



## LJMU Research Online

Wilding, CS, Weetman, D, Rippon, EJ, Steen, K, Mawejje, HD, Barsukov, I and Donnelly, MJ

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

<http://researchonline.ljmu.ac.uk/2636/>

### Article

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

**Wilding, CS, Weetman, D, Rippon, EJ, Steen, K, Mawejje, HD, Barsukov, I and Donnelly, MJ (2015) Parallel evolution or purifying selection, not introgression, explains similarity in the pyrethroid detoxification linked GSTE4 of *Anopheles gambiae* and *An. arabiensis*. MOLECULAR GENETICS**

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](mailto:researchonline@ljmu.ac.uk)

<http://researchonline.ljmu.ac.uk/>



1 Parallel evolution or purifying selection, not introgression, explain similarity in the pyrethroid  
2 detoxification linked GSTE4 of *Anopheles gambiae* and *An. arabiensis*  
3

4  
5 Wilding, C.S.<sup>1\*†‡</sup>, Weetman, D.<sup>1‡</sup>, Rippon, E.J.<sup>1</sup>, Steen, K.<sup>1</sup>, Mawejje, H.D.<sup>2</sup>, Barsukov, I.<sup>3</sup> and Donnelly,  
6 M.J.<sup>1,4</sup>  
7

8  
9  
10  
11 <sup>1</sup>Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool,  
12 UK.  
13

14  
15  
16 <sup>2</sup>Infectious Diseases Research Collaboration, Makerere University, Kampala, Uganda  
17

18  
19  
20 <sup>3</sup>Department of Structural and Chemical Biology, Institute of Integrative Biology, Biosciences  
21 Building, University of Liverpool, Crown Street, Liverpool, UK.  
22

23  
24  
25 <sup>4</sup>Malaria Programme, Wellcome Trust Sanger Institute, Hinxton, Cambridge.  
26

27  
28  
29 † Present address: School of Natural Sciences and Psychology, Liverpool John Moores University,  
30 Liverpool, L3 3AF.  
31

32  
33  
34 ‡ Contributed equally to this work  
35

36  
37  
38 \*Corresponding author:  
39

40  
41 Craig S Wilding, School of Natural Sciences and Psychology, Liverpool John Moores University,  
42 Liverpool, L3 3AF, UK  
43

44  
45  
46 Tel +44 (0)151 2312500  
47

48  
49  
50 Email [c.s.wilding@ljmu.ac.uk](mailto:c.s.wilding@ljmu.ac.uk)  
51

1  
2  
3 ABSTRACT  
4  
5

6 Insecticide resistance is a major impediment to the control of vectors and pests of public health  
7 importance and is a strongly selected trait capable of rapid spread, sometimes even between  
8 closely-related species. Elucidating the mechanisms generating insecticide resistance in mosquito  
9 vectors of disease, and understanding the spread of resistance within and between populations and  
10 species are vital for the development of robust resistance management strategies. Here we studied  
11 the mechanisms of resistance in two sympatric members of the *Anopheles gambiae* species complex  
12 – the major vector of malaria in sub-Saharan Africa – in order to understand how resistance has  
13 developed and spread in eastern Uganda, a region with some of the highest levels of malaria.  
14  
15

16 In eastern Uganda, where the mosquitoes *Anopheles arabiensis* and *An. gambiae* can be found  
17 sympatrically, low levels of hybrids (0.4%) occur, offering a route for introgression of adaptively  
18 important variants between species. In independent microarray studies of insecticide resistance,  
19 *Gste4*, an insect-specific glutathione S-transferase, was among the most significantly up-regulated  
20 genes in both species. To test the hypothesis of interspecific introgression, we sequenced 2.3kbp  
21 encompassing *Gste4*. Whilst this detailed sequencing ruled out introgression, we detected strong  
22 positive selection acting on *Gste4*. However, these sequences, followed by haplotype-specific qPCR,  
23 showed that the apparent up-regulation in *An. arabiensis* is a result of allelic variation across the  
24 microarray probe binding sites which artefactually elevates the gene expression signal. Thus, face-  
25 value acceptance of microarray data can be misleading and it is advisable to conduct a more detailed  
26 investigation of the causes and nature of such signal.  
27  
28

29 The identification of positive selection acting on this locus led us to functionally express and  
30 characterise allelic variants of GSTE4. Although the *in vitro* data do not support a direct role for  
31 GSTE4 in metabolism, they do support a role for this enzyme in insecticide sequestration. Thus, the  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 demonstration of a role for an up-regulated gene in metabolic resistance to insecticides should not  
2 be limited to simply whether it can metabolise insecticide; such a strict criterion would argue against  
3  
4 the involvement of GSTE4 despite the weight of evidence to the contrary.  
5  
6

7  
8 **Keywords:**  
9

10 Insecticide resistance; *Anopheles gambiae*; introgression; microarray; gene expression; qPCR;  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## INTRODUCTION

1  
2  
3 Resistance to the insecticides employed in public health is a major challenge to the control of insect-  
4  
5 borne disease including malaria. Insects have evolved a diverse and impressive array of mechanisms  
6  
7 to counteract insecticide-based control measures (Hemingway and Ranson 2000). For mosquito  
8  
9 vectors of malaria, current controls rely mainly on pyrethroid-treated bednets or spraying of  
10  
11 insecticide onto surfaces where mosquitoes rest postprandially. Resistance-associated mutations in  
12  
13 the voltage-gated sodium channel, the target of pyrethroids, are well known and have evolved  
14  
15 repeatedly in *Anopheles gambiae sensu stricto* (Pinto et al. 2007; Donnelly et al. 2009). However,  
16  
17 resistance can also arise due to elevated expression of, or allelic variants in, metabolic genes, which  
18  
19 can act with target-site mutations to increase resistance (Mitchell et al. 2014). The identification of  
20  
21 the mechanisms underpinning resistance is a vital first step for the development of assays which can  
22  
23 be used to understand and predict how resistance spreads within and between populations, and  
24  
25 sometimes species.  
26  
27  
28  
29  
30

31  
32 Detoxification of xenobiotics such as insecticides requires either metabolism (sometimes through  
33  
34 intermediary compounds, which require processing and may be more toxic than the original  
35  
36 xenobiotic) or transformation through conjugation for subsequent sequestration and elimination. In  
37  
38 addition to the metabolic processes required to remove insecticide from within the insect, exposure  
39  
40 to toxic compounds can also trigger discrete, non-specific physiological reactions *e.g.* pyrethroid  
41  
42 exposure induces oxidative stress and lipid peroxidation (Vontas et al. 2001). Thus, the ability of a  
43  
44 mosquito to survive insecticide exposure may require multiple metabolic pathways, potentially  
45  
46 mediated by a wide range of enzymes. Identifying those genes underpinning such resistance can aid  
47  
48 not only in understanding potential cross-resistance to alternative insecticides but potentially lead to  
49  
50 diagnostic assays to aid resistance-monitoring. Whole genome microarrays have been used  
51  
52 extensively to study insecticide resistance phenotypes in *An. gambiae s.s.* and *An. coluzzii* and in  
53  
54 such studies it is typical to detect up-regulation of transcripts representing a wide-range of pathways  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 (e.g. Mitchell et al. 2012; Fossog Tene et al. 2013; Kwiatkowska et al. 2013). This suggests that  
2 metabolism is a complex, multigenic process, and is consistent with the sigmoidal distribution of  
3  
4 dose-responses often seen in field populations (e.g. Müller et al. 2008; Mawejje et al. 2013) which  
5  
6 imply a broad distribution of resistant phenotypes. Though large numbers of genes often appear  
7  
8 differentially regulated, microarray datasets can be littered with false positive hits (e.g. see Aubert et  
9  
10 al. 2004; Pawitan et al. 2005). However, confidence in identification of differentially-regulated genes  
11  
12 increases if a gene is identified in independent studies of the same phenotype. Repeated  
13  
14 identification of particular cytochrome P450s, including *Cyp6p3* and *Cyp6m2* in microarray studies of  
15  
16 resistant *An. gambiae* (Müller et al. 2007; Djouaka et al. 2008; Müller et al. 2008; Mitchell et al.  
17  
18 2012), and of *Cyp6p4* and *Cyp6p9* in *An. funestus* (Wondji et al. 2009; Riveron et al. 2013) has been  
19  
20 important in identifying these genes as worthy of the expense and time-consuming  
21  
22 enzymatic/biochemical characterization which has subsequently confirmed the role of these  
23  
24 enzymes in resistance (Müller et al. 2008; Stevenson et al. 2011; Mitchell et al. 2012; Riveron et al.  
25  
26 2013)

27  
28  
29  
30  
31  
32  
33  
34 In Uganda, a country with high levels of malaria transmission (Yeka et al. 2012), resistance to  
35  
36 pyrethroid insecticides is present in the three main malaria vectors; *An. gambiae* and *An. arabiensis*  
37  
38 (Ramphul et al. 2009; Verhaeghen et al. 2010; Mawejje et al. 2013) and *An. funestus* (Morgan et al.  
39  
40 2010). The relative frequency of *An. arabiensis* has risen in neighbouring Kenya (Lindblade et al.  
41  
42 2006; Bayoh et al. 2010; Mwangangi et al. 2013) and Tanzania (Derua et al. 2012) following  
43  
44 insecticidal control measures and there is now some evidence of elevated frequencies in eastern  
45  
46 Uganda (Mawejje et al. 2013) suggesting an increasing role in malaria transmission. Resistance to  
47  
48 pyrethroids is present, and apparently increasing, in *An. arabiensis* from Jinja, eastern Uganda  
49  
50 (Mawejje et al. 2013) but is not mediated by known 'knockdown resistance' target-site mechanisms  
51  
52 (*L1014F* and *L1014S*) in the voltage-gated sodium channel, which are extremely rare (*L1014S*  
53  
54 frequency <0.1% (Mawejje et al. 2013)). In the absence of a known target-site mechanism, metabolic  
55  
56 mechanisms are strongly implicated in the resistance phenotype.  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 Although *An. arabiensis* has an increasing role in malaria transmission, *An. gambiae* s.s. remains the  
2 major vector in some locations in Uganda such as Tororo (Weetman et al. unpublished), a region  
3 with extremely high rates of malaria infection (Kilama et al. 2014), wherein malaria infections have  
4 increased recently despite widespread bednet usage (Jagannathan et al. 2012), and the Northern  
5 Ugandan district of Apac, where insecticidal interventions have impacted upon clinical malaria  
6 indicators (Kigozi et al. 2012). Here we characterise the resistance mechanisms circulating in *An.*  
7 *arabiensis* from Jinja, and *An. gambiae* s.s. from Tororo and use recombinant protein expression  
8 followed by functional validation to examine the role of an up-regulated gene (*Gste4*) in the  
9 resistance phenotype. We show that *Gste4* shows a strong signature of selective importance, and  
10 that the signature, and gene expression of *Gste4*, is haplotype-specific.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



## 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*

1 We have previously described the pattern of insecticide resistance in *An. arabiensis* from Jinja  
2 (00°25'51" N 033°13'44" E) (Mawejje et al. 2013). Samples were collected as larvae (for details of  
3 collection locales see Mawejje et al. 2013) and raised to adulthood prior to bioassaying. Resistance  
4 to pyrethroids (permethrin and deltamethrin) in this population is more moderate than Tororo with  
5 an LT<sub>50</sub> to both insecticides of ≈ 50 minutes. For this second microarray experiment, resistant female  
6 samples surviving 60 min exposure to permethrin as per the WHO protocol (WHO 2013) and control  
7 samples, treated in an identical fashion except exposures were to untreated control papers, were  
8 stored in RNAlater. Colony samples were drawn from the Dongola (origin Dongola, Sudan, Ng'habi et  
9 al. 2007) and Moz (origin Chokwe, southern Mozambique, Witzig et al. 2013) colonies, both of which  
10 are susceptible to pyrethroids.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

#### 24 *RNA extraction and microarray analysis*

25  
26  
27 All individuals used were 3-5 day old females. RNA was extracted from pools of 10 mosquitoes using  
28 the PicoPure (Arcturus) kit for *An. gambiae* samples or RNAqueous4PCR kit (Ambion) for *An.*  
29 *arabiensis* samples following the manufacturer's recommendations and including a DNase step.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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

1 Software (Agilent Technologies). Analysis was undertaken using custom R-scripts and the MAANOVA  
2 package for R (Wu et al. 2009).  
3

#### 4 5 *Sequencing of the region around Gste4* 6

7  
8 Primers were designed to amplify *Gste4* and adjacent 5' and 3' regions (see Supplementary Table 1  
9 for these and all subsequent primer sequences). Primers GSTe5\_seq and GSTe2\_seq amplified a  
10 2245bp section of genomic DNA (chromosome 3R: 28,595,701-28,597,945) inclusive of sections of  
11 *Gste2* and *Gste5*, the entirety of *Gste4* and intergenic regions between *Gste2-Gste4* and *Gste4-Gste5*  
12 (Figure 1). PCRs were undertaken on DNA taken from resistance phenotyped sympatric *An.*  
13 *arabiensis* and *An. gambiae* from Jinja, and a single sample from each of the Dongola, Moz and  
14 Sennar (origin Sennar, Sudan, Du et al. 2005) colonies of *An. arabiensis*. Amplified products were  
15 cloned into pJET (Fermentas) and individual colonies picked for sequencing. Only single products  
16 from each specimen were sequenced unless intra-individual length variation was noted on agarose  
17 gels in which case both alleles were sequenced. Amplification primers and an internal sequencing  
18 primer Gste4\_seq were used in sequencing reactions (Figure 1). Sequences were manually edited  
19 and aligned in CodonCode Aligner (CodonCode Corporation), and Maximum Likelihood phylogenetic  
20 trees constructed in MEGA v5.2 (Tamura et al. 2011) using the appropriate model as determined by  
21 Model Test (Posada and Crandall 1998) with bootstrapping (500 replicates). The nucleotide  
22 sequences of *Gste4* from *An. quadriannulatus* and the outgroup *An. chrysti* (within and without the  
23 *An. gambiae* complex, respectively) were obtained from VectorBase (Megy et al. 2012) (supercontig  
24 KB667655: 1004768-1005118 (exon 1), 1005183-1005509 (exon 2) and contig APCM01015419:  
25 2842-3190 (exon 1), 3256-3582 (exon 2) respectively) and translated. Haplotype diversity and  
26 McDonald-Kreitman tests of selection were conducted in DnaSP (Librado and Rozas 2009) with the  
27 neutrality index (NI) calculated from this output where  $NI = [(P_N/D_N)/(P_S/D_S)]$  (Li et al. 2008) and –  
28  $\log_{10}(NI) > 0$  is indicative of positive selection. Due to zero values in the McDonald-Kreitman test we  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 followed the recommendation of Li *et al.* (2008) by adding a pseudocount of 1 to each cell before  
2 calculation of the NI.  
3

#### 4 5 *qPCR and haplotype-specific qPCR* 6

7  
8 cDNA was produced from  $\approx 2.5\mu\text{g}$  RNA samples (see above) using oligo dT<sub>20</sub> and superscript III  
9 (Invitrogen) as per the manufacturer's instructions. qPCR was undertaken on 1/50 dilutions of cDNA  
10 (Invitrogen) as per the manufacturer's instructions. qPCR was undertaken on 1/50 dilutions of cDNA  
11 using exon-crossing *Gste4* qPCR primers (GSTe4qPCR\_F1 and GSTe4qPCR\_R1) and haplotype specific  
12 qPCR primers designed to amplify group specific haplotypes of *Gste4* (GSTe4\_Hap8 and  
13 GSTe4\_Hap12 for group  $\alpha$ ; GSTe4\_Hap8 and GSTe4\_Hap9 for clade  $\beta$ ) which differed in the presence  
14 of large indels in the 3' UTR (see results). Three normalising genes, ribosomal protein S7  
15 (AGAP010592), ubiquitin (AGAP007927) and elongation factor (AGAP005128) were run on the same  
16 sample aliquots. qPCR was undertaken in triplicate in 20 $\mu\text{l}$  volumes containing 1x Agilent Brilliant III  
17 SYBR qPCR mastermix, 300nM each primer and 1 $\mu\text{l}$  cDNA (1/50 dilution) on an Agilent MX3005 with  
18 cycling conditions of 3min at 95°C followed by 40 cycles of 10s at 95°C and 10s at 60°C. Analysis  
19 used the  $\Delta\Delta\text{Ct}$  method (Livak and Schmittgen 2001).  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

#### 35 36 *Cloning and expression of GSTe4* 37

38  
39 Primers (GSTe4cDNA\_RE\_F and GSTe4cDNA\_RE\_R) were designed to amplify the full length  
40 sequence of *Gste4* incorporating a 5' *NdeI* site (CATATG where ATG is the translation initiation  
41 codon) and a 3' *BamHI* site, based on the *Gste4* sequence in VectorBase ([www.vectorbase.org\\_gene](http://www.vectorbase.org_gene)  
42 identifier AGAP009193; [Refseq accession XM\\_319967](#)).  
43  
44  
45  
46  
47  
48

49 cDNA was prepared from RNA extracted from pyrethroid resistant *An. arabiensis* using Superscript III  
50 (Invitrogen) following the manufacturer's recommendations and full-length *Gste4* amplified using  
51 high-fidelity Phusion polymerase (Fermentas). Products of the correct size were cloned into pJET  
52 (Fermentas) and sequenced. Inserts from plasmids containing confirmed *Gste4* were excised with  
53  
54  
55  
56  
57  
58  
59 *NdeI* and *BamHI* and ligated into pET15b (Novagen). pET15b contains an IPTG inducible T7 promoter,  
60  
61  
62  
63  
64  
65

1 a 6 x HIS tag and a thrombin cleavage site. GSTE4 expression vector was then transformed into  
2 BL21(DE3) (NEB), grown at 37°C in LB until an OD of 0.8 was reached, then expression was induced  
3  
4 with 1mM IPTG and cultures incubated at 25°C overnight. Cells were harvested at 10,000 rpm for  
5  
6 10min at 4°C and the supernatant discarded. The pellet was re-suspended in 20ml of low imidazole  
7  
8 buffer (25mM imidazole, 20mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5M NaCl, pH 7.4) and frozen at -80°C. After thawing,  
9  
10 lysozyme (0.5mg/ml) and DNase (0.05mg/ml) were added to the cell suspension and the solution  
11  
12 incubated on ice for 10min. Cells were disrupted by French press homogeniser (Stansted Fluid Power  
13  
14 Ltd) at 20,000psi and centrifuged at 18,000rpm for 20min to remove cell debris. Supernatant was  
15  
16 filtered through a 0.45µm filter and loaded manually on a 5mL His-trap column (GE Healthcare), pre-  
17  
18 equilibrated with low imidazole buffer. The column was washed with 25ml of low imidazole buffer,  
19  
20 followed by 25ml of medium imidazole buffer (50mM imidazole, 20mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5M NaCl, pH  
21  
22 7.4). The protein was then eluted manually with high imidazole buffer (0.5M imidazole, 20mM  
23  
24 Na<sub>2</sub>HPO<sub>4</sub>, 0.5M NaCl, pH7.4) and concentrated with a Vivaspin 20 concentrator (Sartorius) then  
25  
26 exchanged into 20mM Tris, 150mM NaCl pH 7.4 buffer using a PD-10 column (GE Healthcare).  
27  
28  
29  
30  
31

32  
33  
34 Protein concentration was determined using a NanoDrop (NanoDrop Technologies) and by the  
35  
36 Bradford assay (Bradford 1976). Activity of purified protein was checked by a colorimetric activity  
37  
38 assay measuring conjugation of reduced glutathione (GSH) to the model substrate 1-chloro-2, 4-  
39  
40 dinitrobenzene (CDNB) at 340nm ( $\epsilon=9.6\text{mM}^{-1}\cdot\text{cm}^{-1}$ ) (Habig et al. 1974) at a constant 22.5°C in a Cary  
41  
42 300 Bio UV-Vis spectrophotometer over 1min. Reactions contained 0.1M sodium phosphate, 500ng  
43  
44 enzyme, 1mM CDNB, 5mM GSH and 3.3% methanol in 1ml total volume.  
45  
46  
47  
48

#### 49 *Determination of optimal pH of GSTe4 variants*

50  
51

52 The optimal pH for each variant was determined using the CDNB activity assay over the pH range  
53  
54 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.  
55  
56 Kinetic constants ( $V_{\text{max}}$  and  $K_m$ ) for both CDNB and GSH were also determined for both variants at  
57  
58 pH6.5 and the optimal pHs as determined above.  
59  
60  
61  
62  
63  
64  
65

### *Determination of temperature optima*

1  
2  
3 Aliquots of both variants of GSTE4 were incubated for 30min over a range of temperatures (30°C-  
4  
5 65°C in 5°C increments). Following incubation, CDNB activity was measured as above.  
6  
7

### *Interaction of recombinant GSTE4 with permethrin and deltamethrin*

8  
9  
10  
11 Inhibition by permethrin and deltamethrin was determined by change in CDNB activity following  
12  
13 addition of 0µM, 25µM, 50µM, 75µM, 100µM deltamethrin or permethrin with a saturating  
14  
15 concentration of GSH (5mM) and CDNB (1mM) in triplicate reactions in 0.1M sodium phosphate  
16  
17 buffer with 500ng enzyme. Inhibition was measured at pH6.5, pH7 and pH7.8 (the optimal pHs  
18  
19 determined above plus neutrality).  
20  
21  
22  
23

### *In vitro permethrin and deltamethrin metabolism assays*

24  
25  
26  
27 Metabolism was undertaken in 0.1M sodium phosphate (pH6.5, pH7, pH7.8) in 0.5ml volumes  
28  
29 containing 5mM GSH, 10µM insecticide (DDT, permethrin or deltamethrin) and 50µg recombinant  
30  
31 enzyme (with 80°C 30min heat inactivated GSTE4 enzyme in negative control reactions). Reactions  
32  
33 were incubated at 25°C for 2h with shaking. Following incubation, bifenthrin (as spike-in extraction  
34  
35 control) was added to 10µM then reactions extracted twice with 1 volume *tert*-butyl methyl ether.  
36  
37 Extractions were pooled and dried under a constant stream of N<sub>2</sub> then resuspended in 150µl  
38  
39 methanol prior to analysis by reverse-phase HPLC (Chromleon, Dionex) with a monitoring  
40  
41 absorbance of 232nm. Reactions (100µl) were loaded into an isocratic mobile phase (90% methanol:  
42  
43 10% water) with a 1ml/min flow rate through a 250mm C18 column (Acclaim 120, Dionex) at 23°C.  
44  
45  
46  
47  
48  
49

### *Analysis of peroxidase function*

50  
51  
52  
53 Determination of Se-independent peroxidase function followed Vontas *et al.* (2001). In brief  
54  
55 reactions contained 1mM EDTA, 200µM NADPH, 1mM GSH, 0.3U glutathione reductase, 2µg enzyme  
56  
57 (removed from control reactions) and either 1.5mM cumene hydroperoxide or 1.5mM *t*-butyl  
58  
59  
60  
61  
62  
63  
64  
65

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).

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## 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 =  $4 \times 10^{-5}$ ). 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/miameexpress/>) 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



1 was applied and a conservative significance threshold applied (FDR-corrected significance level set at  
2  $\log_{10}(q \text{ value}) > 4$  ( $q < 0.0001$ )) (see Supplementary table S3 for results). When these 4,094  
3 significantly up-regulated probes were ranked by fold-change (FC), three separate probes targeting  
4 *Gste4* were within the top 25 significant probes and were the highest FCs of known detoxification  
5 family members (average FC for *Gste4* = 16.6). In pairwise comparisons between Jinja resistant and  
6 sympatric controls we applied a standard, multiple test-corrected threshold ( $q < 0.05$ ) more  
7 appropriate for within-population comparisons where expected differential expression between  
8 groups is likely to be lower. Here 1851 probes were significant (only 22 probes were significant, with  
9 the strict FDR-corrected significance level set at  $\log_{10}(q \text{ value}) > 4$  and these were mainly serine  
10 proteases). For comparisons of Jinja resistant to either Dongola or Moz susceptible colony samples  
11 1641 and 673 probes respectively were significantly differentially regulated at the strict  $\log_{10}(q$   
12 value)  $> 4$  level. All microarray data have been submitted to ArrayExpress with accession number E-  
13 MTAB-1873.

#### 31 *Haplotype analysis and SNP genotyping*

32 Sequencing of 2319bp around *Gste4* from both *An. arabiensis* ( $N = 10$  from Jinja plus one sequence  
33 from each of the Dongola, Moz and Sennar colonies) and *An. gambiae* ( $N = 10$ ) revealed marked  
34 variability, with higher variability in *An. arabiensis* from Jinja (haplotype diversity = 0.982, number of  
35 segregating sites = 98 (of 2319),  $\pi = 0.01719$ ) than *An. gambiae* (haplotype diversity = 0.682, number  
36 of segregating sites = 83,  $\pi = 0.00974$ ). Sequences have been submitted to Genbank with accession  
37 numbers KF733184-KF733209. Maximum likelihood phylogenetic reconstruction of these sequences  
38 shows two monophyletic clades composed of either *An. gambiae* or *An. arabiensis* haplotypes (Fig  
39 3A). When *Gste4* coding sequence alone is used as input the species-specific clades are still apparent  
40 (though with low bootstrap support; Supp. Fig 3). However, when amino acid-based trees are  
41 constructed, two groupings (labelled Group  $\alpha$  and Clade  $\beta$ ) are evident: these are not species-  
42 specific and the majority of sequences fall into group  $\alpha$  which is composed of both *An. arabiensis*  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 and *An. gambiae* sequences (Fig 3B). Thus, these sequences differ in nucleotide sequence in a  
2 species-specific manner, indicative that *Gste4* has not introgressed between these species, but are  
3 near-identical in amino acid sequence. The amino acid sequence of GSTE4 from *An. quadriannulatus*  
4 falls in clade  $\beta$ , suggesting that group  $\alpha$  may be more derived, although there is insufficient sampling  
5 to be conclusive. We also note that from our sequencing of this region there is no evidence of  
6 haplotypes containing the 42 amino acid deletion exhibited by cDNA clones 1 and 7 (see  
7 recombinant protein expression section) indicating that these may be the result of PCR errors or PCR  
8 recombination and not genuine variants segregating in the population. However, haplotype  
9 sequences exhibiting a 20 amino acid deletion were present (samples labelled Jinja *An. arabiensis* 1 &  
10 2) and by using primers GSTe4qPCRF1 and GSTe4qPCRR2 on genomic DNA we confirmed this  
11 deletion (see Supplementary Figure 4) suggesting this is a genuine variant segregating in the  
12 population. The correct splice donor and acceptor sites are present in these sequences adding  
13 weight to the interpretation that this is a genuine coding variant present in this population.  
14 However, we have not expressed these variants in our *E. coli* system.

15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34 From the haplotype sequences it was apparent that the 3'UTR region displays large differences in  
35 presence/absence of large indels. The multiple probes designed by the Agilent eArray microarray  
36 design software targeted this region and although multiple probes interrogate this region, they  
37 overlap by just 1bp and hence target the same portion of the 3' UTR (Fig. 4). Given the size of the  
38 indels it is likely these probes will hybridise with only one of the UTR variants (Sub-clade  $\beta'$  of Fig. 4  
39 and Fig. 3B).

#### 40 41 42 43 44 45 46 47 48 49 *McDonald-Kreitman Tests*

50  
51  
52 Utilising only sequences from sympatric *An. gambiae s.l.* from Jinja, based on the total sequenced  
53 coding region (inclusive of partial coding sequences of *Gste2* and *Gste5*) and comparing group  $\alpha$   
54 sequences to clade  $\beta$  sequences,  $D_S = 0$ ,  $P_S = 20$ ,  $D_N = 4$ ,  $P_N = 7$  yielding  $-\log_{10}(NI) = 1.12$  (following  
55 addition of pseudocount) and Fisher's exact test  $p = 0.0105$ . For GSTE4 alone  $D_S = 0$ ,  $P_S = 10$ ,  $D_N = 4$ ,  $P_N$   
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 = 7 yielding  $-\log_{10}(\text{NI}) = 0.837$  and  $p=0.055$ . The positive values of the NI are strongly indicative of  
2 the action of positive selection (Li et al. 2008).  
3

#### 4 5 *qPCR validation of gene expression results* 6

7  
8 Owing to the cross-species microarray hit for *Gste4*, qPCR focussed on this gene for *An. gambiae*  
9 from Tororo and also the two most significant genes (AGAP007187 and AGAP007188).  
10 Unfortunately, owing to extremely high sequence similarity between these latter genes and  
11 paralogues (98-99%) within the cluster AGAP007187-AGAP007190), it proved impossible to obtain  
12 efficient, specific qPCR primers. However, *Gste4* showed significant differences in gene expression  
13 between resistant and susceptible *An. gambiae* families, albeit at a lower fold change than observed  
14 in the microarray experiment (t-test: FC = 1.54;  $t_{34}=2.18$ ,  $P=0.034$ ).  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

25  
26 For *An. arabiensis*, qPCR did not fully validate the microarray results (Table 1). Permethrin resistant  
27 *An. arabiensis* showed significantly higher expression of *Gste4* (1.33-1.49  $p = 0.003$  where Bonferroni  
28 corrected  $\alpha = 0.017$ ) than samples from the two colonies. The difference between resistant and  
29 control samples was not significant after multiple testing correction ( $p = 0.047$  where Bonferroni  
30 corrected  $p = 0.017$ ). Due to the likely differential hybridisation of the microarray probes with  
31 different *Gste4* haplotypes we further examined *Gste4* expression using haplotype specific qPCR (see  
32 below).  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

#### 43 *Haplotype-specific qPCR* 44

45  
46 The two groups of GSTE4 haplotypes ( $\alpha$  and  $\beta$ ) are differentiable by large indels in the 3' UTR. We  
47 designed qPCR primers to measure haplotype specific expression of group members through  
48 placement of clade specific primers across an indel region that differed between group  $\alpha$  and clade  
49  $\beta$ . When gene expression was measured separately for each group there were large differences in  
50 fold-change, particularly in comparisons of permethrin resistant versus either Dongola or Moz  
51 colony samples with the clade  $\beta$  qPCR identifying FC>6000 (an artefactual consequence of no  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 measurable gene expression in Dongola/Moz) in both comparisons but group  $\alpha$  qPCR showing a  
2 significant 1.45 fold over expression for permethrin resistant versus Dongola and no significant  
3  
4 difference for permethrin resistant versus Moz (Table 1).  
5  
6

### 7 *Recombinant protein expression*

8  
9

10 In order to capture representative *Gste4* sequences for heterologous expression we sequenced nine  
11 separate *Gste4* full-length clones. From sequences of these nine clones of *Gste4* amplified from  
12 cDNA of permethrin resistant *An. arabiensis* five different protein-coding variants were identified  
13 (Figure 4) differing at 3-6 amino acids from the reference genome sequence of *An. gambiae*. In  
14 addition, two clones (1 and 7) exhibited a 42 amino acid deletion compared to the reference  
15 sequence. Whilst this coding-sequence does appear unlikely to be functional, it was isolated from  
16 two separate cDNA pools in two separate PCRs suggesting that it has not arisen through PCR error.  
17 cDNA sequences have been submitted to Genbank with accession numbers KF733210-KF733214.  
18 Three GSTE4 variants (variants 1, 4 and 9 of Figure 5) were taken forward to expression. Variant 1,  
19 which had the 42 amino acid deletion, exhibited no activity with the model substrate CDNB and no  
20 further work was undertaken on this variant. We note that this variant had a full-length open-  
21 reading frame and therefore was not obviously pseudogenic (*c.f.* the pseudogene of *An. stephensi*  
22 *Gste2* in Ayres et al. 2011). In expression of variants 4 (from clade  $\beta$  and henceforth labelled  
23 GSTE4Beta) and 9 (from group  $\alpha$  and henceforth labelled GSTE4Alpha), chosen as being the most  
24 divergent and representative of the two groups of *Gste4* ( $\alpha$  and  $\beta$  – see Fig 3b) we isolated 3-6 ml of  
25 10-12mg/ml of both variants. Both variants showed activity with the model substrate CDNB  
26 indicating that the recombinant enzyme was functional.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

51 We note that recombinant protein GSTE4Alpha is nearly identical in sequence to the majority of the  
52 haplotypes in group  $\alpha$ , but differed by two amino acids T222S and N223K that are not evident in any  
53 group  $\alpha$  sequence (all sequences are 222T and 223N). These are within the 3' primer site; since  
54 primers were designed based upon the VectorBase sequence these non-synonymous changes are  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 likely to result from incorporation of primers into the amplicon (hence are primer-induced amino  
2 acid changes rather than real variants present in these haplotypes).  
3

#### 4 5 *Characterization of activity* 6

7  
8 We characterised activity across a range of pHs – GSTE4Alpha and GSTE4Beta exhibited very  
9 different pH activity profiles and optima with GSTE4Beta showing optimal activity at pH7.8 and  
10 GSTE4Alpha at pH7 (Fig 6). We studied enzyme kinetics at three different pHs – 6.5 (the pH used for  
11 study of GSTe2 (Dowd et al. 2010), 7 and 7.8. Enzyme kinetics showed the differing activity profiles  
12 of these two variants with pH (Table 2). GSTE4Alpha displayed a consistently lower Km for CDNB  
13 than GSTE4Beta at all three pHs, suggesting it has a higher affinity for this substrate. Affinities for  
14 GSH were similar for both variants except at pH7.8 where the affinity of GSTE4Beta was low (high  
15 Km) and that of GSTE4Alpha was not measurable since the reaction did not plateau over the range  
16 measured.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

31 Whilst both GSTE4 variants showed similar patterns of temperature dependent activity: 100%  
32 activity at 35°C and 0% activity at 45°C, at 40°C there was a significant difference in activity with  
33 GSTE4Alpha variant more stable than GSTE4Beta (92% activity versus 66% activity – see Fig 7).  
34  
35  
36  
37  
38

#### 39 *Inhibition by and metabolism of insecticides in vitro* 40

41  
42 Activity against CDNB of both variants GSTE4Alpha and GSTE4Beta was strongly inhibited by  
43 permethrin and deltamethrin with the lowest inhibition at pH7 (Figure 8). GSTE4Alpha showed  
44 significantly higher inhibition than GSTE4Beta for both insecticides and for all pHs indicating that it  
45 has a higher affinity for pyrethroids. Although both insecticides inhibit the enzymes there was no  
46 evidence of actual metabolism of pyrethroids (results not shown).  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## DISCUSSION

1  
2  
3 Resistance to pyrethroid insecticides in *An. gambiae s.l.* in eastern Uganda is extensive and appears  
4  
5 to be increasing (Ramphul et al. 2009; Verhaeghen et al. 2010; Mawejje et al. 2013). There is some  
6  
7 evidence that the role of *An. arabiensis* in malaria transmission in the region may also be on the  
8  
9 increase (Mawejje et al. 2013) as has been seen in neighbouring countries (Lindblade et al. 2006;  
10  
11 Bayoh et al. 2010; Derua et al. 2012; Mwangangi et al. 2013). Here, we have undertaken microarray  
12  
13 analysis of the pyrethroid resistant phenotype in both *Anopheles gambiae* and *An. arabiensis* from  
14  
15 the same geographical region using two very different experimental microarray designs and have  
16  
17 detected the same gene – *Gste4* up-regulated in both studies. Repeatability across studies adds  
18  
19 weight to the interpretation of likely involvement of this enzyme in the resistance phenotype. We  
20  
21 see an obvious disparity in the number of significantly up-regulated probes in the two microarray  
22  
23 designs – 57 for the comparison of ‘resistant’ versus ‘susceptible’ *An. gambiae* families compared to  
24  
25 >4,000 for the comparison of insecticide resistant *An. arabiensis* with colonised resistant strains. This  
26  
27 illustrates the effect of very different designs. The much greater number of probes detected in the  
28  
29 latter design may reflect geographic confounding or the effects of inbreeding and colonisation (see  
30  
31 (Kristensen et al. 2005)).

32  
33  
34 The identification of the same up-regulated gene (*Gste4*) in two closely-related species from the  
35  
36 same region might have been a result of introgressive hybridization. However, we find clear, well-  
37  
38 supported species-specific clustering of *An. gambiae* and *An. arabiensis* *Gste4* haplotypes based  
39  
40 upon >2kbp of DNA sequence spanning *Gste4* indicating that introgression definitely does not  
41  
42 underlie this observation. In fact, the genomic region containing the *Gste4* locus is in a region of the  
43  
44 genome where *An. arabiensis* and *An. gambiae* show high levels of divergence (Weetman et al.  
45  
46 2014). In contrast to the results based on genomic DNA sequence, when GSTE4 amino acid  
47  
48 sequences are studied the most common protein sequence is shared by both species. This, despite  
49  
50 the clear separation of the whole haplotype sequence suggests that these species have converged  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 on the identical protein sequence or that the functional constraints have prevented divergence from  
2 ancestral sequence. The McDonald-Kreitman test result strongly supports the action of positive  
3 selection on these sequences indicative of either convergence or constraints on evolutionary  
4 change. Evolutionary convergence is a strong indication of adaptive evolution (Zhang and Kumar  
5 1997) and is highly suggestive of an important functional role for this enzyme.  
6  
7  
8  
9

10  
11  
12 Members of the glutathione-S transferase class of enzymes have been demonstrated to have roles in  
13 metabolism, detoxification and excretion of xenobiotics, coping with oxidative stress, and in  
14 processing odorant signals (Ranson and Hemingway 2005a; Ranson and Hemingway 2005b). Within  
15 *Anopheles gambiae s.l.* 28 GSTs are recognised (Ranson and Hemingway 2005a) with one class – the  
16 epsilon GSTs – being insect-specific (Ayres et al. 2011). At least one epsilon-class member, GSTE2,  
17 has DDTase activity and a demonstrated role in insecticide resistance ((Ranson et al. 1997; Wang et  
18 al. 2008; Mitchell et al. 2014). Whilst there is no direct evidence of a role for GSTs in pyrethroid  
19 resistance, GSTs have been implicated in the pyrethroid resistance phenotype through detoxification  
20 of pyrethroid-induced lipid peroxidation products (Vontas et al. 2001) and through potential  
21 sequestration of insecticide through binding of pyrethroid molecules to GSTs (Jirajaroenrat et al.  
22 2001; Kostaropoulos et al. 2001).  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

39 Characterisation of the role of GSTE4 in pyrethroid resistance requires heterologous expression and  
40 *in vitro* assays. Whilst a recombinant GSTE4 variant has been expressed previously (Ortelli et al.  
41 2003) this came from a susceptible colony of *An. gambiae* (with identical amino acid sequence to the  
42 reference PEST genome GSTE4 sequence). We have not identified this particular cDNA sequence in  
43 our (limited) sequencing of *Gste4* in pyrethroid resistant *An. arabiensis* from Jinja. There is high  
44 variability in *Gste4* coding sequences in *An. arabiensis* from this region – from just nine clones  
45 sequenced we identified five different amino acid variants (although two of these contained a 42  
46 amino acid deletion causing a loss of function). We have now biochemically characterised two of  
47 these variants from *An. arabiensis* which differ by five amino acids. One of these two variants falls  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 within clade  $\beta$  and the other is from group  $\alpha$  for which the amino acid sequence is conserved across  
2 *An. gambiae* and *An. arabiensis*. Note that we are aware that the design of primers for cloning of  
3  
4 full-length *Gste4* likely resulted in primer-induced changes in two amino acids in the C-terminus of  
5  
6 this protein. Whilst we do not know the functional significance of these alterations, and residues in  
7  
8 this C-terminal domain may contribute to substrate specificity (Sheehan et al. 2001), since these are  
9  
10 primer-induced changes affecting both variants equally, these are likely to have suppressed any  
11  
12 variant associated differences, not to have caused them.  
13  
14  
15

16  
17 Our enzyme kinetic data show differences in reaction kinetics, in pH optima and in inhibition by  
18  
19 insecticides between these two variants. Typically, enzyme characterisation studies on *An. gambiae*  
20  
21 *s.l.* study just one variant (usually from the susceptible Kisumu strain e.g. Ortelli et al. 2003). The  
22  
23 variants studied here are segregating in field-collected samples and the differences in kinetics may  
24  
25 be of functional importance. Indeed there is evidence from the paralogous GSTE2 that different  
26  
27 allelic variants can have very different kinetic and metabolic activities (Mitchell et al. 2014). Whilst  
28  
29 metabolism studies did not show clear evidence for metabolic activity of either variant with  
30  
31 pyrethroids, inhibition of GST variants has been taken as suggestive of binding and potentially  
32  
33 sequestration (Jirajaroenrat et al. 2001; Kostaropoulos et al. 2001). Our inhibition assays conducted  
34  
35 with co-incubated insecticide suggest pyrethroids may be capable of occupying either the active site  
36  
37 or the GSH binding site of *Gste4* and the differential inhibition we have seen indicates that GSTE4  
38  
39 encoded by different haplotypes have differing sequestration abilities. It is interesting that in *An.*  
40  
41 *arabiensis* two variants with different pH optima, reaction kinetics and inhibition by insecticides are  
42  
43 found in the population at similar frequencies; suggesting a role for balancing selection maintaining  
44  
45 alleles with differing functions or organ specificity.  
46  
47  
48  
49  
50  
51

52  
53 Whilst both the biochemical data suggest at present that a link to insecticide resistance is unclear,  
54  
55 our assays are not comprehensive and GSTE4 may have a role in some other pathway of importance  
56  
57 for the insecticide resistance phenotype. GSTs have known roles as catalysers of secondary  
58  
59  
60  
61



1 metabolism products of reactions involving cytochrome P450s (Ranson and Hemingway 2005a) and  
2 hence we may not have utilised the appropriate substrate. Further work on this awaits identification  
3 and isolation of insecticide metabolites. We did not detect activity with either cumene  
4 hydroperoxide or *t*-butyl hydroperoxide indicating that GSTE4 does not have a Se-independent  
5 peroxidase function (Vontas et al. 2001) which is in line with Ortelli et al. (2003) who found no  
6 activity with cumene hydroperoxide for the Kisumu variant.  
7  
8  
9  
10  
11  
12

13  
14 The up-regulation of *Gste4* detected by microarray in *An. gambiae* was validated through qPCR.  
15 Although *Gste4* was up-regulated in microarray comparisons of *An. arabiensis*, qPCR validation  
16 indicated some discrepancies – fold changes in comparisons of resistant samples to the two colonies  
17 were much lower with qPCR than microarray, and no significant difference in *Gste4* expression was  
18 seen in comparison of resistant samples to sympatric controls through qPCR. The sequencing of this  
19 region in field samples demonstrated that the microarray probes are unlikely to adequately  
20 hybridise to some *Gste4* haplotypes and this may have potentially lead to erroneous conclusions.  
21 Our sequencing of *Gste4* encompassed the full-length of the gene, untranslated regions (UTRs) and  
22 flanking intergenic regions. Sequences of the 3' UTR showed that large indels segregating in the *An.*  
23 *arabiensis* population co-localise with the binding sites for the whole genome array probes targeting  
24 this gene. In fact, the microarray probes are likely to only work on members of sub-clade  $\beta'$  and not  
25 to hybridize at all to other members of clade  $\beta$  or any member of the  $\alpha$  group. To address this, we  
26 designed haplotype-specific 3'UTR qPCR primers which differentiate members of group  $\alpha$  (the group  
27 exhibiting signs of sequence convergence) from clade  $\beta$ . Clade  $\beta$  expression is absent (or at  
28 extremely low levels) in the Dongola and Moz colonies, though present in the Jinja samples and this  
29 inflates the Log Q-value disproportionately in comparisons of resistant *An. arabiensis* to colony  
30 samples. Expression of members of group  $\alpha$ , whilst at higher levels in permethrin resistant samples  
31 to the Dongola colony, is not significantly up-regulated versus the Moz colony or sympatric controls.  
32 Thus, there is a haplotype-specific component to the *Gste4* up-regulation we inadvertently detected  
33 through microarray in *An. arabiensis* but little evidence of true gene expression differences when  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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 TFIIIE $\alpha$  transcription initiation factors.

1 Since we could not validate these results nor develop a functional assay in the absence of known  
2 function we did not pursue these hits further. For *An. arabiensis* two P450s showed evidence of up-  
3 regulation. *Cyp6m2*, up-regulated in many microarray comparisons of *An. gambiae* (Djouaka et al.  
4 2008; Stevenson et al. 2011; Mitchell et al. 2012) was not identified as up-regulated in the Jinja  
5 microarray using our strict criteria, however in qPCR there is significant up-regulation when  
6 permethrin resistant samples are compared to either of the two colony samples. This discrepancy  
7 between microarray and qPCR requires further investigation but may also indicate allelic differences  
8 in primer/probe binding sequences. The differential regulation of *Cyp6m3* seen in microarray  
9 comparisons seems to be completely driven by extremely low level expression in the two colony  
10 samples and shows no evidence of differential regulation in sympatric comparisons. We note that  
11 for this population of *An. arabiensis*, prior exposure to piperonyl butoxide (PBO) in diagnostic  
12 bioassays partially restored the susceptible phenotype (Mawejje et al. 2013). This partial restoration  
13 does indeed indicate that cytochrome P450s likely have some additional role in the resistance  
14 phenotype and serves to remind of the complexity of mechanisms underpinning insecticide  
15 resistance.

16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36 Whilst *Gste4* was up-regulated and demonstrated to be the subject of strong selection in two  
37 sympatric species capable of hybridising (Weetman et al. 2014) introgression does not explain this  
38 shared mechanism. Whilst our data do not support introgression of *Gste4* between these species,  
39 the identification of the same gene in two independent microarray studies, and the demonstration  
40 of strong selection on this gene is highly suggestive of an important function. The *in vitro* data  
41 indicates that GSTE4 is involved in sequestration of pyrethroids and is worthy of further study to  
42 elucidate the sequestration mechanism.

## 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

- 1  
2  
3 Andreassi C and Riccio A (2009) To localize or not to localize: mRNA fate is in 3'UTR ends. Trends Cell  
4  
5 Biol 19(9): 465-474.  
6  
7  
8 Aubert J, Bar-Hen A, Daudin J-J and Robin S (2004) Determination of the differentially expressed  
9  
10 genes in microarray experiments using local FDR. BMC Bioinformatics 5(1): 125.  
11  
12 Ayres CF, Muller P, Dyer N, Wilding CS, Rigden DJ and Donnelly MJ (2011) Comparative genomics of  
13  
14 the anopheline glutathione S-transferase epsilon cluster. PLoS ONE 6(12).  
15  
16  
17 Barrett LW, Fletcher S and Wilton SD (2012) Regulation of eukaryotic gene expression by the  
18  
19 untranslated gene regions and other non-coding elements. Cell Mol Life Sci 69(21): 3613-  
20  
21 3634.  
22  
23  
24 Bass C, Nikou D, Donnelly MJ, Williamson MS, Ranson H, Ball A, Vontas J and Field LM (2007)  
25  
26 Detection of knockdown resistance (*kdr*) mutations in *Anopheles gambiae*: a comparison of  
27  
28 two new high-throughput assays with existing methods. Malar J 6: e111.  
29  
30  
31 Bayoh MN, Mathias D, Odiere M, Mutuku F, Kamau L, Gimnig J, Vulule J, Hawley W, Hamel M and  
32  
33 Walker E (2010) *Anopheles gambiae*: historical population decline associated with regional  
34  
35 distribution of insecticide-treated bed nets in western Nyanza Province, Kenya. Malar J 9(1):  
36  
37 62.  
38  
39  
40  
41 Bradford MM (1976) Rapid and sensitive method for quantitation of microgram quantities of  
42  
43 protein utilizing principle of protein-dye binding. Anal Biochem 72(1-2): 248-254.  
44  
45  
46 Derua Y, Alifrangis M, Hosea K, Meyrowitsch D, Magesa S, Pedersen E and Simonsen P (2012) Change  
47  
48 in composition of the *Anopheles gambiae* complex and its possible implications for the  
49  
50 transmission of malaria and lymphatic filariasis in north-eastern Tanzania. Malar J 11(1): 188.  
51  
52  
53 Djouaka RF, Bakare AA, Coulibaly ON, Akogbeto MC, Ranson H, Hemingway J and Strode C (2008)  
54  
55 Expression of the cytochrome P450s, CYP6P3 and CYP6M2 are significantly elevated in  
56  
57 multiple pyrethroid resistant populations of *Anopheles gambiae* s.s. from Southern Benin  
58  
59 and Nigeria. BMC Genomics 9: e538.  
60  
61  
62  
63  
64  
65

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- Donnelly MJ, Corbel V, Weetman D, Wilding CS, Williamson MS and Black WC (2009) Does *ldr* genotype predict insecticide-resistance phenotype in mosquitoes? Trends Parasitol 25(5): 213-219.
- Dowd AJ, Morou E, Steven A, Ismail HM, Labrou N, Hemingway J, Paine MJ and Vontas J (2010) Development of a colourimetric pH assay for the quantification of pyrethroids based on glutathione-S-transferase. Int J Environ Anal Chem 90(12): 922-933.
- Du W, Awolola TS, Howell P, Koekemoer LL, Brooke BD, Benedict MQ, Coetzee M and Zheng L (2005) Independent mutations in the *Rdl* locus confer dieldrin resistance to *Anopheles gambiae* and *An. arabiensis*. Insect Mol Biol 14(2): 179-183.
- Fossog Tene B, Poupardin R, Costantini C, Awono-Ambene P, Wondji CS, Ranson H and Antonio-Nkondjio C (2013) 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 8(4): e61408.
- Habig WH, Pabst MJ and Jakoby WB (1974) Glutathione-s-transferases - first enzymatic step in mercapturic acid formation. J Biol Chem 249(22): 7130-7139.
- Hemingway J and Ranson H (2000) Insecticide resistance in insect vectors of human disease. Annu Rev Entomol 45: 371-391.
- Jagannathan P, Muhindo M, Kakuru A, Arinaitwe E, Greenhouse B, Tappero J, Rosenthal P, Kaharuza F, Kanya M and Dorsey G (2012) Increasing incidence of malaria in children despite insecticide-treated bed nets and prompt anti-malarial therapy in Tororo, Uganda. Malar J 11(1): 435.
- Jirajaroenrat K, Pongjaroenkit S, Krittanai C, Prapanthadara L-a and Ketterman AJ (2001) Heterologous expression and characterization of alternatively spliced glutathione S-transferases from a single *Anopheles* gene. Insect Biochem Mol Biol 31(9): 867-875.

- 1 Kigozi R, Baxi SM, Gasasira A, Sserwanga A, Kakeeto S, Nasr S, Rubahika D, Dissanayake G, Kanya  
2 MR, Filler S and Dorsey G (2012) Indoor residual spraying of insecticide and malaria  
3 morbidity in a high transmission intensity area of Uganda. PLoS ONE 7(8): e42857.  
4  
5  
6  
7 Kilama M, Smith DL, Hutchinson R, Kigozi R, Yeka A, Lavoy G, Kanya MR, Staedke SG, Donnelly MJ,  
8  
9 Drakeley C, Greenhouse B, Dorsey G and Lindsay SW (2014) Estimating the annual  
10 entomological inoculation rate for *Plasmodium falciparum* transmitted by *Anopheles*  
11 *gambiae* s.l. using three sampling methods in three sites in Uganda. Malaria J 13(1): 111.  
12  
13  
14  
15  
16 Kostaropoulos I, Papadopoulos AI, Metaxakis A, Boukouvala E and Papadopoulou-Mourkidou E  
17  
18 (2001) Glutathione S-transferase in the defence against pyrethroids in insects. Insect  
19 Biochem Mol Biol 31(4-5): 313-319.  
20  
21  
22  
23  
24 Kristensen TN, Sørensen P, Kruhøffer M, Pedersen KS and Loeschcke V (2005) Genome-wide analysis  
25 on inbreeding effects on gene expression in *Drosophila melanogaster*. Genetics 171(1): 157-  
26 167.  
27  
28  
29  
30  
31 Kwiatkowska RM, Platt N, Poupardin R, Irving H, Dabire RK, Mitchell S, Jones CM, Diabaté A, Ranson  
32 H and Wondji CS (2013) Dissecting the mechanisms responsible for the multiple insecticide  
33 resistance phenotype in *Anopheles gambiae* s.s., M form, from Vallée du Kou, Burkina Faso.  
34 Gene 519(1): 98-106.  
35  
36  
37  
38  
39  
40 Li J, Ribeiro JMC and Yan G (2010) Allelic gene structure variations in *Anopheles gambiae*  
41 mosquitoes. PLoS ONE 5(5): e10699.  
42  
43  
44  
45 Li YF, Costello JC, Holloway AK and Hahn MW (2008) "Reverse ecology" and the power of population  
46 genomics. Evolution 62(12): 2984-2994.  
47  
48  
49  
50 Librado P and Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism  
51 data. Bioinformatics 25(11): 1451-1452.  
52  
53  
54  
55 Lindblade KA, Gimnig JE, Kamau L, Hawley WA, Odhiambo F, Olang G, Ter Kuile FO, Vulule JM and  
56 Slutsker L (2006) Impact of sustained use of insecticide-treated bednets on malaria vector  
57 species distribution and culicine mosquitoes. J Med Entomol 43(2): 428-432.  
58  
59  
60  
61  
62  
63  
64  
65

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- Livak KJ and Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 25(4): 402-408.
- Mawejje HD, Wilding CS, Rippon EJ, Hughes A, Weetman D and Donnelly MJ (2013) Insecticide resistance monitoring of field-collected *Anopheles gambiae s.l.* populations from Jinja, eastern Uganda, identifies high levels of pyrethroid resistance. *Med Vet Entomol* 27(3): 276-283.
- 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 and the VectorBase C (2012) VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Res* 40(D1): D729-D734.
- Mitchell S, Stevenson B, Müller P, Wilding C, Yawson A, Field S, Hemingway J, Paine M, Ranson H and Donnelly M (2012) Identification and validation of a gene causing cross-resistance between insecticide classes in *Anopheles gambiae* from Ghana. *Proc Natl Acad Sci U S A* 109: 6147-6152
- 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 and Donnelly MJ (2014) Metabolic and target-site mechanisms combine to confer strong DDT resistance in *Anopheles gambiae*. *PLoS ONE* 9(3): e92662.
- Morgan JC, Irving H, Okedi LM, Steven A and Wondji CS (2010) Pyrethroid resistance in an *Anopheles funestus* population from Uganda. *PLoS ONE* 5(7): e11872.
- Müller P, Donnelly MJ and Ranson H (2007) Transcription profiling of a recently colonised pyrethroid resistant *Anopheles gambiae* strain from Ghana. *BMC Genomics* 8: e36.
- Müller P, Warr E, Stevenson BJ, Pignatelli PM, Morgan JC, Steven A, Yawson AE, Mitchell SN, Ranson H, Hemingway J, Paine MJI and Donnelly MJ (2008) Field-caught permethrin-resistant *Anopheles gambiae* overexpress CYP6P3, a P450 that metabolises pyrethroids. *PLoS Genet* 4(11): e1000286.



- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- Mwangangi J, Mbogo C, Orindi B, Muturi E, Midega J, Nzovu J, Gatakaa H, Githure J, Borgemeister C, Keating J and Beier J (2013) Shifts in malaria vector species composition and transmission dynamics along the Kenyan coast over the past 20 years. *Malar J* 12(1): 13.
- Ng'habi KR, Horton A, Knols BGJ and Lanzaro GC (2007) 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* 77(3): 485-487.
- Ortelli F, Rossiter LC, Vontas J, Ranson H and Hemingway J (2003) Heterologous expression of four glutathione transferase genes genetically linked to a major insecticide-resistance locus from the malaria vector *Anopheles gambiae*. *Biochem J* 373: 957-963.
- Pawitan Y, Michiels S, Koscielny S, Gusnanto A and Ploner A (2005) False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics* 21(13): 3017-3024.
- Pinto J, Lynd A, Vicente JL, F. S, Randle NP, Caccone A, Gentile G, Moreno M, Simard F, Charlwood JD, do Rosário VE, della Torre A and Donnelly MJ (2007) Origins and distribution of knockdown resistance mutations in the afrotropical mosquito vector *Anopheles gambiae*. *PLoS ONE* 11: e1243.
- Posada D and Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14(9): 817-818.
- Ramphul U, Boase T, Bass C, Okedi LM, Donnelly MJ and Muller P (2009) Insecticide resistance and its association with target-site mutations in natural populations of *Anopheles gambiae* from eastern Uganda. *Trans R Soc Trop Med Hyg* 103(11): 1121-1126.
- Ranson H and Hemingway J (2005a) 5.11 - Glutathione Transferases. In: Editors-in-Chief: Lawrence IG, Kostas I and Sarjeet SG (eds) *Comprehensive Molecular Insect Science*. Elsevier, Amsterdam, pp. 383-402.
- Ranson H and Hemingway J (2005b) Mosquito glutathione transferases. In: Helmut S and Lester P (eds) *Methods Enzymol*. Academic Press, pp. 226-241.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- Ranson H, Prapanthadara LA and Hemingway J (1997) Cloning and characterization of two glutathione S-transferases from a DDT-resistant strain of *Anopheles gambiae*. *Biochem J* 324: 97-102.
- Riveron JM, Irving H, Ndula M, Barnes KG, Ibrahim SS, Paine MJ and Wondji CS (2013) 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* 110(1): 252-257.
- Scott JA, Brogdon WG and Collins FH (1993) Identification of single specimens of the *Anopheles gambiae* complex by the Polymerase Chain Reaction. *Am J Trop Med Hyg* 49(4): 520-529.
- Sheehan D, Meade G, Foley VM and Dowd CA (2001) Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem J* 360(1): 1-16.
- Stevenson BJ, Bibby J, Pignatelli P, Muangnoicharoen S, O'Neill PM, Lian LY, Muller P, Nikou D, Steven A, Hemingway J, Sutcliffe MJ and Paine MJ (2011) Cytochrome P450 6M2 from the malaria vector *Anopheles gambiae* metabolizes pyrethroids: sequential metabolism of deltamethrin revealed. *Insect Biochem Mol Biol* 41(7): 492-502.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10): 2731-2739.
- Verhaeghen K, Van Bortel W, Roelants P, Okello PE, Talisuna A and Coosemans M (2010) Spatio-temporal patterns in *kdr* frequency in permethrin and DDT resistant *Anopheles gambiae s.s.* from Uganda. *Am J Trop Med Hyg* 82(4): 566-573.
- Vontas JG, Small GJ and Hemingway J (2001) Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in *Nilaparvata lugens*. *Biochem J* 357: 65-72.
- Wang Y, Qiu L, Ranson H, Lumjuan N, Hemingway J, Setzer W, Meehan E and Chen L (2008) Structure of an insect epsilon class glutathione S-transferase from the malaria vector *Anopheles*

1 *gambiae* provides an explanation for the high DDT-detoxifying activity. J Struct Biol 164(2):  
2 228-235.  
3

4 Weetman D, Steen K, Rippon EJ, Mawejje HD, Donnelly MJ and Wilding CS (2014) Contemporary  
5 gene flow between wild *An. gambiae s.s.* and *An. arabiensis*. Parasites and Vectors 7: 345.  
6

7 Weetman D, Wilding CS, Müller P, Steen K, Rippon EJ, Morgan JC, Mawejje HD, Rigden D, Okedi LM  
8 and Donnelly MJ (unpublished) Metabolic gene polymorphisms contribute to class I and II  
9 pyrethroid resistance in East African *Anopheles gambiae*.  
10

11 WHO (2013) Test procedures for insecticide resistance monitoring in malaria vector mosquitoes  
12 World Health Organisation, Geneva.  
13

14 Wilding CS, Weetman D, Steen K and Donnelly MJ (2009) High, clustered, nucleotide diversity in the  
15 genome of *Anopheles gambiae* revealed by SNP discovery through pooled-template  
16 sequencing: implications for high-throughput genotyping protocols. BMC Genomics 10:  
17 e320.  
18

19 Witzig C, Parry M, Morgan JC, Irving H, Steven A, Cuamba N, Keraf-Hinzoumbe C, Ranson H and  
20 Wondji CS (2013) Genetic mapping identifies a major locus spanning P450 clusters  
21 associated with pyrethroid resistance in *kdr*-free *Anopheles arabiensis* from Chad. Heredity  
22 110(4): 389-397.  
23

24 Wondji CS, Irving H, Morgan J, Lobo NF, Collins FH, Hunt RH, Coetzee M, Hemingway J and Ranson H  
25 (2009) Two duplicated P450 genes are associated with pyrethroid resistance in *Anopheles*  
26 *funestus*, a major malaria vector. Genome Res 19: 452-459.  
27

28 Wu H, Yang H and Churchill GA (2009). <http://churchill.jax.org/software/rmaanova/maanova.pdf>  
29

30 Yeka A, Gasasira A, Mpimbaza A, Achan J, Nankabirwa J, Nsobya S, Staedke SG, Donnelly MJ,  
31 Wabwire-Mangen F, Talisuna A, Dorsey G, Kanya MR and Rosenthal PJ (2012) Malaria in  
32 Uganda: challenges to control on the long road to elimination: I. Epidemiology and current  
33 control efforts. Acta Trop 121(3): 184-195.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Zhang J and Kumar S (1997) Detection of convergent and parallel evolution at the amino acid  
sequence level. *Mol Biol Evol* 14(5): 527-536.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Table 1. Fold-changes and *P*-values from qPCR validation of microarray hits.

	<i>Gste4</i>										
	Microarray		exon-crossing <i>Gste4</i> primers			Clade β			Group α		
	<i>P</i>	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			<b>0.49</b>	<b>0.41</b>	<b>0.58</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.58</b>	<b>0.33</b>	<b>0.82</b>
Moz vs Control			<b>0.55</b>	<b>0.41</b>	<b>0.70</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.70</b>	<b>0.53</b>	<b>0.87</b>
Perm Resistant vs Control	0.801	1.04	<b>0.74</b>	<b>0.67</b>	<b>0.81</b>	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	<b>1.49</b>	<b>1.35</b>	<b>1.63</b>	<b>6795.69</b>	<b>4966.25</b>	<b>8625.12</b>	<b>1.45</b>	<b>1.07</b>	<b>1.83</b>
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	<b>1.33</b>	<b>1.21</b>	<b>1.46</b>	<b>7680.12</b>	<b>5612.59</b>	<b>9747.65</b>	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			GSTe4 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 ( $\mu\text{mol}/\text{min}/\text{mg}$ )	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 ( $\mu\text{mol}/\text{min}/\text{mg}$ )	12.071	15.597	34.607	29.148	35.207	1.92E+08

Figure 1

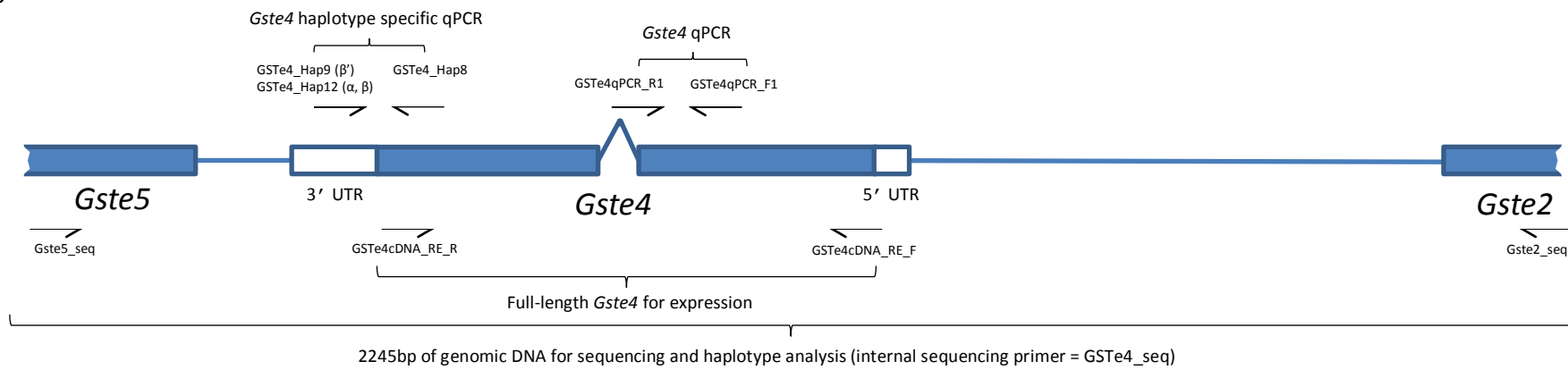


Figure 1. Genomic context of the *Gste4* gene on chromosome 3R of *Anopheles gambiae*. The locations of primers designed for sequencing and qPCR are indicated on the figure.

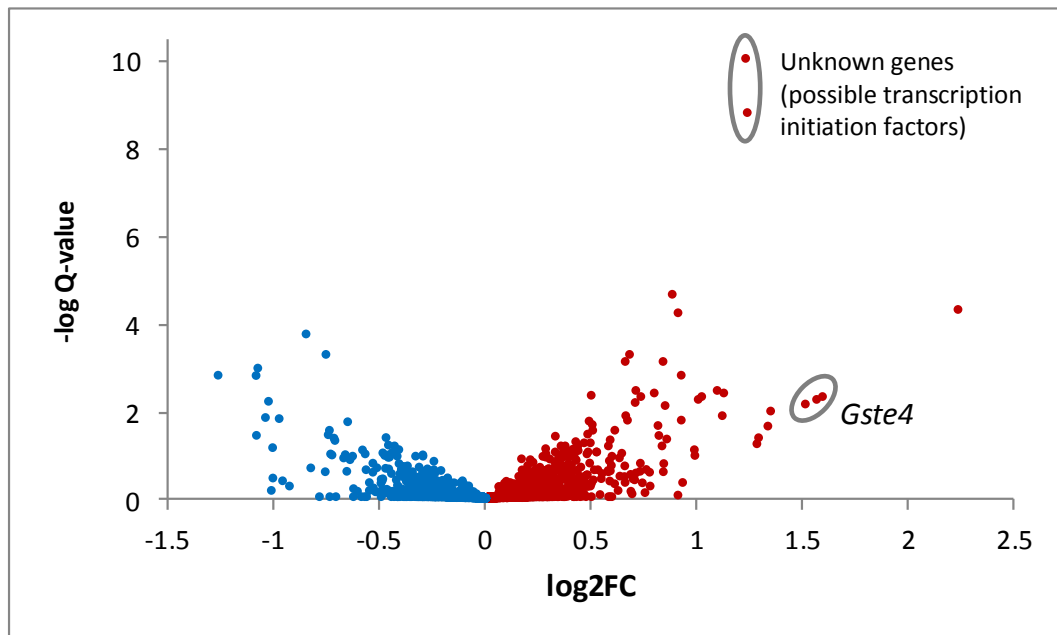
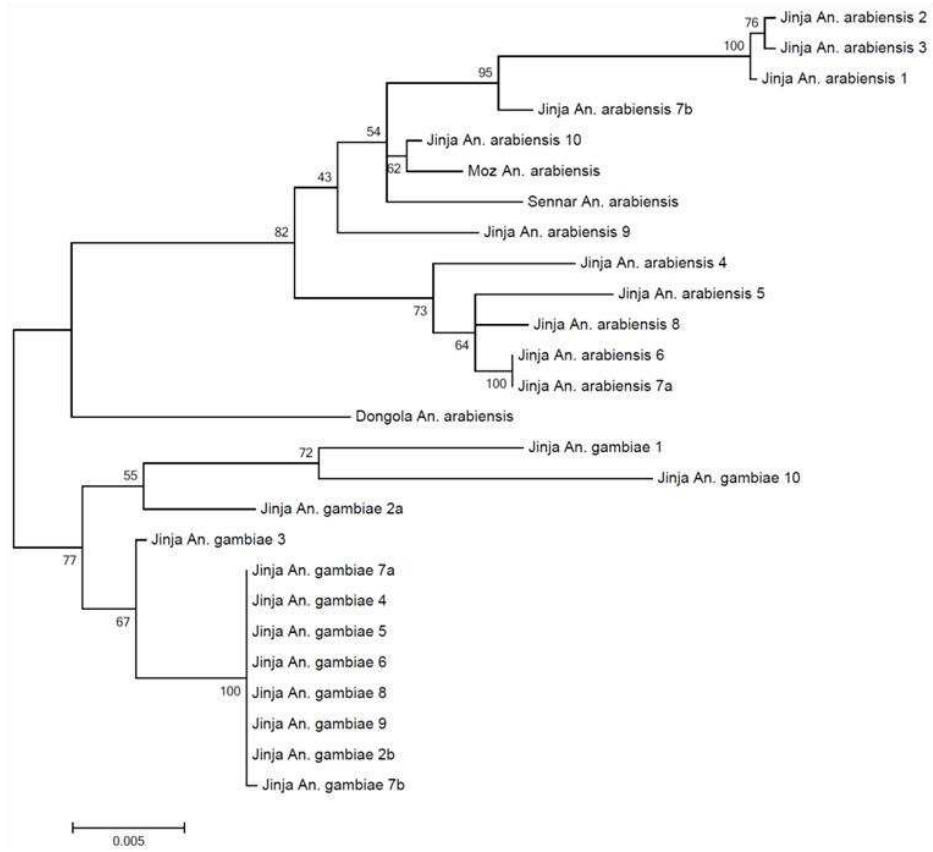


Fig 2. Volcano plot summarising  $\log_2$  fold changes ( $\log_2FC$ ) plotted against multiple testing corrected probability ( $-\log Q$ -value) for 20 resistant vs. 20 susceptible *An. gambiae* s.s. families from Tororo



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.

A.



B.

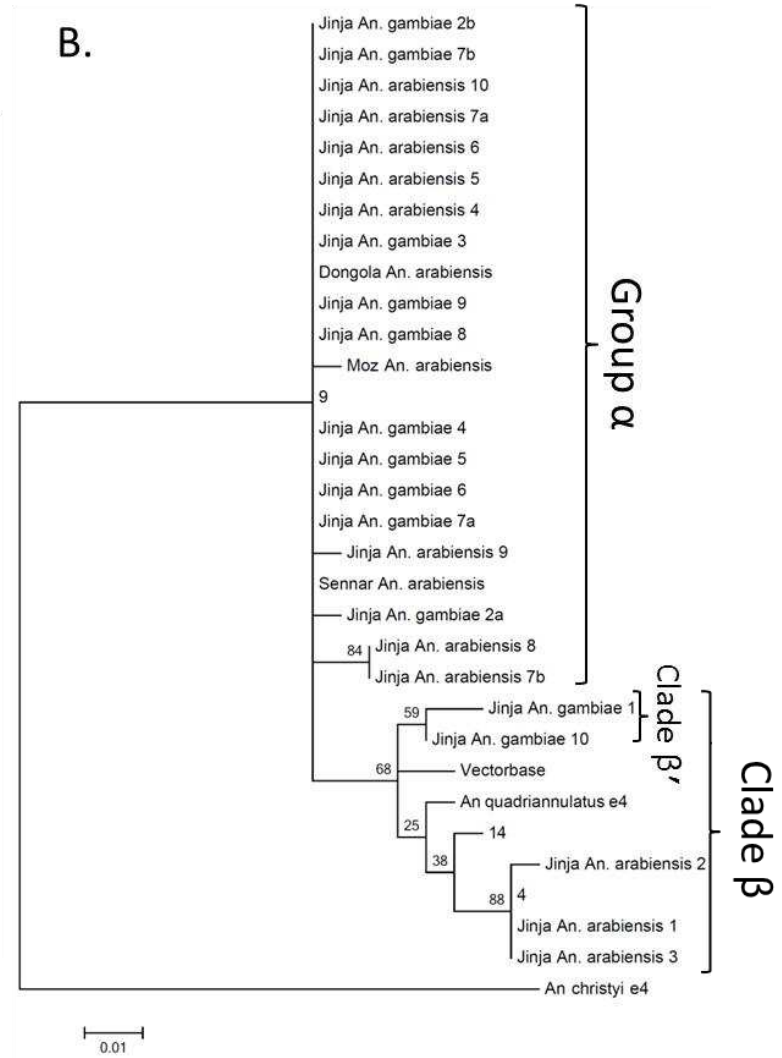


Figure 4. Alignment of the three Agilent whole genome microarray probes (60bp sequence; DETOX\_622\_PI422610884, DETOX\_623\_PI422610884, DETOX\_624\_PI422610884) designed to interrogate *Gste4* to the 3'UTR of *Gste4* in Ugandan *Anopheles* sequences. PEST = PEST reference sequence from VectorBase [55]. Representative haplotypes of this region are shown for members of the  $\alpha$ ,  $\beta$  and  $\beta'$  groups of Fig. 2B).

```

DETOX_622_PI422610884   GCATA-----GGCACCGAAA-----TAC---AACAAAATGATGCAAATTGAGAGAGTATATTTGGTAGCTGTT
DETOX_623_PI422610884   CATA-----GGCACCGAAA-----TAC---AACAAAATGATGCAAATTGAGAGAGTATATTTGGTAGCTGTTT
DETOX_624_PI422610884   ATA-----GGCACCGAAA-----TAC---AACAAAATGATGCAAATTGAGAGAGTATATTTGGTAGCTGTTTG
PEST                    GCATA-----GGCACCGAAA-----TCC---AACAAAATGATGCAAATTGAGAGAGTATATTTGGTAGCTGTTTG
Jinja An. gambiae 1( $\beta'$ ) GCATA-----GGCACCGAAA-----TCC---AACAAAATGATGCAAATTGAGAGAGTATATTTGGTAGCTGTTT
Jinja An.arabiensis 1( $\beta$ ) GCATACATTGAGCATTACAAAATTGTGACGTCGGTACTAAAAGTACTATTTTCGCAAAGAAAATGATGCAAATTGAGAGAGTATATTTGGTAGCTGTTTG
Jinja An.arabiensis 10( $\alpha$ ) GCATACATTGAGCATTACAAAATTGTGACGTCGGCACTAAAAGTACTATTACGCAAAGAAAAGTATGATGCAAATTGAGAGAGCATATTTGGTAGCTGTTT

```

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](http://www.vectorbase.com)). Residues differing from the VectorBase sequence are highlighted. Variant 9 has been subsequently characterized as GSTE4ALPHA and variant 4 as GSTE4BETA.

```

GSTe4_VB      MPNIKLYTAKLSPPGRSVELTAKALGLELDIVPINLLAQEHLTEAFRKLNPQHTIPLIDD
14            MPNIKLYTAKLSPPGRSVELTAKALGLELDIVPINLLAQEHLTEAFRKLNPQHTIPLIDD
9            MPNIKLYTAKLSPPGRSVELTAKALGLELDIVPINLLAQEHLTEAFRKLNPQHTIPLIDD
4            MPNIKLYTAKLSPPGRSVELTAKALGLELDIVPINLLAQEHLTEAFRKLNPQHTIPLIDD
1            MPNIKLYTAKLSPPGRSVELTAKALGLELDIVPINLLAQEHLTEAFRKLNPQHTIPLIDD
7            MPNIKLYTAKLSPPGRSVELTAKALGLELDIVPINLLAQEHLTEAFRKLNPQHTIPLIDD
*****:*** *****:***

GSTe4_VB      NGTIVWDSHAINVYLVSKYGKPEGDSLYPSDVVQRAKVNAALHFDSGVLFARFRFYLEPI
14            NGTIVWDSHAINVYLVSKYGKPEGDSLYPSDVVQRAKVNAALHFDSGVLFARFRFYLEPI
9            NGTIVWDSHAINVYLVSKYGKPEGDSLYPSDVVQRAKVNAALHFDSGVLFARFRFYLEPI
4            NGTIVWDSHAINVYLVSKYGKPEGDSLYPSDVVQRAKVNAALHFDSGVLFARFRFYLEPI
1            NGTIVWDSHAINVY-----LEPI
7            NGTIVWDSHAINVY-----LEPI
*****      *****

GSTe4_VB      LYYGATEPQEKIDNLYRAYELLNDTLVDEYIVGNEMTLADLSCIASIASMHAIFPIDAG
14            LYYGATEPQEKIDNLYRAYELLNDTLVDEYIVGNEMTLADLSCIASIASMHAIFPIDAG
9            LYYGATEPQEKIDNLYRAYELLNDTLVDEYIVGNEMTLADLSCIASIASMHAIFPIDAG
4            LYYGAAETPQEKIDNLYRAYELLNDTLVDEYIVGNEMTLADLSCIASIASMHAIFPIDAG
1            LYYGAAETPQEKIDNLYRAYELLNDTLVDEYIVGNEMTLADLSCIASIASMHAIFPIDAG
7            LYYGAAETPQEKIDNLYRAYELLNDTLVDEYIVGNEMTLADLSCIASIASMHAIFPIDAG
*****:*****

GSTe4_VB      KYPRLAGWVVERLAKLPYYEATNRAGAEELAQLYRAKLEQNRRTNAK
14            KYPRLAGWVVERLAKLPYYEATNRAGAEELAQLYRAKLEQNRRTNAK
9            KYPRLAGWVVERLAKLPYYEATNRAGAEELAQLYRAKLEQNRRTNAK
4            KYPRLAGWVVERLAKLPYYEATNRAGAEELAQLYRAKLEQNRRTNAK
1            KYPRLAGWVVERLAKLPYYEATNRAGAEELAQLYRAKLEQNRRTNAK
7            KYPRLAGWVVERLAKLPYYEATNRAGAEELAQLYRAKLEQNRRTNAK
*****:*****

```

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

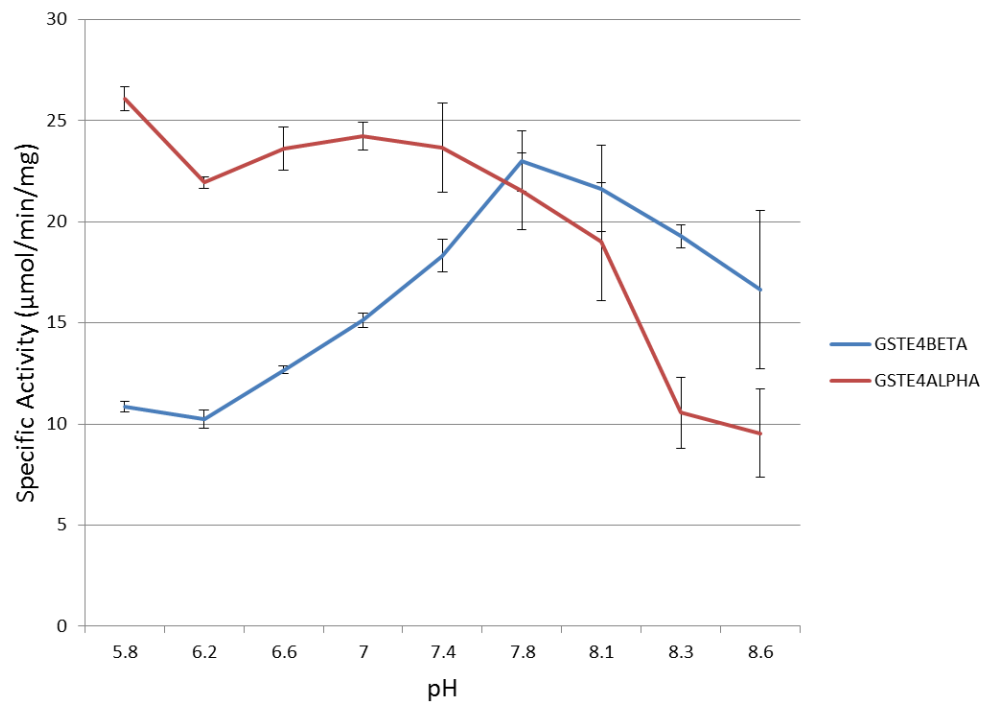


Figure 7. Temperature stability of two variants of GSTE4 (GSTE4Alpha and GSTE4Beta).

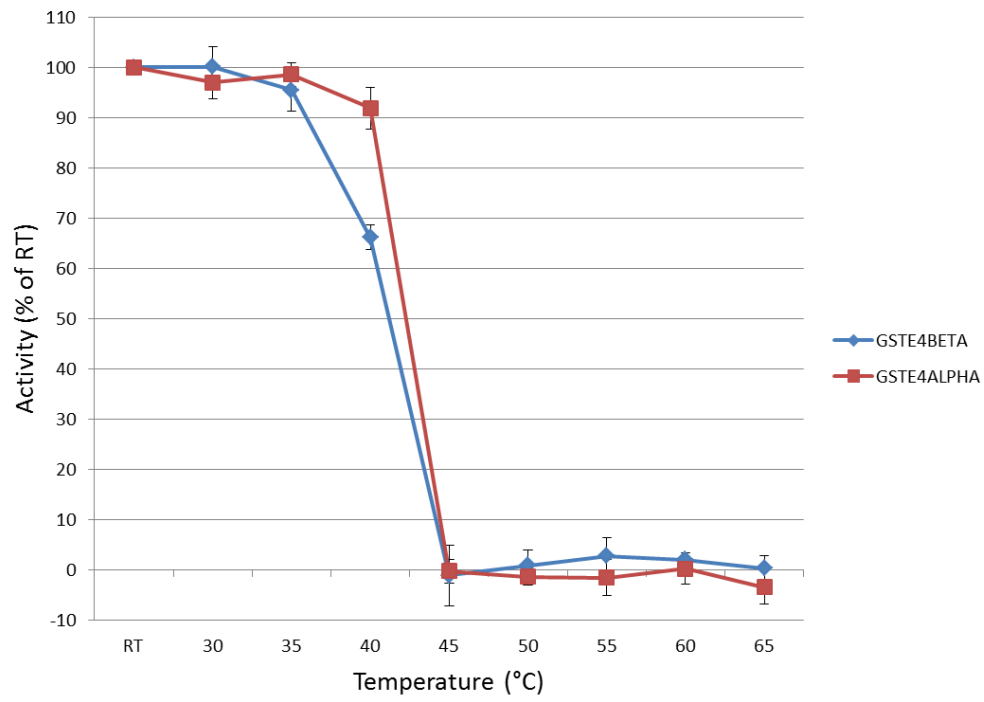


Figure 8. Inhibition of GSTE4 variants GSTE4Alpha (A - dark blue) and GSTE4Beta (B - light blue) by various concentrations of insecticide (0-100µM). Values are % of activity of the 0µM insecticide point (± 95% C.I.). Note that at higher concentrations of insecticide, activity in the blank samples was > experimental likely due to precipitation of insecticide. The activity (Y-axis) in the absence of insecticide has been set at 100% for clarity.

