Published in final edited form as:
Mol Genet Genomics. 2015 February ; 290(1): 201–215. doi:10.1007/s00438-014-0910-9.

Parallel evolution or purifying selection, not introgression, explain similarity in the pyrethroid detoxification linked GSTE4 of *Anopheles gambiae* and *An. arabiensis*

C.S. Wilding[1,*,†,‡], **D. Weetman**[1,‡], **E.J. Rippon**[1], **K. Steen**[1], **H.D. Mawejje**[2], **I. Barsukov**[3], and **M.J. Donnelly**[1,4]

[1]Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, UK [2]Infectious Diseases Research Collaboration, Makerere University, Kampala, Uganda [3]Department of Structural and Chemical Biology, Institute of Integrative Biology, Biosciences Building, University of Liverpool, Crown Street, Liverpool, UK [4]Malaria Programme, Wellcome Trust Sanger Institute, Hinxton, Cambridge

Abstract

Insecticide resistance is a major impediment to the control of vectors and pests of public health importance and is a strongly selected trait capable of rapid spread, sometimes even between closely-related species. Elucidating the mechanisms generating insecticide resistance in mosquito vectors of disease, and understanding the spread of resistance within and between populations and species are vital for the development of robust resistance management strategies. Here we studied the mechanisms of resistance in two sympatric members of the *Anopheles gambiae* species complex – the major vector of malaria in sub-Saharan Africa – in order to understand how resistance has developed and spread in eastern Uganda, a region with some of the highest levels of malaria.

In eastern Uganda, where the mosquitoes *Anopheles arabiensis* and *An. gambiae* can be found sympatrically, low levels of hybrids (0.4%) occur, offering a route for introgression of adaptively important variants between species. In independent microarray studies of insecticide resistance, *Gste4*, an insect-specific glutathione S-transferase, was among the most significantly up-regulated genes in both species. To test the hypothesis of interspecific introgression, we sequenced 2.3kbp encompassing *Gste4*. Whilst this detailed sequencing ruled out introgression, we detected strong positive selection acting on *Gste4*. However, these sequences, followed by haplotype-specific qPCR, showed that the apparent up-regulation in *An. arabiensis* is a result of allelic variation across the microarray probe binding sites which artefactually elevates the gene expression signal. Thus, face-value acceptance of microarray data can be misleading and it is advisable to conduct a more detailed investigation of the causes and nature of such signal.

[*]Corresponding author: Craig S Wilding, School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, L3 3AF, UK, Tel +44 (0)151 2312500, c.s.wilding@ljmu.ac.uk.
[†]Present address: School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, L3 3AF.
[‡]Contributed equally to this work

The identification of positive selection acting on this locus led us to functionally express and characterise allelic variants of GSTE4. Although the *in vitro* data do not support a direct role for GSTE4 in metabolism, they do support a role for this enzyme in insecticide sequestration. Thus, the demonstration of a role for an up-regulated gene in metabolic resistance to insecticides should not be limited to simply whether it can metabolise insecticide; such a strict criterion would argue against the involvement of GSTE4 despite the weight of evidence to the contrary.

Keywords

Insecticide resistance; *Anopheles gambiae*; introgression; microarray; gene expression; qPCR;

INTRODUCTION

Resistance to the insecticides employed in public health is a major challenge to the control of insect-borne disease including malaria. Insects have evolved a diverse and impressive array of mechanisms to counteract insecticide-based control measures (Hemingway and Ranson 2000). For mosquito vectors of malaria, current controls rely mainly on pyrethroid-treated bednets or spraying of insecticide onto surfaces where mosquitoes rest postprandially. Resistance-associated mutations in the voltage-gated sodium channel, the target of pyrethroids, are well known and have evolved repeatedly in *Anopheles gambiae sensu stricto* (Pinto et al. 2007; Donnelly et al. 2009). However, resistance can also arise due to elevated expression of, or allelic variants in, metabolic genes, which can act with target-site mutations to increase resistance (Mitchell et al. 2014). The identification of the mechanisms underpinning resistance is a vital first step for the development of assays which can be used to understand and predict how resistance spreads within and between populations, and sometimes species.

Detoxification of xenobiotics such as insecticides requires either metabolism (sometimes through intermediary compounds, which require processing and may be more toxic than the original xenobiotic) or transformation through conjugation for subsequent sequestration and elimination. In addition to the metabolic processes required to remove insecticide from within the insect, exposure to toxic compounds can also trigger discrete, non-specific physiological reactions *e.g.* pyrethroid exposure induces oxidative stress and lipid peroxidation (Vontas et al. 2001). Thus, the ability of a mosquito to survive insecticide exposure may require multiple metabolic pathways, potentially mediated by a wide range of enzymes. Identifying those genes underpinning such resistance can aid not only in understanding potential cross-resistance to alternative insecticides but potentially lead to diagnostic assays to aid resistance-monitoring. Whole genome microarrays have been used extensively to study insecticide resistance phenotypes in *An. gambiae s.s.* and *An. coluzzii* and in such studies it is typical to detect up-regulation of transcripts representing a wide-range of pathways (e.g. Mitchell et al. 2012; Fossog Tene et al. 2013; Kwiatkowska et al. 2013). This suggests that metabolism is a complex, multigenic process, and is consistent with the sigmoidal distribution of dose-responses often seen in field populations (e.g. Müller et al. 2008; Mawejje et al. 2013) which imply a broad distribution of resistant phenotypes. Though large numbers of genes often appear differentially regulated, microarray datasets

can be littered with false positive hits (e.g. see Aubert et al. 2004; Pawitan et al. 2005). However, confidence in identification of differentially-regulated genes increases if a gene is identified in independent studies of the same phenotype. Repeated identification of particular cytochrome P450s, including *Cyp6p3* and *Cyp6m2* in microarray studies of resistant *An. gambiae* (Müller et al. 2007; Djouaka et al. 2008; Müller et al. 2008; Mitchell et al. 2012), and of *Cyp6p4* and *Cyp6p9* in *An. funestus* (Wondji et al. 2009; Riveron et al. 2013) has been important in identifying these genes as worthy of the expense and time-consuming enzymatic/biochemical characterization which has subsequently confirmed the role of these enzymes in resistance (Müller et al. 2008; Stevenson et al. 2011; Mitchell et al. 2012; Riveron et al. 2013)

In Uganda, a country with high levels of malaria transmission (Yeka et al. 2012), resistance to pyrethroid insecticides is present in the three main malaria vectors; *An. gambiae* and *An. arabiensis* (Ramphul et al. 2009; Verhaeghen et al. 2010; Mawejje et al. 2013) and *An. funestus* (Morgan et al. 2010). The relative frequency of *An. arabiensis* has risen in neighbouring Kenya (Lindblade et al. 2006; Bayoh et al. 2010; Mwangangi et al. 2013) and Tanzania (Derua et al. 2012) following insecticidal control measures and there is now some evidence of elevated frequencies in eastern Uganda (Mawejje et al. 2013) suggesting an increasing role in malaria transmission. Resistance to pyrethroids is present, and apparently increasing, in *An. arabiensis* from Jinja, eastern Uganda (Mawejje et al. 2013) but is not mediated by known 'knockdown resistance' target-site mechanisms (*L1014F* and *L1014S*) in the voltage-gated sodium channel, which are extremely rare (*1014S* frequency <0.1% (Mawejje et al. 2013)). In the absence of a known target-site mechanism, metabolic mechanisms are strongly implicated in the resistance phenotype.

Although *An. arabiensis* has an increasing role in malaria transmission, *An. gambiae s.s.* remains the major vector in some locations in Uganda such as Tororo (Weetman et al. unpublished), a region with extremely high rates of malaria infection (Kilama et al. 2014), wherein malaria infections have increased recently despite widespread bednet usage (Jagannathan et al. 2012), and the Northern Ugandan district of Apac, where insecticidal interventions have impacted upon clinical malaria indicators (Kigozi et al. 2012). Here we characterise the resistance mechanisms circulating in *An. arabiensis* from Jinja, and *An. gambiae s.s.* from Tororo and use recombinant protein expression followed by functional validation to examine the role of an up-regulated gene (*Gste4*) in the resistance phenotype. We show that *Gste4* shows a strong signature of selective importance, and that the signature, and gene expression of *Gste4*, is haplotype-specific.

METHODS

Sampling of pyrethroid resistant *An. gambiae*

For gene expression profiling we used a novel family-line approach to classify isofemale families of *An. gambiae* ($N = 80$ families) as 'resistant' and 'susceptible' to the class II pyrethroid insecticide lambda-cyhalothrin based on their relative position on an intra-population continuum of resistance (percentage survival in WHO bioassays - see below). Whilst the methodology is laborious, this approach has three main advantages (1) expression profiles are measured in sympatric individuals, thus no susceptible colonies (subject to

geographical confounding) are used; (2) resistant samples are not compared to unexposed control samples, which inevitably contain a proportion of resistant individuals (Müller et al. 2008)); (3) none of the samples for which profiles are obtained have been exposed to insecticide, so any differential expression can be considered constitutive, rather than induced.

Isofemale lines of *An. gambiae* were established from resting *Anopheles* collected in 2009 in Ngelechom, Abwanget, Angorom, Aburi and Amoni, all villages in Tororo District, Uganda close to the National Livestock Resources Research Institute (NaLiRRi 00°61'64.6"N, 34°14'53.2" E). Individual family-line phenotypes were established by exposing 10–20 (mean = 15) 3–5 day old F1 females to lambda-cyhalothrin following the WHO protocol (WHO 2013) modified to have a 90 minute exposure in order to approximate the population specific LT_{50} (time to kill 50% of the population). Ten unexposed, age matched females from each family were also stored in RNAlater (Sigma Aldrich). Mothers were identified to species using the PCR of Scott (1993) and typed for the *L1014S kdr* mutation using the TaqMan protocol of Bass (2007). RNAlater-preserved samples from the 20 most resistant and 20 most susceptible family lines (see Suppl. Fig. 1) were used for gene expression analysis.

Sampling of pyrethroid resistant *An. arabiensis*

We have previously described the pattern of insecticide resistance in *An. arabiensis* from Jinja (00°25'51" N 033°13'44" E) (Mawejje et al. 2013). Samples were collected as larvae (for details of collection locales see Mawejje et al. 2013) and raised to adulthood prior to bioassaying. Resistance to pyrethoids (permethrin and deltamethrin) in this population is more moderate than Tororo with an LT_{50} to both insecticides of ≈ 50 minutes. For this second microarray experiment, resistant female samples surviving 60 min exposure to permethrin as per the WHO protocol (WHO 2013) and control samples, treated in an identical fashion except exposures were to untreated control papers, were stored in RNAlater. Colony samples were drawn from the Dongola (origin Dongola, Sudan, Ng'habi et al. 2007) and Moz (origin Chokwe, southern Mozambique, Witzig et al. 2013) colonies, both of which are susceptible to pyrethroids.

RNA extraction and microarray analysis

All individuals used were 3–5 day old females. RNA was extracted from pools of 10 mosquitoes using the PicoPure (Arcturus) kit for *An. gambiae* samples or RNAqueous4PCR kit (Ambion) for *An. arabiensis* samples following the manufacturer's recommendations and including a DNase step. Total RNA quantity was checked using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, USA) and integrity measured using an Agilent RNA 6000 Nano assay on an Agilent 2100 Bioanalyser. Labelling (both Cy3 and Cy5) was undertaken on 100ng total RNA using the Agilent Low Input Quick Amp Labelling kit (Agilent Technologies) with labelled RNA purified using the Qiagen RNeasy mini kit and eluted in 30μl water. Quantity and quality of labelled RNA was performed as above. Cy3- and Cy5-labelled RNA (300ng each) were combined and hybridised to a custom *Anopheles gambiae* whole genome microarray (AGAM_15K; full details provided at http://www.ebi.ac.uk/arrayexpress: A-MEXP-2196, see Mitchell et al. 2012). Experimental

designs are shown in Suppl. Fig. 2. Hybridisations were undertaken for 17 hours at 65°C at 10 rpm rotation following the manufacturer's protocol (Agilent Technologies). Scanning of each microarray slide was performed with the Agilent G2565 Microarray Scanner System using the Agilent Feature Extraction Software (Agilent Technologies). Analysis was undertaken using custom R-scripts and the MAANOVA package for R (Wu et al. 2009).

Sequencing of the region around *Gste4*

Primers were designed to amplify *Gste4* and adjacent 5' and 3' regions (see Supplementary Table 1 for these and all subsequent primer sequences). Primers GSTe5_seq and GSTe2_seq amplified a 2245bp section of genomic DNA (chromosome 3R: 28,595,701–28,597,945) inclusive of sections of *Gste2* and *Gste5*, the entirety of *Gste4* and intergenic regions between *Gste2-Gste4* and *Gste4-Gste5* (Figure 1). PCRs were undertaken on DNA taken from resistance phenotyped sympatric *An. arabiensis* and *An. gambiae* from Jinja, and a single sample from each of the Dongola, Moz and Sennar (origin Sennar, Sudan, Du et al. 2005) colonies of *An. arabiensis*. Amplified products were cloned into pJET (Fermentas) and individual colonies picked for sequencing. Only single products from each specimen were sequenced unless intra-individual length variation was noted on agarose gels in which case both alleles were sequenced. Amplification primers and an internal sequencing primer Gste4_seq were used in sequencing reactions (Figure 1). Sequences were manually edited and aligned in CodonCode Aligner (CodonCode Corporation), and Maximum Likelihood phylogenetic trees constructed in MEGA v5.2 (Tamura et al. 2011) using the appropriate model as determined by Model Test (Posada and Crandall 1998) with bootstrapping (500 replicates). The nucleotide sequences of *Gste4* from *An. quadriannulatus* and the outgroup *An. chrysti* (within and without the *An. gambiae* complex, respectively) were obtained from VectorBase (Megy et al. 2012) (supercontig KB667655: 1004768-1005118 (exon 1), 1005183-1005509 (exon 2) and contig APCM01015419: 2842-3190 (exon 1), 3256-3582 (exon 2) respectively) and translated. Haplotype diversity and McDonald-Kreitman tests of selection were conducted in DnaSP (Librado and Rozas 2009) with the neutrality index (NI) calculated from this output where $NI = [(P_N/D_N)/(P_S/D_S)]$ (Li et al. 2008) and $-\log_{10}(NI) > 0$ is indicative of positive selection. Due to zero values in the McDonald-Kreitman test we followed the recommendation of Li *et al*. (2008) by adding a pseudocount of 1 to each cell before calculation of the NI.

qPCR and haplotype-specific qPCR

cDNA was produced from ≈2.5μg RNA samples (see above) using oligo dT_{20} and superscript III (Invitrogen) as per the manufacturer's instructions. qPCR was undertaken on 1/50 dilutions of cDNA using exon-crossing *Gste4* qPCR primers (GSTe4qPCR_F1 and GSTe4qPCR_R1) and haplotype specific qPCR primers designed to amplify group specific haplotypes of *Gste4* (GSTe4_Hap8 and GSTe4_Hap12 for group α; GSTe4_Hap8 and GSTe4_Hap9 for clade β) which differed in the presence of large indels in the 3' UTR (see results). Three normalising genes, ribosomal protein S7 (AGAP010592), ubiquitin (AGAP007927) and elongation factor (AGAP005128) were run on the same sample aliquots. qPCR was undertaken in triplicate in 20μl volumes containing 1× Agilent Brilliant III SYBR qPCR mastermix, 300nM each primer and 1μl cDNA (1/50 dilution) on an Agilent

MX3005 with cycling conditions of 3min at 95°C followed by 40 cycles of 10s at 95°C and 10s at 60°C. Analysis used the ΔΔCt method (Livak and Schmittgen 2001).

Cloning and expression of GSTe4

Primers (GSTe4cDNA_RE_F and GSTe4cDNA_RE_R) were designed to amplify the full length sequence of *Gste4* incorporating a 5' *Nde*I site (CATATG where ATG is the translation initiation codon) and a 3' *Bam*HI site, based on the *Gste4* sequence in VectorBase (www.vectorbase.org_gene identifier AGAP009193; Refseq accession XM_319967).

cDNA was prepared from RNA extracted from pyrethroid resistant *An. arabiensis* using Superscript III (Invitrogen) following the manufacturer's recommendations and full-length *Gste4* amplified using high-fidelity Phusion polymerase (Fermentas). Products of the correct size were cloned into pJET (Fermentas) and sequenced. Inserts from plasmids containing confirmed *Gste4* were excised with *Nde*I and *Bam*HI and ligated into pET15b (Novagen). pET15b contains an IPTG inducible T7 promoter, a 6 x HIS tag and a thrombin cleavage site. GSTE4 expression vector was then transformed into BL21(DE3) (NEB), grown at 37°C in LB until an OD of 0.8 was reached, then expression was induced with 1mM IPTG and cultures incubated at 25°C overnight. Cells were harvested at 10,000 rpm for 10min at 4°C and the supernatant discarded. The pellet was re-suspended in 20ml of low imidazole buffer (25mM imidazole, 20mM Na_2HPO_4, 0.5M NaCl, pH 7.4) and frozen at −80°C. After thawing, lysozyme (0.5mg/ml) and DNAse (0.05mg/ml) were added to the cell suspension and the solution incubated on ice for 10min. Cells were disrupted by French press homogeniser (Stansted Fluid Power Ltd) at 20,000psi and centrifuged at 18,000rpm for 20min to remove cell debris. Supernatant was filtered through a 0.45μm filter and loaded manually on a 5mL His-trap column (GE Healthcare), pre-equilibrated with low imidazole buffer. The column was washed with 25ml of low imidazole buffer, followed by 25ml of medium imidazole buffer (50mM imidazole, 20mM Na_2HPO_4, 0.5M NaCl, pH 7.4). The protein was then eluted manually with high imidazole buffer (0.5M imidazole, 20mM Na_2HPO_4, 0.5M NaCl, pH7.4) and concentrated with a Vivaspin 20 concentrator (Sartorius) then exchanged into 20mM Tris, 150mM NaCl pH 7.4 buffer using a PD-10 column (GE Healthcare).

Protein concentration was determined using a NanoDrop (NanoDrop Technologies) and by the Bradford assay (Bradford 1976). Activity of purified protein was checked by a colorimetric activity assay measuring conjugation of reduced glutathione (GSH) to the model substrate 1-chloro-2, 4-dinitrobenzene (CDNB) at 340nm (ε=9.6mM^{-1}.cm^{-1}) (Habig et al. 1974) at a constant 22.5°C in a Cary 300 Bio UV-Vis spectrophotometer over 1min. Reactions contained 0.1M sodium phosphate, 500ng enzyme, 1mM CDNB, 5mM GSH and 3.3% methanol in 1ml total volume.

Determination of optimal pH of GSTe4 variants

The optimal pH for each variant was determined using the CDNB activity assay over the pH range 5.8–8.6 (5.8, 6.2, 6.6, 7.0, 7.4, 7.8, 8.1, 8.3, 8.6). Reactions were undertaken as above,

in triplicate. Kinetic constants (V_{max} and K_m) for both CDNB and GSH were also determined for both variants at pH6.5 and the optimal pHs as determined above.

Determination of temperature optima

Aliquots of both variants of GSTE4 were incubated for 30min over a range of temperatures (30°C–65°C in 5°C increments). Following incubation, CDNB activity was measured as above.

Interaction of recombinant GSTE4 with permethrin and deltamethrin

Inhibition by permethrin and deltamethrin was determined by change in CDNB activity following addition of 0μM, 25μM, 50μM, 75μM, 100μM deltamethrin or permethrin with a saturating concentration of GSH (5mM) and CDNB (1mM) in triplicate reactions in 0.1M sodium phosphate buffer with 500ng enzyme. Inhibition was measured at pH6.5, pH7 and pH7.8 (the optimal pHs determined above plus neutrality).

In vitro permethrin and deltamethrin metabolism assays

Metabolism was undertaken in 0.1M sodium phosphate (pH6.5, pH7, pH7.8) in 0.5ml volumes containing 5mM GSH, 10μM insecticide (DDT, permethrin or deltamethrin) and 50μg recombinant enzyme (with 80°C 30min heat inactivated GSTE4 enzyme in negative control reactions). Reactions were incubated at 25°C for 2h with shaking. Following incubation, bifenthrin (as spike-in extraction control) was added to 10μM then reactions extracted twice with 1 volume *tert*-butyl methyl ether. Extractions were pooled and dried under a constant stream of N_2 then resuspended in 150μl methanol prior to analysis by reverse-phase HPLC (Chromeleon, Dionex) with a monitoring absorbance of 232nm. Reactions (100μl) were loaded into an isocratic mobile phase (90% methanol: 10% water) with a 1ml/min flow rate through a 250mm C18 column (Acclaim 120, Dionex) at 23°C.

Analysis of peroxidase function

Determination of Se-independent peroxidase function followed Vontas *et al*. (2001). In brief reactions contained 1mM EDTA, 200μM NADPH, 1mM GSH, 0.3U glutathione reductase, 2μg enzyme (removed from control reactions) and either 1.5mM cumene hydroperoxide or 1.5mM *t*-butyl hydroperoxide in 31.5mM potassium phosphate pH7. Reactions were incubated at 25°C for 5min prior to addition of peroxide reagent then absorbance was measured for 4min at 25°C and 340nm in a Versamax plate reader (Molecular Devices, Sunnyvale, CA, USA).

RESULTS

Resistance to lambda-cyhalothrin in *An. gambiae* from Tororo

In the *An. gambiae s.s*. population from Tororo we found wide variation in resistance across families (0–100% mortality following 90 min exposure to the pyrethroid lambda cyhalothrin in individual families – see Suppl. Fig. 1) yet the *1014S kdr* mutation approaches fixation (99.5% in Nagongera, Tororo in October 2012 (unpublished data) and in Jinja, 120km distant from Tororo, *1014S* is at 95% frequency in *An. gambiae* (Mawejje et al. 2013)).

Thus, whilst this target-site mechanism may contribute to population-level resistance, it cannot explain the variation in survival following a 90 min exposure to lambda-cyhalothrin.

Microarray analysis - *An. gambiae*

In comparisons of the 20 most highly resistant and 19 most susceptible families (see Supplementary Figure 1 for details of family resistance levels – note that a single susceptible family, incorrectly identified as *An. gambiae* was excluded from analyses) 57 probes representing 50 genes were significantly differentially regulated with $q < 0.05$ (Supplementary table S2). The most statistically-significant probes (Fig. 2) targeted two genes within a cluster of very closely-related, unannotated genes on chromosome 2L (AGAP007187, AGAP007188). Of the significant probes, the most strongly up-regulated in the resistant families were *Gste4* (mean Fold Change (FC) = 2.8; mean *q value* = 0.006 Benjamini-Hochberg FDR adjusted) and a single probe for chymotrypsin 1 (FC = 4.7; mean *q value* = 4×10^{-5}). Only one other known detoxification gene (*Cyp9j4*) was represented among the significantly differentially-expressed probes. All microarray data have been submitted to ArrayExpress (http://www.ebi.ac.uk/miamexpress/) with accession number E-MTAB-1874.

Microarray analysis - *An. arabiensis*

In comparisons of Jinja permethrin-resistant *An. arabiensis* versus sympatric controls and two colonies (Dongola and Moz), 4,094/15,164 probes were significant when an ANOVA F-test approach was applied and a conservative significance threshold applied (FDR-corrected significance level set at log10 (q value) > 4 ($q < 0.0001$)) (see Supplementary table S3 for results). When these 4,094 significantly up-regulated probes were ranked by fold-change (FC), three separate probes targeting *Gste4* were within the top 25 significant probes and were the highest FCs of known detoxification family members (average FC for *Gste4* = 16.6). In pairwise comparisons between Jinja resistant and sympatric controls we applied a standard, multiple test-corrected threshold ($q < 0.05$) more appropriate for within-population comparisons where expected differential expression between groups is likely to be lower. Here 1851 probes were significant (only 22 probes were significant, with the strict FDR-corrected significance level set at $\log_{10}$ (q value) > 4 and these were mainly serine proteases). For comparisons of Jinja resistant to either Dongola or Moz susceptible colony samples 1641 and 673 probes respectively were significantly differentially regulated at the strict $\log_{10}$ (q value) > 4 level. All microarray data have been submitted to ArrayExpress with accession number E-MTAB-1873.

Haplotype analysis and SNP genotyping

Sequencing of 2319bp around *Gste4* from both *An. arabiensis* ($N = 10$ from Jinja plus one sequence from each of the Dongola, Moz and Sennar colonies) and *An. gambiae* ($N = 10$) revealed marked variability, with higher variability in *An. arabiensis* from Jinja (haplotype diversity = 0.982, number of segregating sites = 98 (of 2319), $\pi = 0.01719$) than *An. gambiae* (haplotype diversity = 0.682, number of segregating sites = 83, $\pi = 0.00974$). Sequences have been submitted to Genbank with accession numbers KF733184-KF733209. Maximum likelihood phylogenetic reconstruction of these sequences shows two

monophyletic clades composed of either *An. gambiae* or *An. arabiensis* haplotypes (Fig 3A). When *Gste4* coding sequence alone is used as input the species-specific clades are still apparent (though with low bootstrap support; Supp. Fig 3). However, when amino acid-based trees are constructed, two groupings (labelled Group α and Clade β) are evident: these are not species-specific and the majority of sequences fall into group α which is composed of both *An. Arabiensis* and *An. gambiae* sequences (Fig 3B). Thus, these sequences differ in nucleotide sequence in a species-specific manner, indicative that *Gste4* has not introgressed between these species, but are near-identical in amino acid sequence. The amino acid sequence of GSTE4 from *An. quadriannulatus* falls in clade β, suggesting that group α may be more derived, although there is insufficient sampling to be conclusive. We also note that from our sequencing of this region there is no evidence of haplotypes containing the 42 amino acid deletion exhibited by cDNA clones 1 and 7 (see recombinant protein expression section) indicating that these may be the result of PCR errors or PCR recombination and not genuine variants segregating in the population. However, haplotype sequences exhibiting a 20 amino acid deletion were present (samples labelled Jinja An arabiensis 1 & 2) and by using primers GSTe4qPCRF1 and GSTe4qPCRR2 on genomic DNA we confirmed this deletion (see Supplementary Figure 4) suggesting this is a genuine variant segregating in the population. The correct splice donor and acceptor sites are present in these sequences adding weight to the interpretation that this is a genuine coding variant present in this population. However, we have not expressed these variants in our *E. coli* system.

From the haplotype sequences it was apparent that the 3'UTR region displays large differences in presence/absence of large indels. The multiple probes designed by the Agilent eArray microarray design software targeted this region and although multiple probes interrogate this region, they overlap by just 1bp and hence target the same portion of the 3' UTR (Fig. 4). Given the size of the indels it is likely these probes will hybridise with only one of the UTR variants (Sub-clade β' of Fig. 4 and Fig. 3B).

McDonald-Kreitman Tests

Utilising only sequences from sympatric *An. gambiae s.l.* from Jinja, based on the total sequenced coding region (inclusive of partial coding sequences of *Gste2* and *Gste5*) and comparing group α sequences to clade β sequences, $D_S = 0$, $P_S = 20$, $D_N = 4$, $P_N = 7$ yielding $-\log_{10}(NI) = 1.12$ (following addition of pseudocount) and Fisher's exact test $p = 0.0105$. For GSTE4 alone $D_S = 0$, $P_S = 10$, $D_N = 4$, $P_N = 7$ yielding $-\log_{10}(NI) = 0.837$ and $p=0.055$. The positive values of the NI are strongly indicative of the action of positive selection (Li et al. 2008).

qPCR validation of gene expression results

Owing to the cross-species microarray hit for *Gste4*, qPCR focussed on this gene for *An. gambiae* from Tororo and also the two most significant genes (AGAP007187 and AGAP007188). Unfortunately, owing to extremely high sequence similarity between these latter genes and paralogues (98–99%) within the cluster AGAP007187-AGAP007190), it proved impossible to obtain efficient, specific qPCR primers. However, *Gste4* showed significant differences in gene expression between resistant and susceptible *An. gambiae*

families, albeit at a lower fold change than observed in the microarray experiment (t-test: FC = 1.54; t_{34}=2.18, P=0.034).

For *An. arabiensis*, qPCR did not fully validate the microarray results (Table 1). Permethrin resistant *An. arabiensis* showed significantly higher expression of *Gste4* (1.33–1.49 p = 0.003 where Bonferroni corrected $\alpha = 0.017$) than samples from the two colonies. The difference between resistant and control samples was not significant after multiple testing correction ($p = 0.047$ where Bonferroni corrected $p = 0.017$). Due to the likely differential hybridisation of the microarray probes with different *Gste4* haplotypes we further examined *Gste4* expression using haplotype specific qPCR (see below).

Haplotype-specific qPCR

The two groups of GSTE4 haplotypes (α and β) are differentiable by large indels in the 3' UTR. We designed qPCR primers to measure haplotype specific expression of group members through placement of clade specific primers across an indel region that differed between group α and clade β. When gene expression was measured separately for each group there were large differences in fold-change, particularly in comparisons of permethrin resistant versus either Dongola or Moz colony samples with the clade β qPCR identifying FC>6000 (an artefactual consequence of no measurable gene expression in Dongola/Moz) in both comparisons but group α qPCR showing a significant 1.45 fold over expression for permethrin resistant versus Dongola and no significant difference for permethrin resistant versus Moz (Table 1).

Recombinant protein expression

In order to capture representative *Gste4* sequences for heterologous expression we sequenced nine separate *Gste4* full-length clones. From sequences of these nine clones of *Gste4* amplified from cDNA of permethrin resistant *An. arabiensis* five different protein-coding variants were identified (Figure 4) differing at 3–6 amino acids from the reference genome sequence of *An. gambiae*. In addition, two clones (1 and 7) exhibited a 42 amino acid deletion compared to the reference sequence. Whilst this coding-sequence does appear unlikely to be functional, it was isolated from two separate cDNA pools in two separate PCRs suggesting that it has not arisen through PCR error. cDNA sequences have been submitted to Genbank with accession numbers KF733210-KF733214. Three GSTE4 variants (variants 1, 4 and 9 of Figure 5) were taken forward to expression. Variant 1, which had the 42 amino acid deletion, exhibited no activity with the model substrate CDNB and no further work was undertaken on this variant. We note that this variant had a full-length open-reading frame and therefore was not obviously pseudogenic (*c.f.* the pseudogene of *An. stephensi Gste2* in Ayres et al. 2011). In expression of variants 4 (from clade β and henceforth labelled GSTE4Beta) and 9 (from group α and henceforth labelled GSTE4Alpha), chosen as being the most divergent and representative of the two groups of *Gste4* (α and β - see Fig 3b) we isolated 3–6 ml of 10–12mg/ml of both variants. Both variants showed activity with the model substrate CDNB indicating that the recombinant enzyme was functional.

We note that recombinant protein GSTE4Alpha is nearly identical in sequence to the majority of the haplotypes in group α, but differed by two amino acids T222S and N223K that are not evident in any group α sequence (all sequences are 222T and 223N). These are within the 3' primer site; since primers were designed based upon the VectorBase sequence these non-synonymous changes are likely to result from incorporation of primers into the amplicon (hence are primer-induced amino acid changes rather than real variants present in these haplotypes).

Characterization of activity

We characterised activity across a range of pHs – GSTE4Alpha and GSTE4Beta exhibited very different pH activity profiles and optima with GSTE4Beta showing optimal activity at pH7.8 and GSTE4Alpha at pH7 (Fig 6). We studied enzyme kinetics at three different pHs – 6.5 (the pH used for study of GSTe2 (Dowd et al. 2010), 7 and 7.8. Enzyme kinetics showed the differing activity profiles of these two variants with pH (Table 2). GSTE4Alpha displayed a consistently lower Km for CDNB than GSTE4Beta at all three pHs, suggesting it has a higher affinity for this substrate. Affinities for GSH were similar for both variants except at pH7.8 where the affinity of GSTE4Beta was low (high Km) and that of GSTE4Alpha was not measurable since the reaction did not plateau over the range measured.

Whilst both GSTE4 variants showed similar patterns of temperature dependent activity: 100% activity at 35°C and 0% activity at 45°C, at 40°C there was a significant difference in activity with GSTE4Alpha variant more stable than GSTE4Beta (92% activity versus 66% activity – see Fig 7).

Inhibition by and metabolism of insecticides *in vitro*

Activity against CDNB of both variants GSTE4Alpha and GSTE4Beta was strongly inhibited by permethrin and deltamethrin with the lowest inhibition at pH7 (Figure 8). GSTE4Alpha showed significantly higher inhibition than GSTE4Beta for both insecticides and for all pHs indicating that it has a higher affinity for pyrethroids. Although both insecticides inhibit the enzymes there was no evidence of actual metabolism of pyrethroids (results not shown).

DISCUSSION

Resistance to pyrethroid insecticides in *An. gambiae s.l.* in eastern Uganda is extensive and appears to be increasing (Ramphul et al. 2009; Verhaeghen et al. 2010; Mawejje et al. 2013). There is some evidence that the role of *An. arabiensis* in malaria transmission in the region may also be on the increase (Mawejje et al. 2013) as has been seen in neighbouring countries (Lindblade et al. 2006; Bayoh et al. 2010; Derua et al. 2012; Mwangangi et al. 2013). Here, we have undertaken microarray analysis of the pyrethroid resistant phenotype in both *Anopheles gambiae* and *An. arabiensis* from the same geographical region using two very different experimental microarray designs and have detected the same gene – *Gste4* up-regulated in both studies. Repeatability across studies adds weight to the interpretation of likely involvement of this enzyme in the resistance phenotype. We see an obvious disparity

in the number of significantly up-regulated probes in the two microarray designs – 57 for the comparison of 'resistant' versus 'susceptible' *An. gambiae* families compared to >4,000 for the comparison of insecticide resistant *An. arabiensis* with colonised resistant strains. This illustrates the effect of very different designs. The much greater number of probes detected in the latter design may reflect geographic confounding or the effects of inbreeding and colonisation (see (Kristensen et al. 2005)).

The identification of the same up-regulated gene (*Gste4*) in two closely-related species from the same region might have been a result of introgressive hybridization. However, we find clear, well-supported species-specific clustering of *An. gambiae* and *An. arabiensis Gste4* haplotypes based upon >2kbp of DNA sequence spanning *Gste4* indicating that introgresson definitely does not underlie this observation. In fact, the genomic region containing the *Gste4* locus is in a region of the genome where *An. arabiensis* and *An. gambiae* show high levels of divergence (Weetman et al. 2014). In contrast to the results based on genomic DNA sequence, when GSTE4 amino acid sequences are studied the most common protein sequence is shared by both species. This, despite the clear separation of the whole haplotype sequence suggests that these species have converged on the identical protein sequence or that the functional constraints have prevented divergence from ancestral sequence. The McDonald-Kreitman test result strongly supports the action of positive selection on these sequences indicative of either convergence or constraints on evolutionary change. Evolutionary convergence is a strong indication of adaptive evolution (Zhang and Kumar 1997) and is highly suggestive of an important functional role for this enzyme.

Members of the glutathione-S transferase class of enzymes have been demonstrated to have roles in metabolism, detoxification and excretion of xenobiotics, coping with oxidative stress, and in processing odorant signals (Ranson and Hemingway 2005a; Ranson and Hemingway 2005b). Within *Anopheles gambiae s.l.* 28 GSTs are recognised (Ranson and Hemingway 2005a) with one class – the epsilon GSTs – being insect-specific (Ayres et al. 2011). At least one epsilon-class member, GSTE2, has DDTase activity and a demonstrated role in insecticide resistance ((Ranson et al. 1997; Wang et al. 2008; Mitchell et al. 2014). Whilst there is no direct evidence of a role for GSTs in pyrethroid resistance, GSTs have been implicated in the pyrethroid resistance phenotype through detoxification of pyrethroid-induced lipid peroxification products (Vontas et al. 2001) and through potential sequestration of insecticide through binding of pyrethroid molecules to GSTs (Jirajaroenrat et al. 2001; Kostaropoulos et al. 2001).

Characterisation of the role of GSTE4 in pyrethroid resistance requires heterologous expression and *in vitro* assays. Whilst a recombinant GSTE4 variant has been expressed previously (Ortelli et al. 2003) this came from a susceptible colony of *An. gambiae* (with identical amino acid sequence to the reference PEST genome GSTE4 sequence). We have not identified this particular cDNA sequence in our (limited) sequencing of *Gste4* in pyrethroid resistant *An. arabiensis* from Jinja. There is high variability in *Gste4* coding sequences in *An. arabiensis* from this region – from just nine clones sequenced we identified five different amino acid variants (although two of these contained a 42 amino acid deletion causing a loss of function). We have now biochemically characterised two of these variants from *An. arabiensis* which differ by five amino acids. One of these two variants falls within

clade β and the other is from group α for which the amino acid sequence is conserved across *An. gambiae* and *An. arabiensis*. Note that we are aware that the design of primers for cloning of full-length *Gste4* likely resulted in primer-induced changes in two amino acids in the C-terminus of this protein. Whilst we do not know the functional significance of these alterations, and residues in this C-terminal domain may contribute to substrate specificity (Sheehan et al. 2001), since these are primer-induced changes affecting both variants equally, these are likely to have suppressed any variant associated differences, not to have caused them.

Our enzyme kinetic data show differences in reaction kinetics, in pH optima and in inhibition by insecticides between these two variants. Typically, enzyme characterisation studies on *An. gambiae s.l.* study just one variant (usually from the susceptible Kisumu strain e.g. Ortelli et al. 2003). The variants studied here are segregating in field-collected samples and the differences in kinetics may be of functional importance. Indeed there is evidence from the paralogous GSTE2 that different allelic variants can have very different kinetic and metabolic activities (Mitchell et al. 2014). Whilst metabolism studies did not show clear evidence for metabolic activity of either variant with pyrethroids, inhibition of GST variants has been taken as suggestive of binding and potentially sequestration (Jirajaroenrat et al. 2001; Kostaropoulos et al. 2001). Our inhibition assays conducted with co-incubated insecticide suggest pyrethroids may be capable of occupying either the active site or the GSH binding site of *Gste4* and the differential inhibition we have seen indicates that GSTE4 encoded by different haplotypes have differing sequestration abilities. It is interesting that in *An. arabiensis* two variants with different pH optima, reaction kinetics and inhibition by insecticides are found in the population at similar frequencies; suggesting a role for balancing selection maintaining alleles with differing functions or organ specificity.

Whilst both the biochemical data suggest at present that a link to insecticide resistance is unclear, our assays are not comprehensive and GSTE4 may have a role in some other pathway of importance for the insecticide resistance phenotype. GSTs have known roles as catalysers of secondary metabolism products of reactions involving cytochrome P450s (Ranson and Hemingway 2005a) and hence we may not have utilised the appropriate substrate. Further work on this awaits identification and isolation of insecticide metabolites. We did not detect activity with either cumene hydroperoxide or *t*-butyl hydroperoxide indicating that GSTE4 does not have a Se-independent peroxidase function (Vontas et al. 2001) which is in line with Ortelli et al. (2003) who found no activity with cumene hydroperoxide for the Kisumu variant.

The up-regulation of *Gste4* detected by microarray in *An. gambiae* was validated through qPCR. Although *Gste4* was up-regulated in microarray comparisons of *An. arabiensis*, qPCR validation indicated some discrepancies - fold changes in comparisons of resistant samples to the two colonies were much lower with qPCR than microarray, and no significant difference in *Gste4* expression was seen in comparison of resistant samples to sympatric controls through qPCR. The sequencing of this region in field samples demonstrated that the microarray probes are unlikely to adequately hybridise to some *Gste4* haplotypes and this may have potentially lead to erroneous conclusions. Our sequencing of *Gste4* encompassed the full-length of the gene, untranslated regions (UTRs) and flanking intergenic regions.

Sequences of the 3' UTR showed that large indels segregating in the *An. arabiensis* population co-localise with the binding sites for the whole genome array probes targeting this gene. In fact, the microarray probes are likely to only work on members of sub-clade β' and not to hybridize at all to other members of clade β or any member of the α group. To address this, we designed haplotype-specific 3'UTR qPCR primers which differentiate members of group α (the group exhibiting signs of sequence convergence) from clade β. Clade β expression is absent (or at extremely low levels) in the Dongola and Moz colonies, though present in the Jinja samples and this inflates the Log Q-value disproportionately in comparisons of resistant *An. arabiensis* to colony samples. Expression of members of group α, whilst at higher levels in permethrin resistant samples to the Dongola colony, is not significantly up-regulated versus the Moz colony or sympatric controls. Thus, there is a haplotype-specific component to the *Gste4* up-regulation we inadvertently detected through microarray in *An. arabiensis* but little evidence of true gene expression differences when this is accounted for. In fact, when *Gste4* exon-crossing qPCR primers are used (which are not haplotype-specific) there seems to be slightly lower expression of *Gste4* in resistant samples compared to control *An. arabiensis*. Although the up-regulation of *Gste4* was not validated, it did lead us to further study of this gene and the evidence of sequence convergence is not reliant on gene expression data and stands as evidence of an important functional role. This haplotype specific component to the expression argues strongly for robust, replicated microarray experimental design to ensure type 1 errors are minimised. The *An. gambiae* genome is particularly variable (Wilding et al. 2009) and even though the 3' UTR is less variable than other regions of the gene (Li et al. 2010) the impact of length variation in this region on measures of gene expression could be great. If microarray probes are designed to this region rather than placed in exons where length variation is less likely, then the effects of large differences in length/sequence should be considered, especially if comparisons are not with sympatric samples where this is less likely to be an issue. It should be noted that such variation is also likely to impact upon RNASeq experiments since divergent reads will not adequately map to the reference genome. Whilst the 3' UTR variation does cause technical problems for microarray work, and potentially for RNASeq, it may be of biological interest: 3' UTRs sequence have important roles in directing tissue- and cellular compartment-specific expression (Andreassi and Riccio 2009; Barrett et al. 2012) and the very different UTR sequences of *Gste4* indicate that research into tissue specific expression may be fruitful.

We note that although *Gste4* was identified as up-regulated in both microarray studies, other loci are potentially involved in the resistance phenotype. However, there were no other loci identified as up-regulated across both studies. Whilst the most strongly up-regulated probes in the Tororo *An. gambiae* microarray were multiple probes targeting *Gste4* the most significantly over-expressed probes targeted a cluster of closely related genes of unknown function on chromosome 2L. Due to the very high sequence similarity of these genes it was not possible to design locus specific qPCR primers and we were unable to validate these results. We are also not able to ascribe a function to these genes although they bear some resemblance to human TFIIEα transcription initiation factors. Since we could not validate these results nor develop a functional assay in the absence of known function we did not pursue these hits further. For *An. arabiensis* two P450s showed evidence of up-regulation.

Cyp6m2, up-regulated in many microarray comparisons of *An. gambiae* (Djouaka et al. 2008; Stevenson et al. 2011; Mitchell et al. 2012) was not identified as up-regulated in the Jinja microarray using our strict criteria, however in qPCR there is significant up-regulation when permethrin resistant samples are compared to either of the two colony samples. This discrepancy between microarray and qPCR requires further investigation but may also indicate allelic differences in primer/probe binding sequences. The differential regulation of *Cyp6m3* seen in microarray comparisons seems to be completely driven by extremely low level expression in the two colony samples and shows no evidence of differential regulation in sympatric comparisons. We note that for this population of *An. arabiensis*, prior exposure to piperonyl butoxide (PBO) in diagnostic bioassays partially restored the susceptible phenotype (Mawejje et al. 2013). This partial restoration does indeed indicate that cytochrome P450s likely have some additional role in the resistance phenotype and serves to remind of the complexity of mechanisms underpinning insecticide resistance.

Whilst *Gste4* was up-regulated and demonstrated to be the subject of strong selection in two sympatric species capable of hybridising (Weetman et al. 2014) introgression does not explain this shared mechanism. Whilst our data do not support introgression of *Gste4* between these species, the identification of the same gene in two independent microarray studies, and the demonstration of strong selection on this gene is highly suggestive of an important function. The *in vitro* data indicates that GSTE4 is involved in sequestration of pyrethroids and is worthy of further study to elucidate the sequestration mechanism.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The project described was supported by Award Numbers U19AI089674 and R01AI082734 from the National Institute of Allergy and Infectious Diseases (NIAID). HDM was supported by the Uganda Malaria Clinical Operational and Health Services (COHRE) Training Program at Makerere University, Grant #D43-TW00807701A1, from the Fogarty International Center (FIC) at the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAID, FIC or NIH. We wish to thank John Morgan and Loyce Okedi (NaLiRi, Tororo) for assistance with mosquito collections in Tororo. CSW is grateful for advice on heterologous expression and enzyme characterisation from Andrew Dowd and Mark Paine. Samples for the Dongola colony were obtained through the MR4 as part of the BEI Resources Repository, NIAID, NIH: *Anopheles arabiensis* DONGOLA, MRA-856, deposited by M.Q. Benedict.

REFERENCES

Andreassi C, Riccio A. To localize or not to localize: mRNA fate is in 3′ UTR ends. Trends Cell Biol. 2009; 19(9):465–474. [PubMed: 19716303]

Aubert J, Bar-Hen A, Daudin J-J, Robin S. Determination of the differentially expressed genes in microarray experiments using local FDR. BMC Bioinformatics. 2004; 5(1):125. [PubMed: 15350197]

Ayres CF, Muller P, Dyer N, Wilding CS, Rigden DJ, Donnelly MJ. Comparative genomics of the anopheline glutathione S-transferase epsilon cluster. PLoS ONE. 2011; 6(12)

Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell Mol Life Sci. 2012; 69(21):3613–3634. [PubMed: 22538991]

Bass C, Nikou D, Donnelly MJ, Williamson MS, Ranson H, Ball A, Vontas J, Field LM. Detection of knockdown resistance (*kdr*) mutations in *Anopheles gambiae*: a comparison of two new high-throughput assays with existing methods. Malar J. 2007; 6:e111.

Bayoh MN, Mathias D, Odiere M, Mutuku F, Kamau L, Gimnig J, Vulule J, Hawley W, Hamel M, Walker E. *Anopheles gambiae*: historical population decline associated with regional distribution of insecticide-treated bed nets in western Nyanza Province, Kenya. Malar J. 2010; 9(1):62. [PubMed: 20187956]

Bradford MM. Rapid and sensitive method for quantitation of microgram quanitities of protein utilizing principle of protein-dye binding. Anal Biochem. 1976; 72(1–2):248–254. [PubMed: 942051]

Derua Y, Alifrangis M, Hosea K, Meyrowitsch D, Magesa S, Pedersen E, Simonsen P. Change in composition of the *Anopheles gambiae* complex and its possible implications for the transmission of malaria and lymphatic filariasis in north-eastern Tanzania. Malar J. 2012; 11(1):188. [PubMed: 22681999]

Djouaka RF, Bakare AA, Coulibaly ON, Akogbeto MC, Ranson H, Hemingway J, Strode C. Expression of the cytochrome P450s, CYP6P3 and CYP6M2 are significantly elevated in multiple pyrethroid resistant populations of *Anopheles gambiae s.s.* from Southern Benin and Nigeria. BMC Genomics. 2008; 9:e538.

Donnelly MJ, Corbel V, Weetman D, Wilding CS, Williamson MS, Black WC. Does *kdr* genotype predict insecticide-resistance phenotype in mosquitoes? Trends Parasitol. 2009; 25(5):213–219. [PubMed: 19369117]

Dowd AJ, Morou E, Steven A, Ismail HM, Labrou N, Hemingway J, Paine MJI, Vontas J. Development of a colourimetric pH assay for the quantification of pyrethroids based on glutathione-S-transferase. Int J Environ Anal Chem. 2010; 90(12):922–933.

Du W, Awolola TS, Howell P, Koekemoer LL, Brooke BD, Benedict MQ, Coetzee M, Zheng L. Independent mutations in the *Rdl* locus confer dieldrin resistance to *Anopheles gambiae* and *An. arabiensis*. Insect Mol Biol. 2005; 14(2):179–183. [PubMed: 15796751]

Fossog Tene B, Poupardin R, Costantini C, Awono-Ambene P, Wondji CS, Ranson H, Antonio-Nkondjio C. Resistance to DDT in an urban setting: common mechanisms implicated in both M and S forms of *Anopheles gambiae* in the city of Yaoundé Cameroon. PLoS ONE. 2013; 8(4):e61408. [PubMed: 23626680]

Habig WH, Pabst MJ, Jakoby WB. Glutathione-s-transferases - first enzymatic step in mercapturic acid formation. J Biol Chem. 1974; 249(22):7130–7139. [PubMed: 4436300]

Hemingway J, Ranson H. Insecticide resistance in insect vectors of human disease. Annu Rev Entomol. 2000; 45:371–391. [PubMed: 10761582]

Jagannathan P, Muhindo M, Kakuru A, Arinaitwe E, Greenhouse B, Tappero J, Rosenthal P, Kaharuza F, Kamya M, Dorsey G. Increasing incidence of malaria in children despite insecticide-treated bed nets and prompt anti-malarial therapy in Tororo, Uganda. Malar J. 2012; 11(1):435. [PubMed: 23273022]

Jirajaroenrat K, Pongjaroenkit S, Krittanai C, Prapanthadara L-a, Ketterman AJ. Heterologous expression and characterization of alternatively spliced glutathione S-transferases from a single *Anopheles* gene. Insect Biochem Mol Biol. 2001; 31(9):867–875. [PubMed: 11439246]

Kigozi R, Baxi SM, Gasasira A, Sserwanga A, Kakeeto S, Nasr S, Rubahika D, Dissanayake G, Kamya MR, Filler S, Dorsey G. Indoor residual spraying of insecticide and malaria morbidity in a high transmission intensity area of Uganda. PLoS ONE. 2012; 7(8):e42857. [PubMed: 22880123]

Kilama M, Smith DL, Hutchinson R, Kigozi R, Yeka A, Lavoy G, Kamya MR, Staedke SG, Donnelly MJ, Drakeley C, Greenhouse B, Dorsey G, Lindsay SW. Estimating the annual entomological inoculation rate for *Plasmodium falciparum* transmitted by *Anopheles gambiae s.l.* using three sampling methods in three sites in Uganda. Malaria J. 2014; 13(1):111.

Kostaropoulos I, Papadopoulos AI, Metaxakis A, Boukouvala E, Papadopoulou-Mourkidou E. Glutathione S-transferase in the defence against pyrethroids in insects. Insect Biochem Mol Biol. 2001; 31(4–5):313–319. [PubMed: 11222940]

Kristensen TN, Sørensen P, Kruhøffer M, Pedersen KS, Loeschcke V. Genome-wide analysis on inbreeding effects on gene expression in *Drosophila melanogaster*. Genetics. 2005; 171(1):157–167. [PubMed: 15944359]

Kwiatkowska RM, Platt N, Poupardin R, Irving H, Dabire RK, Mitchell S, Jones CM, Diabaté A, Ranson H, Wondji CS. Dissecting the mechanisms responsible for the multiple insecticide resistance phenotype in *Anopheles gambiae* s.s., M form, from Vallée du Kou, Burkina Faso. Gene. 2013; 519(1):98–106. [PubMed: 23380570]

Li J, Ribeiro JMC, Yan G. Allelic gene structure variations in *Anopheles gambiae* mosquitoes. PLoS ONE. 2010; 5(5):e10699. [PubMed: 20502664]

Li YF, Costello JC, Holloway AK, Hahn MW. "Reverse ecology" and the power of population genomics. Evolution. 2008; 62(12):2984–2994. [PubMed: 18752601]

Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25(11):1451–1452. [PubMed: 19346325]

Lindblade KA, Gimnig JE, Kamau L, Hawley WA, Odhiambo F, Olang G, Ter Kuile FO, Vulule JM, Slutsker L. Impact of sustained use of insecticide-treated bednets on malaria vector species distribution and culicine mosquitoes. J Med Entomol. 2006; 43(2):428–432. [PubMed: 16619629]

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods. 2001; 25(4):402–408. [PubMed: 11846609]

Mawejje HD, Wilding CS, Rippon EJ, Hughes A, Weetman D, Donnelly MJ. Insecticide resistance monitoring of field-collected *Anopheles gambiae s.l.* populations from Jinja, eastern Uganda, identifies high levels of pyrethroid resistance. Med Vet Entomol. 2013; 27(3):276–283. [PubMed: 23046446]

Megy K, Emrich SJ, Lawson D, Campbell D, Dialynas E, Hughes DST, Koscielny G, Louis C, MacCallum RM, Redmond SN, Sheehan A, Topalis P, Wilson D, the VectorBase C. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Res. 2012; 40(D1):D729–D734. [PubMed: 22135296]

Mitchell S, Stevenson B, Müller P, Wilding C, Yawson A, Field S, Hemingway J, Paine M, Ranson H, Donnelly M. Identification and validation of a gene causing cross-resistance between insecticide classes in *Anopheles gambiae* from Ghana. Proc Natl Acad Sci U S A. 2012; 109:6147–6152. [PubMed: 22460795]

Mitchell SN, Rigden DJ, Dowd AJ, Lu F, Wilding CS, Weetman D, Dadzie S, Jenkins AM, Regna K, Boko P, Djogbenou L, Muskavitch MAT, Ranson H, Paine MJI, Mayans O, Donnelly MJ. Metabolic and target-site mechanisms combine to confer strong DDT resistance in *Anopheles gambiae* . PLoS ONE. 2014; 9(3):e92662. [PubMed: 24675797]

Morgan JC, Irving H, Okedi LM, Steven A, Wondji CS. Pyrethroid resistance in an *Anopheles funestus* population from Uganda. PLoS ONE. 2010; 5(7):e11872. [PubMed: 20686697]

Müller P, Donnelly MJ, Ranson H. Transcription profiling of a recently colonised pyrethroid resistant *Anopheles gambiae* strain from Ghana. BMC Genomics. 2007; 8:e36.

Müller P, Warr E, Stevenson BJ, Pignatelli PM, Morgan JC, Steven A, Yawson AE, Mitchell SN, Ranson H, Hemingway J, Paine MJI, Donnelly MJ. Field-caught permethrin-resistant *Anopheles gambiae* overexpress CYP6P3, a P450 that metabolises pyrethroids. PLoS Genet. 2008; 4(11):e1000286. [PubMed: 19043575]

Mwangangi J, Mbogo C, Orindi B, Muturi E, Midega J, Nzovu J, Gatakaa H, Githure J, Borgemeister C, Keating J, Beier J. Shifts in malaria vector species composition and transmission dynamics along the Kenyan coast over the past 20 years. Malar J. 2013; 12(1):13. [PubMed: 23297732]

Ng'habi KR, Horton A, Knols BGJ, Lanzaro GC. A new robust diagnostic polymerase chain reaction for determining the mating status of female *Anopheles gambiae* mosquitoes. The American Journal of Tropical Medicine and Hygiene. 2007; 77(3):485–487. [PubMed: 17827364]

Ortelli F, Rossiter LC, Vontas J, Ranson H, Hemingway J. Heterologous expression of four glutathione transferase genes genetically linked to a major insecticide-resistance locus from the malaria vector *Anopheles gambiae* . Biochem J. 2003; 373:957–963. [PubMed: 12718742]

Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A. False discovery rate, sensitivity and sample size for microarray studies. Bioinformatics. 2005; 21(13):3017–3024. [PubMed: 15840707]

Pinto J, Lynd A, Vicente JL, FS, Randle NP, Caccone A, Gentile G, Moreno M, Simard F, Charlwood JD, do Rosário VE, della Torre A, Donnelly MJ. Origins and distribution of knockdown resistance mutations in the afrotropical mosquito vector *Anopheles gambiae* . PLoS ONE. 2007; 11:e1243. [PubMed: 18043750]

Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. 1998; 14(9):817–818. [PubMed: 9918953]

Ramphul U, Boase T, Bass C, Okedi LM, Donnelly MJ, Muller P. Insecticide resistance and its association with target-site mutations in natural populations of *Anopheles gambiae* from eastern Uganda. Trans R Soc Trop Med Hyg. 2009; 103(11):1121–1126. [PubMed: 19303125]

Ranson, H.; Hemingway, J. 5.11 - Glutathione Transferases. In: Lawrence, IG.; Kostas, I.; Sarjeet, SG., editors. Comprehensive Molecular Insect Science. Amsterdam: Elsevier; 2005a. p. 383-402.Editors-in-Chief

Ranson, H.; Hemingway, J. Mosquito glutathione transferases. In: Helmut, S.; Lester, P., editors. Methods Enzymol. Academic Press; 2005b. p. 226-241.

Ranson H, Prapanthadara LA, Hemingway J. Cloning and characterization of two glutathione S-transferases from a DDT-resistant strain of *Anopheles gambiae* . Biochem J. 1997; 324:97–102. [PubMed: 9164846]

Riveron JM, Irving H, Ndula M, Barnes KG, Ibrahim SS, Paine MJ, Wondji CS. Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus* . Proc Natl Acad Sci U S A. 2013; 110(1):252–257. [PubMed: 23248325]

Scott JA, Brogdon WG, Collins FH. Identification of single specimens of the *Anopheles gambiae* complex by the Polymerase Chain Reaction. Am J Trop Med Hyg. 1993; 49(4):520–529. [PubMed: 8214283]

Sheehan D, Meade G, Foley VM, Dowd CA. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. Biochem J. 2001; 360(1):1–16. [PubMed: 11695986]

Stevenson BJ, Bibby J, Pignatelli P, Muangnoicharoen S, O'Neill PM, Lian LY, Muller P, Nikou D, Steven A, Hemingway J, Sutcliffe MJ, Paine MJ. Cytochrome P450 6M2 from the malaria vector *Anopheles gambiae* metabolizes pyrethroids: sequential metabolism of deltamethrin revealed. Insect Biochem Mol Biol. 2011; 41(7):492–502. [PubMed: 21324359]

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28(10):2731–2739. [PubMed: 21546353]

Verhaeghen K, Van Bortel W, Roelants P, Okello PE, Talisuna A, Coosemans M. Spatio-temporal patterns in *kdr* frequency in permethrin and DDT resistant *Anopheles gambiae s.s.* from Uganda. Am J Trop Med Hyg. 2010; 82(4):566–573. [PubMed: 20348500]

Vontas JG, Small GJ, Hemingway J. Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in *Nilaparvata lugens* . Biochem J. 2001; 357:65–72. [PubMed: 11415437]

Wang Y, Qiu L, Ranson H, Lumjuan N, Hemingway J, Setzer W, Meehan E, Chen L. Structure of an insect epsilon class glutathione S-transferase from the malaria vector *Anopheles gambiae* provides an explanation for the high DDT-detoxifying activity. J Struct Biol. 2008; 164(2):228–235. [PubMed: 18778777]

Weetman D, Steen K, Rippon EJ, Mawejje HD, Donnelly MJ, Wilding CS. Contemporary gene flow between wild *An. gambiae s.s.* and *An. arabiensis* . Parasites and Vectors. 2014; 7:345. [PubMed: 25060488]

Weetman, D.; Wilding, CS.; Müller, P.; Steen, K.; Rippon, EJ.; Morgan, JC.; Mawejje, HD.; Rigden, D.; Okedi, LM.; Donnelly, MJ. Metabolic gene polymorphisms contribute to class I and II pyrethroid resistance in East African *Anopheles gambiae*. (unpublished)

WHO. Test procedures for insecticide resistance monitoring in malaria vector mosquitoes. Geneva: World Health Organisation; 2013.

Wilding CS, Weetman D, Steen K, Donnelly MJ. High, clustered, nucleotide diversity in the genome of *Anopheles gambiae* revealed by SNP discovery through pooled-template sequencing: implications for high-throughput genotyping protocols. BMC Genomics. 2009; 10:e320.

Witzig C, Parry M, Morgan JC, Irving H, Steven A, Cuamba N, Kerah-Hinzoumbe C, Ranson H, Wondji CS. Genetic mapping identifies a major locus spanning P450 clusters associated with pyrethroid resistance in *kdr*-free *Anopheles arabiensis* from Chad. Heredity. 2013; 110(4):389–397. [PubMed: 23299100]

Wondji CS, Irving H, Morgan J, Lobo NF, Collins FH, Hunt RH, Coetzee M, Hemingway J, Ranson H. Two duplicated P450 genes are associated with pyrethroid resistance in *Anopheles funestus*, a major malaria vector. Genome Res. 2009; 19:452–459. [PubMed: 19196725]

Wu, H.; Yang, H.; Churchill, GA. 2009. http://churchill.jax.org/software/rmaanova/maanova.pdf

Yeka A, Gasasira A, Mpimbaza A, Achan J, Nankabirwa J, Nsobya S, Staedke SG, Donnelly MJ, Wabwire-Mangen F, Talisuna A, Dorsey G, Kamya MR, Rosenthal PJ. Malaria in Uganda: challenges to control on the long road to elimination: I. Epidemiology and current control efforts. Acta Trop. 2012; 121(3):184–195. [PubMed: 21420377]

Zhang J, Kumar S. Detection of convergent and parallel evolution at the amino acid sequence level. Mol Biol Evol. 1997; 14(5):527–536. [PubMed: 9159930]

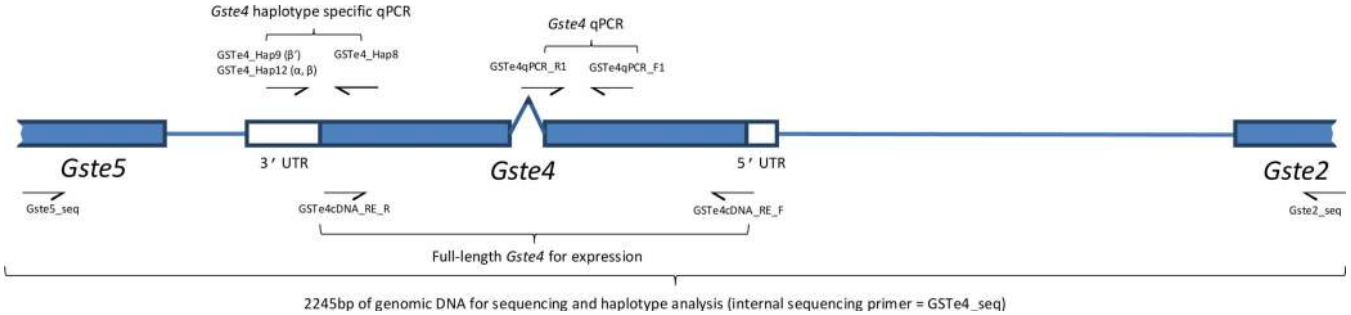

Figure 1.

Genomic context of the *Gste4* gene on chromosome 3R of *Anopheles gambiae*. The locations of primers designed for sequencing and qPCR are indicated on the figure.

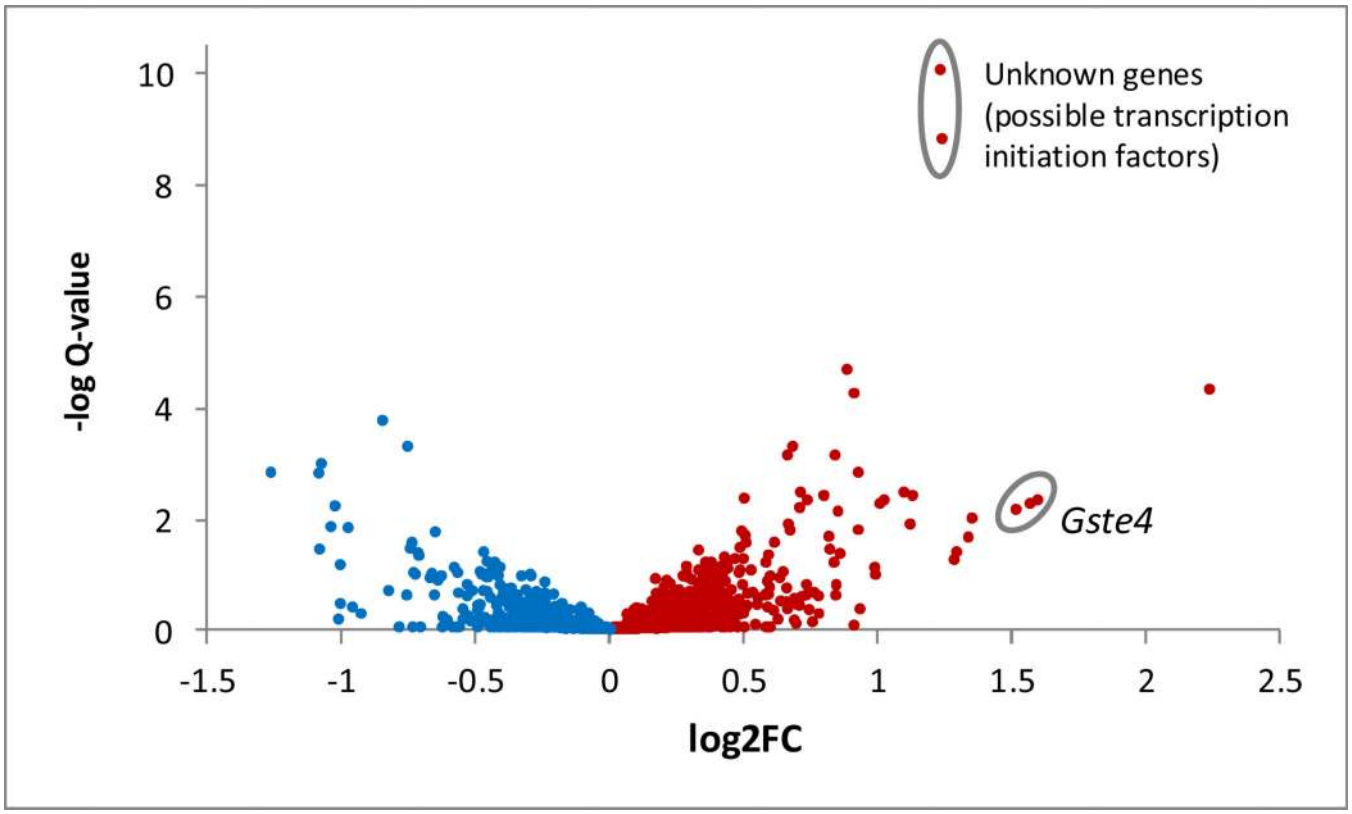

Fig 2.

Volcano plot summarising $\log_2$ fold changes (log2FC) plotted against multiple testing corrected probability (−log Q-value) for 20 resistant vs. 20 susceptible *An. gambiae s.s.* families from Tororo

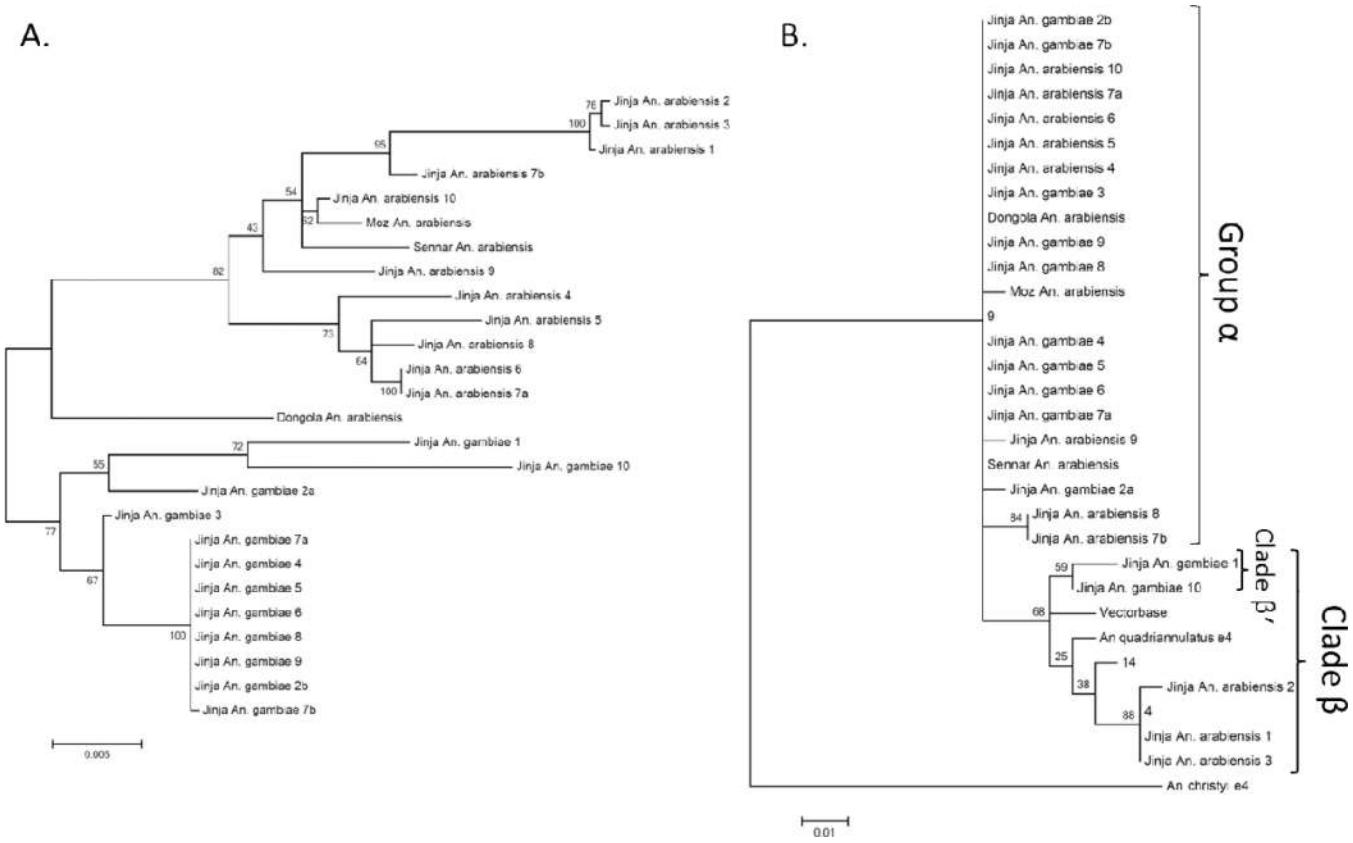

Figure 3.

a) ML phylogeny of 2319bp of sequence spanning *Gste4*_using best fit model (Tamura and Nei with invariant sites (TN93+I)). Values at nodes are bootstrap support values (% of 500 bootstraps). B) ML phylogeny of amino acid sequences of GSTE4 using best fit model Whelan and Goldman with uniform sites. Sequences of cloned cDNAs 4, 9 and 14, the amino acid sequence from the reference PEST genome *An. christyi* GSTE4 and *An. quadriannulatus* GSTE4 are also included. Note that cDNA 9 contained two primer induced amino acid sequence changes. For clarity, the native sequence is included in Fig 3b.

Figure 4.

Alignment of the three Agilent whole genome microarray probes (60bp sequence; DETOX_622_PI422610884, DETOX_623_PI422610884, DETOX_624_PI422610884) designed to interrogate *Gste4* to the 3'UTR of *Gste4* in Ugandan *Anopheles* sequences. PEST = PEST reference sequence from VectorBase [55]. Representative haplotypes of this region are shown for members of the α, β and β' groups of Fig. 2B).

Figure 5.

Amino acid alignment of full length GSTe4 sequences for expression. GSTe4_VB is the sequence from the *Anopheles gambiae* PEST genome sequence (Gene identifier AGAP0091913 on www.vectorbase.com). Residues differing from the VectorBase sequence are highlighted. Variant 9 has been subsequently characterized as GSTE4ALPHA and variant 4 as GSTE4BETA.

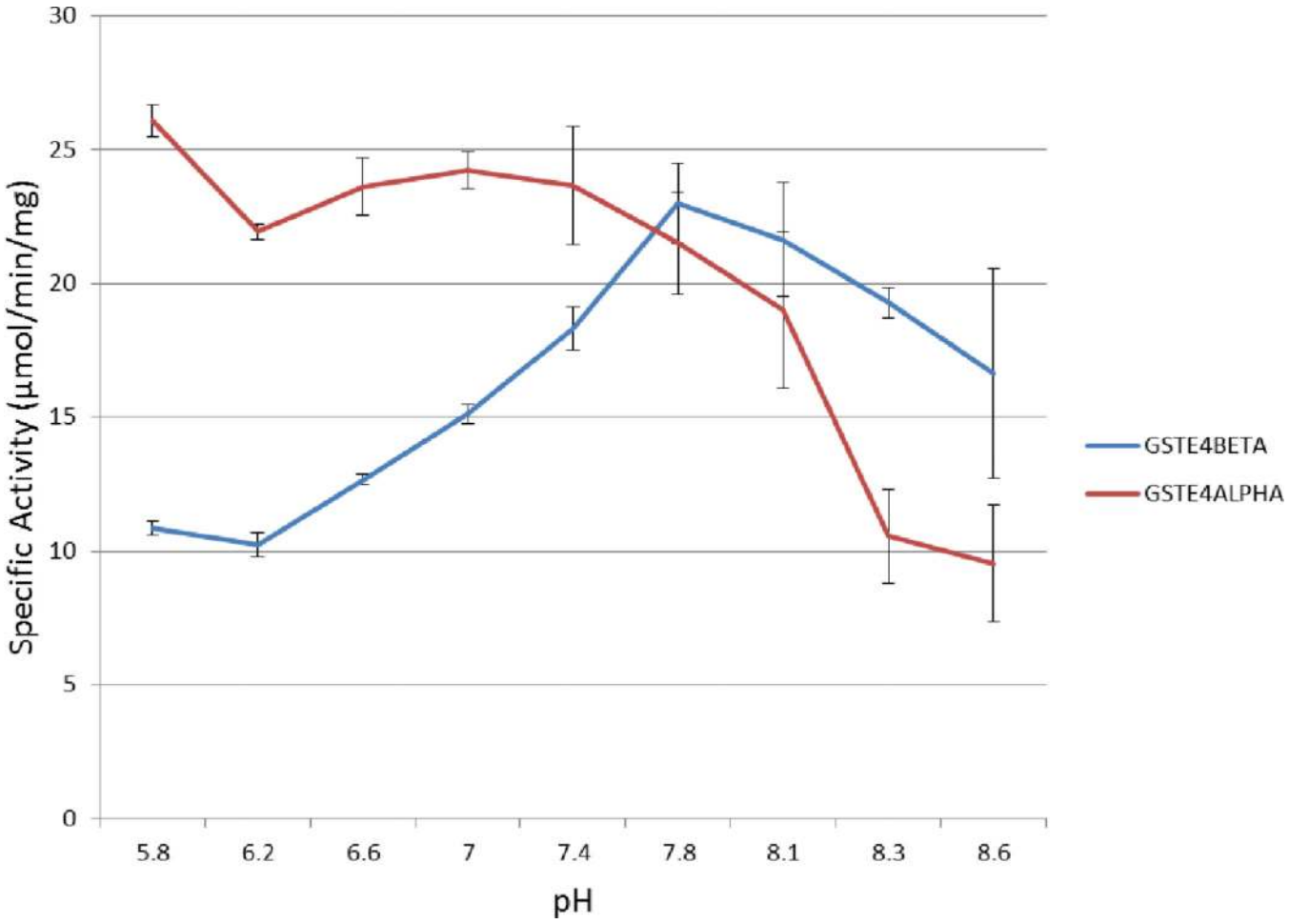

Figure 6.
Determination of pH optima for two variants of GSTE4 (GSTE4Alpha and GSTE4Beta).

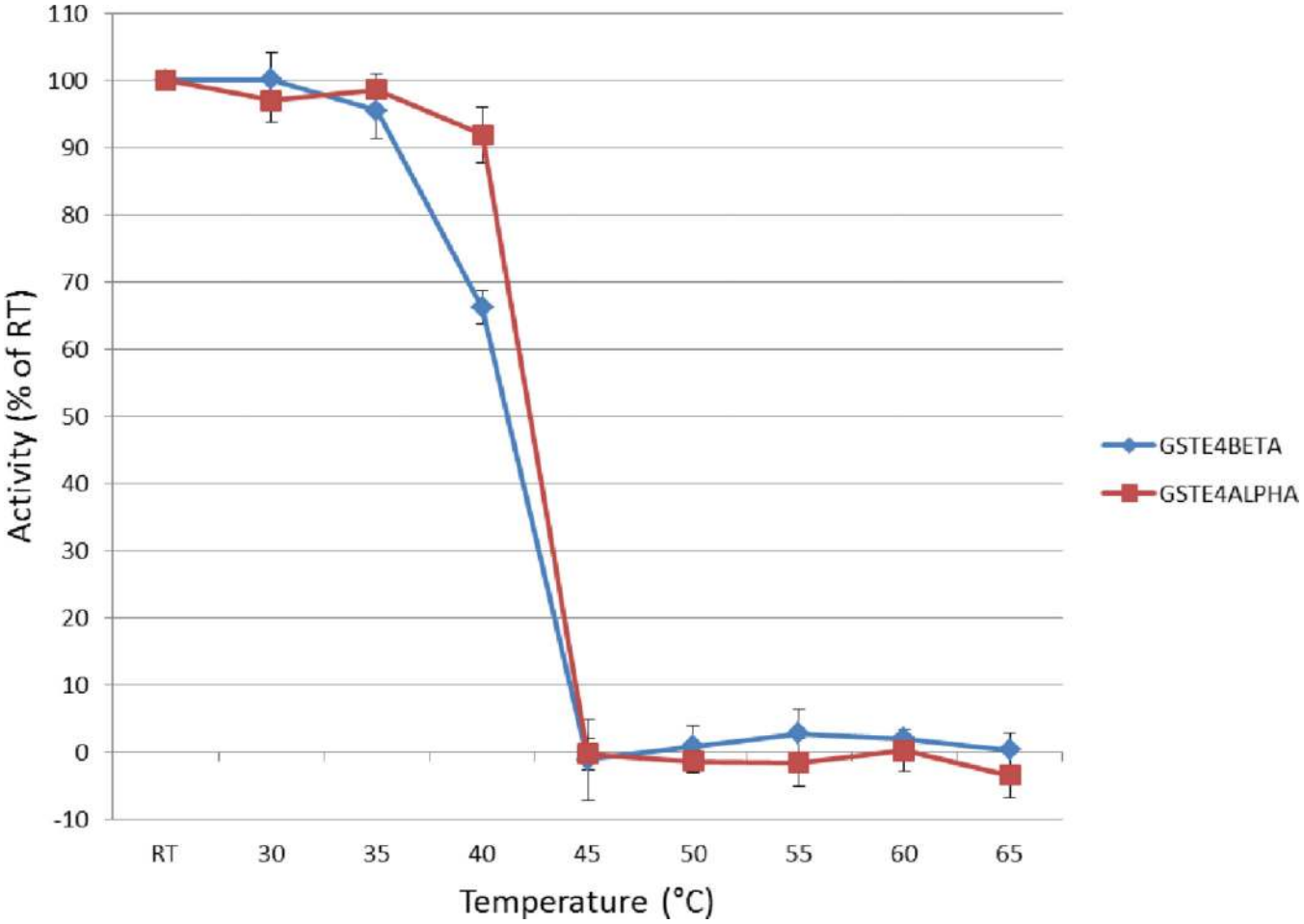

Figure 7.
Temperature stability of two variants of GSTE4 (GSTE4Alpha and GSTE4Beta).

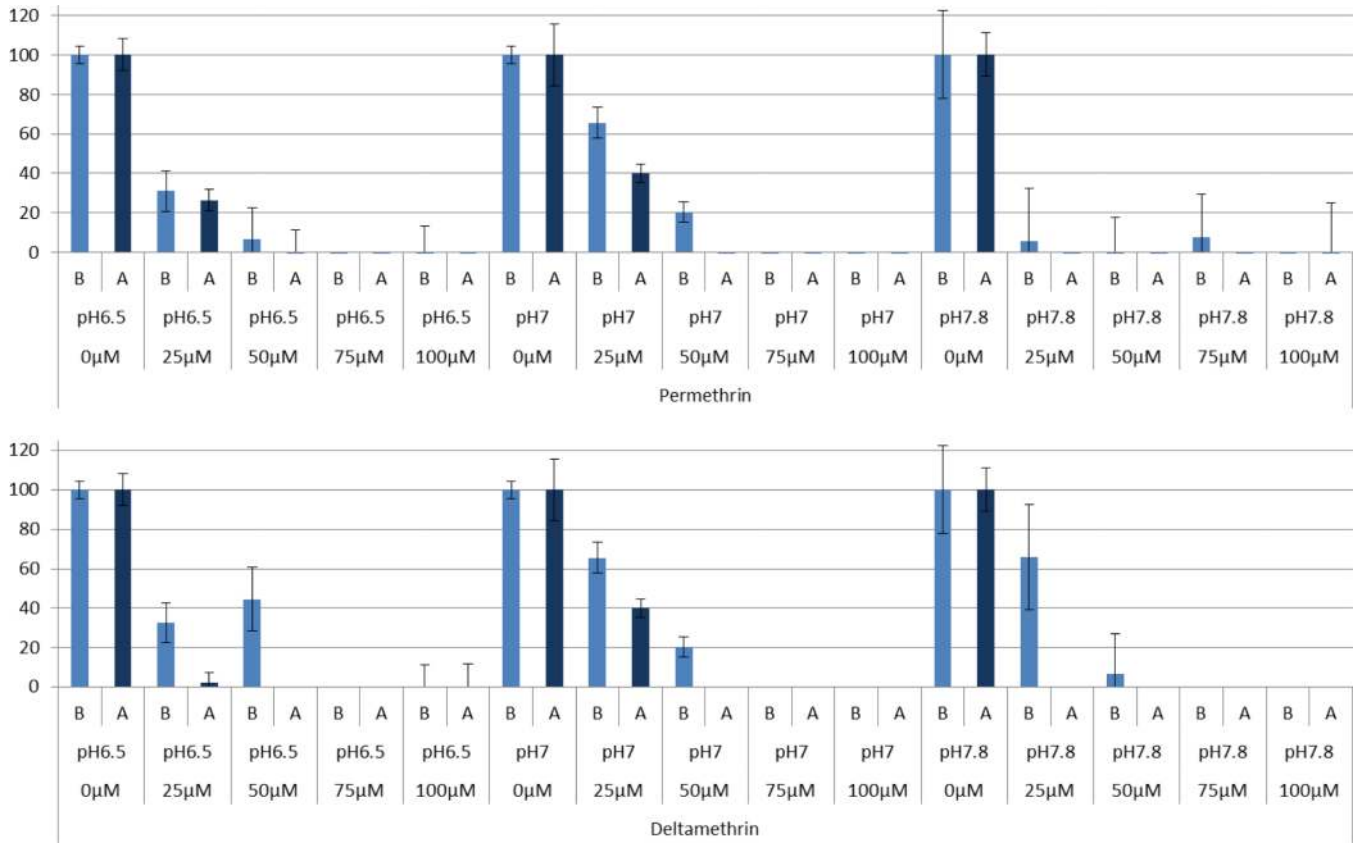

Figure 8.

Inhibition of GSTE4 variants GSTE4Alpha (A - dark blue) and GSTE4Beta (B - light blue) by various concentrations of insecticide (0–100μM). Values are % of activity of the 0μM insecticide point (± 95% C.I.). Note that at higher concentrations of insecticide, activity in the blank samples was > experimental likely due to precipitation of insecticide. The activity (Y-axis) in the absence of insecticide has been set at 100% for clarity.

Table 1

Fold-changes and *P*-values from qPCR validation of microarray hits.

	Gste4										
	Micro array		**exon-crossing *Gste4* primers**			**Clade β**			**Group α**		
	P	**FC**	**FC**	**LCI**	**UCI**	**FC**	**LCI**	**UCI**	**FC**	**LCI**	**UCI**
Control vs Control			1.00	0.78	1.22	1.00	0.70	1.30	1.00	0.37	1.63
Dongola vs Control			**0.49**	**0.41**	**0.58**	**0.00**	**0.00**	**0.00**	**0.58**	**0.33**	**0.82**
Moz vs Control			**0.55**	**0.41**	**0.70**	**0.00**	**0.00**	**0.00**	**0.70**	**0.53**	**0.87**
Perm Resistant vs Control	0.801	1.04	**0.74**	**0.67**	**0.81**	1.07	0.78	1.36	0.84	0.62	1.05
Dongola vs Dongola			1.00	0.83	1.17	1.00	−0.03	2.03	1.00	0.58	1.42
Perm Resistant vs Dongola	0.000	8.51	**1.49**	**1.35**	**1.63**	**6795.69**	**4966.25**	**8625.12**	**1.45**	**1.07**	**1.83**
Moz vs Moz			1.00	0.74	1.26	1.00	0.16	1.84	1.00	0.75	1.25
Perm Resistant vs Moz	0.000	6.13	**1.33**	**1.21**	**1.46**	**7680.12**	**5612.59**	**9747.65**	1.20	0.89	1.51

Table 2

Kinetic constants for the two variants of GSTE4 over 3 pH values: 6.5, 7 (experimentally determined optimum for GSTE4Alpha) and 7.8 (experimentally determined optimum for GSTE4Beta)

		GSTe4 v4			GST e4 v9		
		pH6.5	**pH7**	**pH7.8**	**pH6.5**	**pH7**	**pH7.8**
[CDNB]	CDNB Km (mM)	0.021	0.072	0.055	0.010	0.012	0.002
	CDNB Vmax (μmol/min/mg)	8.442	11.575	11.816	20.277	20.955	11.815
[GSH]	GSH Km (mM)	1.227	1.395	7.435	1.975	3.820	7.66E+07
	GSH Vmax (μmol/min/mg)	12.071	15.597	34.607	29.148	35.207	1.92E+08