Wilding, CS, Smith, I, Lynd, A, Yawson, AE, Weetman, D, Paine, MJI and Donnelly, MJ

A cis-regulatory sequence driving metabolic insecticide resistance in mosquitoes: Functional characterisation and signatures of selection

http://researchonline.ljmu.ac.uk/id/eprint/2702/

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)


LJMU has developed LJMU Research Online for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

http://researchonline.ljmu.ac.uk/
A cis-regulatory sequence driving metabolic insecticide resistance in mosquitoes: functional characterisation and signatures of selection

Wilding C.S.¹, Smith I¹, Lynd A.¹, Yawson A.E.², Weetman D.¹, Paine M.J.I.¹ and Donnelly M.J.¹

1. Vector Group, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, U.K.
2. Biotechnology and Nuclear Agriculture Research Institute, Ghana Atomic Energy Commission, Kwabenya, Accra, Ghana

Corresponding author:

Dr Craig Wilding, Vector Group, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, U.K.

Tel: +44(0)151 7053225
Fax: +44(0)151 7053369
Email: c.s.wilding@liverpool.ac.uk

Keywords:

Transposable element, MITE, *Culex quinquefasciatus*, insecticide resistance, promoter, gene expression
Abstract

Although cytochrome P450 (CYP450) enzymes are frequently up-regulated in mosquitoes resistant to insecticides, no regulatory motifs driving these expression differences with relevance to wild populations have been identified. Transposable elements (TEs) are often enriched upstream of those CYP450s involved in insecticide resistance, leading to the assumption that they contribute regulatory motifs that directly underlie the resistance phenotype. A partial CuRE1 (Culex Repetitive Element 1) transposable element is found directly upstream of CYP9M10, a cytochrome P450 implicated previously in larval resistance to permethrin in the ISOP450 strain of Cx. quinquefasciatus, but is absent from the equivalent genomic region of a susceptible strain. Via expression of CYP9M10 in E.coli we have now demonstrated time- and NADPH-dependant permethrin metabolism, prerequisites for confirmation of a role in metabolic resistance, and through qPCR shown that CYP9M10 is >20-fold over-expressed in ISOP450 compared to a susceptible strain. In a fluorescent reporter assay the region upstream of CYP9M10 from ISOP450 drove 10x expression compared to the equivalent region (lacking CuRE1) from the susceptible strain. Close correspondence with the gene expression fold-change implicates the upstream region including CuRE1 as a cis-regulatory element involved in resistance. Only a single CuRE1 bearing allele, identical to the CuRE1 bearing allele in the resistant strain, is found throughout Sub-Saharan Africa, in contrast to the diversity encountered in non-CuRE1 alleles. This suggests a single origin and subsequent spread due to selective advantage. CuRE1 is detectable using a simple diagnostic. When applied to Cx. quinquefasciatus larvae from Ghana we have demonstrated a significant association with permethrin resistance in multiple field sites (mean Odds Ratio = 3.86) suggesting this marker has relevance to natural populations of vector mosquitoes. However, when CuRE1 was excised from the allele used in the reporter assay through fusion PCR, expression was unaffected, indicating that the TE has no direct role in resistance and hence that CuRE1 is acting only as a marker of an as yet unidentified regulatory motif in the association analysis. This suggests that a re-evaluation of the assumption that TEs contribute regulatory motifs involved in gene expression may be necessary.
1. Introduction

The control of a number of pernicious mosquito-borne diseases such as malaria, lymphatic filariasis and dengue is reliant upon the application of insecticides. Insecticide exposure is a potent selective force and resistance can emerge over a relatively short period of time, threatening the efficacy of control programs. Resistance commonly arises as a consequence of mutations in the neuronal target sites of the insecticides or through copy number variation, elevated expression of, or allelic variants in, metabolic detoxification enzymes (Ranson et al., 2011). Resistance mutations in target-sites are commonly non-synonymous changes which can be identified relatively readily and screened for in natural populations. Indeed, for the African malaria vector *Anopheles gambiae*, screening for mutations in the voltage gated sodium channel that confer resistance to DDT and pyrethroid insecticides, e.g. the \(1014F\) and \(1014S\) mutations commonly referred to as knock down resistance or \(kdr\) mutations, has become a routine component of many disease vector control campaigns (e.g. Chanda et al., 2011; Himeidan et al., 2011; Mathias et al., 2011; Yewhalaw et al., 2011). Given that many resistance mutations are thought to be recessive in expression, development of DNA diagnostics enables disease and agricultural control programmes to identify resistance mechanisms when they are at low frequency and take ameliorating action before they cause economic or health impacts (Hemingway, 2009; Kelly-Hope et al., 2008). As shown in a number of recent studies, resistance alleles can rise towards fixation very rapidly (Garcia et al., 2009; Lynd et al., 2010; Mathias et al., 2011), thus advanced warning of segregation of low-frequency resistance alleles in a population could be immensely beneficial.

In whole genome microarray studies of pyrethroid resistant mosquitoes, members of the cytochrome P450 class of metabolic enzymes are frequently up-regulated (Djouaka et al., 2008; Mitchell et al., 2012; Muller et al., 2008; Wondji et al., 2009) suggestive of a predominant role in the resistance phenotype. Confirmatory evidence of a direct role in resistance must come from *in vitro* metabolism of insecticides with heterologously expressed CYP450, and for a limited number of
CYP450s this evidence has been attained (Chiu et al., 2008; Mitchell et al., 2012; Muller et al., 2008; Stevenson et al., 2011). However, whilst gene expression differences coupled with \textit{in vitro} metabolic activities have confirmed the roles of a number of CYP450s in metabolic resistance to insecticides, to date, the causal mutations that result in this over-expression have remained elusive.

Motifs in the region directly upstream of genes responsible for the binding of transcription factors (\textit{cis}-elements), and particularly those in the promoter region, are likely candidate regions for resistance associated variants. However, identification of \textit{cis}-regulatory elements poses a particular challenge. Whilst some progress has been made in our knowledge of these (Lenhard et al., 2012; Wittkopp and Kalay, 2012), and conserved motifs upstream of genes are identifiable in mosquitoes which may represent \textit{cis}-regulatory elements (Sieglaff et al., 2009), validated statistical methods for identification of motifs in sequence data are lacking for many species (Down et al., 2007; Saminadin-Peter et al., 2012). Transposable elements (TEs) are ubiquitous features of mosquito genomes (Boulesteix and Biemont, 2005; Fernandez-Medina et al., 2011; Tu and Coates, 2004) with some 29% of the assembled \textit{Culex quinquefasciatus} genome composed of TEs (Arensburger et al., 2010). TEs can contribute regulatory motifs affecting gene expression when inserted upstream of genes (Ganko et al., 2003; Jordan et al., 2003; Thornburg et al., 2006; van de Lagemaat et al., 2003). TEs with roles in regulating expression levels would therefore represent readily identifiable regulatory elements.

Numerous examples exist of TEs located in close proximity to genes involved in metabolic resistance in insects (Chen and Li, 2007; Chen and Li, 2008; Li et al., 2007). Perhaps the best described example of a TE inserted close to a functional CYP450 is \textit{CYP6g1} in \textit{Drosophila melanogaster} which confers resistance to DDT and for which the allele containing the TE now has a worldwide distribution (Brookfield, 2004; Catania et al., 2004; Daborn et al., 2002; Schlenke and Begun, 2004; Schmidt et al., 2010). In the CYP450 family generally, transposable elements are over-represented in proximity to insect CYP450s involved in metabolic detoxification whilst being largely absent from CYP450s involved in essential processes such as ecdysone biosynthesis and developmental regulation (Chen
and Li, 2007). This suggests not only that TEs are selectively retained upstream of P450s putatively involved in insecticide resistance but that they are likely to have a functional role, perhaps contributing regulatory motifs that control or contribute to the resistance phenotype.

*Culex quinquefasciatus* has a widespread tropical and sub-tropical distribution, vectors West Nile virus and lymphatic filariasis, and is a major biting nuisance in the urban tropics. The JPAL strain of *Cx. quinquefasciatus* was established from Saudi Arabia (Amin and Hemingway, 1989) and exhibits high resistance in the aquatic larval stage to permethrin (a class I pyrethroid) but relatively low resistance to class II pyrethroids such as deltamethrin (resistance ratios of 2,500 and 56 respectively, compared to the susceptible Ogasawara lab strain (Weerasinghe et al., 2001)). In a series of experiments Tomita and colleagues implicated mutations in the sodium channel (*L1014F*), the target-site of permethrin, and increased expression of cytochrome P450s as the likely causes of the permethrin resistant phenotype (Itokawa et al., 2011; Itokawa et al., 2010; Komagata et al., 2010) in JPAL. Using a custom microarray incorporating probes for 197 genes including 62 CYP450s, three genes, *CYP9M10, CYP4H34* and *CYP6Z10*, were detected as up-regulated (39.2, 5.6 and 2.6 fold respectively) in JPAL relative to the Ogasawara strain (Komagata et al., 2010). Two of the three genes, *CYP9M10* and *CYP4H34*, showed much higher transcription in larvae, implicating these as potential mediators of the larval resistance phenotype.

However these studies were not able to determine the relative importance of the target site (*L1014F*) mutation and the CYP450 over-expression in conferring the resistance phenotype. To address this, Hardstone et al. (2007) developed the ISOP450 strain of *Cx. quinquefasciatus* through back-crossing JPAL to the standard susceptible S-LAB strain and selecting larvae at the L4 stage with permethrin, so embedding the pyrethroid resistance locus or loci in the S-LAB genetic background. Via crossing and selection, kdr-mediated resistance was removed such that permethrin resistance in ISOP450 is mediated solely by metabolic resistance. Only *CYP9M10* is highly over-expressed in ISOP450 relative to S-LAB (Hardstone et al., 2010) indicating that this locus, and not *CYP4H34*,
underlies the resistance in this strain and additionally suggesting it may also be the causal mechanism in the parent JPAL strain.

Crosses between the JPAL and Ogasawara strains, followed by allele-specific qPCR were used to infer that the over expression of *CYP9M10* in JPAL was under the control of a *cis*-acting element (Itokawa et al., 2010). A likely regulatory element is *CuRE1* (*Culex* Repetitive Element 1) a *Miniature Inverted-repeat Transposable Element* (MITE) which is found upstream of *CYP9M10* in the JPAL but not the susceptible Ogasawara strain (Itokawa et al., 2010). *CYP9M10* is also duplicated in JPAL (Itokawa et al., 2010) with the duplicate gene copies (*CYP9M10* and *CYP9M10v2*) having identical amino acid sequences and sharing the same upstream sequence, including the *CuRE1* element, up to 1.1kb 5’ of *CYP9M10*. However, this duplication alone is insufficient to explain the resistance patterns. Whether *CuRE1* is an important marker for permethrin resistance in wild populations or contributes regulatory motifs resulting in gene expression differences has not yet been answered.

In this study we demonstrate the ability of heterologously-expressed *CYP9M10* to metabolise permethrin and confirm the findings of Hardstone et al. (2010) that *CYP9M10* is up-regulated in the ISOP450 strain. Using a reporter assay we demonstrate that the genomic region upstream of *CYP9M10* from the resistant strain contains regulatory motifs, but, crucially, that the *CuRE1* element is not the origin of these. This suggests that a re-evaluation of the assumption that TEs inserted upstream of CYP450s consistently contribute functional motifs is necessary. Additionally, we detect evidence of strong selection on this genomic region, the consequence of which is that whilst *CuRE1* is not causal, it is in tight linkage disequilibrium (LD) with the causal motif and can hence act as a marker of metabolic resistance in wild-caught mosquitoes.
2. Materials and Methods

2.1 Strains of Cx. quinquefasciatus used

The ISOP450 strain of *Culex quinquefasciatus* was supplied by Dr J. Scott (Cornell University) and exhibits larval resistance to permethrin (Hardstone et al., 2007). The CqSF strain from Recife, Brazil was supplied by Dr C. Fontes (Centro de Pesquisa Aggeu Magalhães, Recife, Brazil) and is classified as susceptible. To confirm the CqSF phenotype we performed standard larval bioassays following WHO protocols (W.H.O., 2005). Briefly, batches of 25 larvae were exposed to permethrin concentrations (ranging from 0.0001 to 10 ppm) in 100 ml water and mortality determined after 24 hrs. ‘Dead’ individuals were those unable to mount an escape response to capture with forceps. Log-concentration vs mortality curves were generated from these bioassay data and LC$_{50}$ calculated in R (R Development Core Team, 2011).

2.2 RNA isolation and qPCR

RNA was isolated from pools of 30 L4 larvae using either the Picopure kit (Arcturus) or the RNAqueous4PCR kit (Ambion) and cDNA produced from 2.5 μg total RNA using Superscript III (Invitrogen) and an oligo dT$_{20}$ primer. The cDNA was cleaned using a Qiagen PCR purification kit prior to use in qPCR experiments. A five-fold dilution series (1 to 1/15,625) of both CqSF and ISOP450 cDNAs was used to produce standard curves. Primers for qPCR of *CYP9M10* were Cx_9M10qPCR-1 and Cx_9M10qPCR-4, and for the normalising 40S ribosomal protein S7 (CPIJ006763) Cx_S7qPCR-1 and Cx_S7qPCR-2 (see supplementary material for these and all subsequent primer sequences). Both sets of primers span an exon-exon junction. Primers for *CYP4H34* were CYP4H34-1 and CYP4H34-2 of Komagata et al. (2010). Since *CYP4H34* is intron-less these do not span an exon-exon junction.

qPCR reactions were conducted in triplicate in 20 μl volumes containing 1x Agilent Brilliant III SYBR qPCR mastermix, 300 nM each primer (with the exception of primer Cx_9M10qPCR-4 which was used at 600 nM following optimisation) and 1 μl template RNA (1/625 dilution) on an Agilent MX3005 qPCR
system with cycling conditions of 3min at 95°C and 40 cycles of 10s at 95°C and 10s at 60°C. The ΔΔCt method was used for calculation of fold changes (Livak and Schmittgen, 2001).

2.3 Cloning and expression of P450s

Full length cDNAs of CYP9M10 and CYP4H34 from both resistant (ISOP450) and susceptible (CqSF) strains were amplified from cDNA by RT-PCR using primer pairs CYP9M10expFOR1 and CYP9M10expREV1 or CYP4H34expFOR1 and CYP4H34expREV1. Purified products were cloned using the cloneJET PCR cloning system (Fermentas) and sequenced. Full length sequences (Accession numbers JQ001925 and JQ001926 for CqSF and ISOP450 CYP9M10 sequences respectively and JQ001927 and JQ001928 for ISOP450 and CqSF CYP4H34 sequences) are provided in the supplementary material.

For expression, a bacterial ompA leader sequence was fused to both CYP9M10 and CYP4H34 and ligated into the expression vector pCW-ori+ plasmid, pB13 (Pritchard et al., 1998) to produce pB13::CYP9M10 and pB13::CYP4H34, as described previously (McLaughlin et al., 2008; Pritchard et al., 1997). Further details are provided in the supplementary material.

Competent JM109 cells (Promega) were co-transformed with pB13::CYP9M10 or pB13::CYP4H34 and pACYC-AgCPR (McLaughlin et al., 2008) for co-expression of P450 and its redox partner An. gambiae cytochrome P450 reductase (CPR). Expression, membrane isolation and determination of P450 content was carried out as described previously (McLaughlin et al., 2008; Muller et al., 2008; Stevenson et al., 2011) – see supplementary material. P450 activity and concentration was estimated through CO difference spectrum (Omura and Sato, 1964), total protein through Bradford assay with BSA standards and CPR by cytochrome c reduction (Strobel et al., 1978).

2.4 Metabolism assays

Metabolism assays were conducted by incubating the recombinant P450 complex with two pyrethroid insecticides. Permethrin (46% cis:52% trans) and deltamethrin (Greyhound
Chromatography and Allied Chemicals, Birkenhead, U.K.) were dissolved in 20% ethanol before addition to the reaction with a final ethanol content of 2% in metabolism reactions. Reaction conditions were 100μl 0.2M Tris-HCl pH 7.4, 1mM glucose-6-phosphate, 0.25mM MgCl₂, 0.1mM NADP+, 1U/ml glucose-6-phosphate dehydrogenase (G6PDH), 0.1μM CYP9M10 or CYP4H34, 0.8μM b5 and 20μM deltamethrin or permethrin. A negative NADPH control omitted NADP+ and G6PDH. Reactions were incubated at 30°C with shaking (1200rpm) with membranes kept separate from reaction buffer for 5min before the reaction was initiated by vortexing. A tenth of a ml of acetonitrile containing 20μM bifenthrin (as internal HPLC standard) was added to stop reactions. Following shaking (1200rpm) for 20min at 30°C in order to ensure dissolution of all pyrethroids, samples were centrifuged at 20,000g for 5min and the substrate transferred to HPLC vials then analysed by reverse-phase HPLC with a monitoring absorbance of 232nm (Chromeleon, Dionex). 100μl was loaded into an isocratic mobile phase of 90% acetonitrile and 10% water with a flow-rate of 1ml/min and substrate peaks separated with a 250mm C18 column (Acclaim 120, Dionex) at 23°C. Deltamethrin eluted at ≈11.5min and cis- and trans- permethrin at 15.2 and 18.5 min respectively. Time course metabolism assays of permethrin metabolism using CqSF CYP9M10 membranes were conducted as above but for 0, 5, 10, 20, 30 and 60 minutes before reactions were quenched.

2.5 Luciferase reporter assay of CYP9M10 promoter region

The region immediately 5’ of CYP9M10 from both ISOP450 and CqSF strains were amplified using primers 9M10UP5 and 9M10UP8. The closest gene 5’ of CYP9M10 is the CYP450 CPU014219 (encoded on the opposite strand) and the primer 9M10UP5 sequence is situated within CPU014219. Primer 9M10UP8 is sited within the 5’ UTR of CYP9M10. These primers gave a ≈ 1.5kb region from ISOP450 and 834bp from CqSF. Because the breakpoint 5’ of the duplicated CYP9M10 in JPAL, the parent strain of ISOP450, is 1.1kb upstream, the amplified region is from CYP9M10 and not CYP9M10v2 (see Fig. 4 of Itokawa et al. 2010). Products were cloned into pJET1.2 (Fermentas) and sequenced. From these sequences two primer pairs 9M10UP5-KPNI and 9M10UP6-KPNI were
designed sharing a common reverse primer 9M10UP8-HINDIII terminating 3bp from the start of the CYP9M10 coding sequence. Primers incorporated either HindIII or KpnI to facilitate cloning into the pGL3-Basic reporter vector. Products (1504bp and 1315bp for ISOP450, and 834bp and 647bp for CqSF for 9M10UP5-KPNI-9M10UP8-HINDIII and 9M10UP6-KPNI-9M10UP8-HINDIII respectively) were amplified using Phusion high fidelity polymerase (Fermentas) and cloned into PJet1.2. Following excision with HindIII and KpnI, these were ligated into pGL3-Basic (Promega) and transformed into XL1-Blue (Agilent). To remove the CuRE1 element whilst retaining all additional differences between the susceptible and resistant alleles, fusion PCR with primers 9M10UP5-KPNI or 9M10UP6-KPNI and Cure1_fusion_3, and 9M10UP8-HINDIII and Cure1_fusion_4 was undertaken. Products of both PCRs were diluted 1/100 and 1μl used in PCR with 9M10UP5-KPNI or 9M10UP6-KPNI and 9M10UP8-HINDIII to knit the fragments together. Products were cloned into PJet1.2 excised with HindIII and KpnI, ligated into pGL3-Basic and sequenced. Plasmids were extracted with GeneJet plasmid miniprep kit (Fermentas) and concentrations adjusted to 100ng/μl. Dual luciferase assays were undertaken using An. gambiae cell line Sua5.1 (Müller et al., 1999) maintained in Schneider’s Drosophila medium supplemented with 10% foetal calf serum and 1% penicillin/streptomycin. Approximately 5 x 10^5 cells per well were plated into 24-well plates 1 day prior to transfection and allowed to reach 60-70% confluency.

Qiagen effectene transfection reagent was used for transfection of constructs and the Dual-Luciferase Reporter Assay (Promega) used for promoter activity measurements. 600ng reporter constructs (CYP9M10 upstream sequences in pGL3-Basic) and pGL3 without insert or 600ng shuttle construct pSLfa1180fa as controls (http://flybase.org/reports/FBmc0002761.html) were co-transfected with 1ng actin-Renilla internal control in 60μl DNA condensation buffer, 4.8μl enhancer and 6μl effectene in triplicate. Following incubation at 27°C for 48hrs and washing of cells with PBS, cells were harvested in 100μl passive lysis buffer (Promega) and luciferase activity measured on a luminometer (EG&G Berthold). Construct luciferase activity was normalised to Renilla luciferase activity.
2.6 Association of kdr or the CuRE1 insertion upstream of CYP9M10 with permethrin resistance in Ghana

Culex larvae were collected in May to early June 2011 from four sites in the Greater Accra region of southern Ghana (Ashaiman, Kpone, Labadi and Madina) and one site 80km distant (Koforidua) (GPS co-ordinates are provided in the supplementary material). Larval permethrin bioassays were performed on third to fourth instar larvae using the approach detailed above with the exception that assays were conducted in 125ml volumes and at permethrin concentrations ranging from 0.000015 to 3ppm. One hundred larvae were tested at each concentration point. Given that these were wild caught Culex from mixed species collections it was necessary to identify the larvae using a combination of morphological keys (Hopkins, 1952) and molecular assays (Smith and Fonseca, 2004). In addition an 830bp fragment of the 3’ region of the mitochondrial cytochrome oxidase one (COI) gene was also amplified using the primer pair C1-J-2183 and TL2-N-3014 (Simon et al., 1994) and phylogenetic trees constructed to ensure that specimens clustered with known Cx. quinquefasciatus. Only confirmed Cx. quinquefasciatus were included in subsequent analyses. To perform an allele:phenotype association test a sample site-specific definition of resistant and susceptible was made (Table 1). Susceptible larvae were those killed at the lowest concentrations and resistant larvae those that survived exposure to the highest concentrations. These individuals were then screened for the genotype at codon 1014 of the VGSC using the pyrosequencing assay of Wondji et al. (2008) and for the presence of the CuRE1 insertion upstream of CYP9M10 using the protocol of Itokawa et al. (2010). Since the duplication of CYP9M10 could be a cryptic genotypic variable a subset of samples were screened with the duplication diagnostic PCR of Itokawa et al. (2010). Association between the resistance phenotype and genotype was analysed using a chi-squared test in PopTools (Hood, 2008).

2.7 Screening for the CuRE1 insertion in field populations of Culex quinquefasciatus from Africa
Screens for the presence of the CuRE1 element upstream of CYP9M10 were conducted on additional populations using the assay of Itokawa et al. (2010). Confirmed Cx. quinquefasciatus from Nkhotakota Malawi (N = 12), Salima, Malawi (N = 7), Quelimane, Mozambique (N = 6), Lusaka, Zambia (N = 1) and Tororo, Uganda (N = 3 – n.b. these are siblings). A subset of samples were also amplified with primer pair 9M10UP3 and 9M10a, cloned into PJET1.2 (Fermentas) and sequenced.
3. Results

3.1 Bioassays

The LC$_{50}$ for permethrin in the CqSF strain was 0.019 ppm (95% C.I.s 0.012-0.022) giving a resistance ratio (RR) for ISOP450 (LC$_{50}$ 2.1ppm; Hardstone et al. 2009) relative to the susceptible CqSF strain of 112.

3.2 Gene expression and sequencing of CYP9M10 and CYP4H34

Differential expression of CYP9M10 in the ISOP450 strain was confirmed by qPCR. Standard curves for CYP9M10, CYP4H34 and S7 exhibited efficiencies of 91.05, 99.75 and 106.75% respectively. Relative fold-changes of CYP9M10 and CYP4H34 ($\Delta\Delta C_t$ method corrected for efficiency) in the ISOP450 strain relative to CqSF were (Mean (95% CL) 22.94 (16.67-29.21) and 0.6 (0.31-0.88) respectively. Thus, the ISOP450 strain has >20 fold expression of CYP9M10 relative to the CqSF strain but marginally lower (0.6x) CYP4H34 expression.

Cloning and sequencing of CYP9M10 and CYP4H34 was undertaken to investigate the role of allelic variation in the resistance phenotype. Cloning of CYP9M10 revealed that the ISOP450 strain and CqSF strain differ in amino acid sequence from the Johannesburg strain genome reference strain (Vectorbase gene ID CPIJ014218) by four and seven amino acids respectively. The ISOP450 sequence differs from the CqSF strain at 10 amino acids (R143G, M156T, L232V, I245F, V279M, M288I, E289D, K356M, S422T, D445N). Critically, CYP9M10 from both ISOP450 and CqSF encode full-length transcripts as does the Johannesburg reference strain (Itokawa et al. 2011). Therefore relatively greater susceptibility cannot be ascribed simply to a non-functional version of CYP9M10 in CqSF, unlike the Ogasawara susceptible strain studied previously (Itokawa et al., 2010), which contains a deletion resulting in a premature stop codon.
CYP4H34 from ISOP450 strain differs from that of the Johannesburg strain (CPIJ011127) at 6 amino acids, CqSF CYP4H34 differs from CPIJ011127 at 3 amino acids and ISOP450 differs from the CqSF strain at 9 amino acids (T118S, K211M, T257A, V342L, V347D, T357I, V414A, E418K, K480N).

Alignments of CYP9M10 and CYP4H34 are given in Supplementary Figs. 1 and 2; Genbank accession numbers are JQ001925-JQ001928 for CqSF CYP9M10, ISOP450 CYP9M10, ISOP450 CYP4H34 and CqSF CYP4H34 respectively.

3.3 Recombinant expression and functional validation of CYP9M10 and CYP4H34

Through co-expression of both ISOP450 and CqSF CYP9M10 or CYP4H34 with AgCPR we isolated E.coli membranes with 0.08-0.51nmol P450/mg protein and CPR activity of 40-150 nmol cytochrome c reduced/min/mg protein. The majority of the P450 was functional as adjudged by the prominence of the P450 peak relative to the P420 peak in CO-difference spectra (see Supplementary Fig. 3). CYP9M10 and CYP4H34 membranes from both ISOP450 and CqSF clearly metabolized permethrin in an NADP+ dependent fashion (see Fig. 1 and Supplementary Fig. 4a-d). Three metabolite peaks (eluting at ≈7.62min, 8.88min and 10.2min) were evident following 60 min reactions in +NADPH reactions for both allelic variants of CYP9M10 (Fig. S4a-b), indicating that both are able to metabolise permethrin and that allelic variation does not contribute to resistance in the ISOP450 strain. In time-course experiments, these CYP9M10 metabolites increased in a NADPH- and time-dependant manner (Fig. 1A and 1B respectively). Weak metabolism of deltamethrin by CYP9M10 was noted, with minor metabolite peaks at 5.3min, 5.98min and 6.62min seen following 60 min reactions in +NADPH reactions only (Supplementary Fig. 4e-f).

Metabolite peaks for CYP4H34 eluted at 5.54min, 5.91min, 7.63min and 8.76min (Supplementary Fig. 4c-d) indicating that metabolism may involve a different reaction course for this enzyme. Both ISOP450 and CqSF CYP4H34 alleles were able to metabolise permethrin but there was no evidence for metabolism of deltamethrin by CYP4H34 expressed from either allele (Supplementary Fig. 4g-h).
3.4 **Luciferase assay of region upstream of CYP9M10**

The region upstream of *CYP9M10* from both the ISOP450 and CqSF was cloned and sequenced. The upstream region of ISOP450 amplified with primers 9M10UP5 and 9M10UP8 is 1502bp in length whilst that amplified from CqSF is 834bp. A number of SNPs and indels differentiate the upstream region of ISOP450 from CqSF (Supplementary Fig. 5) but the main difference between these regions is the 672bp sequence incorporating a fragment of a MITE (*CuRE1*) in the region upstream of ISOP450 *CYP9M10* that is absent from the region upstream of CqSF *CYP9M10*. In order to localise regulatory motifs, two different sized fragments immediately upstream of the *CYP9M10* start codon were amplified and cloned into the reporter gene vector pGL3. When used in reporter gene assays, inserts from both strains drove reporter gene expression, but expression driven by inserts derived from ISOP450 was 10x greater than that from CqSF (Fig. 2). Using fusion PCR we removed the *CuRE1* element whilst retaining all additional variation. Sequences lacking *CuRE1* drove expression at a similar level to those with the TE (Fig. 2) indicating that regulatory motifs are not located in this sequence.

3.5 **Association of the MITE insertion upstream of CYP9M10 with permethrin resistance in Ghana**

A total of 264 definitively confirmed *Cx. quinquefasciatus* larvae obtained from larval breeding sites across Greater Accra were assayed for permethrin susceptibility. LC$_{50}$ based population-specific resistance ratios relative to the susceptible S-Lab strain are shown in Table 1 and ranged from 38-1,443. Due to differences in the population specific LC$_{50}$ values, diagnostic concentrations varied for each population but were always chosen to establish a clearly distinct resistant and susceptible phenotype, whilst giving appropriate sample sizes for an association test ($N \approx 50$ per site).

There was a significant association between the *CuRE1* insertion upstream of *CYP9M10* and a permethrin resistant phenotype across all five sample sites combined ($p < 0.0001$; Table 1) and for
four of the five collection sites when analysed individually. Where significant, population specific odds ratios (ORs) for the CuRE1 insertion ranged from 2.90 (Labadi) to 6.15 (Koforidua).

The duplication of CYP9M10 described previously in the JPAL strain (Itokawa et al., 2011; Itokawa et al., 2010) could be a cryptic variable in studies of resistance association. To investigate this, a subset of Ghanaian samples (N = 24) was analysed for the presence of the duplicated CYP9M10 (CYP9M10v2) using the Gen2F PCR of Itokawa et al. (2010). All positive samples exhibited the standard CYP9M10. Two samples were positive for CYP9M10v2 indicating the duplicated copy is segregating in Ghana at low frequency.

Two SNPs in codon 1014 of the VGSC (TTA-TTT and TTA-TTC both L-F) were detected in these populations through pyrosequencing. There was no positive association of 1014F with resistance although a negative association was detected for the Madina population (Table 1).

3.6 Presence of CuRE1 in additional African populations

In screening of field-collected adult Cx. quinquefasciatus samples from Malawi, Mozambique, Uganda and Zambia we detected the presence of the inserted CuRE1 allele in samples from Accra, Ghana, Lusaka, Zambia and Nkhotakota and Salima, Malawi, Quelimane, Mozambique and Tororo, Uganda. Following cloning and sequencing of representative alleles no variation was found in the 1107bp fragment containing the CuRE1 allele from any location and all sequences were identical to that from ISOP450 (Supplementary Fig. 6). Thus, whilst the CuRE1 element does not itself contain regulatory motifs, it can act as an easily scorable marker for the as yet unidentified regulatory regions in the ISOP450 upstream region. Non-CuRE1 alleles exhibited variation, particularly in the vicinity of two poly-A tracts (Supplementary Fig. 6). Samples positive for the CuRE1 insertion were also screened by PCR to determine the presence of CYP9M10 or CYP9M10v2 (duplicated CYP9M10) (Itokawa et al., 2010). Whilst the majority of samples were positive for CYP9M10, some were
positive for both CYP9M10 and CYP9M10v2 thus the duplication of CYP9M10 is polymorphic in African populations.
4. Discussion

Using an E.coli system to express CYP9M10 and CYP4H34 from both the ISOP450 strain and the susceptible CqSF strain followed by in vitro metabolism, we have demonstrated the ability of both enzymes to metabolise permethrin. The evidence for in vitro metabolism and up-regulation presented here (see also Komagata et al., 2010), implicate CYP9M10 as a causal gene underlying resistance in lab strains. CYP4H34 does not metabolise deltamethrin and CYP9M10 displays only very weak activity, correlating with the known phenotypic resistance profile of the ISOP450 strain (Hardstone et al., 2007). Differential metabolism of class I and class II pyrethroids has been seen for other CYP450s e.g. CYP6AA3 from An. minimus (Duangkaew et al., 2011). Improvements in our understanding of the structural mechanisms underpinning such differential activity may be gained from modelling of these P450s and we are currently pursuing this line of enquiry. Cytochrome P450s display high levels of nucleotide variability (Wilding et al., 2009) and allelic variants may underlie resistance. Both CYP9M10 and CYP4H34 exhibit alleles from the resistant and susceptible strains that differ at 10 and 9 amino acids respectively. However, we conclusively support the hypothesis of Hardstone et al. (2010) and Komagata et al. (2010) that resistance in the ISOP450 strain (and hence likely in the parental strain JPAL) is due to over-expression of CYP9M10 and not allelic variants since both ISOP450 and CqSF derived alleles were capable of metabolising permethrin to the same extent. Interestingly, both CYP9M10 and CYP4H34 can metabolise permethrin although they appear to have differing actions, as evidenced by the differences in metabolite profiles. Whilst CYP4H34 is capable of metabolising permethrin, it does appear to have a much lesser role in the resistance phenotype.

Although there is no difference in the metabolic activity of divergent alleles for both genes, unlike CYP9M10, CYP4H34 is not over-expressed in the resistant (ISOP450) strain. Our estimate of >20-fold higher expression of CYP9M10 in ISOP450 relative to CqSF (from qPCR) is similar in magnitude to the estimate of 11-fold over expression (vs. S-LAB) in a North American strain selected for permethrin resistance (Liu et al., 2011) but less than the 264-fold change found for JPAL vs Ogasawara (Komagata et al., 2010). It is however incompatible with the 1800-fold over-expression reported by
Hardstone et al. (2010) for ISOP450 vs S-LAB. Whilst CYP9M10 can be induced by recent exposure (Liu et al., 2011) this cannot explain the levels estimated by Hardstone et al. (2010) which are much higher even than for tandemly duplicated genes e.g. esterase E4 in aphids which exhibit fold changes of >60 (Puinean et al., 2010). Nevertheless, the over-expression encountered in our study is compatible with gene expression underlying the observed phenotype.

In the JPAL strain of Culex and the ISOP450 strain derived from JPAL, a partial miniature inverted transposable element (MITE), CuRE1, is inserted directly upstream of CYP9M10 (Itokawa et al., 2010). The presence of this MITE correlates with survivorship in crosses of JPAL and a susceptible strain (Itokawa et al., 2010), but no direct evidence that this MITE contributes functional motifs has been generated previously. Using a luminescent reporter assay we have now shown that this upstream region from the resistant ISOP450 strain drives approximately 10-fold expression when compared to the region from CqSF which lacks the CuRE1 element. When accounting for the CYP9M10 duplication in the ISOP450 strain, this value is strikingly similar to the 20x fold-change we estimated by qPCR for the over-expression of CYP9M10. Whilst there are numerous SNPs/indels differentiating the upstream regions of ISOP450 and CqSF, the most notable difference is the presence of the 672bp region derived from the MITE suggesting that, as for the Accord element of D. melanogaster (Chung et al., 2007), this TE may contribute regulatory sequences. However, when the MITE is excised through fusion PCR, whilst retaining the SNPs and indels which differentiate the upstream regions of ISOP450 and CqSF, the 10-fold reporter expression is maintained, indicating that CuRE1 does not provide regulatory elements. Whilst we cannot discount the possibility that CuRE1 provides regulatory motifs active in other tissues (Sua5.1 are a haemocyte cell line and regulatory motifs may be tissue specific) it seems likely that the important regulatory motifs are present within the SNPs and indels which differentiate these sequences.

Genomic regions subjected to strong, recent selection pressure would be expected to show evidence of a selective sweep with reduced (or obliterated) diversity extending outward from the site of
selection to a distance dependent upon levels of recombination, strength of selection and time since emergence of the mutation e.g. (Catania et al., 2004; Karasov et al., 2010; Lynd et al., 2010; Saminadin-Peter et al., 2012). Through cloning and sequencing of the region upstream of CYP9M10 in field populations we have shown that sequence identical to that found in the ISOP450 strain can be found in Ghana, Malawi, Zambia and Uganda. This suggests not only a ubiquitous pan-African distribution of the resistant allele but also a recent selective sweep (Lynd et al., 2010), which mirrors the situation for Drosophila CYP6g1 where a selective sweep is indicated from worldwide sampling (Catania et al., 2004; Schlenke and Begun, 2004). A consequence of the selective sweep is that CuRE1 is in perfect linkage disequilibrium with regulatory motifs present within the remainder of the upstream sequence. Thus, whilst CuRE1 is not causal, it can act as a marker of the causative cis-regulatory elements. Itokawa et al. (2010) developed a simple diagnostic for the CuRE1 insertion which has not been applied previously to wild populations. Through screening for the presence/absence of the TE insertion we have now demonstrated a strong association between presence of CuRE1 and permethrin resistance in field-collected mosquitoes from 4 of 5 separate populations collected from Accra, Ghana. Since populations are separated by up to 80km, these are likely to represent at least partially independent populations. Effect sizes (odds ratios) are high – up to 6.15, comparable to the OR (≈ 6) for the known pyrethroid target site L1014F kdr mutation in a study of insecticide resistance in An. gambiae (Weetman et al., 2010). Whilst presence of CuRE1 is associated with resistance, there is no association between resistance and genotype at codon 1014 of the voltage-gated sodium channel, and, indeed a negative association for the Madina population. It is possible that the phenotypic effect of these kdr mutations is relevant only in adult populations which were not the target of this study.

These data indicate that the assay of Itokawa et al. (2010) has relevance in natural populations. This therefore represents a marker for metabolic based resistance to insecticide that has proven applicability to wild collected populations. However, whilst this assay detects the presence of the
CuRE1 TE, it is not the TE that is causal. Hence, further work is necessary to dissect the cis-element in order to identify the causal regulatory motifs.

5. Conclusion

In this study, through heterologous expression of a cytochrome P450 from Culex quinquefasciatus, we implicate CYP9M10 as the enzyme responsible for metabolic resistance in the ISOP450 strain. Results of reporter assays demonstrate that whilst a transposable element is inserted upstream of CYP9M10 in resistant mosquitoes, this does not provide regulatory motifs underlying expression suggesting that a re-evaluation of the assumption that TEs inserted upstream of CYP450s contribute functional motifs may be necessary. There is evidence of strong selection acting on the region upstream of CYP9M10 such that only a single resistant haplotype is found throughout Sub-Saharan Africa. Using a simple diagnostic to detect this haplotype we are able to demonstrate a strong association with resistance.

Acknowledgements

We would like to thanks Ms Folasade Ajibade for assistance with operation of the HPLC. The assistance of John Essandoh and Alessi Kwawukume with Ghanaian field work is gratefully acknowledged. We are very grateful to Gareth Lycett, Hilary Ranson and Chris Jones for critical comments on the manuscript. This work was supported by the Innovative Vector Control Consortium.

Appendix. Supporting material

Supplementary material related to this article can be found online

References


Chen, S., Li, X., 2007. Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome P450 genes. BMC Evol Biol 7, 46.


Djouaka, R.F., Bakare, A.A., Coulibaly, O.N., Akogbeto, M.C., Ranson, H., Hemingway, J., Strode, C., 2008. Expression of the cytochrome P450s, CYP6P3 and CYP6M2 are significantly elevated in multiple pyrethroid resistant populations of Anopheles gambiae s.s. from Southern Benin and Nigeria. BMC Genomics 9, 538.


Figure Legends:

Figure 1. A) Metabolism of permethrin by CYP9M10 from ISOP450 with (P+) and without (P-) NADPH. *Cis*-permethrin elutes at ≈18min and *trans*-permethrin at ≈15min. B) Time series of *in vitro* metabolism of permethrin by ISOP450 CYP9M10 membranes. Peak area of peaks identified in 2A (7.8, 9.2 and 10.2minutes) are shown from +NADPH reactions. All peaks in –NADPH reactions did not increase with time but only the 7.8min peak is indicated for clarity.

Figure 2. Reporter gene activity (normalised to renilla fluorescence) in large fragments (amplified with 9M10UP5 & 9M10UP8) and small fragments (amplified with 9M10UP6 & 9M10UP8) from ISOP450 (+/- CuRE1) and CqSF, pGL3 and shuttle constructs (both controls). Error bars show +/- S.E. mean.
<table>
<thead>
<tr>
<th>Population</th>
<th>LC₅₀ in ppm (95% C.I.)</th>
<th>RR³</th>
<th>Phenotype</th>
<th>Phenotype definition</th>
<th>CuRE1 alleles⁴</th>
<th>kdr alleles⁵</th>
<th>P</th>
<th>A</th>
<th>T or C</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashaiman</td>
<td>0.0651 (0.041 - 0.105)</td>
<td>38</td>
<td>R</td>
<td>Survived exposure ≥ 0.15ppm</td>
<td>36 12 &lt;0.0001</td>
<td>20 10 0.355</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S</td>
<td>Killed by exposure ≤ 0.015ppm</td>
<td>17 31</td>
<td>11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koforidua</td>
<td>0.3666 (0.215 - 0.625)</td>
<td>215</td>
<td>R</td>
<td>Survived exposure ≥ 1.5ppm</td>
<td>45 9 &lt;0.0001</td>
<td>12 14 0.969</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S</td>
<td>Killed by exposure ≤ 0.15ppm</td>
<td>26 32</td>
<td>10 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kpone</td>
<td>0.9564 (0.516 - 1.772)</td>
<td>563</td>
<td>R</td>
<td>Survived exposure ≥ 3ppm</td>
<td>46 8 0.484 24</td>
<td>24 0.985</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S</td>
<td>Killed by exposure ≤ 0.015ppm</td>
<td>40 10</td>
<td>19 21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labadi</td>
<td>2.4537 (1.788 - 3.367)</td>
<td>1443</td>
<td>R</td>
<td>Survived exposure ≥ 3ppm</td>
<td>48 12 0.0183 26</td>
<td>22 0.682</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S</td>
<td>Killed by exposure ≤ 1.5ppm</td>
<td>22 16</td>
<td>17 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Madina</td>
<td>0.2272 (0.184 - 0.281)</td>
<td>134</td>
<td>R</td>
<td>Survived exposure ≥ 3ppm</td>
<td>57 7 0.00038 15</td>
<td>41 1.74x10⁻⁴</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S</td>
<td>Killed by exposure ≤ 0.15ppm</td>
<td>33 21</td>
<td>1 61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Association of CuRE1 or kdr allele count with population specific permethrin resistance phenotype. ³ Resistance ratios relative to the S-Lab susceptible strain based upon S-LAB LC₅₀ of 0.0017ppm (Hardstone et al., 2009). R = resistant. S = susceptible. ⁴ allelic counts of CuRE1 positive (+CuRE1) and CuRE1 negative samples (-CuRE1). ⁵ allelic counts for nucleotide at position 3 of codon 1014 of the voltage gated sodium channel (numbering according to the housefly para sequence, GenBank accession no. X96668). TTA = leucine TTC/TTT = phenylalanine.
Figure 2

Reporter gene activity (normalised to Renilla)