a variety of habitats such as pools, puddles, drainage channels, irrigations and vegetable fields,

and transported to the laboratory in partly-filled labelled plastic containers. Larvae were reared in the insectaries at the Animal Science department of the Biotechnology and Nuclear Agricultural Research Institute (BNARI), Accra. Insecticide resistance assays were performed on 3–5 day-old females following standard WHO tube assay protocols (0.05% deltamethrin, 0.75% permethrin or 4% DDT; overall bioassay results for the population shown in Supplementary Table S3).

DNA was extracted from single mosquitoes using the Nexttec 96-well plate DNA Isolation kit. Species identification was performed on the whole-genome-amplified DNA (Section 2.2) using standard species identification PCR[26] followed by SINE PCR to further distinguish *An. gambiae* from *An. coluzzii*[21].

Samples from Kenya. Mosquitoes were sampled from four malaria-endemic regions in Western Kenya (Teso: 0.656 N, 34.353 E, Bondo: −0.098 S, 34.273 E, Rachuonyo: −0.353 S, 34.656 E and Nyando: −0.162 S, 34.921 E). These regions were under distinct vector control interventions, with Teso and Bondo having insecticide treated nets and Nyando and Rachuonyo having both insecticide treated nets and indoor residual spraying. *Anopheles* larvae were collected from aquatic habitats using the standard dipping method and transferred to plastic tins using wide-mouth pipette for transportation to KEMRI, Kisumu laboratories for rearing. Larvae were reared on a mixture of fish food and brewer's yeast provided daily. Upon pupation, individuals were transferred to cages to emerge as adults and provided with 10% sucrose in cotton pledgets.

Three-day-old adult females were exposed to either permethrin (0.75%) or deltamethrin (0.05%) using WHO impregnated papers for 1 hour following WHO guidelines[27] (overall bioassay results for the population shown in Supplementary Table S4). Mortality was recorded and all mosquitoes were placed in individual tubes and frozen at −20 °C for molecular analysis. DNA was extracted from whole samples using ethanol precipitation[28].

A newly-developed melt-curve based species identification assay[29] was used because existing techniques are prone to interpretation errors as a result of similar sized PCR bands for *An. arabiensis* vs *An. gambiae* on agarose gels. The assay uses SYBR-green with a universal forward primer (5′-ATTGCTACCACCAAAATACATGAAA -3), a reverse primer matching both *An. arabiensis* and *An. gambiae* with G-8 extension (5′-GGGGGGGGGAATAATAAGGAACTGCATTTAAT-3′) to slightly increase the amplicon melting temperature, and an *An. arabiensis* specific reverse primer (5′- GGATGTCTAATAGTCTCAATAGATG -3′).

SNP genotyping. We developed a multiplex panel of 28 SNP markers (Supplementary Data S1) based on the AgamP3 reference genome[30] and information on variable sites identified by the *An. gambiae* 1000 Genomes (Ag1000 G) project[31]. Primers for these SNPs were developed using the MassARRAY Assay Design software (version 4.0.0.2; Agena Biosciences, Hambrug, Germany).

The SNPs that we chose fell into three categories. First, we included eight SNPs that have previously been associated with insecticide resistance. These include the two *kdr* mutations (*Vgsc*-995F and *Vgsc*-995S), as well as a further *Vgsc* mutation (*Vgsc*-1570Y) known to exist on a *Vgsc*-995F background and conferring increased resistance to pyrethroids[9], two mutations associated with resistance to dieldrin (*Rdl*-296G and *Rdl*-296S[8]), a mutation in the gene *Gste2* associated with metabolic resistance to DDT (*Gste2*-114T[11]) and a mutation associated with resistance to carbamates and organophosphates (*Ace1*-280S, previously referred to as 119S[15]). The eighth SNP in this set was a mutation in codon 119 of *Gste2* (*Gste2*-119V), which was included because a mutation in the same codon has been shown to strongly increase resistance to DDT in *An. funestus*[13], raising the prospect that the mutation in *An. gambiae* may have a similar function.

Second, we included eight non-synonymous SNPs (*Vgsc*-1868T, *Vgsc*-1874S, *Vgsc*-1874L, *Vgsc*-1853I, *Vgsc*-1934V, *Vgsc*-1746S, *Vgsc*-791M and *Vgsc*-1597G) that have been found to be strongly associated with the *Vgsc*-995F mutation[31] and are thus regarded as potential candidates influencing resistance either by enhancing the impact of *Vgsc*-995F or compensating for suspected fitness effects.

Third, we included twelve SNPs that can differentiate the five haplotype backgrounds of the *Vgsc*-995F mutation (haplotype backgrounds F1-F5) and the five haplotype backgrounds of the *Vgsc*-995S mutation (haplotype backgrounds S1-S5) (Anopheles gambiae 1000 Genomes Consortium 2017). These SNPs were identified using the Ag1000G data by searching for SNPs that differentiated each haplotype background from the other haplotypes with the same *kdr* mutation. Within each of the two *kdr* mutations (*Vgsc*-995F and *Vgsc*-995S), the SNPs always provided perfect separation between the five possible haplotype backgrounds, although a few haplotypes that were not assigned any background in the Ag1000G data could be incorrectly assigned a background based on these SNPs (Supplementary Data S2).

DNA extractions were whole-genome amplified and then genotyped with the Agena Biosciences iPLEX platform as described[32], using our panel of 28 SNPs. Genotype calls were obtained from the raw data using the TYPER software (version 4, Agena Biosciences) and raw data as cluster plots were inspected and manually curated to remove ambiguous calls. To assess the quality of the assays and the genotype calling process, we also used our panel to genotype 287 samples previously used in phase 2 of Ag1000G (https://www.malariagen.net/data/ag1000g-phase-2-ar1) and compared the two sets of genotype calls. Raw data and genotype calls are provided in Supplementary Data S3, and genotype calls are also provided along with sample information in Supplementary Data S4. Samples that failed genotyping at more than 50% of loci were excluded from the analysis (this filter removed 39 samples from Ghana and two samples from Kenya).

Deviations from Hardy-Weinberg equilibrium. Tests for deviations from Hardy-Weinberg equilibrium were performed using the *HWExact* function from the R package *HardyWeinberg*. These tests were performed for every SNP × location combination in which the SNP was segregating (i.e.: where both alleles were present in the sampling location). This resulted in 72 separate tests, which were adjusted for multiple testing using a Bonferroni correction, with a resulting threshold α value of 0.0007.

Identification of *kdr* haplotype backgrounds. The process of identifying the *kdr* haplotype background of a sample based on its genotype was implemented using custom R functions that has been made available as Supplementary Materials and which is described in Supplementary Methods S1.

Statistical analysis of resistance. The two SNPs in *kdr* (*Vgsc*-995F and *Vgsc*-995S) are at different nucleotide positions but affect the same codon and are mutually exclusive. We therefore re-coded these two SNPs as a single variant with three alleles: L (wild-type), F and S. Similarly, we also re-coded the two *Rdl* SNPs as a single variant with three alleles.

Associations between genotype and insecticide resistance were tested using generalised linear models (GLM) with binomial errors and a logit link function, implemented in R using the package *lme4*[33]. Genotypes were included as categorical fixed effects. Three candidate resistance markers (*Vgsc*-1853I, *Vgsc*-1934V and *Vgsc*-1874S) were not included in the models due to the scarcity of the mutant allele ($n < 5$) in the combined data, and the marker for *Vgsc*-791M was excluded as it was almost perfectly associated with *Vgsc*-1746S (Supplementary Table S5). For the data from Ghana, where all samples came from the same location, no random effects were included and the analysis was performed using the *glm* function. For the three other countries, sampling location was included as a random effect using generalised linear mixed models (GLMM) implemented by the function *glmer*. In one case (permethrin phenotype from DRC) where the mixed models failed to converge, preventing the inclusion of random effects, the modelling was repeated by including location as a fixed effect, using the function *glm*. In Burkina Faso and DRC, the majority of samples were of the species *An. gambiae*, with very few *An. coluzzii* (7 from Burkina Faso, 2 from the DRC), and *An. coluzzii* samples were thus excluded from the analyses. Similarly, only 6 *An. arabiensis* samples were found among the samples tested for resistance to deltamethrin in Kenya and were thus excluded. *An. arabiensis* were better represented in Kenya among the samples tested for resistance against permethrin (18 *An. arabiensis* samples), and species was thus included as a random factor.

Minimal significant models were obtained by a stepwise backward elimination process using a custom R function provided in the Supplementary Materials and described in Supplementary Methods S2. Briefly, resistance markers were included together into the full model, and the significance of each marker was obtained using the *anova* function to compare the full model to the model without the marker. Stepwise, the least significant marker was removed until only significant markers were left, providing the minimal model. The *P*-values reported for the significant markers are the result of the ANOVA comparing the minimal model against the model with the marker removed.

Detection of *Gste2* duplication. The Gstue_Dup7 duplication was detected by PCR using primers designed on either side of the duplication breakpoint. The position of the breakpoint was known from previous data[34] and primers were designed using NCBI Primer Blast[35]. Further details on the primers are provided in Supplementary Methods S3.

No human subjects were involved in this research.

Results

High concordance between iPLEX MassARRAY and Ag1000G genotype calls. In total, 29,148 assays were run using the iPLEX MassARRAY panel (1,041 samples × 28 SNPs). Of these, 1,408 (4.8%) failed to produce a genotype call, with none of the individual assays having a failure rate greater than 10% (Supplementary Fig. S1). The failure rate was lowest in *An. coluzzii*, where 24 out of 7,980 assays failed (0.3%, Supplementary Fig. S2), followed by *An. gambiae*, where 274 out of 18,788 assays failed (1.46%, Supplementary Fig. S3), and highest in *An. arabiensis*, where there were 59 failures out of 672 assays (8.8%, Supplementary Fig. S4). The remaining samples were of unknown species as the DNA was not of high enough quality for the species ID assay.

Of the 8,036 individual assays performed on samples from Ag1000G (287 samples × 28 SNPs), only 5 (0.06%) iPLEX calls were divergent from the Ag1000G callset. The rest either gave concordant calls (98.2%) or failed calling by the iPLEX platform (1.7%). The high levels of concordance with the Ag1000G calls indicate that the results of the iPLEX assays can be used with confidence.

Deviation from Hardy-Weinberg equilibrium implies heterologous duplication of *Gste2* in Ghana. The SNPs *Ace1*-280S and *Gste2*-114T showed significant deviation from Hardy-Weinberg expectations in *An. coluzzii* from Ghana, caused by an excess of heterozygotes at these loci ($P < 10^{-13}$ in both cases, Table 1). Such heterozygote excess can be caused by the presence of heterologous amplifications, which can create functional heterozygotes by combining wild-type and mutant alleles on a single chromosome. For both *Gste2*-114 T and *Ace1*-280 S, the raw intensities for heterozygous samples are split into several clusters rather than the normal single cluster (Fig. 1), which is also consistent with the presence of duplicate copies of one or both alleles. The heterologous *Ace1* duplication in Ghana is already well-documented[36,37], but no such duplication has yet been reported in *Gste2*.

To confirm whether the heterozygote excess in *Gste2*-114T is caused by a duplication, we tested our samples for the presence of the duplication Gstue_Dup7, the only *Gste2* duplication known to occur in *An. coluzzii* from Ghana[34]. Gstue_Dup7 was present in 185 out of 238 (78%) of samples in our *An. coluzzii* samples from Ghana, a figure seven times higher than was found in Ag1000G, where 6 out of 55 (11%) *An. coluzzii* from Ghana had the duplication[34].

Gstue_Dup7 was strongly associated with heterozygosity at *Gste2*-114T (Fig. 1d). Out of 185 Ghanaian *An. coluzzii* samples that carried Gstue_Dup7, 183 were heterozygote at the 114T locus and 2 were homozygote mutants, while none were homozygote wild-type. In contrast, out of 53 samples that did not carry Gstue_Dup7, 31 were 114T heterozygotes, 19 were homozygote mutants and 3 were homozygote wild-type. Furthermore, the

	Gste2-114T				Ace1-280S			
	AA	AG	GG	Hardy-Weinberg P	AA	AG	GG	Hardy-Weinberg P
Burkina Faso An. gambiae	71	8	1	0.26	0	33	56	0.04
DRC An. gambiae	148	0	0	1	0	0	186	1
Ghana An. coluzzii	3	214	21	1.6×10^{-41}	2	153	83	1.4×10^{-14}
Kenya An. gambiae	139	0	0	1	0	0	155	1

Table 1. Allele frequencies for *Gste2*-114T and *Ace1*-280S show large excesses of heterozygotes in Ghana. *Gste2*-114T: wild-type allele = A, mutant = G. *Ace1*-280S: wild-type allele = G, mutant = A.

114T heterozygotes without the duplication did not show the same scatter of raw intensities as the heterozygotes that carried the duplication (Fig. 1d).

High diversity of *kdr* haplotypes in the DRC. The *Vgsc*-995 mutations were at high frequency in all four countries (Fig. 2, Table 2). The wild-type allele was found at 2% frequency in *An. gambiae* from Kenya, but was completely absent in *An. gambiae* from Burkina Faso and the DRC. In *An. coluzzii*, the wild-type codon was found at 27% frequency in Ghana, and at similar frequencies in the few samples from Burkina Faso and the DRC.

In the West African populations (Burkina Faso and Ghana), all *kdr* mutant alleles were *Vgsc*-995F of the F1 haplotype background (Fig. 2, Supplementary Data S6). In Kenya, all *kdr* mutants in *An. gambiae* carried the *Vgsc*-995S mutant allele from either of two backgrounds (S1 or S3; Supplementary Data S6). In contrast, most *kdr* haplotypes in Kenyan *An. arabiensis* were wild-type (34 out of 48), while the rest were predominantly *Vgsc*-995F (12 out of 48). These *Vgsc*-995F samples did not carry any of the SNPs associated with the five known haplotype background, possibly indicating that the *Vgsc*-995F haplotype background in *An. arabiensis* is distinct from those found to date in *An. gambiae*. The two *Vgsc*-995S haplotypes in this population could not be assigned a background due to failure of most of the haplotype background assays.

In the DRC, all five *Vgsc*-995F haplotype backgrounds were detected, the most frequent background being F5. Similarly, four of the five *Vgsc*-995S backgrounds were found in the DRC (Fig. 2, Supplementary Data S6), the most frequent background being S5, and haplotype S2 being absent. Previously, eight *kdr* haplotypes were found in Cameroon, with haplotypes S1 and S3 being absent[31].

Association with insecticide resistance. In Burkina Faso, both *Gste2* mutations were associated with increased resistance to permethrin, demonstrating for the first time an association between *Gste2*-119V and insecticide resistance (Table 3). In contrast, the 1874L *Vgsc* mutation was negatively associated with resistance to permethrin.

In the DRC, where the wild-type *Vgsc*-995 codon was absent, the *Vgsc*-995F mutation was more strongly associated with resistance to both permethrin and deltamethrin than *Vgsc*-995S, with a slight tendency for heterozygotes to have lower resistance to deltamethrin than *Vgsc*-995S homozygotes (Table 3). Seemingly contrary to the Burkina Faso data, there was a negative association between the *Gste2*-119V mutation and resistance to deltamethrin, although heterozygotes showed slightly higher resistance than homozygote wild-types.

In Ghana, no SNPs were associated with resistance to deltamethrin, while the *Vgsc*-995F mutation was associated with resistance to permethrin and the *Rdl*-296G mutation was associated with resistance to DDT (Table 3). There were only three homozygous wild-type samples at the *Vgsc*-995F locus, thus the significant difference is driven by the difference between the homozygote mutant and heterozygotes.

In Kenya, there were no significant associations between any of the SNPs and resistance to either permethrin or deltamethrin, although it is notable that *Vgsc*-995 was the only resistance candidate SNP found in this population and was nearly fixed in *An. gambiae* (Table 2 & Supplementary Data S5).

Discussion

We have developed a multiplex SNP panel for quick, high-throughput screening of large numbers of *An. gambiae s.l.* at multiple markers useful for detecting insecticide resistance, testing the importance of putative resistance diagnostic markers and tracking the spread of resistance mutations. We have allied this with a formal analytical approach for analysing phenotypic association of these markers, made available as an R function. By screening 713 mosquitoes from four different countries, we have identified a heterologous duplication in a resistance-conferring SNP, confirmed an association with resistance in a candidate SNP, and found the DRC to be an area in which nearly all evolutionary origins of the *kdr* resistance mutation co-exist.

The presence of a heterologous duplication in *Gste2* in *An. gambiae* from Ghana was first indicated by a very high excess in *Gste2*-114T heterozygotes and the characteristic multiple clusters of signal intensities in the raw data, both of which bear a striking resemblance to the pattern seen in Ghanaian *Ace1*-280S data, known to be caused by a heterologous duplication. We confirmed the role of a duplication by showing that the heterozygosity is associated with the presence of the *Gste2* duplication Gstue_Dup7[34]. While the vast majority of Gstue_Dup7 alleles are heterologous, we also found evidence of a rare variant of Gstue_Dup7 in two samples where the duplication was homozygous for the 114T SNP. So far, eleven different duplications in the cluster of genes around *Gste2*
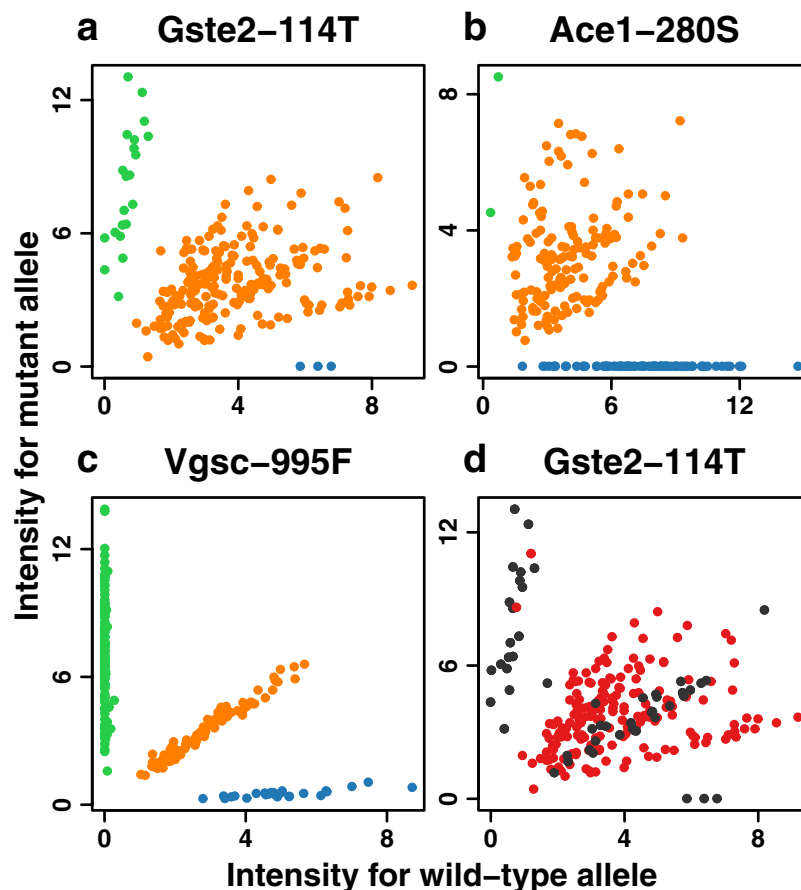
Figure 1. Raw intensities from the iPLEX MassARRAY assays in Ghana show multiple clusters among heterozygote samples (orange points) for *Gste2*-114T (**a**) and *Ace1*-280S (**b**). This is typically seen when duplications create ratios of mutant/wild-type alleles that differ from the usual 1/1 in heterozygotes. In comparison, equivalent data for *Vgsc*-995F (**c**) shows the normal pattern, with heterozygotes consistently falling in a single cluster along a straight diagonal line. Homozygote wild-type and homozygote mutant samples are shown in blue and green respectively. Colour-coding the *Gste2*-114T data according to whether samples carry Dup7 (red) or not (black) shows that heterozygotes with no duplication predominantly fall along a straight diagonal line (**d**). The X and Y axes respectively show intensities for the wild-type and mutant alleles.

have been reported in *An. gambiae* and *An. coluzzii* from across SSA[34], but only one of these, Gstue_Dup1, was shown to be associated with the 114T mutation, being homologous for the mutant allele. The evolution of duplications in *Gste2* may therefore mirror what has been found in *Ace1*, where both homologous and heterologous duplications exist. If the 114T mutation carries a physiological cost in the absence of insecticide pressure, the heterologous duplication discovered here may serve to reduce this cost, as has been found in the case of the *Ace1* duplication[38].

Gstue_Dup7 was found in 78% of our *An. coluzzii* samples from Ghana collected in 2015 from Keta (5.9175 N, 0.9916 E), which is seven times higher than was found in samples collected in 2012, albeit from different sites further West in Ghana (Koforidua (6.0945 N, 0.2609 W), Madina (5.6685 N, 0.2193 W), Takoradi (4.9122 N, 1.7740 W), Twifo Praso (5.6086 N, 1.5493 W)[34]). While we cannot exclude the possibility that this increase reflects differences between sites, the high frequency of this duplication in Keta, combined with the striking difference in frequency between samples collected in 2012 and 2015, strongly suggest that this duplication is rapidly spreading through the *An. coluzzii* population.

The high diversity of *kdr* haplotype backgrounds in the DRC points to this area of central Africa as a key to understanding the spread of the most well-characterised insecticide resistance allele in *An. gambiae*. All five *Vgsc*-995F backgrounds[31] were found in this population, as well as four of the five *Vgsc*-995S backgrounds, combining haplotypes previously found in Uganda, (S1 and S3), Kenya (S3) and Cameroon (S4 and S5). The only background not confirmed in the DRC was S2, which was previously found in Gabon and Cameroon[31]. This striking geographical co-existence of *kdr* haplotype backgrounds could have two explanations. First, most of the main *kdr* origins could have occured in central Africa, and subsequently spread from there to other countries, with different haplotypes becoming established in different regions. Second, The different *kdr* haplotypes could have spread from their respective origins and converged in the DRC, where most then persisted without being entirely replaced (although F5 is by far the most prevalent, with a frequency of over 50%). Differentiating between these
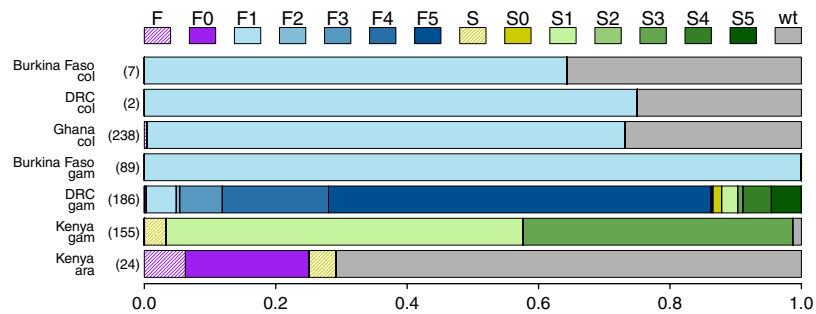
Figure 2. Frequencies of *kdr* haplotypes in the four countries from the study (total number of individuals, i.e. half the number of haplotypes, shown in brackets). The F1 haplotype is the only one present in Ghana and Burkina Faso, while haplotypes S1 and S3, along with a possible new *Vgsc*-995F haplotype, are found in Kenya. In the DRC, at least nine out of the ten *kdr* haplotypes co-exist. F and S indicate samples for which the *kdr* mutation was known, but the haplotype could not be determined because of failed calls in at least some background markers; F0 and S0 indicate haplotypes where the *kdr* mutation was known and all of the background markers produced wild-type calls, suggesting a novel haplotype background; wt = wild-type; ara = *An. arabiensis;* col = *An. coluzzii;* gam = *An. gambiae.*

Country	Species	Number of samples	*Vgsc*-995F	*Vgsc*-995S	*Rdl*-296G	*Rdl*-296S	*Ace1*-280S	*Gste2*-114T	*Gste2*-119V	*Vgsc*-1570Y
BF	*An. coluzzii*	7	0.64	0	0	0.43	0	0.57	0	0.29
DRC	*An. coluzzii*	2	0.75	0	0	0.25	0	0.50	0	0
Ghana	*An. coluzzii*	238	0.73	0	0.54	0	0.33	0.54	0	0
BF	*An. gambiae*	89	1	0	0.35	0	0.19	0.06	0.65	0.20
DRC	*An. gambiae*	186	0.86	0.14	0.05	0	0	0	0.35	0.01
Kenya	*An. gambiae*	155	0	0.98	0	0	0	0	0	0
Kenya	*An. arabiensis*	24	0.25	0.04	0	0	0	0	0	0

Table 2. Allele frequencies for the eight SNPs with known associations with insecticide resistance. We note that the sample size for *An. coluzzii* in Burkina Faso and the DRC are small and the allele frequencies will thus be imprecise. The full table showing all 28 markers in the panel is provided in Supplementary Data S5.

Country	Insecticide	Codon	P	Intercept genotype	Effect of het. vs intercept	Effect of hom. mutant vs intercept	Effect of hom. mutant vs het.
Ghana	Perm	*Vgsc*-995	0.023	LL	LF: ≥4* ↑	FF: ≥4* ↑	1.58 ↑
	Delt	NA					
	DDT	*Rdl*-296	7.1×10^{-7}	AA	AG: 2.6 ↑	GG: 2.95 ↑	0.35 ↑
Burkina Faso	Perm	*Vgsc*-1874L	0.041	PP	LP: −1.18 ↓	LL: −2.4 ↓	−1.22 ↓
		Gste2-114T	0.022	II	IT: 3.01 ↑	TT: ≥4* ↑	≥ 4* ↑
		Gste2-119V	0.026	LL	LV: ≥4* ↑	VV: ≥4* ↑	0.35 ↑
Kenya	Perm	NA					
	Delt	NA					
DRC	Perm	*Vgsc*-995	0.004	SS	SF: ≥4* ↑	FF: ≥4* ↑	1.1 ↑
	Delt	*Gste2*-119V	0.0078	LL	LV: 0.36 ↑	VV: −1.54 ↓	−1.91 ↓
		Vgsc-995	0.0038	SS	SF: −0.55 ↓	FF: 1.5 ↑	2.06 ↑

Table 3. Significant associations between genotype and insecticide resistance. Effects are shown as log odds ratio. Direction of effect compared to the reference (intercept) genotype shown by arrows (↑ indicates increase in resistance compared to the intercept; ↓ indicates decrease in resistance). In Ghana, *kdr* was either wild-type (L) or 995F. In the DRC, *kdr* was either 995S or 995F. *Effects ≥4 (corresponding to odds ratio > 50) were usually seen where there were few samples in a category, leading to all samples having the same phenotype and making odds ratios impossible to calculate.

two hypotheses will require more in-depth analysis of the sequence variability within each haplotype sequence for each background in the DRC and elsewhere.

The *Vgsc*-995F haplotype found in *An. arabiensis* from Kenya did not conform to any of the known haplotype backgrounds for this mutation, suggesting that a sixth *Vgsc*-995F origin exists, which may have originated in East

Africa. The Kenyan samples in the present study are from Western Kenya, where *Vgsc*-995F has previously been reported in both *An. arabiensis* and *A. gambiae*[39]. In contrast, the Kenyan samples included in the Ag1000G data, where the haplotype backgrounds were defined, came from an eastern Kenyan population of *An. gambiae* where *Vgsc*-995F was not found. The Kenyan *Vgsc*-995F background was therefore not present in Ag1000G and could not be included in the definition of haplotype backgrounds. Whether the *Vgsc*-995F background present in *An. gambiae* from western Kenya corresponds to one of the five known 995F backgrounds, or to the one found in *An. arabiensis* from the same area remains to be resolved.

Associations of genetic markers with insecticide resistance revealed several unexpected associations. The *Gste2*-114T marker, which has been previously associated with resistance to DDT[11] was associated with resistance to permethrin in Burkina Faso. To our knowledge, this is the first report of a role for this mutation in permethrin resistance, although one previous study found that it was also associated with resistance to another pyrethroid, deltamethrin[40]. The evidence is therefore increasing for a role of *Gste2*-114T in cross-resistance to multiple insecticides. In Ghana, there was no significant association between *Gste2*-114T and resistance to pyrethroids or DDT, but this is likely due to the very high frequency of heterozygotes, making any association difficult to detect. In the same gene, the *Gste2*-119V mutation was associated with increased permethrin resistance in Burkina Faso, but with decreased deltamethrin resistance in the DRC. These contrasting effects may reflect different changes in binding affinity to these two insecticides caused by this mutation. While the metabolic properties induced by the *Gste2*-119V mutation have not been studied, research on the *Gste2*-119F mutation in *An. funestus* showed that the mutant form was able to metabolise permethrin, but not deltamethrin[13]. The negative association with deltamethrin resistance found here is, however, novel, and may be explained by a metabolic cost imposed by the mutation, or by reduced affinity for deltamethrin in the case of the *An. gambiae* mutation. If the opposing effects on permethrin and deltamethrin resistance detected here are confirmed, it will be important to routinely screen for this mutation when making a choice of insecticides in control campaigns.

Intriguingly, *Rdl*-296G was associated with resistance to DDT in Ghana. This mutation is in the target site of cyclodiene insecticides such as dieldrin, which was used in malaria control programmes in Africa during the 1950s and 1960s, but has since been discontinued. Resistance to dieldrin has somehow been maintained in natural mosquito populations[41–43], although the selective forces causing its persistence remain a puzzle. One possibility is that cyclodienes continue to be used as insecticides in private agriculture, but another is that the *Rdl*-296 mutations confer cross-resistance to other insecticides. For example, insertion of the *An. gambiae Rdl* gene into *Xenopus* oocytes demonstrated that fipronil, deltamethrin and imidacloprid all inhibited the GABA receptor, of which RDL is a subunit, and that this effect was reduced by the *Rdl*-296G mutation[44]. However, the same experiment found no effect of DDT on RDL. Our results may point to an as-yet unrecognised example of cross-resistance between dieldrin and DDT, through mechanisms that remain to be elucidated.

The *Vgsc*-1874L mutation, exclusively found on the background of the *Vgsc*-995 F1 haplotype, was negatively associated with resistance to permethrin in Burkina Faso, where the F1 haplotype is fixed in the population. The mutation's continued presence in the population despite its cost to permethrin resistance might be due to one of two factors. First, the mutation may provide improved resistance to other insecticides. However, since *kdr* is a target site resistance mechanism, the insecticides that this could apply to are limited to the pyrethroids and DDT. Second, the *Vgsc*-1874L mutation may serve to compensate the costs of *kdr*, serving to increase the fitness of the haplotype in the absence of insecticides. This would be a worrying development in the evolution of insecticide resistance, as it would favour the persistence of the *kdr* mutation during cycles of insecticide rotation in control programmes.

We do not propose that genotypic screening should completely replace phenotypic measurements of resistance in monitoring programmes. Genetic markers cannot yet accurately predict the resistance of a population, although some can be highly diagnostic[45], and further studies such as this one are required to address this shortcoming. Furthermore, as and when mosquitoes develop new resistance mutations, phenotypic assays will be required to discover these mutations and quantify their effects. However, once the range and effects of genetic markers are properly understood and can form the basis of predictive models, we hope that it will be possible to scale down the laborious phenotypic assays in favour of large scale genotyping. This will of course require high-throughput, cost-effective genotyping methods. We chose the iPLEX MassARRAY technology in our study due to its ability for multiplexing markers, making it more cost-effective at scale than assays with only one or a few markers. Still greater multiplexing can be achieved by other methods such as amplicon sequencing, which offer a prospective avenue for further development of our panel as new markers are discovered and need to be added.

In conclusion, our results highlight the importance and value of high-throughput genetic screening for the study and monitoring of insecticide resistance in malaria vectors. The three main steps involved in developing new markers of insecticide resistance are 1. discovery of new mutations, 2. confirmation of association with resistance and 3. tracking the mutation across populations (although the order of steps 1 and 2 can be reversed if a forward genetics approach is used to identify mutations responsible for an observed difference in phenotype). Our panel of 28 genetic markers achieved all three of these steps, albeit on different markers. We found evidence for a heterologous duplication in *Gste2*, we reported the first evidence of resistance association for *Gste2*-119V, bringing the total number of resistance-associated SNPs in *An. gambiae s.l.* up to 10, and we investigated the distribution of *Vgsc*-995 mutations across East, Central and West Africa. Our results also highlight gene and mutations that should be specifically taken into account when investigating resistance to certain insecticides. For example, at least two mutations in *Gste2* are involved in resistance to permethrin, while a role for *Rdl* or its associated markers in DDT resistance would be an interesting topic for further investigation. Work on our genotypic resistance panel should now expand in two directions. First, the panel should be applied to more populations, building a more complete picture of the distribution of these genotypes and their importance in resistance. In particular, more accurate measures of resistance can be obtained by using a range of insecticide concentrations and identifying the mosquitoes killed or surviving at each concentration, thus discerning between slightly resistant and highly

resistant phenotypes. Second, more markers should be gradually added to the panel as their importance becomes recognised. Once routine monitoring of important mutations become common-place, our understanding of the genetic underpinnings of insecticide resistance will increase dramatically.

Data Availability

All of the raw data for this study are included as Supplementary Materials.

References

1. Bhatt, S. *et al.* The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).
2. Implications of Insecticide Resistance Consortium. Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: trends in pyrethroid resistance during a WHO-coordinated multi-country prospective study. *Parasites & vectors* **11**, 550 (2018).
3. Kafy, H. T. *et al.* Impact of insecticide resistance in *Anopheles arabiensis* on malaria incidence and prevalence in Sudan and the costs of mitigation. *Proceedings of the National Academy of Sciences* **114**, E11267–E11275 (2017).
4. Protopopoff, N. *et al.* Effectiveness of a long-lasting piperonyl butoxide-treated insecticidal net and indoor residual spray interventions, separately and together, against malaria transmitted by pyrethroid-resistant mosquitoes: a cluster, randomised controlled, two-by-two factorial design trial. *The Lancet* **391**, 1577–1588 (2018).
5. Ranson, H. & Lissenden, N. Insecticide resistance in African *Anopheles* mosquitoes: a worsening situation that needs urgent action to maintain malaria control. *Trends in Parasitology* **32**, 187–196 (2016).
6. Kleinschmidt, I. *et al.* Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: a WHO-coordinated, prospective, international, observational cohort study. *The Lancet infectious diseases* **18**, 640–649 (2018).
7. Miotto, O. *et al.* Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nature genetics* **45**, 648 (2013).
8. Du, W. *et al.* Independent mutations in the Rdl locus confer dieldrin resistance to *Anopheles gambiae* and *An. arabiensis*. *Insect Molecular Biology* **14**, 179–183 (2005).
9. Jones, C. M. *et al.* Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of *Anopheles gambiae*. *Proceedings of the National Academy of Sciences* **109**, 6614–6619 (2012).
10. Martinez-Torres, D. *et al.* Molecular characterization of pyrethroid knockdown resistance (*kdr*) in the major malaria vector *Anopheles gambiae* ss. *Insect molecular biology* **7**, 179–184 (1998).
11. Mitchell, S. N. *et al.* Metabolic and target-site mechanisms combine to confer strong DDT resistance in *Anopheles gambiae*. *PLoS One* **9**, e92662 (2014).
12. Ranson, H. *et al.* Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. *Insect molecular biology* **9**, 491–497 (2000).
13. 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 Biology* **15**, R27 (2014).
14. 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*. *Scientific reports* **8**, 2920 (2018).
15. Weill, M. *et al.* Insecticide resistance in mosquito vectors. *Nature* **423**, 136–137 (2003).
16. Malaria Genomic Epidemiology Network *et al.* Reappraisal of known malaria resistance loci in a large multicenter study. *Nature genetics* **46**, 1197–1204 (2014).
17. Volkman, S. K. *et al.* A genome-wide map of diversity in Plasmodium falciparum. *Nature genetics* **39**, 113 (2007).
18. Norris, L. C. *et al.* Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proceedings of the National Academy of Sciences* **112**, 815–820 (2015).
19. Brogdon, W. G. & McAllister, J. C. Simplification of adult mosquito bioassays through use of time-mortality determinations in glass bottles. *Journal of the American Mosquito Control Association* **14**, 159–164 (1998).
20. Livak, K. J. Organization and mapping of a sequence on the Drosophila melanogaster X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* **107**, 611–634 (1984).
21. Santolamazza, F. *et al.* Insertion polymorphisms of SINE200 retrotransposons within speciation islands of *Anopheles gambiae* molecular forms. *Malaria journal* **7**, 163 (2008).
22. Lynd, A. *et al.* Insecticide resistance in *Anopheles gambiae* from the northern Democratic Republic of Congo, with extreme knockdown resistance (kdr) mutation frequencies revealed by a new diagnostic assay. *Malaria journal* **17**, 412 (2018).
23. Vazquez-Prokopec, G. M., Galvin, W. A., Kelly, R. & Kitron, U. A new, cost-effective, battery-powered aspirator for adult mosquito collections. *Journal of medical entomology* **46**, 1256–1259 (2009).
24. Gillies, M. T. & Coetzee, M. A supplement to the Anophelinae of Africa south of the Sahara. *Publications of the South African Institute for Medical Research* **55**, 1–143 (1987).
25. Geneva: World Health Organization. Test procedures for insecticide resistance monitoring in malaria vector mosquitoes. (2016).
26. Scott, J. A., Brogdon, W. G. & Collins, F. H. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *The American journal of tropical medicine and hygiene* **49**, 520–529 (1993).
27. Geneva: World Health Organisation. Test procedures for insecticides monitoring in malaria vector mosquitoes. (2013).
28. Collins, F. H. *et al.* A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *The American journal of tropical medicine and hygiene* **37**, 37–41 (1987).
29. Chabi, J. *et al.* Rapid high throughput SYBR green assay for identifying the malaria vectors *Anopheles arabiensis*, *Anopheles coluzzii* and *Anopheles gambiae* s.s. Giles. *bioRxiv*, (2018).
30. Sharakhova, M. V. *et al.* Update of the *Anopheles gambiae* PEST genome assembly. *Genome biology* **8**, R5 (2007).
31. Anopheles gambiae 1000 Genomes Consortium. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature* **552**, 96-100 (2017).
32. Fabrigar, D. J., Hubbart, C., Miles, A. & Rockett, K. High-throughput genotyping of *Anopheles* mosquitoes using intact legs by Agena Biosciences iPLEX. *Molecular ecology resources* **16**, 480–486 (2016).
33. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1–48 (2015).
34. Lucas, E. R., *et al.* Whole genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. *Genome Research* In press, (2019).
35. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics* **13**, 134 (2012).
36. Assogba, B. S. *et al.* The *ace-1* locus is amplified in all resistant *Anopheles gambiae* mosquitoes: fitness consequences of homogeneous and heterogeneous duplications. *PLoS Biology* **14**, e2000618 (2016).
37. Weetman, D. *et al.* Contemporary evolution of resistance at the major insecticide target site gene Ace-1 by mutation and copy number variation in the malaria mosquito *Anopheles gambiae*. *Molecular Ecology* **24**, 2656–2672 (2015).
38. Assogba, B. S. *et al.* An *ace-1* gene duplication resorbs the fitness cost associated with resistance in *Anopheles gambiae*, the main malaria mosquito. *Scientific Reports* **5**, 14529 (2015).

39. Ochomo, E. *et al.* Presence of the knockdown resistance mutation, Vgsc-1014F in *Anopheles gambiae* and *An. arabiensis* in western Kenya. *Parasites & vectors* **8**, 616 (2015).
40. Opondo, K. O. *et al.* Does insecticide resistance contribute to heterogeneities in malaria transmission in The Gambia? *Malaria journal* **15**, 166 (2016).
41. Asih, P. B. S. *et al.* Existence of the *Rdl* mutant alleles among the *Anopheles* malaria vector in Indonesia. *Malaria journal* **11**, 57 (2012).
42. Kwiatkowska, R. M. *et al.* Dissecting the mechanisms responsible for the multiple insecticide resistance phenotype in *Anopheles gambiae* ss, M form, from Vallée du Kou, Burkina Faso. *Gene* **519**, 98–106 (2013).
43. Wondji, C. S. *et al.* Identification and distribution of a GABA receptor mutation conferring dieldrin resistance in the malaria vector *Anopheles funestus* in Africa. *Insect biochemistry and molecular biology* **41**, 484–491 (2011).
44. Taylor-Wells, J., Brooke, B. D., Bermudez, I. & Jones, A. K. The neonicotinoid imidacloprid, and the pyrethroid deltamethrin, are antagonists of the insect Rdl GABA receptor. *Journal of neurochemistry* **135**, 705–713 (2015).
45. Essandoh, J., Yawson, A. E. & Weetman, D. Acetylcholinesterase (*Ace-1*) target site mutation 119S is strongly diagnostic of carbamate and organophosphate resistance in *Anopheles gambiae* ss and *Anopheles coluzzii* across southern Ghana. *Malaria journal* **12**, 404 (2013).

Acknowledgements

Author Contributions

E.R.L., K.A.R., E.O., D.P.K., D.W. and M.J.D. designed the study. A.L., J.E., N.G., B.K., J.M. and E.O. collected the samples. K.A.R., A.L., J.E., N.G., B.K., H.N., C.H., E.J.R. and A.E.V.H. performed the lab work. E.R.L. and K.A.R. carried out the analysis. E.R.L., K.A.R., A.L., J.E., N.G., B.K., E.J.R., A.E.V.H., D.W. and M.J.D. wrote the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-49892-6.

Competing Interests: The authors declare no competing interests.

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