Identification of a rapidly-spreading triple mutant for high-level metabolic insecticide resistance in *Anopheles gambiae* provides a real-time molecular diagnostic for anti-malarial intervention deployment.

Harun Njoroge[1,2], Arjen van't Hof[1], Ambrose Oruni[1,3], Dimitra Pipini[1], Sanjay C. Nagi[1], Amy Lynd[1], Eric R. Lucas[1], Sean Tomlinson[1], Xavi Grau-Bove[1], Daniel McDermott[1], Francis T. Wat'senga[4], Emile Z. Manzambi[4], Fiacre R. Agossa[5], Arlette Mokuba[5], Bilali Kabula[6], Charles Mbogo[7,8], Joel Bargul[9,10], Mark J.I. Paine[1], David Weetman[1] and Martin J. Donnelly[1,11,*]

[1]Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK.

[2]Kenya Medical Research Institute (KEMRI) Centre for Geographic Medicine Coast, KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

[3]College of Veterinary Medicine, Animal Resources & Bio-security, Makerere University, Kampala, Uganda

[4]Institut National de Recherche Biomédicale, PO Box 1192 Kinshasa, Democratic Republic of Congo.

[5]USAID President's Malaria Initiative, VectorLink Project, Abt Associates, 6130 Executive Blvd, Rockville, MD 20852, USA.

[6]Amani Research Centre, National Institute for Medical Research, Muheza, Tanzania

[7]Population Health Unit, KEMRI-Wellcome Trust Research Programme, Nairobi

[8]KEMRI-Centre for Geographic Medicine Research Coast, Kilifi

[9]Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya

[10]The Animal Health Department, International Centre of Insect Physiology and Ecology, Nairobi, P.O. Box 30772-00100, Kenya

[11]Parasites and Microbes Programme, Wellcome Sanger Institute, Hinxton, Cambridge, CB10 1RQ, UK.

[*]Corresponding author:

Department of Vector Biology, Liverpool Tropical School of Medicine, Pembroke Place, Liverpool, L3 5QA, UK. Telephone +44 151 705 3296, E-mail martin.donnelly@lstmed.ac.uk

Abstract

Insecticide resistance provides both an increasingly pressing threat to the control of vector-borne diseases and insights into the remarkable capacity of natural populations to show rapid evolutionary responses to contemporary selection. Malaria control remains heavily dependent on deployment of pyrethroid insecticides, primarily in long lasting insecticide treated nets (LLINs), but resistance in the major malaria vectors has increased over the last 15 years in concert with dramatic expansion of LLIN distributions. Identifying genetic mechanisms causing high-level resistance in mosquitoes, which may almost entirely overcome pyrethroid efficacy, is crucial for the development and deployment of potentially resistance-breaking tools. Using the *Anopheles gambiae* 1000 genomes (Ag1000g) data we identified a very recent selective sweep in Ugandan mosquitoes which localized to a cluster of cytochrome P450 genes, including some commonly implicated in resistance. Further interrogation revealed a haplotype involving a trio of mutations, a nonsynonymous point mutation in *Cyp6p4* (I236M), an upstream insertion of a partial Zanzibar-like transposable element (TE) and a duplication of the *Cyp6aa1* gene. The mutations appear to have originated recently in *An. gambiae* from the Kenya-Uganda border region around Lake Victoria, with stepwise replacement of the double-mutant (Zanzibar-like TE and *Cyp6p4-236M*) with the triple-mutant haplotype (including *Cyp6aa1* duplication), which has spread into the Democratic Republic of Congo and Tanzania. The triple-mutant haplotype is strongly associated with increased expression of genes able to metabolise pyrethroids; is strongly predictive of resistance to pyrethroids most notably deltamethrin, a commonly-used LLIN insecticide, but importantly, appears less effective against nets co-treated with the synergist piperonyl butoxide (PBO). Frequencies of the triple-mutant haplotype remain spatially variable within countries, suggesting an effective marker system to guide deployment decisions for limited supplies of PBO-pyrethroid co-treated LLINs across African countries. Duplications of *Cyp6aa1* gene are common in *An. gambiae* across Africa and are likely to be a useful diagnostic for high levels of pyrethroid resistance.

Introduction

Insecticide resistance in disease vectors has become an influential model for understanding rapid contemporary evolution but, more importantly, identifying how resistance arises and spreads is crucial for disease control. Resistance to pyrethroid insecticides in African malaria mosquitoes has spread to near ubiquity [1, 2] and, though it is often difficult to demonstrate its impact on malaria infections [3], in some cases has reached levels that threaten the effectiveness of anti-vector interventions programmes [4, 5]. A better understanding of resistance genetics will permit a more informed selection and deployment of insecticides to combat evolving mosquito populations. Whilst our understanding of the genetic basis of insecticide resistance in mosquitoes has advanced substantially [6], especially in the important vector *Anopheles funestus* [7], molecular diagnostics for the major vectors in the *An. gambiae* complex, remain limited to a handful of mutations [8] which explain a relatively small fraction of the variance in phenotype [9, 10] or which are now at such high frequency as to provide limited diagnostic resolution [11].

Long-lasting insecticidal nets (LLINs) are the principal tool for vector control to combat malaria, especially in sub-Saharan Africa [12]. The majority of LLINs are treated only with pyrethroid insecticides, to which resistance is now widespread [13]. Though behavioural variation and physical or physiological modifications affecting insecticide uptake may sometimes play a role, pyrethroid resistance is caused predominantly by two distinct mechanisms. The first is target-site resistance via point mutations which results in decreased sensitivity to the insecticide [14]; the second is metabolic resistance due to over-expression or altered activity of detoxification enzymes of which the cytochrome P450 family is commonly considered most important [15, 16]. Cytochrome P450 activity is inhibited by the synergist piperonyl-butoxide (PBO), and bed nets incorporating PBO are effective against P450-mediated resistance, as demonstrated by large-scale field trials [4, 17]. Given the continued operational use of pyrethroid-containing nets, it is vital that we understand the genetic mechanisms that may impact their efficacy, to optimise bednet deployment, preferably using information from rapidly-applied DNA markers. In advance of a randomized control trial of PBO-LLINs [4] we sought to characterise pyrethroid resistance mechanisms in the primary malaria vector *An. gambiae s.s.* [18] in Uganda and Kenya .

The recent development of the *Anopheles gambiae* 1000 genomes project (Ag1000g) has led to a step change in our ability to identify DNA variation driven by selection pressure. We have been able to perform genome-wide searches for regions under recent natural selection in insecticide resistant populations

85 across Africa, and work has shown that the strongest selective sweeps in the genome are all found around
86 genes known to be important for resistance [6, 19]. Furthermore, a whole-genome scan of copy number
87 variants (CNVs) in the Ag1000g data revealed that increases in gene copy number were highly enriched in
88 clusters of detoxification genes, pointing to a potentially widespread mechanism for increased gene
89 expression [20], which, in some cases, may elevate resistance. A number of gene duplications were
90 observed around the *Cyp6aa*/*Cyp6p* gene family cluster on chromosome 2R, and the majority of these
91 duplications included the gene *Cyp6aa1* (Figure 1). *Cyp6aa1* has been found to be overexpressed in
92 pyrethroid resistant populations in congeneric species [21-23], but it has received very little attention
93 compared to known insecticide-metabolizing genes such as *Cyp6m2*[24, 25], *Cyp6p3* [15] and *Cyp9k1* [26]
94 and its importance in resistance in *An. gambiae* remains unknown.
95
96 In this study, we examine a strong selective sweep detected in the *Cyp6aa/Cyp6p* genomic region in
97 samples of *An. gambiae* from Uganda and Western Kenya. We find that the sweep is closely associated
98 with three mutations (a SNP in *Cyp6p4*, a duplication of *Cyp6aa1* and a partial transposable element
99 insertion) in tight physical and statistical linkage. The triple-mutant haplotype is associated with strong
100 pyrethroid resistance, most notably to deltamethrin. The three mutations appeared sequentially, leading
101 to successive selective sweeps, with the triple-mutant haplotype replacing earlier variants and then
102 spreading rapidly across East Africa and Central Africa. We show that this haplotype is under positive
103 selection and causes increased expression of key cytochrome P450s and through recombinant protein
104 expression using both an *E. coli* and an *Sf9*-baculovirus system we show that both CYP6AA1 and CYP6P4
105 are capable of metabolizing pyrethroid insecticides.
106

107 ## Methods

108 **Interrogation of the Ag1000g dataset and identification of tagging markers**

109 Whole genome sequence data in the Ag1000g data set have previously revealed a strong selective sweep
110 in Ugandan populations around the *Cyp6aa/Cyp6p* cluster [19] (Figure 1). An isoleucine to methionine
111 substitution in codon 236 of CYP6P4 (see extended data Figure 10b in [19]) was identified provisionally as
112 a swept haplotype tagging SNP. Previous work [20] had shown that a duplication of the *Cyp6aa1* gene was
113 also observed in these samples (previously termed *Cyp6aap-Dup1*). To objectively determine how these
114 mutations segregated with the observed selective sweep [19] we grouped the 206 Ugandan haplotypes
115 (n=103 diploid individuals) by similarity using the 1000 SNPs located immediately upstream and
116 downstream of the start of the *Cyp6aa/Cyp6p gene* cluster (500 non-singleton SNPs in each direction from

117 position 2R:28,480,576). Distances were calculated with the *pairwise_distance* function in *scikit-allel* [27]
118 and converted to a nucleotide divergence matrix (*Dxy* statistic) by correcting the distance by the number
119 of sequencing-accessible bases in that region. We defined clusters of highly similar haplotypes by
120 hierarchical clustering with a cutoff distance of 0.001. This resulted in the identification of a cluster of 122
121 highly similar haplotypes. To determine whether the haplotype cluster showed signs of a selective sweep,
122 we estimated the extended haplotype homozygosity (EHH) decay within each of the haplotype groupings,
123 around two different focal loci: (i) the putative sweep SNP marker *Cyp6p4-236M* (2R:28,497,967 +/- 200
124 kbp; total 14,243 phased variants), and (ii) the 5' and 3' breakpoints of the *Cyp6aap-Dup1* duplication
125 (2R:28,480,189 - 200 kbp and 2R:28,483,475 + 200 kbp; total 14,398 phased variants). We used the
126 *ehh_decay* function of *scikit-allel* [27].

127 We also used the haplotype groupings identified above to calculate the Garud *H* statistics [28] and the
128 haplotypic diversity at the *Cyp6aa/Cyp6p* cluster locus (coordinates 2R:28,480,576 to 2R:28,505,816).
129 Specifically, we used the *moving_garud_h* and *moving_haplotype_diversity* functions in *scikit-allel* to
130 obtain a series of estimates for each statistic in blocks of 100 variants located within the cluster, and used
131 a block-jackknife procedure to calculate averages and standard errors of each estimate (*jackknife* function
132 in *scikit-allel misc* module). Python scripts to reproduce these analyses are available
133 https://github.com/xgrau/cyp6-AgUganda together with genomic variation data
134 https://www.malariagen.net/data/ag1000g-phase1-ar3.1.

135 **Molecular screening of colony and wild caught mosquitoes**

136 Locked-nucleic acid (LNA) probe-based PCR diagnostics were designed for all three mutations
137 (Supplementary materials Appendix 1).

138 Colony. Genotype:phenotype association testing was performed using two colonies of *An. gambiae*,
139 BusiaUG (resistant) and Mbita (susceptible). The BusiaUG strain was established in the lab in November
140 2018 from Busia, eastern Uganda and exhibits high resistance to pyrethroids and organochlorines (<10%
141 mortality following WHO exposure test), but full susceptibility to organophosphates and carbamates
142 (Oruni *et al* unpublished). The Mbita strain was first colonised from Mbita Point, Kenya in 1999, and is
143 fully susceptible to pyrethroids. Colonies were reared in insectaries targeted to 25-27°C and 70-80%
144 relative humidity.

145 Freshly emerged females from the Mbita line were mated with 3-5 day-old BusiaUG males and then blood

146 fed. The offspring from this cross were then crossed back to the parental BusiaUG line. This design was
147 chosen as resistance variants are often recessive. The resultant backcrossed 3-5-day-old females were
148 exposed for one hour to deltamethrin, permethrin and α-cypermethrin (the three insecticides most
149 commonly used on LLINs) and DDT (a non-pyrethroid sodium channel antagonist), following WHO
150 standard procedures. Mosquitoes were maintained on a 10% sugar solution after exposure and mortality
151 was recorded 24 hours post exposure.

152 Democratic Republic of Congo. *Anopheles gambiae* were obtained from the President's Malaria Initiative
153 supported entomological surveillance project [29] and from collections conducted by Lynd *et al* [30].
154 Mosquitoes were collected by human landing catch or pyrethrum spray collection from 15 locations
155 between 2013 and 2018. Resistance-phenotyped collections were also obtained from Pwamba, Bassa and
156 Fiwa in Nord Ubangi Province in 2016. These mosquitoes were assessed for susceptibility to deltamethrin
157 or permethrin using a standard WHO tube assay or a cone bioassay where Permanet 3.0 (deltamethrin
158 plus PBO) or Olyset Plus (permethrin plus PBO) were the test nets [23]. Kenya and Uganda. Collection
159 details for specimens from contiguous areas of western Kenya and eastern Uganda have been published
160 previously [20, 24, 25]. Tanzania. Larval mosquito collections were conducted in Geita, Bagamoyo and
161 Muleba districts of Tanzania in 2018.

162 Mosquitoes from all collections were genotyped at the ZZB-TE, *Cyp6p4-236M* and *Cyp6aap-Dup1* loci and
163 genotype:phenotype association testing was performed by Fisher's exact tests.

164 **Analysis of temporal change**

165 The most complete time series of ZZB-TE, *Cyp6p4-236M* and *Cyp6aap-Dup1* allele frequencies were
166 available from Kabondo, DRC; western Kenya and eastern Uganda. To model allele frequency changes
167 over time we estimated the three parameters of the standard recursive population genetic model (allele
168 frequency at time zero, selection coefficient and dominance coefficient) using a maximum likelihood
169 approach assuming a binomial distribution: an approach previously applied to insecticide target-site
170 resistance mutations [31]. The analysis was performed in R (http://www.r-project.org). An estimated
171 generation time of one calendar month was used as in previous studies [31].

172 **Recombinant protein and insecticide metabolism**

173 The CYP6P4-236M variant, and its redox partner cytochrome P450 reductase gene (CPR), were expressed
174 in *E. coli* as per standard protocols [32] (Supplementary materials Appendix 2). Initial efforts to generate

175 recombinant *CYP6AA1* in an *E. coli* system with optimised codon usage failed, and we therefore used an
176 Sf9-baculovirus-based expression system. Since P450 catalytic activity is dependent on electrons supplied
177 by NADPH via CPR, insecticide metabolism was assayed with cell pellets of CYP6P4/CPR or CYP6AA1/CPR
178 in the presence or absence of NADPH. The depletion of the substrate and the appearance of metabolites
179 were monitored by reverse-phase HPLC (Supplementary materials Appendix 3).

180 **Estimation of gene expression of key P450s in triple mutant haplotype**

181 To determine whether the presence of the triple mutant haplotype type was associated with differential
182 expression of genes, we examined individual females from the BusiaUg colony (Triple mutant Freq. 0.297;
183 95%CI 0.233-0.370). Two legs were removed from individual mosquitoes for DNA analysis with the
184 remaining mosquito kept for RNA analysis. DNA was extracted from legs by boiling in STE buffer at 95°C
185 for 90 minutes and individuals were genotyped using the LNA qPCR assays. RNA was then extracted
186 individually from 8 mosquitoes in each genotypic group - homozygotes for the triple mutant haplotype,
187 wild-type homozygotes, and heterozygotes, using the Arcturus Picopure RNA isolation kit (Thermofisher).
188 We then performed SYBR green based qPCR to measure the expression of *Cyp6aa1* and *Cyp6p4* together
189 with the known resistance-linked variant *Cyp6p3* using the housekeeping genes 40s ribosomal protein S7
190 (AGAP010592) and elongation factor Tu (AGAP005128) for normalisation. The ΔΔCT values were tested
191 for normality and homogeneity of variances using the Shapiro-Wilks test, and the Bartlett test,
192 respectively. A significant difference in gene expression between the genotypic groups was determined
193 by a two-tailed two-sample Student's t-test on ΔΔCT values, with a threshold of P=0.05.

194

195 ## Results

196 Hierarchical clustering of 206 haplotypes from Ugandan *An. gambiae* from Ag1000g identified a cluster of
197 122 haplotypes around the *Cyp6aa*/*Cyp6p* gene cluster, putatively representing a swept haplotype in this
198 region. To characterise the signatures of selection in Ugandan haplotypes around this cluster, we
199 examined the profile of extended haplotype homozygosity around the position of the *Cyp6p4-236M SNP*
200 and around the CNV in *Cyp6aa1*. In both cases, we found that the putative swept haplotype had longer
201 stretches of homozygosity than wild-type haplotypes (Figure 2). In addition, we found that Ugandan *An.*
202 *gambiae* had reduced haplotypic diversity along the entire *Cyp6aa/Cyp6p* gene cluster (h = 0.339746 +/-
203 0.005664 standard error) and a combination of Garud's H statistics that was indicative of a hard selective
204 sweep in this region (high H_{12} = 0.821867 +/- 0.006308 SE; low H_2/H_1 = 0.016779 +/- 0.000228 SE)[28].

205 These results confirm that the haplotypes which we have identified have undergone a selective sweep.
206 We then used iterative read mapping of individuals homozygous for the sweep to search for additional
207 mutations that might be distinctive of the haplotype. This revealed that a partial copy of a Ty3/Gypsy
208 Zanzibar transposon insertion (termed ZZB-TE), lacking functional open reading frames, was linked to
209 *Cyp6p4-236M* and *Cyp6aap-Dup1* (Figure 2).

210

211 **The evolution of the ZZB-TE, *Cyp6p4-236M* and *Cyp6aap-Dup1* haplotypes**

212 Based upon Ag1000g data and a time series of collections from Central and East Africa, we were able to
213 trace the sequence of mutational events (ZZB-TE, *Cyp6p4-236M* and *Cyp6aap-Dup1*) and reconstruct the
214 evolutionary history of the swept haplotype. Among the Ag1000g data [19], the *Cyp6p4-236M* mutation
215 was only observed in collections from eastern Uganda (collected in 2012) suggesting that this mutation
216 originated in the eastern Ugandan/western Kenyan region. In a screen of Ugandan and Kenyan collections
217 predating the Ag1000g collections by eight years (2004) (Fig 3) only the ZZB-TE insertion was detected,
218 although the sample size was too small (n=4) to conclude that the *Cyp6p4-236M* allele was absent. The
219 *Cyp6p4-236M* mutation was first observed in this region in 2005 (frequency *Cyp6p4-236M* =0.10) in
220 individuals carrying the ZZB-TE mutation, whilst the *Cyp6aap-Dup1* CNV was first recorded in 2008
221 (Proportion of individuals with Dup1=0.8%). This inferred sequence of events may explain why ZZB-TE and
222 *Cyp6p4-236M* mutations are in tighter statistical linkage with each other than with *Cyp6aap-Dup1* (Figure
223 2), despite the closer proximity of ZZB-TE and *Cyp6aap-Dup1* (Figure 1). Given the very tight association
224 between the ZZB-TE insertion and the *Cyp6p4-236M* SNP, we will henceforth refer to the *Cyp6p4-236M*
225 (double mutant) haplotype and the *Cyp6aap-Dup1* (triple mutant). The double mutant haplotype shows
226 a steady increase in frequency between 2004 and 2011 in Kenya (Figure 3); possibly in response to the
227 introduction and subsequent intensification of bednet distribution programmes [29, 33, 34]. Following its
228 appearance in 2008, the triple mutant haplotype, bearing *Cyp6aap-Dup1* rapidly increased towards
229 fixation in both Ugandan and Kenyan collections, replacing the double mutant. This haplotype
230 replacement and the observation that the triple mutant is the only non-wildtype haplotype observed
231 outside Kenya/Uganda (such as in Tanzania and DRC, Figure 1 and 3) strongly implies an additional
232 selective advantage to the triple mutant. The time series data from across DRC are particularly striking
233 both in terms of the speed of increase of the triple mutant but also the north-south heterogeneity, with
234 very low frequencies in the more southerly provinces (Figure 3).

235

236 **Identifying potential drivers of haplotype frequency increase**

237 The genomic region into which the ZZB-TE inserted does not show histone signals of regulatory activation
238 (H3K27ac, H3K9ac) or repression (H3K9me3), and ATAC-seq data suggests it is not in an open chromatin
239 region [35]. However we took two in silico approaches to determine whether the ZZB-TE insertion (748bp)
240 carried putative regulatory variants. The inserted region had 98.5% sequence identity to a putative
241 enhancer (2R:45966598-45966822) identified by homology with *Drosophila melanogaster* [35]. However
242 despite the similarity, the insertion lacked enhancer-like combination of chromatin marks identified in
243 [35] and its potential regulatory role in nearby genes is unclear. The second approach, involved screening
244 the ZZB-TE inserted sequence for putative enhancers using iEnhancer-2L[36] and iEnhancer-EL [37]. In a
245 windowed analysis of 200bp with a 1bp step across the entire length of ZZB-TE both predicted that some
246 of the windows would have strong enhancer activity, however the windows were not-concordant,
247 precluding further analysis.

248

249 Given that cytochrome P450 mediated resistance is commonly associated with differential gene
250 expression we performed transcription studies within the *Cyp6aa/Cyp6p* cluster between the most
251 contrasting haplotypes, wild-type and triple mutant, present in the BusiaUg colony. The group
252 homozygous for the triple mutant haplotype significantly overexpressed both *Cyp6aa1* (2.23-fold, 95%
253 CIs: 1.73-2.90, P=0.0003) and *Cyp6p4* (2.57-fold, 95% CIs 1.25-5.93, P=0.039) compared to wild-type
254 individuals. The ratio of expression of *Cyp6aa*1 broadly reflected the expected pattern based on genotype
255 ie 2:1.5:1 for triple mutant homozygotes: heterozygotes: wild type genotypes, respectively (Figure 4). As
256 a control we examined a neighbouring, very commonly resistance-associated gene, *Cyp6p3*, but triple
257 mutant and wild type homozygotes did not differ significantly in expression (1.33 fold, 95% CIs 0.64-2.74,
258 P>0.05).

259

260 To investigate whether resistance may be driven at least in part by an effect of the allelic variant on
261 metabolic activity of CYP6P4, we expressed the wild-type (236I) and mutant (236M) forms in an *E. coli*
262 based recombinant protein system (Supplementary materials Appendices 2 and 4). Both alleles were
263 shown to be capable of metabolizing class I (permethrin) and II (deltamethrin) pyrethroids but there was
264 no evidence that the mutant (236M) or wildtype (236I) alleles had different rates of pyrethroid depletion.
265 We also expressed the duplicated P450 CYP6AA1, in an *Sf9*-baculovirus protein expression system. Again,
266 metabolism assays demonstrated that the enzyme was capable of metabolizing both deltamethrin and
267 permethrin (Supplementary materials Appendices 3 and 4). Depletion of deltamethrin was 36.6% greater

268 (SE= 3.79) in the presence of NADPH than in the control (t-test: t=-9.67; d.f. = 8; P= 9.6 $x10^{-6}$),
269 demonstrating that CYP6AA1/CPR is capable of metabolizing deltamethrin *in vitro*. Similarly, permethrin
270 was metabolised by CYP6AA1/CPR, with permethrin being depleted by 22.4% (SE = 0.63) compared to the
271 control without NADPH (t-test: t= -31.08; d.f. = 14; P= 2.55 $x10^{-14}$).
272
273 Given clear evidence of increased expression of both *Cyp6aa1* and *Cyp6p4* in the triple mutant haplotype
274 and the ability of both enzymes to metabolise pyrethroids *in vitro*, we investigated whether the mutations
275 were significantly associated with resistance *in vivo.* Exposure of *An. gambiae* females from Busia, Uganda
276 and Nord Ubangi, DRC in WHO net assays resulted in negligible mortality to the pyrethroid only LLINs,
277 Olyset and Permanent 2.0 (Figure 5). Simultaneous exposure to pyrethroid plus the P450 inhibitor PBO in
278 Olyset + and the top of Permanent 3.0 nets resulted in a marked reduction in resistance, demonstrating
279 that the resistance phenotype is substantially mediated by P450s. We performed laboratory backcrosses
280 between additional mosquitoes from Busia with the pyrethroid susceptible Mbita colony, and found that
281 the triple mutant haplotype was significantly associated with resistance to the most commonly used type
282 II pyrethroids in LLINs: deltamethrin (Fisher's exact test $p=3.2x10^{-6}$) and alphacypermethrin (Fisher's exact
283 test $p=5.9x10^{-7}$) resistance although not to permethrin (Fisher's exact test p=0.06) nor, as a control, DDT
284 (Fisher's exact test p=0.84) (Table 1) in WHO assays. Similarly, specimens collected in 2016 from the DRC
285 showed a strong association between the triple mutant genotype and survival rate 24 hours post-
286 exposure to 0.05% deltamethrin for 1 hour, as well as in 3-minute exposures to deltamethrin-treated sides
287 of a PermaNet 3.0 net (Table 2). No association was found in samples exposed to permethrin (24 hour
288 WHO bioassay) or permethrin-treated Olyset Plus nets (3-minute WHO net bioassay) (Table 2). Complete
289 linkage of the three mutants in the BusiaUG colony and the DRC wild caught collections precludes
290 determination of the relative contribution of each of the three mutations to the resistance phenotype,
291 but taken together these results demonstrate a strong impact of the triple mutant on the efficacy of
292 pyrethroid resistance.
293

294 **Discussion**

295 We have identified a sequential series of fitness-augmenting mutations in *An. gambiae* culminating in a
296 triple-mutant haplotype with a large effect on pyrethroid resistance and which is spreading rapidly across
297 East and Central Africa. The mutation that is probably the oldest in this series, the insertion of ZZB-TE,
298 appears to have originated sometime prior to 2004 in the malaria-endemic area around Lake Victoria,

299 with the *Cyp6p4-236M* SNP evident in 2005 samples and the third, a duplication in *Cyp6aa1* detected in
300 2008. In samples collected only five years later, the triple mutant was detected hundreds of kilometres
301 away in the DRC. These patterns suggest both a large fitness advantage arising from the triple mutant,
302 and a frightening speed at which resistance-conferring mutations are able to spread within and across
303 populations.

304 Second generation nets treated with the synergist PBO were shown to be much more effective in killing
305 mosquitoes carrying the triple mutant genotype. There is therefore a clear incentive for the use of PBO
306 bednets in the regions where this mutation is present. A strong corollary of this finding comes from the
307 cluster randomised control trial conducted in Uganda, where the mutation is at high frequency, which
308 demonstrated that malaria parasite prevalence in children <10 years old (12% vs 14%; Prevalence ratio =
309 0·84, 95% CI 0·72–0·98; P=0·029) and mean number of mosquitoes (Density ratio=0·25, 95% CI 0·18–0·3;
310 P <0·0001) per house were significantly lower in villages that had received PBO LLINs relative to standard
311 LLINs [4]. There is some evidence that the haplotype may be less strongly associated with resistance to
312 permethrin than deltamethrin (and perhaps also alphacypermethrin), although both pyrethroids were
313 metabolised by *Cyp6aa1* and *Cyp6p4*.

314 Our results highlight the importance of gene duplications for the evolution of insecticide resistance. In *An.*
315 *gambiae*, duplications have recently been shown to be concentrated in regions associated with metabolic
316 resistance, and over 40 such duplications have been described across the genome [20]. Thirteen different
317 duplications have so far been described that encompass *Cyp6aa1* (Figure 1), both in West and East Africa
318 and in the two sister-species *An. gambiae* ss and *An. coluzzii* [20]. It seems likely that these other *Cyp6aa1*
319 duplications are also associated with pyrethroid resistance. For example, in *An. coluzzii* sampled from a
320 highly insecticide-resistant population from Cote d'Ivoire, greater than 95% of individuals had Cyp6aap
321 duplications (Supplementary materials Appendix 5). Five different duplications were observed in the
322 collections from 2012 and 2017 and, whilst we detected no association with pyrethroid resistance, two
323 duplications (*Cyp6aap-Dup7* and *Cyp6aap-Dup14*) showed significant increases in frequency over time.
324 Moreover the total number of CNVs per sample (measured as presence of each of the 5 duplications,
325 summed for each sample) increased significantly from an average of 1.59/ individual in 2012 to 2.08 in
326 2017 (Mann Whitney U test; P<0.0001); the mean number of duplications >2 indicates that there are
327 multiple CNVs on the same haplotype. Duplications of the *Cyp6aa1* orthologue have been found in
328 another important malaria vector, *An. funestus*, from west and central Africa [21]. The *Cyp6aa1*
329 orthologue in *An. funestus,* which shares an 87% identity with *An. gambiae,* was also observed to

330 metabolize permethrin and deltamethrin and, when expressed in transformed *Drosophila*, was associated
331 with significant increases in resistance to both permethrin and deltamethrin relative to control
332 mosquitoes [21]. This evolution of multiple *Cyp6aa1* duplications suggests this is an important pan-African
333 resistance mechanism

334 There are now several cases of insecticide resistance evolution where an initial mutation in a genomic
335 region is followed by the spread of additional mutations on the resistant haplotype background [14, 38-
336 42]. It is not yet clear whether the sequence of mutations that we have identified rely on each-other for
337 effect, and thus could only have spread sequentially, or whether each additional mutation coincidentally
338 appeared on the background of an already common mutant haplotype. In the case of *Cyp6p4-236M* and
339 *Cyp6aap-Dup1* it seems that the latter is more likely, although we cannot exclude the possibility that the
340 duplication affects the regulation of *Cyp6p4*. In contrast, the putative enhancer inserted with the ZZB
341 transposon may affect either or both of *Cyp6p4* and *Cyp6aa1*, and may thus interact with the other two
342 mutations in ways that have yet to be determined. Transposable elements can sometimes affect gene
343 expression of neighbouring genes [38] and are abundant in mosquito genomes [43, 44]. Interestingly, in
344 the common house mosquito, *Culex pipiens*, TEs have been found in the flanking regions of the *Ester* locus,
345 a genomic region in which many independent gene duplications have arisen and spread worldwide in
346 response to selection from organophosphates [45]. Clearly, both TEs and gene duplications are an
347 understudied, yet common source of variation that may have important implications for vector control
348 efforts. The appearance and rapid spread of the three mutations described here is broadly coincident with
349 the scale up of LLIN coverage in DRC, Kenya and Uganda. The haplotype is a strongly predictive marker of
350 high-level resistance to pyrethroids, is easily screened with a single diagnostic assay and should be
351 incorporated into regional molecular insecticide resistance monitoring strategies [6, 29, 46].

352

353 **Acknowledgements**

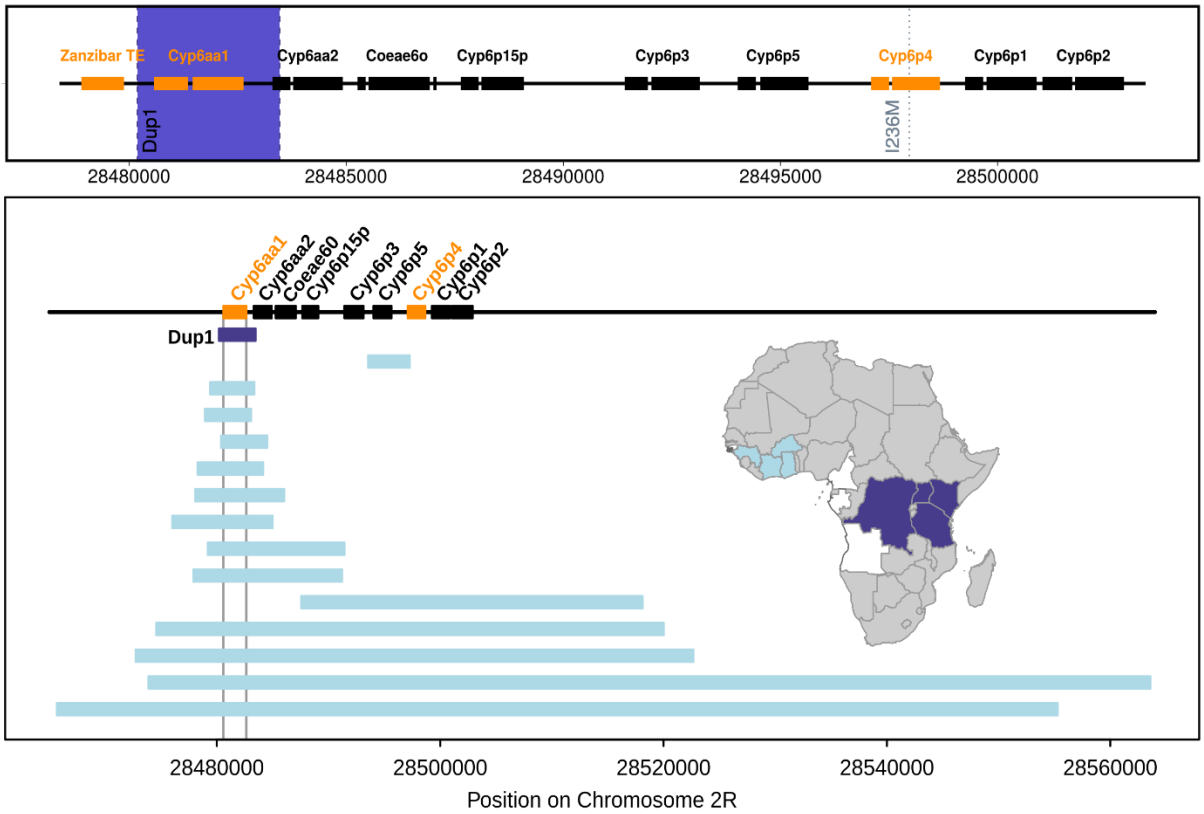
354 This work was supported by the National Institute of Allergy and Infectious Diseases ([NIAID] R01-
355 AI116811) with additional support from the Medical Research Council (MR/P02520X/1). The latter grant
356 is a UK-funded award and is part of the EDCTP2 programme supported by the European Union. HN and
357 AO were supported by Wellcome Trust MSC Training Fellowships in Public Health and Tropical Medicine.
358 SI is funded by the U.S. President's Malaria Initiative. CM was supported by the Bill and Melinda Gates
359 Foundation through the WHO (Award #54497) and by the Biovision Foundation (Grant No.BV HH-07). JLB
360 is supported by DELTAS Africa Initiative grant # DEL-15-011 to THRiVE-2. The DELTAS Africa Initiative is an

independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Welcome Trust grant # 107742/Z/15/Z and the UK government. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIAID, the National Institutes of Health or the other donors.

Author contributions

HN, AvH, MJIP, DW and MJD designed the study; HN, AvT, AO, DP, SCN, AL conducted lab and insectary experiments; HN, AvH, SCN, AL, ERL, ST, XG-B, DM, MJD performed analysis; AO, AL FW, EM, FA, SI, BK, DW, CM conducted field collections and phenotyping; HN,ERL, DW, MJD wrote the manuscript with input from all authors; JLB, CM, MJIP, DW and MJD supervised the study; all authors approved the final version of the manuscript.

376 **Tables and figures**



377

378

379 **Figure 1. Schematic of the mutational events around the *Cyp6aa/Cyp6p* cluster on chromosome 2R in**
380 ***Anopheles gambiae***. Upper panel shows the three mutational events observed in East and Central African
381 samples. Orange indicates the genes/ transposable elements involved and the blue band the extent of the
382 *Cyp6aap-Dup1* CNV. The lower panel, redrawn from [20], shows the genomic extent and geographic
383 distribution of the CNVs that have been observed around the *Cyp6aa/Cyp6p* cluster, 13 out of 15 include
384 *Cyp6aa1*.The dark blue shading on the map indicates the geographic extent of *Cyp6aap-Dup*1. Data from
385 Uganda, Kenya and DRC are reported extensively in the main text, in addition 27.4% (n=84) *of Anopheles*
386 *gambiae* females sampled from northern Tanzania in 2018 were triple mutant carriers.

387

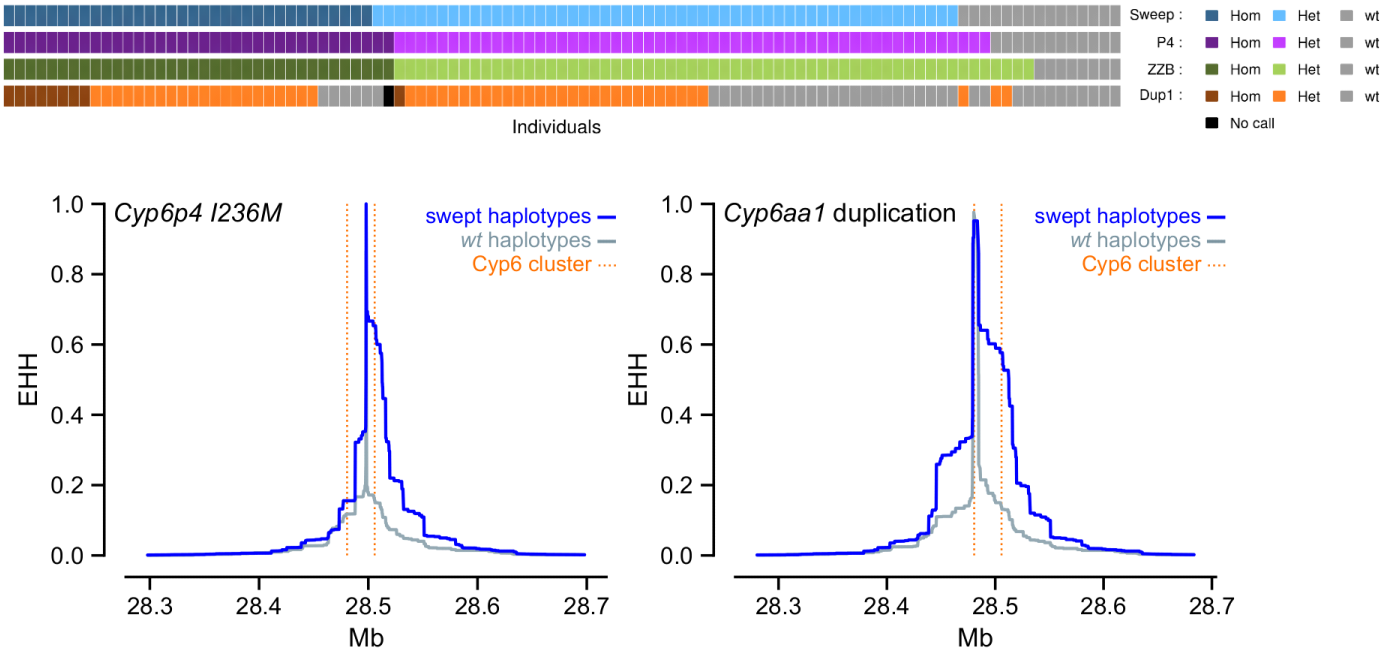


Figure 2. Selective sweep around the *Cyp6aa/Cyp6p* cluster in *Anopheles gambiae* from Uganda.

Upper panel Haplotype contingency for the selective sweep ("Sweep"), *Cyp6p4-236M* SNP ("P4") and the *Cyp6aap-Dup1* ("Dup1"). Each vertical bar represents a single haplotype, colour-coded to show whether it is a copy of the swept haplotype (blue) and whether it carries the SNP (purple), the TE insertion (green) and the duplication (orange). The swept haplotype and the *Cyp6p4-236M* SNP overlap almost completely, while the duplication is found on a subset of these haplotypes. Lower panel Extended Haplotype Homozygosity (EHH) plots around the *CYP6P4-236M* and *CYP6AA1-Dup1* variants show slower loss of homozygosity in the swept haplotypes than in the wild-type.

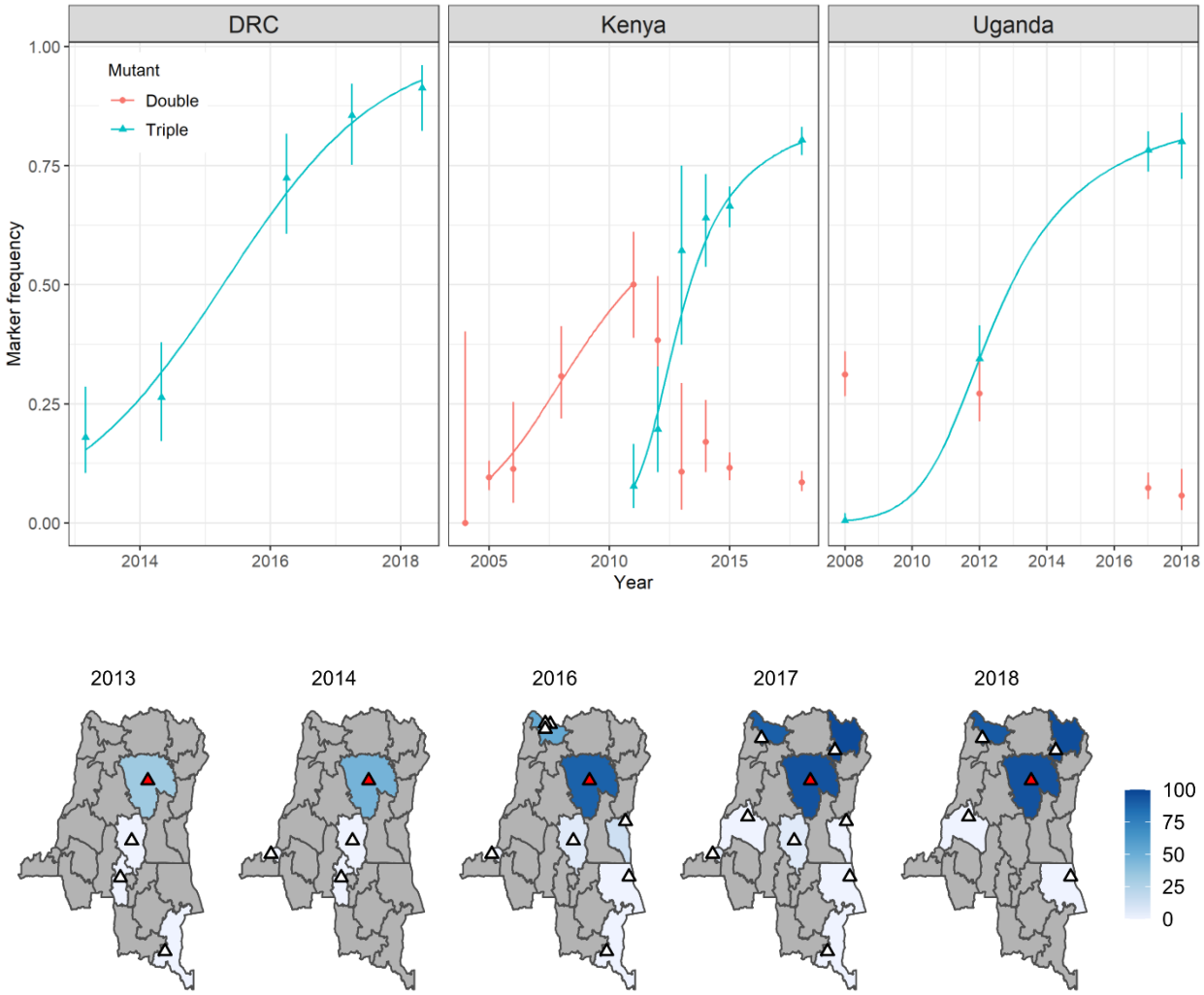
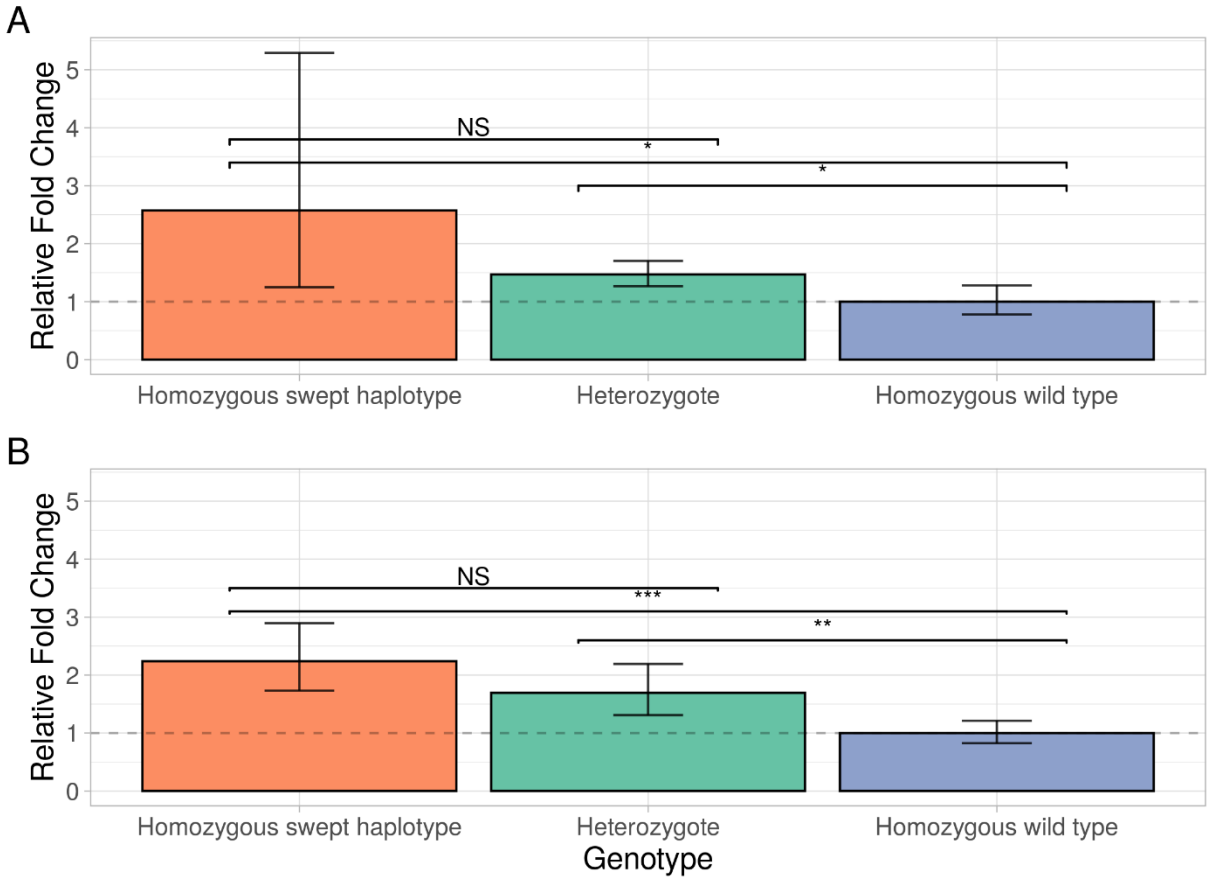


Figure 3. Observed and modelled changes in *Cyp6aa/Cyp6p* haplotype frequencies over time in *Anopheles gambiae s.s.* populations from DRC, Kenya and Uganda. Upper panel shows mutation frequency estimates derived from wild caught individuals from Kabondo, DRC; western Kenya and eastern Uganda, The 95% CIs for each observed data point were calculated according to[47]. Expected data generated from simultaneous maximum likelihood estimates of initial frequency and selection and dominance coefficients which were then used to parameterize standard recursive allele frequency change equation. The lower panel shows data from DRC tracking the emergence and spread of the triple mutant haplotype. Triangles indicate collection locations within each district. The red triangle shows the Kabondo sample site for which there was the most complete time series.



415
416 **Figure 4. Gene expression analysis of *Cyp6p4* and *Cyp6aa1* in genotyped *Anopheles gambiae* females.**
417 Relative fold change for *Cyp6p4* (Panel A) and *Cyp6aa1* (Panel B), comparing individuals homozygous for
418 the swept triple mutant haplotype and heterozygotes in the BusiaUG colony to wild-type individuals.
419 95% confidence intervals are shown. Asterisks indicate statistical significance in two-tailed students t-
420 tests; *** $P \leq 0.001$, ** $P \leq 0.01$, ** $P \leq 0.05$.

421

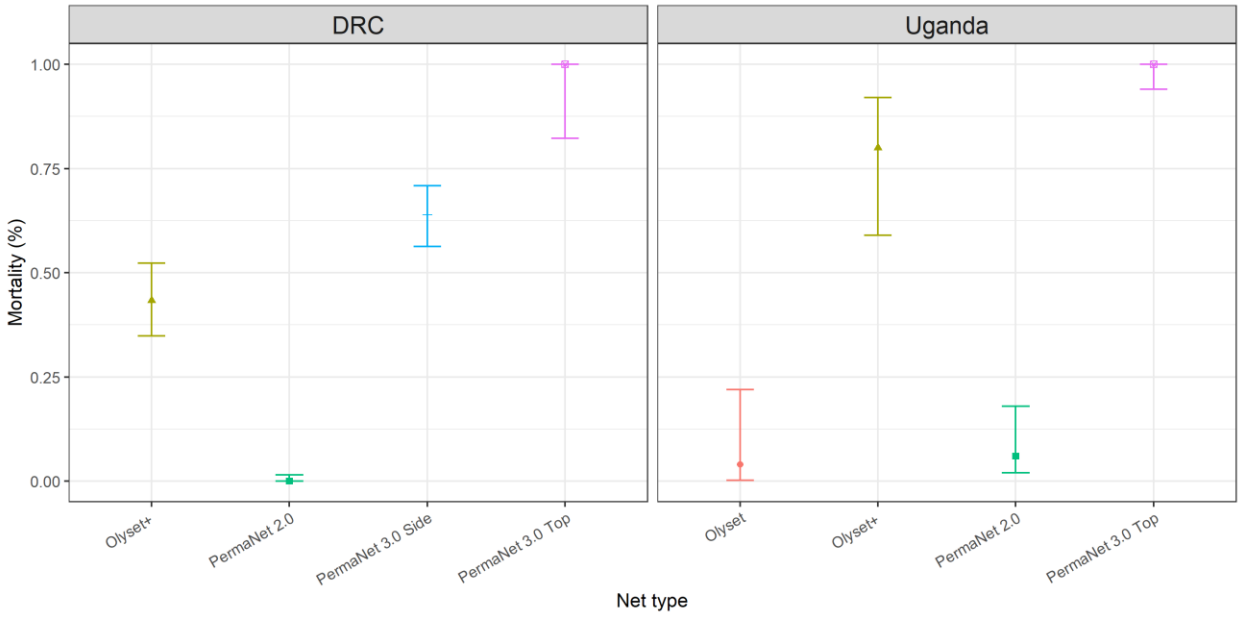


Figure 5. Patterns of insecticide resistance in *Anopheles gambiae* following exposure to permethrin and deltamethrin treated nets. Females from the Democratic Republic of Congo (DRC) and Uganda exhibited very high resistance (low %mortality) in WHO net assays to permethrin (Olyset) and deltamethrin (PermaNet 2.0).There was an increase in mortality following exposure to an increased concentration of deltamethrin (PermaNet 3.0 Side) and very high mortality following exposure to pyrethroid plus the P450 inhibitor piperonyl butoxide (PBO) Olyset+ and PermaNet 3.0 Top. 95% confidence intervals are shown.

Table 1. Association between insecticide susceptibility as determined by WHO tube bioassay and triple mutant genotype in female *Anopheles gambiae* from a cross between BusiaUG and Mbita colonies.

Insecticide	Phenotype	*Wildtype homozygote*	*Heterozygote*	*Triple mutant homozygote*	Test and p value
α-Cypermethrin	**Alive**	1	32	24	Genotypic $p=5.92x10^{-7}$
n=146	**Dead**	14	68	7	Allelic $p=6.34x10^{-5}$; OR 2.74 (1.63-4.69)
Deltamethrin	**Alive**	0	0	8	Genotypic $p=3.21x10^{-6}$
n=60	**Dead**	15	26	11	Allelic- Cannot be calculated
Permethrin	**Alive**	1	18	7	Genotypic p=0.059
n=53	**Dead**	5	20	2	Allelic p=0.08; OR 1.99 (0.86-4.67)
DDT	**Alive**	0	14	1	Genotypic p=1
n=65	**Dead**	3	45	2	Allelic p=0.84; OR 1.19 (0.48-2.94)

Table 2. Association between insecticide susceptibility as determined by WHO tube bioassay or WHO net bioassay and triple mutant genotype in wild-caught, female *Anopheles gambiae* from three locations in Nord Ubangi Province, Democratic Republic of Congo.

Site	Insecticide	Bioassay	Phenotype	*Wildtype homozygote*	*Heterozygote*	*Triple mutant homozygote*	Test and p value
	Permethrin	**Tube**	**Alive**	5	42	30	Genotypic p=0.63
Fiwa	**n=161**		**Dead**	8	49	27	Allelic p=0.42; OR 1.23 (0.76-2.00)
	Deltamethrin	**Tube**	**Alive**	6	14	18	Genotypic $p=6.21x10^{-3}$
Fiwa	**n=78**		**Dead**	7	27	6	Allelic p=0.04; OR 2.01 (1.01-4.06)
	Permethrin	**Tube**	**Alive**	1	19	18	Genotypic p=0.79
Bassa	**n=95**		**Dead**	2	32	23	Allelic p=0.63; OR 1.21 (0.61-2.43)
	Deltamethrin	**Tube**	**Alive**	0	3	16	Genotypic $p=4.02x10^{-3}$
Bassa	**n=47**		**Dead**	1	16	11	Allelic $p=5.71x10^{-3}$; OR 5.44 (1.41-31.28)
	Deltamethrin	**Tube**	**Alive**	7	9	7	Genotypic p=0.066
Pwamba	**n=40**		**Dead**	9	9	1	Allelic p=0.02; OR 2.93 (1.07-8.39)
	Olyset Plus	**Net**	**Alive**	6	21	30	Genotypic p=0.82
Fiwa	**n=106**		**Dead**	6	15	28	Allelic p=0.88; OR 0.93 (0.49-1.77)
	Permanet 3.0	**Net**	**Alive**	5	26	24	Genotypic $p=5.61x10^{-4}$
Fiwa	**n=155**		**Dead**	30	51	19	Allelic $p=1.4x10^{-4}$; OR 2.56 (1.54-4.31)

References

1. Ranson, H. and N. Lissenden, *Insecticide resistance in African Anopheles mosquitoes: a worsening situation that needs urgent action to maintain malaria control.* . Trends in Parasitology, 2016. **32**(3): p. 187-196.
2. WHO, *World malaria report 2019*, W.H. Organization, Editor. 2019, World Health Organization: Geneva. p. 232.
3. 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.* Lancet Infectious Diseases, 2018. **18**(6): p. 640-649.
4. Staedke, S.G., Gonahasa, S., Dorsey, G, Kamya, M.R., Maiteki-Sebuguzi, C., Lynd, A., Katureebe, A., Kyohere, M., Mutungi, P., Kigozi, S.P., Opigo, J., Hemingway, J., and M.J. Donnelly, *Effect of long-lasting insecticidal nets with and without piperonyl butoxide on malaria indicators in Uganda (LLINEUP): a pragmatic, cluster-randomised trial embedded in a national LLIN distribution campaign.* Lancet, 2020(395): p. 1292–303.
5. Killeen, G.F. and H. Ranson, *Insecticide-resistant malaria vectors must be tackled.* Lancet, 2018. **391**(10130): p. 1551-1552.
6. The *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 Research, 2020. **30**(10): p. 1533-1546.
7. Mugenzi, L.M.J., et al., *Cis-regulatory CYP6P9b P450 variants associated with loss of insecticide-treated bed net efficacy against Anopheles funestus.* Nature Communications, 2019. **10**.
8. Donnelly, M.J., A. Isaacs, and D. Weetman, *Identification, validation, and application of molecular diagnostics for insecticide resistance in malaria vectors.* Trends in Parasitology, 2016. **32**(3): p. 197-206.
9. Mitchell, S.N., et al., *Metabolic and Target-Site Mechanisms Combine to Confer Strong DDT Resistance in Anopheles gambiae.* PLoS ONE, 2014. **9**(3).
10. Weetman, D., Wilding, C.S., Neafsey, D.E., Muller, P., Ochomo, E., Isaacs, A.T., Steen, K., Rippon, E.J., Morgan, J.C., Mawejee, H.D., Rigden, D.J., Okedi, L.M., Donnelly, M.J. , *Candidate-gene based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing genetic variation in East African Anopheles gambiae.* Scientific Reports, 2018. **8**: p. e2920.
11. Weetman, D. and M.J. Donnelly, *Evolution of insecticide resistance diagnostics in malaria vectors.* Transactions of The Royal Society of Tropical Medicine and Hygiene, 2015. **109**(5): p. 291-293.
12. Bhatt, S., et al., *The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015.* Nature, 2015. **526**(7572): p. 207-+.
13. Hancock, P.A., et al., *Mapping trends in insecticide resistance phenotypes in African malaria vectors.* PLoS Biology, 2020. **18**(6): p. e3000633.
14. 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.* bioRxiv, 2018(https://doi.org/10.1101/323980).
15. Müller, P., et al., *Field-caught permethrin-resistant Anopheles gambiae overexpress CYP6P3, a P450 that metabolises pyrethroids.* PLoS Genetics, 2008. **4**(11): p. e1000286.
16. Weedall, G.D., et al., *A cytochrome P450 allele confers pyrethroid resistance on a major African malaria vector, reducing insecticide-treated bednet efficacy.* Science Translational Medicine, 2019. **11**(484).

17. 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.* Lancet, 2018. **391**(10130): p. 1577-1588.
18. Lynd, A., et al., *LLIN Evaluation in Uganda Project (LLINEUP) − A cross-sectional survey of species diversity and insecticide resistance in 48 districts of Uganda.* Parasites and Vectors, 2019. **12**: p. e94.
19. The *Anopheles gambiae* 1000 Genomes Consortium, *Natural diversity of the malaria vector Anopheles gambiae.* Nature (London), 2017. **552**: p. 96-100.
20. Lucas, E., et al., *Whole genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes.* Genome Research, 2019. **https://doi.org/10.1101/399568** (29): p. 1250-1261
21. 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, 2018. **9**(3): p. 17.
22. Zhou, D., et al., *Genomic analysis of detoxification supergene families in the mosquito Anopheles sinensis.* PLoS One, 2015. **10**(11): p. e0143387.
23. Kwiatkowska, R.M., et al., *Dissecting the mechanisms responsible for the multiple insecticide resistance phenotype in Anopheles gambiae s.s., M form, from Vallee du Kou, Burkina Faso.* Gene, 2013. **519**(1): p. 98-106.
24. Mitchell, S., et al., *Identification and validation of a gene causing cross-resistance between insecticide classes in Anopheles gambiae from Ghana.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**: p. 6147-6152
25. Edi, C.V., et al., *CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito Anopheles gambiae.* PLoS Genetics, 2014. **10**(3): p. e1004236-e1004236.
26. Vontas, J., et al., *Rapid selection of a pyrethroid metabolic enzyme CYP9K1 by operational malaria control activities.* Proceedings of the National Academy of Sciences of the United States of America, 2018. **115**(18): p. 4619-4624.
27. Miles, A., et al., *cggh/scikit-allel*. 2019, Zenodo.
28. Garud, N.R., et al., *Recent selective sweeps in North American Drosophila melanogaster show signatures of soft Sweeps.* PLOS Genetics, 2015. **11**(2): p. e1005004.
29. Wat'senga, F., et al., *Intensity of pyrethroid resistance in Anopheles gambiae before and after a mass distribution of insecticide-treated nets in Kinshasa and in 11 provinces of the Democratic Republic of Congo.* Malaria Journal, 2020. **19**(1): p. e169.
30. 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, 2018. **17**.
31. Lynd, A., Weetman, D., Barbosa, S., Yawson, A.E., Mitchell, S., Pinto, J., Hastings, I. and Donnelly, M.J., *Field, genetic and modelling approaches show strong positive selection acting upon an insecticide resistance mutation in Anopheles gambiae s.s.* Molecular Biology and Evolution, 2010. **27**: p. 1117-1125.
32. Yunta, C., et al., *Cross-resistance profiles of malaria mosquito P450s associated with pyrethroid resistance against WHO insecticides.* Pesticide Biochemistry and Physiology, 2019. **161**: p. 61-67.
33. Yeka, A., et al., *Malaria in Uganda: Challenges to control on the long road to elimination I. Epidemiology and current control efforts.* Acta Tropica, 2012. **121**(3): p. 184-195.
34. Ministry of Health, R.o.K., *Insecticide Resistance Management Plan, Kenya 2020 – 2024*. 2019, Ministry of Health, Republic of Kenya.

35. Ruiz, J.L., L.C. Ranford-Cartwright, and E. Gómez-Díaz, *The regulatory genome of the malaria vector Anopheles gambiae: integrating chromatin accessibility and gene expression.* bioRxiv, 2020: p. 2020.06.22.164228.
36. Liu, B., et al., *iEnhancer-2L: a two-layer predictor for identifying enhancers and their strength by pseudo k-tuple nucleotide composition.* Bioinformatics, 2015. **32**(3): p. 362-369.
37. Liu, B., et al., *iEnhancer-EL: identifying enhancers and their strength with ensemble learning approach.* Bioinformatics, 2018. **34**(22): p. 3835-3842.
38. Schmidt, J.M., et al., *Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus.* PLoS Genetics, 2010. **6**(6): p. e1000998.
39. Jones, C., et al., *Footprints of positive selection associated with a novel mutation (N1575Y) in the voltage gated sodium channel of Anopheles gambiae.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**: p. 6614-6619.
40. Djogbenou, L., et al., *Ace-I duplication in Anopheles gambiae: a challenge for malaria control.* Malaria Journal, 2009. **8**.
41. Boakye, D.A., et al., *Patterns of household insecticide use and pyrethroid resistance in Anopheles gambiae sensu stricto (Diptera: Culicidae) within the Accra metropolis of Ghana.* African Entomology, 2009. **17**(2): p. 125-130.
42. Grau-Bové, X., et al., *Evolution of the insecticide target Rdl in African Anopheles is driven by interspecific and interkaryotypic introgression.* bioRxiv, 2020: p. 2019.12.17.879775.
43. Neafsey, D.E., et al., *Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes.* Science, 2015. **347**(6217): p. 43-+.
44. Nene, V., et al., *Genome sequence of Aedes aegypti, a major arbovirus vector.* Science, 2007. **316**(5832): p. 1718-1723.
45. Buss, D.S. and A. Callaghan, *Molecular comparisons of the Culex pipiens (L.) complex esterase gene amplicons.* Insect Biochemistry and Molecular Biology, 2004. **34**(5): p. 433-441.
46. Rugnao, S., et al., *LLIN Evaluation in Uganda Project (LLINEUP): factors associated with childhood parasitaemia and anaemia 3years after a national long-lasting insecticidal net distribution campaign: a cross-sectional survey.* Malaria Journal, 2019. **18**: p. e207.
47. Newcombe, R.G., *Two-sided confidence intervals for the single proportion: comparison of seven methods.* Statistics in Medicine, 1998. **17**: p. 857-872.

Supplementary materials

Supplementary materials 1
Appendix 1. LNA assay for haplotype variant detection. 2
Appendix 2. Heterologous expression of *An. gambiae* CYP6P4 and CPR using an *E.coli* expression system 4
Synthesis and site directed mutagenesis 4
Transformation 5
P450 expression 5
Purification of CYP6P4 and AgCPR from bacterial cultures 6
Cytochrome P450 quantification 6
Anopheles gambiae P450 reductase quantification 7
In vitro insecticide metabolism assay 7
Appendix 3. Heterologous expression of *An. gambiae* CYP6AA1 and CPR using bac to bac expression system 9
Generation of Blunt-End PCR products 9
Cloning to pFastBac/CT TOPO donor plasmid and analysis of positive clones 10
Generation of Recombinant Bacmids 15
Transformation of DH10Bac competent cells 15
Analysis of positive clones with colony PCR 15
Isolation of Recombinant Bacmids 16
Production of the recombinant baculovirus 17
Insect cell culture and transfection 17
Virus isolation and amplification 17
Metabolism assays 18
Appendix 4. Metabolism of pyrethroids by heterologously expressed P450s 19
Cyp6aa1 19
Appendix 5. *Cyp6aap* duplication frequency in collections from Cote d'Ivoire 20
References 20
User Guides 20

Appendix 1. LNA assay for haplotype variant detection.

Three, independent, locked-nucleic acid (LNA) probe-based PCR assays were designed to genotype the *Cyp6p4-I236M* SNP, the insertion of the partial ZZB-TE, and the duplication of *Cyp6aa1.* For the *Cyp6p4*-I236M assay, LNA-probes labelled with Fam and Hex target the wild type and mutant alleles respectively, whilst ZZB-TE locus is targeted by Hex and Fam probes complementary to the wild type DNA and the insert (mutant) DNA respectively. *Cyp6p4*-I236M and ZZB-TE assays produce codominant markers. The duplicated region of the haplotype, *Cyp6aap*-Dup1, is detected using a combination of three primers and three probes. Primers produce two amplicons, one within the duplicated region and one spanning the junction site of the duplication. A Hex-labelled probe targets the duplicated region, a Cy5-labelled probe is complementary to DNA external to the duplication to provide a reference value for single copy gDNA and a Fam-labelled probe hybridizes to the junction between duplicated genes to detect mutant duplication genotypes. Wild-type homozygotes are indicated by the absence of Fam signal in the presence of both Cy5 and Hex. Heterozygotes and homozygotes for the duplication are differentiated by analysis of the ratio of the Hex, Fam and Cy5 Ct values: 2x Cy5 - (Fam + Hex). Ratio values are then arranged in ascending order and plotted as a line graph. Heterozygotes and homozygotes can be differentiated by a change in the gradient of the line

All LNA probes were designed (http://biophysics.idtdna.com/) with a Tm difference of at least 16°C between target and non-target sequence. (Supplementary materials Table 1). PCR reactions were carried out in a total reaction volume of 10µl, consisting of 1X Luna® Universal Probe qPCR Master Mix (New England Biolabs, USA), 0.35µM of each primer (Supplementary materials Table 1), 0.2µM of each probe and 1µl of genomic *Anopheles* DNA. Reaction conditions consisted of a 3 minute denature at 95°C; 40 cycles of denaturation for 10 seconds at 95°C, annealing for 20 seconds at 58°C for the Zanzibar-6P assay, or 57°C for the *Cyp6aa1*-Dup1 assay, followed by an extension of 10 seconds at 72°C. The *Cyp6p4*-I236M assay consisted of a 3 minute denature at 95°C; 20 cycles of denaturation for 15 seconds at 95°C, annealing for 30 seconds at 66°C; 23 cycles of denaturation for 10 seconds at 95°C, annealing for 20 seconds at 58°C, and an extension of 10 seconds at 72°C. Reactions were run on an AriaMx Real-Time PCR system (Agilent, USA) equipped with Fam, Hex and Cy5 filters.

Assay	Primer/Probe	5'modification	Sequence	3'modification
Cyp6p4-I236M	CYP6P4_i236M_F		AGTTTATGTTTGCGACCACGTT	
	CYP6P4_i236M_R		TCCACCGTCTCGCGCACAAC	
	CYP6P4_I-wild	FAM	TTC+ATGC+C+G+ATGC	Iowa Black®FQ
	CYP6P4_M-Mutant	HEX	TTC+ATGC+C+C+ATGC	Iowa Black®FQ
ZZB-TE	ZZB_flank_A		CAAAATCAATGKCACRGAGC	
	ZZB_flank_B		CGCTACAATGAAKGGAAAGTC	
	ZZB_int_A		CATTACATGGCGACCGTACCT	
	ZZB_MUT	FAM	A+C+ATTA+CA+CTTTGT+CA+GTA	Iowa Black®FQ
	ZZB_WT	HEX	GATG+TT+CTTTK+T+CA+G+TATT	Iowa Black®FQ
Cyp6aap-Dup1	AA1_Dup1_ins1		CAGTGCGGTACGCTCGTTAA	
	AA1_Dup1_outs		GGATCGGTTTACAGCGGACG	
	AA1_Dup1_ins2		CATCACCTGTGCTCGCAARTT	
	AA1_Dup1_outs	Cy5	CCAT+CA+C+CGAA+CG	Iowa Black®Rq-Sp
	AA1_Dup1_ins	HEX	AGAA+CCTGCA+C+CAA	Iowa Black®FQ
	AA1_Dup1_junct	FAM	A+CA+AT+TAATT+G+CAT+CGG	Iowa Black®FQ

Supplementary table 1.1. Primer and probe sequences for the LNA assays to detect the *Cyp6* cluster variants.

Appendix 2. Heterologous expression of *An. gambiae* CYP6P4 and CPR using an *E.coli* expression system

Synthesis and site directed mutagenesis

Sequences of CYP6P4 variant 1 (236M) were synthesised by Eurofins Genomics. For optimal expression of the variants in bacteria, the gene sequences were modified to contain *E. coli* optimized codons and a bacterial *ompA* leader sequence upstream of the start codon. The leader sequence had an Nde1 restriction site at the start codon while its terminal end was linked to the 5' end of the cytochrome P450 gene with an alanine-proline linker sequence. *Cyp6p4* variants were cloned into a pCY066 plasmid.

The reference sequence (PEST strain) *Cyp6p4*-236I was generated by site directed mutagenesis (SDM). The mutagenic primers were designed in NEBaseChanger and a NEB Q5® site directed mutagenesis kit used for SDM. We designed the mutagenic primers such that the 5' ends of the forward and reverse primers would anneal back to back on the target site to amplify the gene as well as the plasmid. Apart from the single nucleotide change, the primers also introduced restriction sites used for screening SDM mutants. SDM, screening and sequencing primers are shown on Table 2.1. SDM was carried out as per the protocol provided by the manufacturer. The protocol had three steps. The exponential amplification step to introduce mutations, the Kinase, ligation and DpnI treatment step to recyclize the plasmid and the transformation step to introduce plasmids in bacteria for amplification and screening. Agar plates, supplemented with ampicillin (50mg/ml) were used to select transformed bacteria. Screening successful SDM mutants was approached in two ways, a nested PCR that comprised of colony PCR to confirm the colonies had the insert and a second PCR targeting the mutagenized region. Sequencing was performed to confirm successful SDM. The other approach was restriction analysis where single colonies from the selection plate were cultured overnight in Luria-Bertani broth at 37°C and their plasmids purified using a QIAprep® Spin Miniprep kit as per manufacturer's protocol. Restriction analysis with Nde1 and Not1 was done to screen colonies with an intact gene and plasmid sequences. Another digest with enzymes targeting the restriction sites introduced alongside the single nucleotide change was done. In all restriction analysis, 5 µl of plasmids and 5 µl of restriction master mix were incubated at 37°C overnight and visualised by gel electrophoresis. The restriction master mix contained 0. 2 µl of restriction enzyme, 1 µl of NEB buffer 3 for Nde1 and Not1 or NEB cutsmart buffer for Bmt1, Stu1 and BsrG1, and topped up with molecular water to 5 µl. Sequencing was done to confirm the screened successful SDM mutants.

Activity	Orientation	Sequence
SDM primers		
6P4_V1 to 6P4_REF	Forward	aacgcatcGGCATGAAATTAACGGATG;
	Reverse	tggctagcCCTTTAAAGGTCGTAGCAAAC
Screening and sequencing primers		
6P4 Mutants	Forward 1	ACGCGAAAGCCGATCCACTGA
	Forward 2	CGGGAAATGAACAACGTACAA
	Forward 3	ACCGTTCCGGGTACCAAACAT
	Forward 4	CATTGCGATTGCGGTAGCCTTA
	Forward 5	GCTATGTCCTTACCGCATTCGT
	Reverse 1	GAAATTCCCCAGCGACCTGTTG
	Reverse 4	GGTTGACGGGGAAAGGATGAAG
	Reverse 5	ACGTAACAGCGTAATCAGTCCG

Table 2.1. Back to back mutagenic and screening primers used for site directed mutagenesis

Transformation

Each variant was co-transformed with *An. gambiae* P450 reductase (pACYC-AgCPR)[1] in Stellar™ competent *E. coli* cells (Clontech). Fifty microliters of cells thawed on ice were aliquoted into 15 ml loose-cap tubes where 1 µl of the variant and 1 µl of pACYC-AgCPR plasmids were added. The cells were incubated on ice for 30 minutes, heat shocked in a 42°C water bath for 20 seconds and further incubated on ice for 5 minutes. Into the cells, 500 µl of pre-warmed LB broth was added followed by incubation at 37 °C in a shaking incubator set at 200 revolutions per minute (rpm) for 1 hour. The cells were plated in LB agar plates supplemented with 50mg/ml ampicillin and 25 mg/ml chloramphenicol and further incubated overnight at 37°C to select cells that were transformed with both plasmids.

P450 expression

A few colonies from the selection plate were inoculated in 8 ml LB broth supplemented with 8 µl of 50mg/ml ampicillin and 8 µl 25 mg/ml chloramphenicol. The cells were cultured at 37°C in a shaking incubator set at 200 rpm for 12 hours to generate an 8 ml starter culture . Three mls of starter culture used to inoculate 200 ml modified terrific broth supplemented with 50 µl of 4000X trace elements, 200 µl of 1M thiamine hydrochloride, 200 µl of 50mg/ml ampicillin and 200 µl of 25 mg/ml chloramphenicol in a 1 litre Erlenmeyer flask. To prepare 100ml 4000x trace elements stock, 2.45 ferric citrate was added in about 50ml distilled water, stirred over moderate heat until dissolved and then left to cool. To the ferric citrate solution, 10 ml concentrated HCl, 0.131 g of ZnCl2, 0.2 g of CoCl2·6H2O, 0.2 g of Na2MoO4·2H2O, 0.1 g of CaCl2·2H2O, 0.127 g of CuCl2·2H2O, and 0.05 g of H3BO3 were added. The mixture was stirred until dissolved and topped up to 100ml with distilled water. The cells were cultured at 37°C in a shaking incubator set at 200 rpm and their optical density (OD) monitored periodically. At OD 0.4, 200 µl of 1M δ-Aminolevulinic acid (ALA), a

heme precursor was added to the cultures. The incubation temperature and shaking speed were adjusted to 21°C and 120 rpm respectively. The culture monitored periodically until the OD reached 0.6 when expression of P450 was induced with 1mM Isopropyl-β-D-thiogalactopyranoside (IPTG). Periodic monitoring of P450 enzyme concentration by CO-reduced spectrum analysis was done to determine the optimum expression time of the different genes[2].

Purification of CYP6P4 and AgCPR from bacterial cultures

At the optimum expression time, 20-24 hours for CYP6P4, the cultures were stopped by incubation on ice for 10 minutes. They were transferred into centrifuge bottles and centrifuged at 2800g for 20 min at 4°C to pellet the cells. The supernatant was discarded and the pellet re-suspended in 20 ml ice-cold 1X TSE buffer (50mM Tris-acetate, 250mM sucrose and 0.25mM EDTA, pH 7.6). To digest the bacterial cell wall, 250 µl of 20mg/ml lysozyme was added to the re-suspended cells to a final concentration of 0.25 mg/mL and gently stirred at 4°C for one hour. The spheroplasts were pelleted at 2800g for 25 min at 4°C, the supernatant discarded and re-suspended in 8 ml ice-cold spheroplast resuspension buffer (100 mM potassium phosphate, 6 mM magnesium acetate, 20% (v/v) glycerol, and 0.1 mM dithiothreitol (DTT), pH 7.6. To the re-suspended spheroplasts, the following protease inhibitors were added; 10 mg/mL leupeptin, in water; 100 mM phenylmethylsulfonyl fluoride (PMSF), in isopropanol and 10 mg/mL of aprotinin, in 10 mM HEPES, pH 8.0 to a final concentration of 1 µg/mL, 1 mM and 1 µg/mL respectively. To lyse the spheroplasts membranes, sonication in 30 second bursts for a total of 2 minutes was carried out. The suspension was then centrifuged at 30, 000g for 20 minutes at 4°C to separate the membranes from other cell debris. The supernatant containing the membranes was transferred to ultracentrifuge tubes, topped up with buffer and centrifuged at 180, 000g for 1 hour at 4°C to pellet the membranes. The supernatant was discarded and 1ml of ice-cold 1X TSE buffer added to the membranes which were then dispersed using a Dounce tissue grinder (Wheaton). P450 concentration and AgCPR activity in the membrane was measured before being stored in -80°C.

Cytochrome P450 quantification

The concentration of active P450 was determined by comparing the spectra difference of reduced P450-carbon monoxide complex with reduced non-complexed P450. To monitor the concentration of P450 in cell cultures, 1 ml of culture was centrifuged at 16,400 rpm at 4°C to pellet the cells. The cells were gently re-suspended in 1 ml 1X P450 spectrum buffer (100 mM Tris-HCl, 10 mM CHAPS, 40% (v/v) glycerol, and 1 mM EDTA, pH 7.4) to avoid formation of bubbles and a few grains of sodium dithionate (reducing agent) added. The mixture was split into two spectrophotometer cuvettes and a baseline reading taken from 500 to 400 nm with a dual beam Cary 4000 UV-Vis

spectrophotometer. The front cuvette was taken to a fume hood, bubbled with carbon monoxide at the rate of one bubble a minute for 50-80 seconds, mixed gently by inversion and replaced into the machine. The reference cuvette was also mixed by inversion before a second reading from 500 to 400 nm was taken. The concentration of active P450 was calculated by subtracting the peak measurements at 490 nm from that at 450 nm then dividing the difference with the extinction coefficient at 450 nm (0.091 µM/cm). In purified membranes, 50 µl of the membranes are added to 950 µl of 1X P450 spectrum buffer and quantified in a similar manner. However, the quantity of active P450 determined is multiplied by the dilution factor, in our case 20 to get the concentration of the enzyme per ml.

Anopheles gambiae P450 reductase quantification

We measured the quantity of AgCPR by estimating cytochrome c reductase (cyt c) activity. In a microtiter plate, 150 µl of 0.1 mM cyt C prepared in 0.3M potassium phosphate buffer at pH 7.7 was pipetted into six wells. Into each well, 2 µl of the membranes were added. In the first 3 wells, 150 µl of 0.3M potassium phosphate buffer was added to serve as control while 150 µl of 0.1 mM NADPH prepared using the same buffer was added in the other 3 wells to start the reaction. The reaction was monitored by taking readings at 550 nm every minute for 15 minutes with a microplate UV-Vis reader. The slope in mA550 per minute was calculated using a plot derived from OD measured per minute. To determine the specific activity of AgCPR (nmol of cyt c reduced/min/mg), the formula0.0552 * slope / protein concentration assuming a path-length of 0.85cm was used[2]. Protein concentration of the membranes was determined by Bradford assay with bovine serum albumin as the standard [3]. Diethoxyfluorescein (DEF) metabolism assay was also done as described previously to analyze P450 activity on the purified membranes[4].

In vitro insecticide metabolism assay

For each insecticide, 6 assays for the NADPH plus (test) reaction and 6 for the NADPH minus (negative control) were set up per biological replicate. This included 3 technical replicates for reaction with cytochrome b5 (b5) and 3 reactions without b5 in the test and control reactions. In an Eppendorf tube, 2 µl of 0.5 mM insecticide prepared in molecular grade absolute ethanol was added to 48 µl of enzyme mix containing the P450 enzyme with or without b5. To start the reaction, 50 µl of regeneration NADPH mix without NADP and that with NADP were added to the control and test Eppendorf tubes respectively. The tubes were incubated at 30°C on a shaking heating block at 1200 rpm for 2 hours. The reaction was quenched by adding 150 µl of HPLC grade acetonitrile. The tubes were inverted several times to ensure proper mixing of acetonitrile with the reaction mix and then centrifuged at 13,000 rpm for 20 minutes to pellet the membranes to prevent blockage of HPLC

column. We transferred 150 µl of the supernatant into Thermofisher HPLC vials. HPLC run conditions were as previously described[5].Solvent ratio and monitoring wavelength were 85:15 Acetonitrile:water and 226nm for permethrin reactions and 80:20 Acetonitrile:water and 232nm for deltamethrin reactions.

Appendix 3. Heterologous expression of *An. gambiae* CYP6AA1 and CPR using bac to bac expression system

Generation of Blunt-End PCR products

Full length coding sequences of *An. gambiae* CYP6AA1 Ugandan haplotype and *An. gambiae* CPR (transcript AY183375.1, [6] were amplified by PCR from the recombinant plasmid vectors pUAS:AgCYP6AA1 and pCW:AgCPR respectively, using the Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific, Cat. No. F549S) and gene-specific primers (**Table 3.1**). Forward primers include a 5' minimal Kozak sequence (ACC), as in some cases this sequence acts as an enhancer, and reverse primers include the stop codon for expression of the native protein forms. Melting temperatures were calculated with Thermo's Multiple Primer Analyser software (https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html) following polymerase manufacturer's instructions. PCR reactions contained 0.5uM primers, 200uM of each dNTP, 1U polymerase and 1uL plasmid DNA harbouring the template, in a final volume of 50uL. Cycling conditions for CYP6AA1 were: 98°C for 30 seconds followed by 25 cycles of 98°C for 10 seconds, 66°C for 20 seconds and 72°C for 30 seconds and a final extension step at 72°C for 10 minutes. Cycling conditions for AgCPR were: 98°C for 30 seconds followed by 25 cycles of 98°C for 10 seconds, 68°C for 30 seconds and 72°C for 30 seconds and a final extension step at 72°C for 15 minutes. PCR products were separated via gel electrophoresis and excised from gel using Qiagen's QIAquick Gel Extraction Kit (Cat. No. 28704). (**Figure 3.1**).

Table 3.1. Primers used for the cloning of *An. gambiae* genes of interest. In bold the start and end codons of transcription. ACC upstream the start codon is the minimal Kozak seq.

Gene	Primer	Sequence (5'-3')	Prod. size (bp)
AgCYP6AA1 (Ugandan Haplotype)	AgUgCYP6AA1-F AgUgCYP6AA1-R	ACC**ATG**GGTTACGTCAGTGTTGT **TCA**CAGTTTCGTTGCGTTAAGCC	1521
AgCPR (transcript AY183375.1),	AgCPR-F AgCPR-R	ACC**ATG**GACGCCCAGACAGAAA **TTA**GCTCCACACGTCCGCCGA	2043

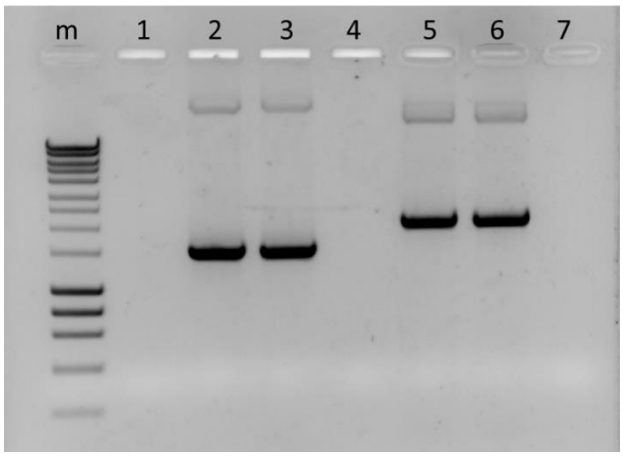


Figure 3.1. Molecular cloning of AgCYP6AA1 and AgCPR. **A.** Agarose gel results, PCR products after amplification with gene specific primers, m: HyperLadder 1kb-Bioline, 1:no sample, 2,3: AgCYP6AA1 (1521bp), 4:Non-template control, 5,6: AgCPR (2043bp), 7:Non-template control. **B** HyperLadder 1kb lanes.

Cloning to pFastBac/CT TOPO donor plasmid and analysis of positive clones

Purified Blunt-End PCR products were cloned into pFastBac/CT-TOPO vector (Invitrogen Cat.No. A11100), which contains a strong polyhedrin (P_H) promoter for high-level baculovirus-based protein expression in insect cells upstream the TOPO Cloning site and a mini-Tn7 elements for site-specific transposition into the bacmid DNA. Cloning reactions, consisting of 1ul of pFastBac/CT-TOPO vector plasmid vector, 1uL of freshly prepared purified PCR product, 1ul of Salt Solution in a final volume of 6uL, were incubated for 5 minutes at room temperature and then used to transform **One Shot Mach1-T1** competent cells. The resulting colonies were selected on ampicillin plates (100ug/mL) and positive clones were analyzed for correct orientation by colony PCR using one internal (gene specific forward primer) and one external primer (SV40 polyA Reverse primer). 10 colonies from each selective plate were analyzed (**Figure 3.2**). The PCR reactions contained 0.2uM primers, 200uM of each dNTP, 1U DreamTaq DNA polymerase and 1 colony as a template, in a final volume of 20uL. The cycling conditions for both CYP6AA1 and CPR were: 95°C for 3 minutes followed by 30 cycles of 95°C for 30 seconds, 42°C for 30 seconds and 72°C for 1 minute and a final extension step at 72°C for 10 minutes, with the annealing temperature calculated according to DreamTaq polymerase guide.

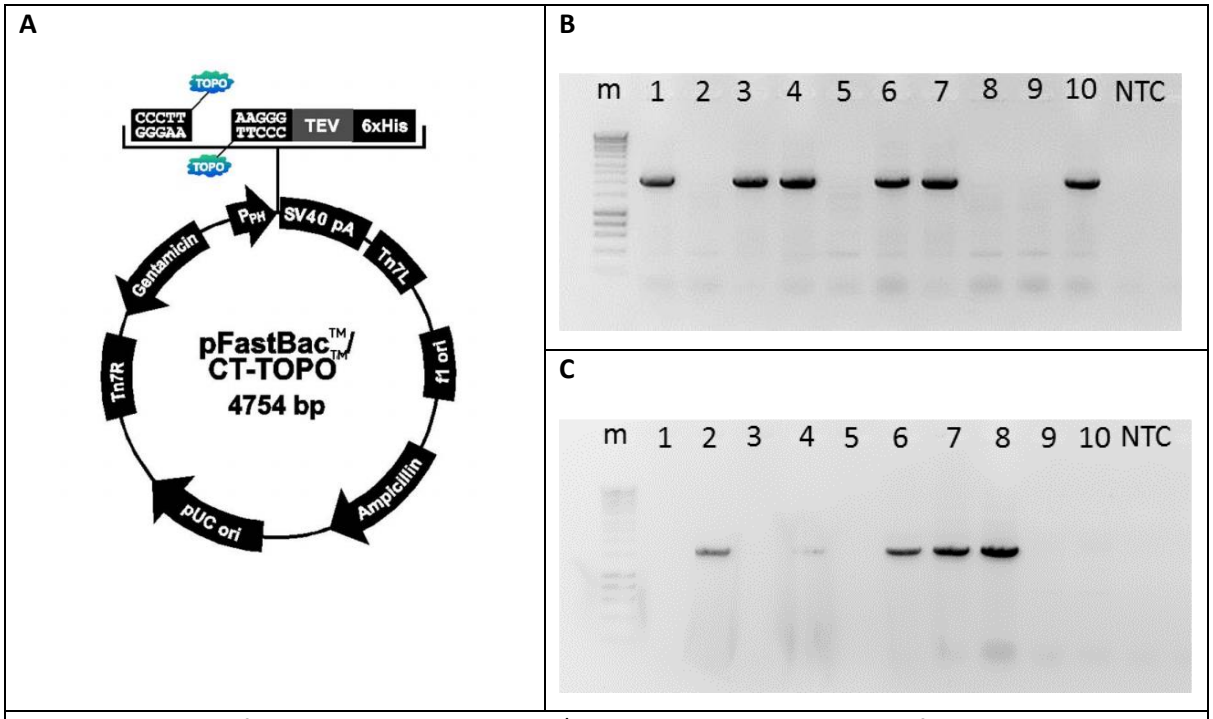


Figure 3.2. A. Map of the donor plasmid pFastBac/CT-TOPO, **B.** Colony PCR results for pFastBac:AgCPR using AgCPR-F and SV40 polyA R primer pair. Lanes 1, 3, 4, 6, 7, and 10 are positive clones. **C.** Colony PCR results for pFastBac:AgCYP6AA1 using AgUgCYP6AA1-F and SV40 polyA R primer pair. Lanes 2, 4, 6, 7 and 8 are positive clones. NTC: No-template control, m: HyperLadder 1kb.

Plasmids from positive clones (one per gene of interest) were grown in LB medium supplemented with 100ug/mL ampicillin, were extracted using QIAprep Spin Miniprep Kit, and were sent for sequencing to validate correct sequences (primers used for sequencing in **Table 3.2**). The sequences from the different reads were processed with BioEdit and were used to make the contigs containing the full sequences which were then aligned with the reference sequences. CYP6AA1 contig has full identity with the reference sequence (Alignment with the reference nucleotide sequence in **Figure 3.3**), whereas CPR contig has two SNPs in the coding sequence corresponding in two missense mutations (V230A and H459Y) in the protein sequence (Nucleotide and amino acid sequence alignment in **Figure 3.4** and 3.**5** respectively). However, these two SNPs correspond to real polymorphisms occurring in the population.

Table 3.2. Primer pairs for the analysis of recombinant pFastBac plasmids and bacmids and internal gene specific primers used for sequencing.

Application	Primer	Sequence (5'-3')
analyzing recombinant plasmids	Polyhedrin Forward primer	AAATGATAACCATCTCGC
	SV40 polyA Reverse primer	GGTATGGCTGATTATGATC
analyzing recombinant bacmids	pUC/M13 Forward primer	CCCAGTCACGACGTTGTAAAACG
	pUC/M13 Reverse primer	AGCGGATAACAATTTCACACAGG
CPR internal primers	CPR 561 F	CGGCATCTACGTGGACAAGC
	CPR 1121 F	TCACCCACTATCTGGAGATAACG
	CPR 1680 R	TCACCCACTATCTGGAGATAACG
CYP6AA1 internal primer	AA1_qPCR_F1	CCACGGTGAATCGGTTCGT

```
CLUSTAL O(1.2.4) multiple sequence alignment

Contig-AA1col2          TAAAAAAACCTATAAATATTCCGGATTATTCATACCGTCCCACCATCGGGCGCGGATCCA      60
AA1_uganda_swept_hap    ------------------------------------------------------------      0

Contig-AA1col2          CCGGTCCCTTACCATGGGTTACGTCAGTGTTGTGTTATTTCTGGTGGTCCCCGCGCTCAC      120
AA1_uganda_swept_hap    -------------ATGGGTTACGTCAGTGTTGTGTTATTTCTGGTGGTCCCCGCGCTCAC      47
                                     **********************************************
Contig-AA1col2          CCTGCTTTACATCTACTTGAAGCAGCACTACCGTCATTGGGCAAATCGGAACCTACCCCA      180
AA1_uganda_swept_hap    CCTGCTTTACATCTACTTGAAGCAGCACTACCGTCATTGGGCAAATCGGAACCTACCCCA      107
                        ************************************************************
Contig-AA1col2          GCTGGAAGCGTCCTTCCCGCTGGGCAACATGAAGGGCGTCGGGTCCGAGATTCACTTCAA      240
AA1_uganda_swept_hap    GCTGGAAGCGTCCTTCCCGCTGGGCAACATGAAGGGCGTCGGGTCCGAGATTCACTTCAA      167
                        ************************************************************
Contig-AA1col2          CGATGTGCTGAACGAAGCGTACGCAAAAGGTAAAGCACAGTCGGCCCCACTCGTCGGGCT      300
AA1_uganda_swept_hap    CGATGTGCTGAACGAAGCGTACGCAAAAGGTAAAGCACAGTCGGCCCCACTCGTCGGGCT      227
                        ************************************************************
Contig-AA1col2          GTACTTTATGCTCAAGCCCATACTGATCGTGACCGAGCTGGACATGGTGAAGCGCATACT      360
AA1_uganda_swept_hap    GTACTTTATGCTCAAGCCCATACTGATCGTGACCGAGCTGGACATGGTGAAGCGCATACT      287
                        ************************************************************
Contig-AA1col2          GGTGAAAGACTTTAGCAGCTTCCACGACCGTGGCCTGTACGTGAACGAGCGGGACGATCC      420
AA1_uganda_swept_hap    GGTGAAAGACTTTAGCAGCTTCCACGACCGTGGCCTGTACGTGAACGAGCGGGACGATCC      347
                        ************************************************************
Contig-AA1col2          CCTGTCCGGGCATCTGTTTGCGCTGGACGGTGAGCGGTGGCGCTATCTGCGCAACAAGCT      480
AA1_uganda_swept_hap    CCTGTCCGGGCATCTGTTTGCGCTGGACGGTGAGCGGTGGCGCTATCTGCGCAACAAGCT      407
                        ************************************************************
Contig-AA1col2          CAGCCCTACGTTCACGTCCGGCAAGATCAAGCTGATGTTCAGCACGATTTGCGAGATCGG      540
AA1_uganda_swept_hap    CAGCCCTACGTTCACGTCCGGCAAGATCAAGCTGATGTTCAGCACGATTTGCGAGATCGG      467
                        ************************************************************
Contig-AA1col2          TGACGAGTTTCTGGCCACGGTGAATCGGTTCGTCGAGCGGGATGAGCCGTTGGATGTGAA      600
AA1_uganda_swept_hap    TGACGAGTTTCTGGCCACGGTGAATCGGTTCGTCGAGCGGGATGAGCCGTTGGATGTGAA      527
                        ************************************************************
Contig-AA1col2          GCTGCTGTCACAGTGCTTCACCTGCGACGTGGTAGGCTCGTGCGCGTTCGGGCTGCAGTG      660
AA1_uganda_swept_hap    GCTGCTGTCACAGTGCTTCACCTGCGACGTGGTAGGCTCGTGCGCGTTCGGGCTGCAGTG      587
                        ************************************************************
Contig-AA1col2          CAACTCGCTCAAAAACGGAGGCTCCAAGCTGCTCGAGATCGGCGACAAGGTGTTCAAACC      720
AA1_uganda_swept_hap    CAACTCGCTCAAAAACGGAGGCTCCAAGCTGCTCGAGATCGGCGACAAGGTGTTCAAACC      647
                        ************************************************************
Contig-AA1col2          ACCTGCGTGGCGAAACATGCTCGCGTTTGCGCTCATTTCGTTTAAGAAGCTGGCGAAACG      780
AA1_uganda_swept_hap    ACCTGCGTGGCGAAACATGCTCGCGTTTGCGCTCATTTCGTTTAAGAAGCTGGCGAAACG      707
                        ************************************************************
Contig-AA1col2          ATTGCGCCTACCCGTGCTACCGGGCGAGGTGACGTCATTCTTTATGCCGCTTGTCACCGA      840
AA1_uganda_swept_hap    ATTGCGCCTACCCGTGCTACCGGGCGAGGTGACGTCATTCTTTATGCCGCTTGTCACCGA      767
                        ************************************************************
Contig-AA1col2          GACGGTGGCGGATCGGGAACGGAACTCGATCGAACGGCCCGATTTTCTCAATCTGCTGAT      900
AA1_uganda_swept_hap    GACGGTGGCGGATCGGGAACGGAACTCGATCGAACGGCCCGATTTTCTCAATCTGCTGAT      827
                        ************************************************************
Contig-AA1col2          ACAGCTCAAGAACAAAGGTACGGTCGAGGACGAATCAACGGAGGGACTGCAAAAGCTAAC      960
AA1_uganda_swept_hap    ACAGCTCAAGAACAAAGGTACGGTCGAGGACGAATCAACGGAGGGACTGCAAAAGCTAAC      887
                        ************************************************************
Contig-AA1col2          GCTCGACGAGGTGGCAGCGCAGGCGTTCGTGTTCTTCTTCGCCGGGTTCGAAACGTCCTC      1020
AA1_uganda_swept_hap    GCTCGACGAGGTGGCAGCGCAGGCGTTCGTGTTCTTCTTCGCCGGGTTCGAAACGTCCTC      947
                        ************************************************************
Contig-AA1col2          GACAACGCTCTCGTTCGCACTGTTCGAGTTGGCCAACAATCCCGACATTCAGGAGCGAGT      1080
AA1_uganda_swept_hap    GACAACGCTCTCGTTCGCACTGTTCGAGTTGGCCAACAATCCCGACATTCAGGAGCGAGT      1007
                        ************************************************************
Contig-AA1col2          GCGGGCGGAGGTGCTGGATAAGCTTCAACTCCACGACGGCAAGATAACGTACGACGCACT      1140
AA1_uganda_swept_hap    GCGGGCGGAGGTGCTGGATAAGCTTCAACTCCACGACGGCAAGATAACGTACGACGCACT      1067
```

```
                          ************************************************************
Contig-AA1col2            GAAGGAGATGATGTACCTCGATCAGGTCATCAATGAGACACTTCGTATGTATCCGCCAGT        1200
AA1_uganda_swept_hap      GAAGGAGATGATGTACCTCGATCAGGTCATCAATGAGACACTTCGTATGTATCCGCCAGT        1127
                          ************************************************************
Contig-AA1col2            TCCGCAGCTGATACGCGTCACCACTCAACCGTACAAGGTGGAGGGAGCGAACGTAACGCT        1260
AA1_uganda_swept_hap      TCCGCAGCTGATACGCGTCACCACTCAACCGTACAAGGTGGAGGGAGCGAACGTAACGCT        1187
                          ************************************************************
Contig-AA1col2            CGAGCCCGACACGATGCTGATGATACCGATCTACGCGATACATCACGATGCCAGCATCTA        1320
AA1_uganda_swept_hap      CGAGCCCGACACGATGCTGATGATACCGATCTACGCGATACATCACGATGCCAGCATCTA        1247
                          ************************************************************
Contig-AA1col2            TCCCGACCCGGAACGTTTCGATCCGGATCGGTTTGCACTGGCCGCGACGCACGCAAGGCA        1380
AA1_uganda_swept_hap      TCCCGACCCGGAACGTTTCGATCCGGATCGGTTTGCACTGGCCGCGACGCACGCAAGGCA        1307
                          ************************************************************
Contig-AA1col2            CACTCACGCGTTCCTGCCGTTCGGCGATGGGCCGCGGAATTGCATCGGCATGCGCTTTGG        1440
AA1_uganda_swept_hap      CACTCACGCGTTCCTGCCGTTCGGCGATGGGCCGCGGAATTGCATCGGCATGCGCTTTGG        1367
                          ************************************************************
Contig-AA1col2            GCTGCTGGAGGTAAAGTTTGGCATCGTGCAGATGGTGAGCAAGCTGCGGTTCACCGTCAG        1500
AA1_uganda_swept_hap      GCTGCTGGAGGTAAAGTTTGGCATCGTGCAGATGGTGAGCAAGCTGCGGTTCACCGTCAG        1427
                          ************************************************************
Contig-AA1col2            CGAGCGTATGCAGATGCCGCTCCGGATGTCGAAAACGGCCTCCATACTGGTGGCGGAAGG        1560
AA1_uganda_swept_hap      CGAGCGTATGCAGATGCCGCTCCGGATGTCGAAAACGGCCTCCATACTGGTGGCGGAAGG        1487
                          ************************************************************
Contig-AA1col2            GGGCATTTGGCTTAACGCAACGAAACTGTGAAAGGGCGAAAACTTGTACTTTCAAGGCCA        1620
AA1_uganda_swept_hap      GGGCATTTGGCTTAACGCAACGAAACTGTGA-----------------------------        1518
                          ********************************
Contig-AA1col2            TCACCATCACCATCACTAGCTCGAGGCATGC   1651
AA1_uganda_swept_hap      ------------------------------    1518
```

Figure 3.3. Nucleotide sequence from CYP6AA1 positive clone aligned with the reference sequence.

```
CLUSTAL O(1.2.4) multiple sequence alignment

Contig_AgCPR_col1       GTTTTGTAATAAAAAAACCTATAAATATTCCGGATTATTCATACCGTCCCACCATCGGGC   60
AY183375.1              ------------------------------------------------------------   0

Contig_AgCPR_col1       GCGGATCCACCGGTCCCTTACCATGGACGCCCAGACAGAAACGGAAGTGCCCGCGGGAAG   120
AY183375.1              ----------------------ATGGACGCCCAGACAGAAACGGAAGTGCCCGCGGGAAG   38
                                              ***************************************
Contig_AgCPR_col1       CGTGAGCGACGAACCGTTCCTCGGCCCGCTTGACATCGTCCTGCTCGTCAGTCTGCTGGC   180
AY183375.1              CGTGAGCGACGAACCGTTCCTCGGCCCGCTTGACATCGTCCTGCTCGTCAGTCTGCTGGC   98
                        ************************************************************
Contig_AgCPR_col1       CGGCACTGCCTGGTATCTGCTCAAGGGCAAGAAAAAGGAAAGCCAAGCTAGTCAGTTTAA   240
AY183375.1              CGGCACTGCCTGGTATCTGCTCAAGGGCAAGAAAAAGGAAAGCCAAGCTAGTCAGTTTAA   158
                        ************************************************************
Contig_AgCPR_col1       ATCCTACTCGATCCAGCCGACGACGGTGAACACGATGACGATGGTGGAGAACTCGTTCAT   300
AY183375.1              ATCCTACTCGATCCAGCCGACGACGGTGAACACGATGACGATGGTGGAGAACTCGTTCAT   218
                        ************************************************************
Contig_AgCPR_col1       CAAGAAGCTACAGTCCTCGGGCCGCCGGCTCGTAGTGTTTTACGGTTCCCAAACAGGCAC   360
AY183375.1              CAAGAAGCTACAGTCCTCGGGCCGCCGGCTCGTAGTGTTTTACGGTTCCCAAACAGGCAC   278
                        ************************************************************
Contig_AgCPR_col1       GGCAGAGGAATTTGCCGGTCGCCTGGCGAAGGAAGGAATCCGCTACCAAATGAAGGGCAT   420
AY183375.1              GGCAGAGGAATTTGCCGGTCGCCTGGCGAAGGAAGGAATCCGCTACCAAATGAAGGGCAT   338
                        ************************************************************
Contig_AgCPR_col1       GGTCGCCGACCCAGAGGAGTGCAATATGGAAGAGCTGCTGATGCTGAAGGACATCGACAA   480
AY183375.1              GGTCGCCGACCCAGAGGAGTGCAATATGGAAGAGCTGCTGATGCTGAAGGACATCGACAA   398
                        ************************************************************
Contig_AgCPR_col1       ATCGTTGGCCGTGTTTTGCTTAGCGACGTACGGCGAGGGCGACCCGACGGACAACTGCAT   540
AY183375.1              ATCGTTGGCCGTGTTTTGCTTAGCGACGTACGGCGAGGGCGACCCGACGGACAACTGCAT   458
                        ************************************************************
Contig_AgCPR_col1       GGAGTTCTACGACTGGATTCAAAACAACGATCTAGATATGACCGGTTTGAATTACGCGGT   600
AY183375.1              GGAGTTCTACGACTGGATTCAAAACAACGATCTAGATATGACCGGTTTGAATTACGCGGT   518
                        ************************************************************
Contig_AgCPR_col1       GTTTGGCCTTGGCAACAAAACGTACGAGCACTACAACAAGGTCGGCATCTACGTGGACAA   660
AY183375.1              GTTTGGCCTTGGCAACAAAACGTACGAGCACTACAACAAGGTCGGCATCTACGTGGACAA   578
                        ************************************************************
Contig_AgCPR_col1       GCGTCTCGAGGAGCTCGGCGCGAACAGAGTGTTTGAGCTCGGTCTGGGTGACGATGATGC   720
AY183375.1              GCGTCTCGAGGAGCTCGGCGCGAACAGAGTGTTTGAGCTCGGTCTGGGTGACGATGATGC   638
                        ************************************************************
Contig_AgCPR_col1       CAACATCGAGGACTACTTCATCACGTGGAAGGAAAAGTTTTGGCCCACGGCTTGCGACTA   780
AY183375.1              CAACATCGAGGACTACTTCATCACGTGGAAGGAAAAGTTTTGGCCCACGGTTTGCGACTA   698
                        ********************************************** *********
Contig_AgCPR_col1       CTTCGGCATCGAGAGCACGGGCGAGGATGTGCTGATGCGGCAGTACCGCCTGCTGGAGCA   840
AY183375.1              CTTCGGCATCGAGAGCACGGGCGAGGATGTGCTGATGCGGCAGTACCGCCTGCTGGAGCA   758
                        ************************************************************
Contig_AgCPR_col1       GCCGGACGTGAGCGCGGACCGCATCTACACCGGCGAGGTGGCCCGGCTCCATTCGCTCCA   900
AY183375.1              GCCGGACGTGAGCGCGGACCGCATCTACACCGGCGAGGTGGCCCGGCTCCATTCGCTCCA   818
                        ************************************************************
Contig_AgCPR_col1       GACGCAGCGGCCACCGTTCGACGCGAAGAACCCGTTCCTCGCCCCGATCAAGGTGAACCG   960
AY183375.1              GACGCAGCGGCCACCGTTCGACGCGAAGAACCCGTTCCTCGCCCCGATCAAGGTGAACCG   878
                        ************************************************************
Contig_AgCPR_col1       GGAGCTGCACAAGGCGGGCGGCCGGTCCTGCATGCACGTCGAGTTCGACATCGAGGGCTC   1020
AY183375.1              GGAGCTGCACAAGGCGGGCGGCCGGTCCTGCATGCACGTCGAGTTCGACATCGAGGGCTC   938
                        ************************************************************
Contig_AgCPR_col1       GAAGATGCGGTACGAGGCGGGCGACCATCTCGCGATGTACCCGGTGAACGATCGCGATCT   1080
```

```
AY183375.1          GAAGATGCGGTACGAGGCGGGCGACCATCTCGCGATGTACCCGGTGAACGATCGCGATCT   998
                    ************************************************************
Contig_AgCPR_col1   GGTCGAGCGGCTCGGCCGGCTGTGCAATGCCGAGCTCGATACGGTCTTCTCGCTCATCAA   1140
AY183375.1          GGTCGAGCGGCTCGGCCGGCTGTGCAATGCCGAGCTCGATACGGTCTTCTCGCTCATCAA   1058
                    ************************************************************
Contig_AgCPR_col1   CACCGACACGGACAGCAGCAAGAAGCATCCGTTCCCGTGCCCCACCACCTACCGGACCGC   1200
AY183375.1          CACCGACACGGACAGCAGCAAGAAGCATCCGTTCCCGTGCCCCACCACCTACCGGACCGC   1118
                    ************************************************************
Contig_AgCPR_col1   GCTCACCCACTATCTGGAGATAACGGCGCTGCCGCGCACCCACATCCTGAAGGAGCTGGC   1260
AY183375.1          GCTCACCCACTATCTGGAGATAACGGCGCTGCCGCGCACCCACATCCTGAAGGAGCTGGC   1178
                    ************************************************************
Contig_AgCPR_col1   CGAGTACTGCGGCGAGGAGAAGGACAAGGAGTTCCTGCGCTTCATCTCGTCGACCGCGCC   1320
AY183375.1          CGAGTACTGCGGCGAGGAGAAGGACAAGGAGTTCCTGCGCTTCATCTCGTCGACCGCGCC   1238
                    ************************************************************
Contig_AgCPR_col1   CGACGGCAAGGCGAAGTACCAGGAGTGGGTGCAGGACAGCTGCCGCAACATCGTGCACGT   1380
AY183375.1          CGACGGCAAGGCGAAGTACCAGGAGTGGGTGCAGGACAGCTGCCGCAACATCGTGCACGT   1298
                    ************************************************************
Contig_AgCPR_col1   GCTCGAGGACATCCCGTCCTGCCATCCGCCGATCGATCACGTGTGCGAGCTGCTGCCCCG   1440
AY183375.1          GCTCGAGGACATCCCGTCCTGCCATCCGCCGATCGATCACGTGTGCGAGCTGCTGCCCCG   1358
                    ************************************************************
Contig_AgCPR_col1   GCTGCAGCCCCGCTACTACTCGATCTCCTCCTCGTCCAAGCTGCACCCGACGACGGTGCA   1500
AY183375.1          GCTGCAGCCCCGCTACCACTCGATCTCCTCCTCGTCCAAGCTGCACCCGACGACGGTGCA   1418
                    **************** *********************************************
Contig_AgCPR_col1   CGTGACCGCGGTGCTGGTGAAGTACGAGACGAAGACGGGCCGGCTGAACAAGGGCGTCGC   1560
AY183375.1          CGTGACCGCGGTGCTGGTGAAGTACGAGACGAAGACGGGCCGGCTGAACAAGGGCGTCGC   1478
                    ************************************************************
Contig_AgCPR_col1   GACGACCTTCCTCGCGGAGAAGCACCCGAACGATGGGGAGCCGGCACCGCGCGTACCAAT   1620
AY183375.1          GACGACCTTCCTCGCGGAGAAGCACCCGAACGATGGGGAGCCGGCACCGCGCGTACCAAT   1538
                    ************************************************************
Contig_AgCPR_col1   CTTCATCCGCAAGAGCCAGTTCCGGTTGCCGCCCAAGCCGGAAACGCCCGTGATCATGGT   1680
AY183375.1          CTTCATCCGCAAGAGCCAGTTCCGGTTGCCGCCCAAGCCGGAAACGCCCGTGATCATGGT   1598
                    ************************************************************
Contig_AgCPR_col1   GGGGCCCGGCACCGGGCTGGCACCGTTCCGGGGCTTCATCCAGGAGCGGGACCACTGCAA   1740
AY183375.1          GGGGCCCGGCACCGGGCTGGCACCGTTCCGGGGCTTCATCCAGGAGCGGGACCACTGCAA   1658
                    ************************************************************
Contig_AgCPR_col1   GCAGGAGGGCAAGGAGATTGGCCAGACGACGCTGTACTTTGGCTGCCGCAAGCGCTCCGA   1800
AY183375.1          GCAGGAGGGCAAGGAGATTGGCCAGACGACGCTGTACTTTGGCTGCCGCAAGCGCTCCGA   1718
                    ************************************************************
Contig_AgCPR_col1   GGACTACATCTACGAGGATGAACTGGAAGACTACTCCAAGCGCGGCATCATCAACCTTCG   1860
AY183375.1          GGACTACATCTACGAGGATGAACTGGAAGACTACTCCAAGCGCGGCATCATCAACCTTCG   1778
                    ************************************************************
Contig_AgCPR_col1   CGTTGCGTTCTCGCGCGACCAGGAGAAGAAGGTGTACGTGACGCACCTGCTCGAGCAGGA   1920
AY183375.1          CGTTGCGTTCTCGCGCGACCAGGAGAAGAAGGTGTACGTGACGCACCTGCTCGAGCAGGA   1838
                    ************************************************************
Contig_AgCPR_col1   CTCGGACCTCATATGGAGCGTGATCGGCGAAAACAAGGGACACTTTTACATCTGCGGTGA   1980
AY183375.1          CTCGGACCTCATATGGAGCGTGATCGGCGAAAACAAGGGACACTTTTACATCTGCGGTGA   1898
                    ************************************************************
Contig_AgCPR_col1   TGCGAAAAATATGGCCACCGATGTGCGAAACATTCTGCTGAAGGTCATCCGCTCGAAGGG   2040
AY183375.1          TGCGAAAAATATGGCCACCGATGTGCGAAACATTCTGCTGAAGGTCATCCGCTCGAAGGG   1958
                    ************************************************************
Contig_AgCPR_col1   TGGGCTCAGCGAAACCGAGGCCCAGCAGTACATCAAAAAGATGGAAGCCCAAAAACGATA   2100
AY183375.1          TGGGCTCAGCGAAACCGAGGCCCAGCAGTACATCAAAAAGATGGAAGCCCAAAAACGATA   2018
                    ************************************************************
Contig_AgCPR_col1   CTCGGCGGACGTGTGGAGCTAAAAGGGCGAAAACTTGTACTTTCAAGGCCATCACCATCA   2160
AY183375.1          CTCGGCGGACGTGTGGAGCTAA-------------------------------------   2040
                    **********************
Contig_AgCPR_col1   CCATCACTAGCTCGAGGCATGC       2182
AY183375.1          ----------------------       2040
```

Figure 3.4. Nucleotide sequence from CPR positive clone aligned with the reference sequence. SNPs are shown in red.

```
CLUSTAL O(1.2.4) multiple sequence alignment

Contig_AgCPR_col1   MDAQTETEVPAGSVSDEPFLGPLDIVLLVSLLAGTAWYLLKGKKKESQASQFKSYSIQPT   60
AY183375.1          MDAQTETEVPAGSVSDEPFLGPLDIVLLVSLLAGTAWYLLKGKKKESQASQFKSYSIQPT   60
                    ************************************************************

Contig_AgCPR_col1   TVNTMTMVENSFIKKLQSSGRRLVVFYGSQTGTAEEFAGRLAKEGIRYQMKGMVADPEEC   120
AY183375.1          TVNTMTMVENSFIKKLQSSGRRLVVFYGSQTGTAEEFAGRLAKEGIRYQMKGMVADPEEC   120
                    ************************************************************

Contig_AgCPR_col1   NMEELLMLKDIDKSLAVFCLATYGEGDPTDNCMEFYDWIQNNDLDMTGLNYAVFGLGNKT   180
AY183375.1          NMEELLMLKDIDKSLAVFCLATYGEGDPTDNCMEFYDWIQNNDLDMTGLNYAVFGLGNKT   180
                    ************************************************************

Contig_AgCPR_col1   YEHYNKVGIYVDKRLEELGANRVFELGLGDDDANIEDYFITWKEKFWPTACDYFGIESTG   240
AY183375.1          YEHYNKVGIYVDKRLEELGANRVFELGLGDDDANIEDYFITWKEKFWPTVCDYFGIESTG   240
                    **************************************************.***********

Contig_AgCPR_col1   EDVLMRQYRLLEQPDVSADRIYTGEVARLHSLQTQRPPFDAKNPFLAPIKVNRELHKAGG   300
AY183375.1          EDVLMRQYRLLEQPDVSADRIYTGEVARLHSLQTQRPPFDAKNPFLAPIKVNRELHKAGG   300
                    ************************************************************

Contig_AgCPR_col1   RSCMHVEFDIEGSKMRYEAGDHLAMYPVNDRDLVERLGRLCNAELDTVFSLINTDTDSSK   360
```

```
AY183375.1              RSCMHVEFDIEGSKMRYEAGDHLAMYPVNDRDLVERLGRLCNAELDTVFSLINTDTDSSK   360
                        ************************************************************

Contig_AgCPR_col1       KHPFPCPTTYRTALTHYLEITALPRTHILKELAEYCGEEKDKEFLRFISSTAPDGKAKYQ   420
AY183375.1              KHPFPCPTTYRTALTHYLEITALPRTHILKELAEYCGEEKDKEFLRFISSTAPDGKAKYQ   420
                        ************************************************************

Contig_AgCPR_col1       EWVQDSCRNIVHVLEDIPSCHPPIDHVCELLPRLQPRYYSISSSSKLHPTTVHVTAVLVK   480
AY183375.1              EWVQDSCRNIVHVLEDIPSCHPPIDHVCELLPRLQPRYHSISSSSKLHPTTVHVTAVLVK   480
                        *************************************:**********************

Contig_AgCPR_col1       YETKTGRLNKGVATTFLAEKHPNDGEPAPRVPIFIRKSQFRLPPKPETPVIMVGPGTGLA   540
AY183375.1              YETKTGRLNKGVATTFLAEKHPNDGEPAPRVPIFIRKSQFRLPPKPETPVIMVGPGTGLA   540
                        ************************************************************

Contig_AgCPR_col1       PFRGFIQERDHCKQEGKEIGQTTLYFGCRKRSEDYIYEDELEDYSKRGIINLRVAFSRDQ   600
AY183375.1              PFRGFIQERDHCKQEGKEIGQTTLYFGCRKRSEDYIYEDELEDYSKRGIINLRVAFSRDQ   600
                        ************************************************************

Contig_AgCPR_col1       EKKVYVTHLLEQDSDLIWSVIGENKGHFYICGDAKNMATDVRNILLKVIRSKGGLSETEA   660
AY183375.1              EKKVYVTHLLEQDSDLIWSVIGENKGHFYICGDAKNMATDVRNILLKVIRSKGGLSETEA   660
                        ************************************************************

Contig_AgCPR_col1       QQYIKKMEAQKRYSADVWS 679
AY183375.1              QQYIKKMEAQKRYSADVWS 679
                        *******************
```

Figure 3.5. Amino acid sequence from CPR positive clone aligned with the reference sequence. The two missense mutations are shown in red.

Generation of Recombinant Bacmids

Transformation of DH10Bac competent cells

pFastBac vectors harbouring the genes of interest, pFastBac:AgCYP6AA1 and pFastBac:AgCPR, were transformed into **Max Efficiency DH10Bac** *E.coli* competent cells containing the bacmid - viral DNA. Cells were left to recover for 4 hours and 10^{-1}, 10^{-2} and 10^{-3} dilutions of the transformants were plated in selective plates containing 50 ug/mL kanamycin (Thermo, 11815-024), 10ug/mL tetracycline (Sigma, 87128), 7ug/mL gentamycin (Thermo, 15750-060) and 100ug/mL X-gal and 40ug/mL IPTG for blue-white selection. The plates were incubated for 48 hours at 37°C.

Analysis of positive clones with colony PCR

Positive white colonies, containing the recombinant bacmids were selected (the insertion disrupts the expression of LacZ peptide), re-streaked to selective plates and analyzed via colony PCR using an external (PUC/M13 F) and an internal primer (gene specific reverse). Successful transposition results in a PCR product longer than 4kb. Due to the big amplicon size, LongAmp Hot Start Taq (M0534S, NEB) was used, and annealing temperatures were calculated using NEBs Tm calculator. PCR reactions contained 0.4uM primers, 300uM of each dNTP, 2.5U DNA polymerase and 1 colony as a template, in a final volume of 25uL. Cycling conditions were: 94°C for 30 seconds followed by 25 cycles of 94°C for 20 seconds, 60°C for 30 seconds and 72°C for 3:30 minutes and a final extension step at 72°C for 10 minutes. PCR products were separated via gel electrophoresis (**Figure 3.6**).

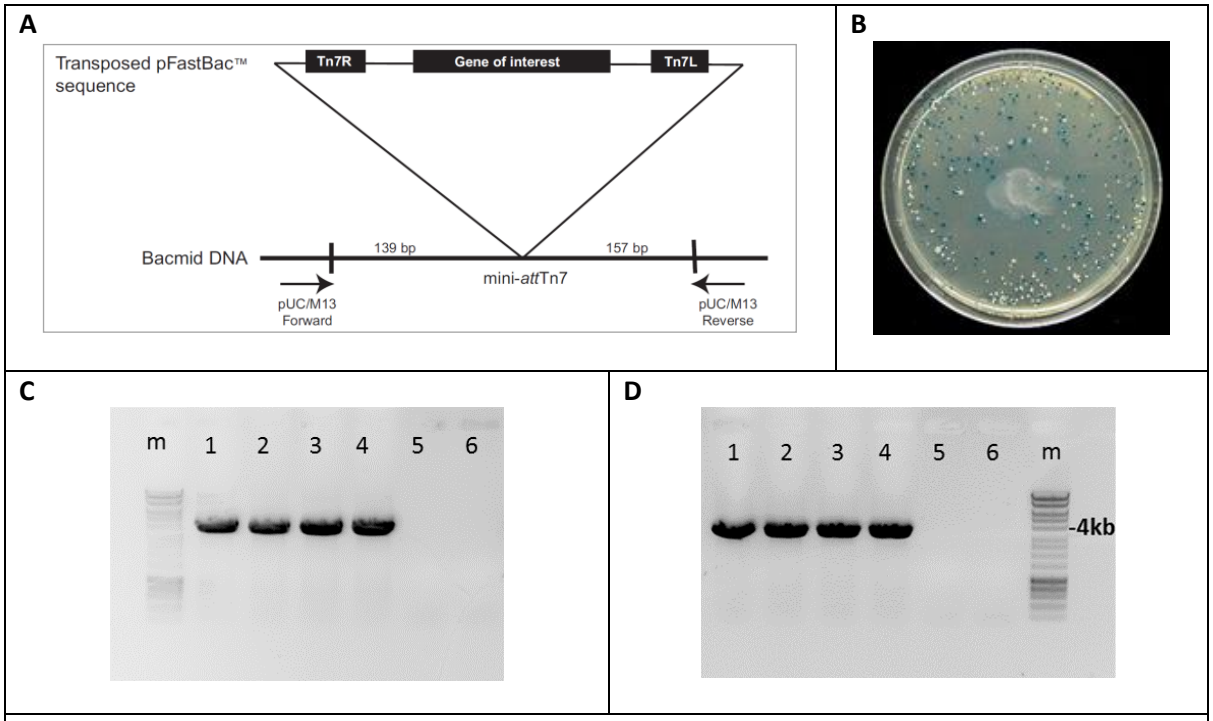


Figure 3.6. A. Mini-Tn7 elements for site-specific transposition into the bacmid DNA. **B.** Selective plate for blue/white selection. White colonies contain recombinant bacmids while blue colonies contain unaltered bacmid. **C.** Colony PCR results for AgCPR using pUC/M13 Forward and AgCPR-R primer pair. Lanes 1-4: positive clones, 5: No template control, 6: empty well. **D.** Colony PCR results for AgCYP6AA1 using pUC/M13 Forward and AgUgCYP6AA1-R primer pair. Lanes 1-4: positive clones, 5: No template control, 6: empty well. m: Bioline HyperLadder 1kb.

Isolation of Recombinant Bacmids

Positive colonies were grown in LB medium supplemented with 50 ug/mL kanamycin, 10ug/mL tetracycline and 7ug/mL gentamycin and recombinant bacmids were isolated using QIAprep Spin Miniprep Kit. Glycerol stocks of the recombinant bacmids in DH10Bac cultures were kept in -80°C. In order to obtain bacmid concentrations sufficient for transfection cultures of 200mL were grown and bacmids were isolated using PureLink HiPure Plasmid Maxiprep Kit (Thermo, K210006) and DNA was resuspended in 200ul TE. Bacmid concentrations are shown in the **Table 3.3** below. The resulting DNA was used to transfect *Sf*9 insect cells.

Table 3.3. Bacmid isolation results

r-bacmid	Concentration (ug/uL)	260/280	260/230
AgCPR bacmid	1.82	2.02	2.55
AgCYP6AA1 bacmid	1.59	2.03	2.55

Production of the recombinant baculovirus

Insect cell culture and transfection

*Spodoptera frugiperda Sf*9 cells were grown in Sf900 III SFM insect cell culture medium (Thermo, Cat.No 12658-019), in the absence of serum with 50ug/mL gentamycin (Thermo, Cat.No 15750-060) in adherent cultures, they were maintained at 27°C and sub-cultured weekly. Cells were established (cultured for 4 passages after thawing) and then they were transfected with the recombinant bacmids using ExpiFectamine Sf Transfection Reagent (Thermo, A38915).

Briefly, in a 6-well plate (Appleton Woods, BC010) 1 x 10^6 cells were cultured per well in 3 mL medium. The next day, antibiotic containing Sf900 III SFM medium was removed and replaced by unsupplemented medium. For each transfection sample, 10uL of ExpiFectamine Sf Transfection Reagent were diluted in 250uL Opti-MEM I Reduced Serum Medium (Thermo, 31985062) and incubated for 5 minutes at room temperature. Subsequently 1ug bacmid DNA was added per sample. No DNA was added in the mock transfection control. See **Fig.3.7** for transfection plate layout. After 5 minutes incubation at room temperature the entire DNA-lipid mixture was added dropwise onto the cells and the plate was incubated at 27°C for 7 days. No signs of infection were obvious at that point.

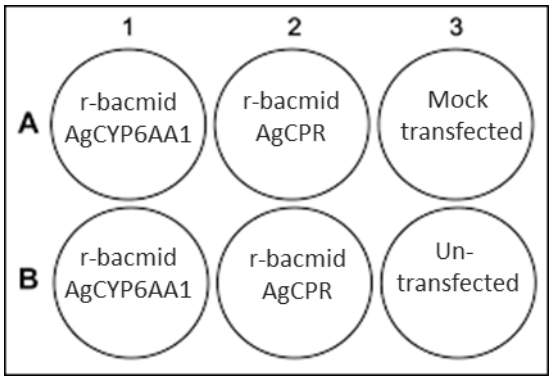


Figure 3.7. Transfection Layout in a 6-well plate.

Virus isolation and amplification

Seven days post transfection the medium containing the P0 viral stocks was collected by centrifugation (2000rpm, 5 minutes), filtered with 0.2um filters and used to infect Sf9 cells in two T25 flasks (Sigma, C6356-200EA) to amplify the viral stocks. One flask was infected with 200uL CYP6AA1-recombinant baculovirus (CYP6AA1rbv) and the other with 200uL CPR-recombinant baculovirus (CPRrbv), and they were inspected daily for visual signs of viral infection in a reverse phase microscope. Signs of successful viral infection include: increased cell diameter, cessation of cell, granular appearance, cell detachment and finally cell lysis. 8 days post infection, the medium containing the P1 viral stocks was collected by centrifugation (2000rpm, 5 minutes) and stored at 4°C. Infected cells were kept for a first expression test. Cells infected with CPRrbv will be checked for CPR expression using the cytochrome c assay and

cells infected with CYP6AA1rbv can be assessed for P450 content and protein formation. In both assays the baseline will be estimated by untransfected cells as a control. P1 viral stocks will be used to identify the virus' multiplicity of infection (MOI) with plague assay, in order to progress with the expression experiments and validate the best ratio for CYP6AA1rbv and CPRrbv co-infection.

Metabolism assays

Deltamethrin and permethrin was dissolved in 100% ethyl alcohol to create 0.5 mM standard solution. The in vitro reactions contained 5 µM deltamethrin or permethrin, 100 pmol of CYP6AA1, 0.25mM $MgCl_2$, 0.1mM NADP+, 1mM glucose-6-phosphate and 1U/mL Glucose-6-phosphate dehydrogenase in 25mM potassium phosphate buffer (pH 7.4) in a total reaction volume of 200uL. NADP+ was not included in the control reactions. Reactions were performed in 5 replicates, and a two-sample T-test of sample reactions (+NADPH) vs the control (−NADPH) was performed for the statistical measurements of substrate depletion. After 2h incubation at 30 °C with 1200 rpm shaking, the reactions were stopped by adding 200µl of acetonitrile to the tubes, and then centrifuged for 20min at 13200xrpm to pellet the proteins. The supernatant was collected and transferred into HPLC glass vials for reverse-phase HPLC analysis. The mobile phase was 80% acetonitrile and 20% H_2O deltamethrin and 85% acetonitrile and 15% H_2O permethrin with flow rate: 1ml/min. The chromatographic analysis was conducted at 23°C and monitored by absorbance at 232nm deltamethrin and 226 nm permethrin.

Appendix 4. Metabolism of pyrethroids by heterologously expressed P450s

Cyp6aa1

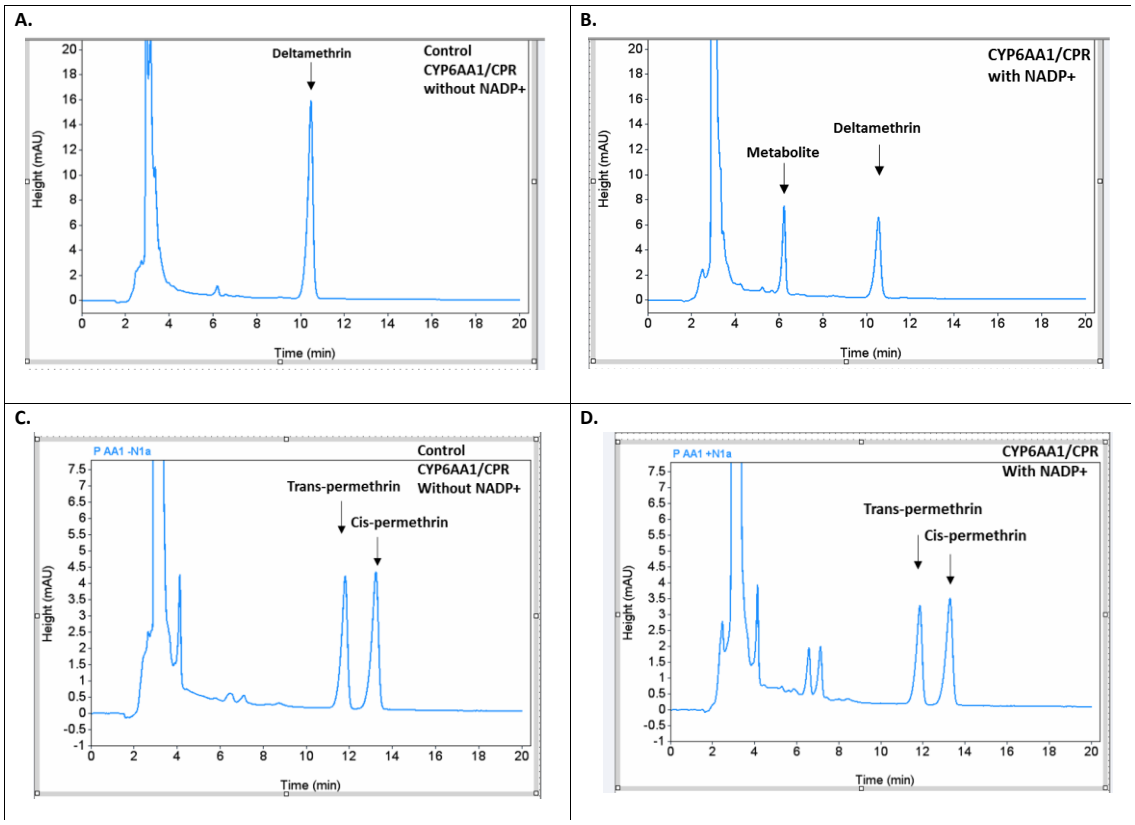


Figure 4.1. Deltamethrin and permethrin metabolism by CYP6AA1/CPR cell pellets. A. Deltamethrin (5 µM) incubated with CYP6AA1/CPR cell pellets for 2h at 30 °C without NADP+. B. Deltamethrin metabolized by Sf9 cell pellets expressing CYP6AA1/CPR with NADP+. Deltamethrin % depletion in the cell extracts expressing CYP6AA1/CPR with and without NADP+. (t-test: t= -9.67; d.f. = 8; P= 9.621E-06). C. Permethrin (5 µM) incubated with CYP6AA1/CPR cell pellets for 2h at 30 °C without NADP+. D. Permethrin metabolized by Sf9 cell pellets expressing CYP6AA1/CPR with NADP+. The black arrows indicate the peaks for cis and trans permethrin D. Permethrin % depletion in the cell extracts expressing CYP6AA1/CPR with and without NADP+. ((t-test: t= -31.085; d.f. = 14; P= 2.554E-14)

Appendix 5. *Cyp6aap* duplication frequency in collections from Cote d'Ivoire

***Cyp6aap* duplication[1]**	**Frequency 2012(n=71)**	**Frequency 2017 (n=164)[2]**	**P value[3]**
Cyp6aap-Dup7	0.324	0.555	0.002
Cyp6aap-Dup10	0	0.05	1
Cyp6aap-Dup11	0.408	0.411	1
Cyp6aap-Dup14	0.465	0.742	0.00008
Cyp6aap-Dup15	0.394	0.331	0.4
Any Cyp6aap-Dup	0.972	0.976	1

Table 5.1. Cyp6aap duplication frequencies in specimens of Anopheles gambiae from Cote d'Ivoire collected in 2012 and 2017. [1]Duplication numbering as per [7]. [2] Sample sizes vary slightly for each Dup due to PCR failures [3]Fisher's exact test

References

1. McLaughlin, L.A., et al., *Characterization of inhibitors and substrates of Anopheles gambiae CYP6Z2.* Insect Molecular Biology, 2008. **17**(2): p. 125-135.
2. Guengerich, F.P., et al., *Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase.* Nature Protocols, 2009. **4**(9): p. 1245-1251.
3. Bradford, M.M., *Rapid and sensitive method for quantitation of microgram quanitities of protein utilizing principle of protein-dye binding.* Analytical Biochemistry, 1976. **72**(1-2): p. 248-254.
4. Yunta, C., et al., *Pyriproxyfen is metabolized by P450s associated with pyrethroid resistance in An. gambiae.* Insect Biochemistry and Molecular Biology, 2016. **78**: p. 50-57.
5. Stevenson, B.J., et al., *Cytochrome P450 6M2 from the malaria vector Anopheles gambiae metabolizes pyrethroids: sequential metabolism of deltamethrin revealed.* Insect Biochemistry and Molecular Biology, 2011. **41**(7): p. 492-502.
6. Nikou, D., H. Ranson, and J.I. Hemingway, *An adult-specific CYP6P450 gene is overexpressed in a pyrethroid-resistant strain of the malaria vector, Anopheles gambiae.* Gene, 2003. **318**: p. 91-102.
7. Lucas, E., et al., *Whole genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes.* Genome Research, 2019. **https://doi.org/10.1101/399568** (29): p. 1250-1261

User Guides

- Bac-to-Bac™ TOPO™ Cloning Kit, Gibco
  - https://assets.thermofisher.com/TFS-Assets/LSG/manuals/bactobac_topo_cloning_kit_man.pdf
- Bac-to-Bac™ TOPO™ Expression System, Gibco
  - https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0000699_BactoBacTOPO_Expression_System_UG.pdf
- Growth and maintenance of insect cell lines, Gibco
  - http://tools.thermofisher.com/content/sfs/manuals/Insect_Cell_Lines_UG.pdf
- Guide to Baculovirus Expression Vector Systems (BEVS) and Insect Cell Culture Techniques
  - http://tools.thermofisher.com/content/sfs/manuals/bevtest.pdf