International Journal of Entomology Research
ISSN: 2455-4758
Impact Factor: RJIF 5.24
www.entomologyjournals.com
Volume 3; Issue 3; May 2018; Page No. 11-18

Isolation of epsilon class glutathione S-transferases GST genes from a DDT resistant laboratory colony of *Anopheles arabiensis* MAT strains

Yayo AM[1], Ado A[2], Habibu UA[3], Mohammed BR[4], Bello MI[5], Hemingway J[6]
[1] Centre for Infectious Disease Research, Bayero University Kano, Nigeria
[2] Sustainable Health and Malaria Control, Gyadi-Gyadi, Kano, Nigeria
[3] Department of Biology, Federal College of Education (Tech), Bichi, Kano, Nigeria
[4] Department of Parasitology and Entomology, University Abuja, Nigeria
[5] Department of Physics, Bayero University Kano, Nigeria
[6] Vector Biology Research Group, Liverpool School of Tropical Medicine, United Kingdom

Abstract
Epsilon class glutathione s-transferases (GSTe) are found specifically in insects and are important in metabolism of insecticides. The position of GSTe genes cluster in the genome of *An. gambiae* has been located and evidence for involvement of individual GSTe genes in DDT resistance demonstrated. Polymerase Chain Reaction (PCR) method was employed using primers which were designed for amplification of the individual members of the Epsilon class *An. gambiae* GSTs to amplify the orthologous genes from *An. arabiensis*. The PCR products were cloned and sequenced. Eight Epsilon GSTe1 – GSTe8 genes were identified in this study. The deduced amino acid sequences shared between 91% - 98.7% similarity with the Epsilon GSTs identified in *An. gambiae*. The high percentage similarity at amino acid level, strongly suggests the inclusion of the *An. arabiensis* GST into the Epsilon class and the phylogenetic analysis revealed the common ancestry of the two species.

Keywords: *Anopheles arabiensis*, Epsilon GST, Isolation, DDT, Resistance

1. Introduction
Glutathione s-transferases (GSTs) are large enzyme family found ubiquitously in most organisms. They are involved in the metabolism, detoxification and excretion of numerous endogenous and exogenous toxic compounds from the cell [1]. Mammalian GSTs have been continuously classified into numerous classes based on various criteria including substrate specificities, immunological properties, amino acid sequence identity and structural characteristics [1, 2]. In insects, the GSTs have been classified into six cytosolic classes: Omega, Sigma, Theta, Zeta, Delta and Epsilon [3, 4]. The Delta and Epsilon class GSTs are specific to insects and in mosquitoes, have been implicated in detoxification of organophosphate [5], pyrethroids [6, 7] and organochlorine insecticides [8, 9]. *An. gambiae* GSTs are of particular interest because of their involvement in resistance to DDT. The pioneer studies on evidence demonstrating the role of *An. gambiae* GSTs in DDT resistance have been reviewed chronologically [3].
Two quantitative trait loci (QTL) associated to DDT resistance rtd1 and rtd2 were identified and cytologically mapped to location at chromosome 3R and a gene associated to the region was over expressed in DDT resistant ZAN/U strain [9, 10]. Several studies on genomic analysis of the epsilon gene cluster in the family Culicidae (mosquitoes) revealed evidence for duplication and independent expansion of epsilon gene family not only between the Culicinae and Anophelinae but also within the genus Anopheles [10, 11, 12]. Eight Epsilon class GST genes (GSTE1 – GSTE8) were identified in the genomes of both *An. gambiae* and Ae aegypti but only four putative orthologs were identifiable between the species [11, 13, 12] using primer walking approach amplified and characterized six GSTE genes from the genomes of the malaria vectors *An. stephensi* and *An. funestus*. They compared the genomic organisation of the six epsilon genes from the two malaria vectors with their putative orthologs from *An. gambiae*. They were not able to amplify GSTE3 and GSTE8 both from *An. stephensi* and *An. funestus.* Three out of the eight *An. gambiae* GSTEs including GSTE1, GSTE3, and GSTE8 have not been identified in *An. plumbeus* [13].The genetic variation observed in epsilon gene cluster between the Anopheline species was thought to be suggestive to their critical role for adaptation to specific environment [14, 15].
Hitherto, research on epsilon GST genes was largely focused on *An. gambiae s.s*. *An. arabiensis* is one of the major malaria vectors in Africa and is mainly responsible for the malaria transmission in South Africa and Sudan [16, 17]. It is a sibling species of *An. gambiae* complex, found sympatrically with *An.gambiae* ss but the two species differ strikingly in their preferences for hosts, oviposition sites, spatial/temporal distributions and tolerance to man-made ecological changes [18, 19]. In particular, ecological niches of the larvae of these sister species vary in intensity of xenobiotic pressure but it is not known how this difference will affect the genes such as GSTEs which enable the species to survive.
In this study, we describe the isolation of GSTEs from *An. arabiensis*, their characterization and genomic comparison

with putative orthologs from *An. gambiae*, *Aedes aegypti* and other dipteran. Degenerative PCR approach, using primers which were designed for amplification of the individual members of Epsilon class *An. gambiae* GSTEs was used to amplify the orthologous genes from *An. arabiensis*. Eight genes were isolated which based on phylogenetic analysis and similarity of deduced amino acid sequences with published sequences of other mosquitoes, were identified as Epsilon GSTs.

2. Methods

2.1 Mosquito strains

Eggs of *An. arabiensis* MAT strain were collected from a field site at Matatuine in Mozambique and a colony was established at Liverpool School of Tropical Medicine in 2002. The eggs were reared to adults in the insectary under standard conditions (temperature, $27^{oc} \pm 1^{oc}$; relative humidity 60 – 80%; 12/12 h L/D cycle). Adults from this strain have been maintained under selection pressure by exposure to What-man filter paper impregnated with 4% DDT according to standard WHO procedure [19].

Species were identified morphologically using taxonomic keys [20] and their identity confirmed through PCR described by [21]. Genomic DNA was extracted using the LIVAK buffer extraction method described by [23]. The DNA was extracted from 3 batches of four individual one-day old adult mosquitoes each from a colony of *An. arabiensis* MAT strain parental line, *An. gambiae* Kisumu and ZAN/U strains. Each batch was homogenised in 100µl of LIVAK buffer (0.13M Tris-HCl-containing 80mM NaCl, 0.16M sucrose, 0.05M EDTA, 0.5% (W/V) SDS, pH 8.0) and the homogenate was incubated at 65ºC for 30 minutes. Potassium acetate (8M) was added to a final concentration of 1 M and the homogenate was mixed and incubated on ice for 10 minutes. The sample was centrifuged at 12000g, 4ºC for 20 minutes and the supernatant was transferred to a new tube. After precipitating DNA with cold absolute ethanol and washing twice with 70% (V/V) ethanol to remove the salt, the pellet was re-suspended in 100µl TE (10mM Tris-HCl, 1mM EDTA, pH 8.0) and the genomic DNA stored at 4ºC.

The species-specific oligonucleotide primers – GA [5' - CTG GTT TGG TCG GCA CGT TT - 3'] and AR [5' - AAG TGT CCT TCT CCA TCC TA - 3'] in combination with a universal primer UN[5' - GTG TGC CCC TTC CTC GAT GT - 3'] (Scott *et al.*, 1993). The primers GA and AR amplify fragments of approximately 350 bp for *An. gambiae* s.s. and 315 bp for *An. arabiensis*. An optimised PCR reaction mix (25µl) was prepared containing the following: 0.2mM dNTPs, 0.2mM $MgCl_2$, 0.0037mM AR, 0.0012mM GA, 0.0025mM UN, 1x manufacturer's reaction buffer (Qiagen), 0.875 units of Taq polymerase (hot start) and 10 ng DNA of the respective template. PCR reactions were incubated at 94ºC for 15 minutes followed by 30 cycles of 94ºC for 30 seconds, 50ºC for 30 seconds, 72ºC for 30 seconds with a final extension step at 72ºC for 10 minutes. Ten microliters of PCR products were run on 2% agarose gel and visualized by ethidium bromide staining. The sizes of bands were estimated based on a 100 bp ladder (Bioline).

2.2 cDNA Synthesis

Total RNA was extracted from a batch of four individual one-day old adult mosquitoes using TRI reagent (Sigma) according to the manufacturer's instructions and described by [24]. The sample was homogenized in 400 µl of TRI reagent (SIGMA) and centrifuged at 12,000 g for 10 min at 4^{oc}. The supernatant was transferred to a new centrifuge tube, homogenized for 15 sec with chloroform and was incubated at 4^{oc}. Two hundred microliter of molecular grade isopropanol was added to precipitate the RNA. The RNA was treated with RQ1 DNase (Promega) to degrade contaminating genomic DNA. Reverse transcription of mRNA to cDNA was achieved using superscript T^{m111} (Invitrogen) and an oligo (dT) adapter primer (5′GACTCGAGTCGACATCGA(dT)17-3'). One microgram of RNA was mixed with 1 µg of the primer and heated to 65^{oc} for five minutes. The reaction mixture was chilled on ice for 1 minute and was then pre-wormed to 50^{oc} for 2 minutes after adding 8 µl of 5x first strand buffer (250Mm Tris-HCL Ph 8.3, 375Mm KCL, 15 Mm $Mgcl_{2)}$, 2 µl of 10 mM of each dNTP and 2 µl of 0.1M DTT. Superscriptase T^{m111}Rnase H $^{-}$ reverse transcriptase was added, reaction incubated at 50^{oc} for 90 minutes and then heated to 70^{oc} for 15 minutes to inactivate the enzyme. The cDNA sample was stored at -20^{oc}.

2.3 PCR amplification of *An. arabiensis* GST genes

Primer pair sequences designed for the amplification of individual members of the *An. gambiae* Epsilon class GST genes were used to amplify full-length orthologous genes from *An. arabiensis* [25] (Table 1). Each gene was amplified using the gDNA and cDNA as templates. The 25-µl reaction contained 1.5 mM dNTPs, 0.5 µl of each primer, 1.25 units of Taq DNA polymerase (Hotstar) and 1X reaction buffer (Qiagen) (50 mM Kcl, 10 mM Tris-Hcl(PH 9.0 at 25^{0C}) 0.1% Triton X-100). Annealing temperatures were determined empirically for each gene using gradient PCR machine PTC-200 (MJ Research). The reaction was incubated at 95^{0C} for 15 minutes, followed by thirty cycles of amplification (95^{0C} for 30 secs, 50-60^{0C} for 30 secs, 72^{0C} for 30 secs and 72^{0C} for 10 minutes). The PCR products were separated on 2% (W/V) agarose gel and visualized using the Gene Genius Bio Imaging System (Syngene).

2.4 Sequencing

PCR products of the expected size were sub cloned into P^{GEM}-T easy vector (Promega, Southampton, UK) and after transformation into Escherichia coli JM 109 competent cells (Invitrogen), plasmids were extracted and purified using Qiagen miniprep kits (Qiagen, Crawley, UK). At least three independent clones were sequenced for each gene. Sequencing reactions were performed using Beckman chemistry analyser and the resultant products analyzed on a Beckman CEQ 800 capillary sequencer. Analysis of the sequences was done using the DNA Star package software (Lasergene). Vector regions were trimmed and sequences assembled using CodonCode Aligner 2.0.4 (default assembly criteria: 70% minimum percent identity and 25 bp minimum overlap length). Identity of the sequences with related sequences in the database of the National Institute of Health, USA (NIH) was obtained by searching using the Blast (http//www.ncbi.nih.org/BLAST)

program.

2.5 Phylogenetic analysis
FASTA files and predicted amino acid sequences of GSTE *Aedes aegypti*, and *An. gambiae* were downloaded from VectorBase (http// www. Vectorbase.org/index.php). GSTE sequences from Drosophila melanogaster were downloaded from FlyBase (http://flybase.org./blast/). Putative amino acid sequences were aligned against the *An. arabiensis* GST sequences using the Clustal W algorithm (gap extension penalty : 1; gap initiation penalty: 3) in BioEdit [26, 27]. After the alignment, evolutionary distances were calculated using the Jukes-Cantor algorithm. Phylogenetic trees were determined by the neighbor-joining method with TREECON software.

3. Results
3.1 Identification of *An.arabiensis* GST genes genes
We amplified and characterized eight GSTE genes including GSTE1, GSTE2, GSTE3, GSTE4, GSTE5, GSTE6, GSTE7 and GSTE8 in *An. arabiensis.* The sizes of PCR products obtained for both genomic and complementary DNA of each of the eight *An. arabiensis* GSTEs are shown (Table 2). GSTEs sequences from *An. arabiensis* showed strong similarity to those of *An. gambiae.* GSTs gene yielded sequences with high percentage identity when compared to *An. gambiae* GSTs. The percentage similarity at amino acid level between the eight *An. arabiensis* GST proteins and their equivalents in *An. gambiae*, ranged from 91% (between arGSTE7 and agGSTE7 to 98.7% (between arGSTE2 and agGSTE2) (Table 2.). GSTE2 – GSTE6 are the most conserved genes between the two species. An alignment of the deduced amino acid sequences of the eight genes identified in this study is shown (Figure 1).

The invariant residues indicated in bold letters are conserved in all the eight proteins and the underlined motifs, which characterised Epsilon class GSTs, are conserved in these and all previously identified orthologous genes in *An. gambiae*. [13].

3.2 Intron positions and sizes
The cDNA sequences of *An. arabiensis* GSTs were aligned with the genomic DNA sequences and the positions and sizes of introns were noted. The coding sequence of all the eight genes is interrupted by an intron or two introns in *An. arabiensis*. The first intron is formed at 48 amino acid residues from the N- terminal in all the eight Epsilon *An. arabiensis* GST genes (Figure 2). Intron sizes were small ranging from 50bp in GSTe8 to 90bp in GSTe6 and GSTe2. The number of introns in each gene is the same as that found in its orthologous counterpart in *An. gambiae*, but there are some slight variations in the sizes (Table 2). The intron –exon boundaries of the genomic sequences, illustrated by GSTe2, are consistent with the canonical GT/AG rule (Figure 2). The positioning of these introns within the coding sequences is highly conserved when compared between *An. gambiae* and *An. arabiensis.*

3.3 Polymorphisms in *An. arabiensis* GSTs
The sequences obtained from the various clones which were derived from each of the *An. arabiensis* GSTe1 – GSTe8, were arranged into respective contigs using the Seqman programme DNA Star package. The DNASP software was used to identify polymorphic sites in the nucleotide and derived amino acid sequences of each gene. Nucleotide ambiguities resulting from sequencing of multiple clones were observed with genomic and cDNA sequences. The number of polymorphic amino acid residues recorded in the *An. arabiensis* GST gene cluster ranged from 3 to 15 (Table 3). Of the eight GST genes, GSTe1 has the highest number of nucleotide substitutions (23) out of which 12 are non-silent mutations, resulting in changes in the amino acid residues. Two polymorphic sites were identified in each of the putative amino acid sequences of *An. arabiensis* GSTe1 and GSTe2 (Table 3).

3.4 Phylogenetics analysis
To examine the relationship between *An. gambiae* and *An. arabiensis* Epsilon class GST genes, multiple sequence alignment of gene clusters from species was carried out prior to molecular Phylogenetics analysis. The *An. gambiae* Epsilon class GST gene sequences were retrieved from the Genebank. Deduced amino acid sequences of the 16 GSTs, eight each from *An. arabiensis* and *An. gambiae*, were aligned using Clustal W and a phylogenetic tree was constructed (Figure 3). The dendogram illustrates the close phylogenetic relationship between the gene families from the two species. Each gene in the cluster is tightly associated with its ortholog.

To further investigate the relationship between *An. arabiensis* GSTs and those from other insects, all the known insect-specific Epsilon GST sequences from *Drosophila melanogester* and *Ae. aegypti*, in addition to those of *An. gambiae*, were retrieved from Genebank and aligned using Clustal W. The alignment, consisting of all the *An. arabiensis* GST subunits and 25 putative Epsilon class GSTs from the three species, were used to construct phylogenetic tree by distance neighbour-joining method. A representative distance tree generated from the alignment is shown (Figure 4).

4. Discussion
4.1 Identification of *An. arabiensis* GST
Given the important role of GSTs in the detoxification of environmental xenobiotics [15], it was postulated that different members of *An. gambiae* complex, which differed in their ecological adaptations may differ in their pattern of GST expressions. In this study the sequence and genomic organisation of the insect specific GST epsilon (GSTE) cluster of *An. arabiensis* were characterized and compared to those of *An. gambiae* and orthologous genes in *Ae. aegypti* and *Drosophila melanogaster*. Eight GSTE genes were identified from *An. arabiensis* MAT the same number as in *An. gambiae*. The *An. arabiensis* GST genes aligned closely with the Epsilon class GST genes, which have been identified in *An. gambiae* [13, 28]. Like in *An. gambiae*, the same number of GSTE genes was recognized but only four putative orthologs were identifiable between the two species [11]. In comparison, six out of the eight episilon GST genes have been identified in the malaria vectors: *An. stephensi*, *An. plumbeus* and *An. funestus* [12]. [12] have not attempted to amplify GSTE8 and were unsuccessful in amplifying GSTE3 from any of the three

malaria vector species. Notably, with the exception of GSTE1 which is absent in *An. plumbeus*, all the GSTEs in are consistent in order and orientation with those in *An. gambiae* [12, 29]. Both An. stephensis and *An. funestus* are phylogentically more related to *An. gambiae* (Cellia subgenus) compared to *An. plumbeus* [30]. The motifs which are characteristics of Epsilon class GSTs, including the semi-conserved motifs shared by other GST classes, are wholly conserved in all the eight *An. arabiensis* GST genes. The high percentage similarity ranging from 91.2% (between ArGSTE7 and AgGSTE7) to 98.7% (between ArGSTE2 and AgGSTE2) at amino acid level, which was observed between *An. arabiensis* GSTs and the identified Epsilon GSTs in *An. gambiae*, is by far above the 40% cut-point criteria which was used for classifying a gene to a particular class [31]. This strongly suggests the inclusion of the *An. arabiensis* GST into the Epsilon class. Likewise, strong similarity of GSTEs sequences from *An. stephensi*, *An. plumbeus* and *An. funestus* to those of *An. gambiae* has also been shown [12].

Here we show that GSTE2 sequence has the highest similarity between the two species which is consistent with [12] who on comparative genomics found no indel in GSTE2 in any of the four malaria vectors mentioned above. It is interesting to note that GSTE2 has been shown by QTL mapping to be associated with resistance to DDT in *An. gambiae* and that AgGSTE2 displayed the highest DDT dehydrochlorinase activity [9, 10]. Recently, we also have reported overexpression of ArGSTE2 in developmental stages of *An. arabiensis* selected for DDT resistance [24]. This may indicate that GSTE2 plays a in particular plays a critical in adaptation of the Anopheline species to selection pressure due to xenobiotic compounds encountered in their various habitats. The introns found in *An. arabiensis* GSTs are similar in respect of size, number and positions to those reported in their corresponding orthologous genes in *An. gambiae*, *An. stephensi*, *An. plumbeus* and *An. funestus* [12, 25]. The considerable conservation of introns in the Epsilon GST gene clusters between the two sibling species is indicative of their common ancestry and supports the evidence shown in a molecular phylogenetic study of the *Anopheles gambiae* complex [32, 33]

4.2 Polymorhisms in *An. arabiensis* GSTs

At chromosomal level, inversion polymorphisms have been identified as the genetic basis for the biological traits characterising *An. gambiae* complex as the most efficient malaria vector system in the world [34]. *An. arabiensis* is polymorphic for inversions a and c on chromosome 2R, which were shown to be associated with non-uniform, indoor resting and feeding behaviour, and this was interpreted to mean decreased chances of exposure to indoor insecticide-treated surfaces [18]. [35] reported an association of DDT resistance with 2Rb inversion, which was explained to be presumably due to presence of a resistance gene in the inverted chromosome. In this study, polymorphisms in the GST sequences were observed within the *An. arabiensis* MAT and between it and the Kisumu and Zanu strains of *An. gambiae*. The majority of the substitutions within *An. arabiensis* strains are silent mutations and, therefore, this is assumed to reflect allelic variations that are maintained in the colony. However, a more detailed analysis of the *An. arabiensis* GST gene sequences is required to determine whether the mutations are attributable to alleilic variation or may be due to recent gene duplication, as was suggested in *An. gambiae* Cyp4 P450 genes [28]. Nevertheless, out of the eight *An. arabiensis* GSTs, GSTe1 is observed to be the most polymorphic. The deduced amino acid and nucleotide sequences of this gene were compared to those of GSTe1K and GSTe1Z, the two highly polymorphic variants of the orthologous gene in *An. gambiae* [13]. Surprisingly, the proteins from the *An. gambiae* differed between themselves by 12 amino acids, while the *An. arabiensis* GSTe1 protein differed by only 7 amino acids with the Kisumu GSTe1K variant. This result supports the finding that the Kisumu variant, unlike a rare allele, exists in equal frequencies in natural population and may confer some fitness advantage in field conditions [13].

4.3 Phylogenetic analysis

Taxonomically, *An. arabiensis* belongs to the *An. gambiae* complex, a group of sibling species which are morphologically indistinguishable but can be separated individually by PCR [20, 21]. A preliminary phylogenetic analysis was conducted to establish the relationship between the GSTs isolated from *An. arabiensis* and all the known Epsilon class GSTs in *An. gambiae*. At this stage the full extent of the GST gene family and their organisation in the genome of *An. arabiensis* is not known. The initial analysis shows the two gene families forming a tight cluster with orthologous genes from both species pairing with one another. The dendogram supports evidence of a close relationship between these two species, as was reported [33]. A further assessment of the phylogenetic relationships of *An. arabiensis* GSTs with other insect GSTs support this view. The topology of the tree in Figure 4, clearly separates the Culidea GST gene lineage from the Drosophilidea GSTs as indicated by the respective bootstrap values. Similarly, within the mosquito GST gene families, the Anopheline GSTs are paired and separated from their evolutionary more distant Aedes counterparts [37]. There are no clear orthologs between these mosquito GSTs and the Drosophilia lineage supporting the earlier suggestion that the Epsilon class GSTs from these dipteran families have radiated independently [28, 36]. The close orthology of *An. arabiensis* GST gene with those of *An. gambiae* might indicate similar *in vivo* biological functions of the individual genes as patterns of orthology was believed to provide clues to physiological functions [27].

5. Tables and figures

Table 1: Primer sequences used for amplification of *An. arabiensis* Epsilon Class GSTs

Gene	Forward Primer (5´ to 3´)	Reverse Primer (5´ to 3´)	Expected band size (bp)
GSTe1	ATC CCG AAA CCT GTG CTA TA	CGC TCT TGC AGC GAG TCA CTT	817
GSTe2	ATC ACC GAG AGC CAC GCA ATC A	GCC ACC	830

GSTe3	ATG GCA CCG ATT GTG TTG TAC	CTA AAC CTT GCC AAG TTT CTG G	741
GSTe4	CGC CAT TCA AAC GAC CAT GCC	CTC AAC CAG ATC AGG TTC AC	743
GSTe5	ATG GCA ACG AAC CCC ATC ATC	CCT TGG AAT CGT GGC TTT AC	76
GSTe6	GTA AGC TCT ACA GCT CGT TCG GC	GCG AAG ATC AAC CGG TAG TGG T	5
GSTe7	CGA GAT TGG TAC TGT ACA CG	G	7
GSTe8	TTA GTA CCG GCC TGT ACT ACG AC	CTC GGA TAG AGA CCG TCG TC	74

Table 2: Comparison of Epsilon class *An. arabiensis* GSTs with orthologous genes in *An. gambiae*

Gene	Product size (bp)		Percentage identity of amino acid sequence with *An. gambiae*	Number of introns (intron size bp) *An. arabiensis*	Number of introns (intron size bp) *An. gambiae*
	cDNA	Genomic			
GSTe1	675	821	94.7	2 (64 and 75)	2 (64 and 78)
GSTe2	624	831	98.7	2 (74 and 90)	2 (74 and 90)
GSTe3	745	750	98.6	1 (69)	1 (69)
GSTe4	693	760	98.6	1 (65)	1 (65)
GSTe5	709	767	97.0	1 (78)	1 (72)
GSTe6	692	776	97.8	1 (90)	1 (91)
GSTe7	681	865	91.2	2 (80 and 60)	2 (69 and 66)
GSTe8	658	757	97.7	2 (50 and 66)	2 (66 and 60)
GSTe1	675	821	94.7	2 (64 and 75)	2 (64 and 78)

The table shows the PCR products sizes of both the genomic and cDNAs from each member of *An. arabiensis* GSTs as well as the number and sizes of introns identified in individual genes. The number and sizes of introns in the orthologous genes identified in *An. gambiae* [24] is indicated in bold.

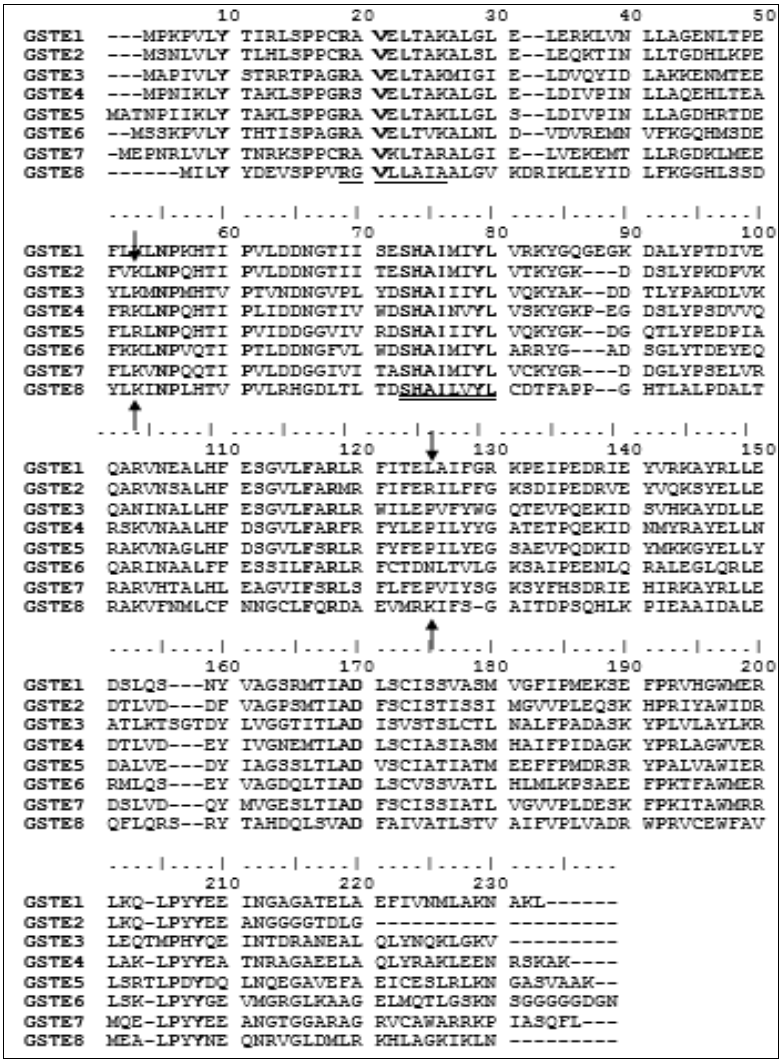

Fig 1: Clustal W alignment of deduced amino acid sequences of *An. arabiensis* GST gene cluster. Residues in bold are conserved in all eight proteins. The two underlined residues are motifs characteristic of Epsilon GSTS (single underlined) or most GST classes (double underlined).The arrows indicate intron positions.

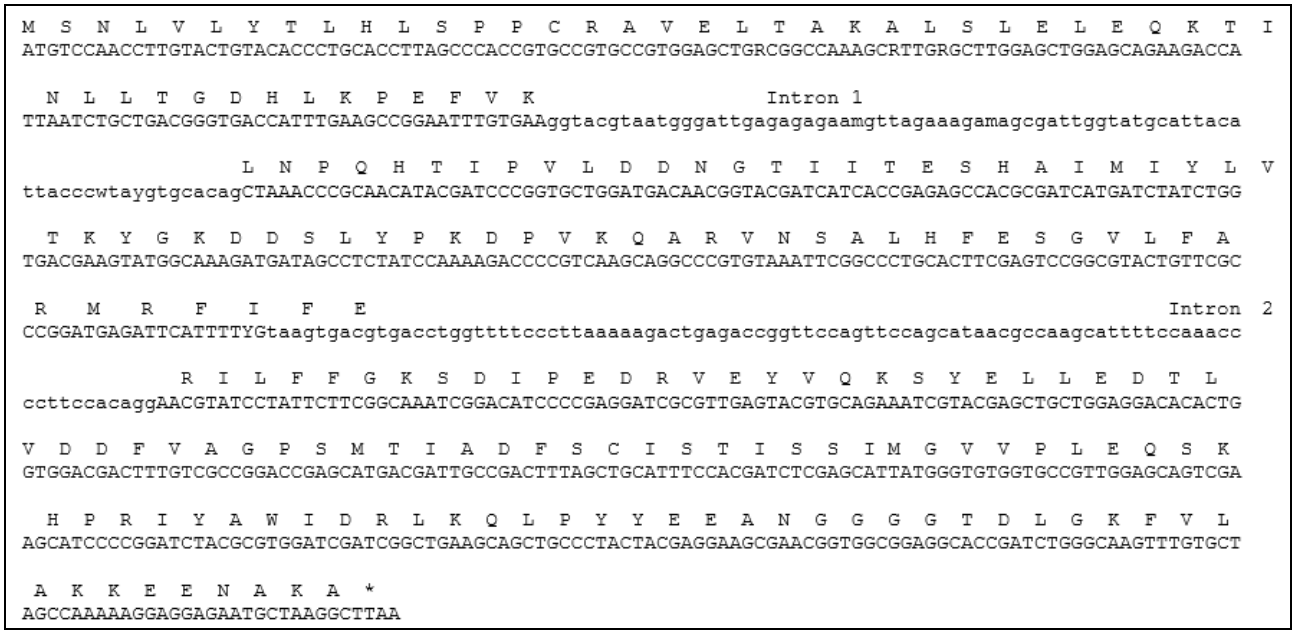

Fig 2: Genomic and deduced amino acid sequences of *An. arabiensis GSTe2* isolated by PCR from MAT strain. The positions of introns are indicated in lower case letters. Intron 1 found 48 amino acid residues from the N – terminal is conserved compared to *An. gambiae*.

Table 3: Comparison of polymorphic sites within and between *An. arabiensis* and *An. gambiae* GSTs

An. arab vs An.gam(Zan)			An.arab vs An.gam.(Kis)		Kis vs Zan		*An arabiensis*	
Genes	**A.acid**	**Nucl**	**A. acid**	**Nucl**	**A. acid**	**Nucl**	**A.acid**	**Nucl**
GSTe1	12	19	7	12	12	23	2	2
GSTe2	2	8	na	na	na	na	2	5
GSTe3	2	3	0	0	2	2		
GSTe4	5	7	1	2	6	8	1	3
GSTe5	5	12	5	18	0	13	1	3
GSTe6	3	6	15	10	0	0	-	-
GSTe7	17	18	15	10	4	10	-	-
GSTe8	4	10	na	na	na	--		-

Analysis of polymorphic sites in the deduced amino acid (A.acid) and cDNA (Nucl) sequences of Epsilon class *GSTe1 – GSTe8* gene clusters from *An. arabiensis* and *An. gambiae.* Consensus sequences from 3 – 5 clones for each of the *An. arabiensis* GSTs and sequences for orthologous genes from *An. gambiae* Zanu (Zan) and Kisumu (Kis) strains (retrieved from the database) (Genebank) were used for the analysis (na = no sequence available).

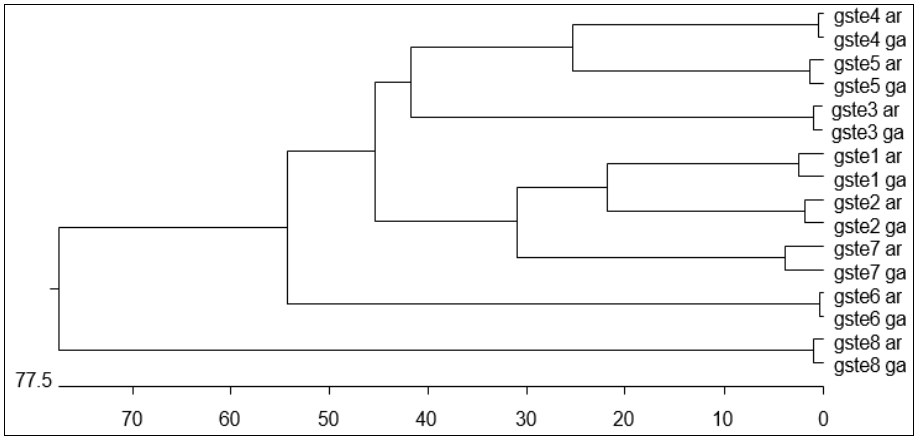

Fig 3: Dendogram illustrating the relationship between *An. arabiensis* and *An. gambiae* Epsilon class GSTs. Amino acid sequences were aligned using Clustal W and the tree was constructed with the neighbour – joining method program from a similarity matrix of pairwise comparisons made by using the jukes – cantor algorithm. The program assumes a constant evolutionary rate, the bar indicates the frequency at which 0.1 amino acid substitution occur at any given site.

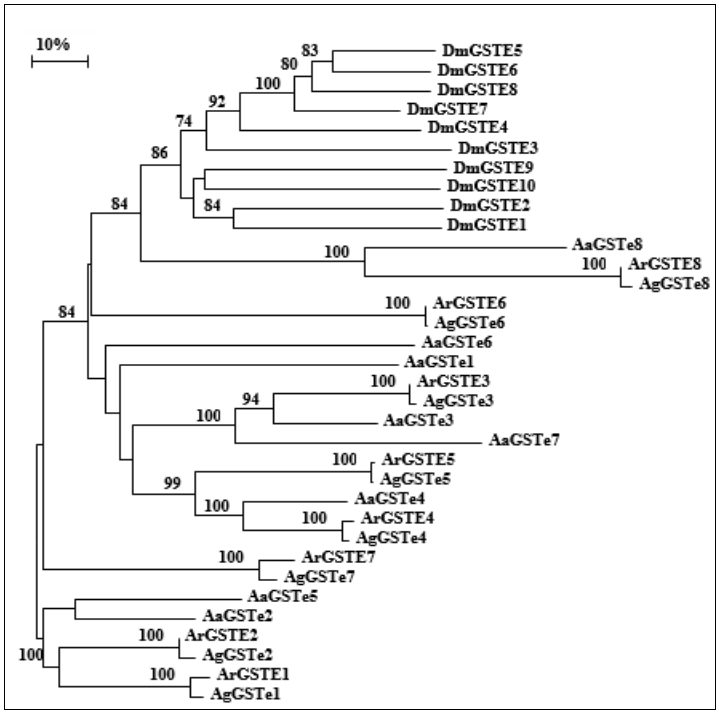

Fig 4: Phylogenetic relationships of Epsilon GST classes in insects (Dm = *Drosophila melanogaster*, Aa = *Aedes aegypti*, Ag = *An. gambiae*, Ar = *An. arabiensis*). Distance neighbour – joining tree illustrating the relationship between *An. arabiensis*, *An. gambiae*, *Aedes aegypti* and *D. melanogaster* Epsilon class GSTs. Clustal W was used to align the putative amino acid sequences of 8 Epsilon class GST subunits in *An. arabiensis* with 25 putative Epsilon class GSTs from *D. melanogaster*, *Ae. Aegypt*i and *An. gambiae*. The tree was constructed by the neighbour – joining method from a similarity matrix of pairwise comparisons made using the jukes – Cantor algorithm. Nodes with distance bootstrap values (500 replicates) of > 70% are shown.

6. Conclusion

This work describes the cloning and identification of eight An. arabiensis GSTs. Comparative sequence and phylogenetic analysis with insect GSTs in the database suggest placement of the An. arabiensis GSTs into the insect Epsilon class of GSTs. However, the full extent of the GST gene family in An. arabiensis can be explored only when the genome sequence becomes available, but the biochemical and immunological properties of the recombinant proteins of some of the genes, particularly ArGSTe1 and Ar GSTe2 identified here, can be studied further to confirm the classification.

7. Acknowledgement

This work was supported by Federal Government of Nigeria through the special scholarship award and supplementary funding from Mc. Auther Foundation, Bayero University Kano, Nigeria.

8. References

1. 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; 360:1-16.
2. Raza H. Dual localization of glutathione – S – transferases in the cytosol and mitochondria:implications in oxidative stress, toxicity and disease. The FEBS Journal, 2011. doc.10.1111/1.1742
3. Che-Mandoza A, Patricia Penilla R, Americo RD. Insecticide resistance and Glutathione S- transferases in mosquitoes: A review. African Journal of Biotechnology. 2009; 8(8):1386-1397.
4. Ranson H, Hemingway J. Mosquito glutathione transferases Review Methods Enzymol. 2005; 401:226-241.
5. Hemingway J, Miyato J, Herath PRJ. A possible novel link between organophosphorus and DDT Insecticide resistance genes in Anopheles: supporting evidence from fenitrothion metabolism studies. Pesticide Biochemistry and Physiology. 1991; 39:49-56
6. Lumjuan N, McCarrol L, Prapanthadara L, Hemingway J, Ranson H. Elevated activity of an Epsilon class glutathione transferase confers DDT resistance in the dengue vector, Aedes aegypti. Insect Biochem. Mol. Biol. 2005; 35:861.
7. Nardini L, Christian RN, Coetzee M, Koekemoer LL. DDT and Pyrethroid resistance in Anopheles arabiensis from South Africa. Parasites and Vectors. 2013; 6:229.
8. Ranson H, Rossiter L, Ortelli F, Jensen B, Wang X, Roth CW, Collins FH, Hemingway J. Identification of a novel class of insect glutathione S-transferases involvedin DDT resistance in the malaria vector, An. gambiae. Biochem. J, 2001; 359:295-304.
9. Ranson H, Jensen B, Wang X, Hemingway J, Collins FH. Genetic mapping of two loci affecting DDT resistance in the malaria vector Anopheles gambiae. Insect Mol. Biol. 2000b; 9:499-507.
10. Lumjuan N, Stevenson BJ, Prapanthadara L.A, Somboon

P, Brophy PM, *et al*. The Aedes aegypti glutathione transferase family. Insect Biochem Mol Biol. 2007; 37:1026 -1035.
11. Ayres C, Muller P, Dyer N, Wilding C, Rigden D *et al*.). Comparartive Genomics of the Anopheline Glutathione S –Transferase Epsilon Cluster. PLoS ONE. 2011; 6(12):e29237.doi: 10.1371/ journal. Pone. 0029237.
12. Ortelli F, Rossiter CL, Vontas J, Ranson, Hemingway J. Heterologous expression of four glutathione stransferase genes genetically linked to a major insecticide-resistancelocus from the malaria vector Anopheles gambiae. Biochem J. 2003; 373:357-963.
13. David JP, Coissac E, Melodelima C, *et al*. Transcriptome response to pollutants and insecticides in the dengue vector Aedes aegypti using next generation sequencing technology: BMC Genomics. 2010; 11:216.
14. Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, *et al*. A deficit of detoxification enzymes pesticide sensitivity and environmental response in Honey bee. Insect Mol. Biol. 2006; 5:615-636.
15. Coetzee M, Craig M, le Suer D. Distribution of African malaria mosquitoes belonging to the Anopheles gambiae complex. Parasit Today 2000, 16:74-77.
16. Abdullah H, Craig SW, Nardini L, Pignatelli P, Koekemoer LL, Ranson H, Coetzee M. Insecticides resistance in Anopheles arabiensis in Sudan:Temporal trends and underlying mechanisms. Parasites & Vectors, 2014; 7:213. http:/www. parasites and vectors.com
17. Cassone BJ, Kamdem C, Cheng C, John CT, Mathew W. Gene expression divergence between malaria vectors sibling species Anopheles gambiae and Anopheles coluzzii from rural and urban Yaounde Cameroun. Mal. Ecol. 2014; 23(9):2242-2259.
18. Colluzzi M, Sabatini A, Petrarca V, Di Deco MA. Chromosomal differentiation and adaptations to human environments in the Anopheles gambiae complex. Trans. Royal Soc. Trop Med. and Hyg. 1979; 73:483-497.
19. WHO. Test procedures for insecticide resistance monitoring in malaria vectors, Bio-efficacy and persistence of insecticides on treated surfaces. World Health Organization, Geneva, 2013.
20. Gillies MT, Coetzee M. A supplement to the Anophelinae of Africa south of Sahara. Publication of the South African Institute for Medical Research, Johannesburg. 1987; 55:143.
21. Scott JA, Brogdon WG, Collins FH. Identification of single specimens of the Anopheles gambiae complex by the polymerase chain reaction. American Journal Trop. Med. Hyg. 1993; 49:520-529.
22. Collins FH, Mencez MA, Rasmussen MO, Mehaffey PC, Besankey NJ, Finnerty V. A ribosomal RNA gene probe differentiates member species of the An. gambiae complex. American J.Trop. Med. and Hyg. 1987; 37:37-41.
23. Yayo AM, Ado A, Habibu UA, Muhammad BR, Ebere N, Hemingway J. Expression patterns of epsilon glutathione S – transferases genes in developmental stages of susceptible and DDT resistant lines of Anopheles arabiensis strains. International Journal of Entomology Research. 2018; 3(2):143-151.
24. Ding Y, Ortelli F, Rossiter L, Hemingway J, Ranson H. The Anopheles gambiae glutathione transferase family: annotation, phylogeny and expression profiles. BMC, Genomics. 2003; 4 (1):35.
25. Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucl Acids Symp. 1999; 41:95-98.
26. Thompson JD, Higgins DG, Gibson TJ Clustal W improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acid Research. 1994; 22(22):4673-80
27. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, *et al*. Evolution of supergene families associated with insecticide resistance. Science. 2002; 298:179-181.
28. Sharakhov IV, Serazin AC, Grushko OG, Dana A, Lobo N, *et al*. Inversion and gene order shuffling in Anopheles gambiae and Anopheles funestus. Science. 2002; 289:182-185.
29. Harbach RE. The classification of the genus Anopheles (Diptera: Culicidae) a working hypothesis of phylogenetic relationships. Bull Entomol Research. 2004; 94:537-553.
30. Chelvanayagam G, Parker MW, oard PG. Fly fishing for GSTs (2001). A unified nomenclature for mammalian and insect glutathione transferases. Chem. Bio. Interact. 2004; 133:256-260.
31. Besansky NJ, Powell JR, Caccone A, Hamm AM, Scot JA, Collins FH. Molecular phylogeny of the Anopheles gambiae complex suggests genetic introgression between principal malaria vectors. Proceedings of the National Academy of Sciences of the United States of America. 1994; 91:6885-688.
32. Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey IV, Sharakov X, *et al*. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science, 2014, doi: 10.1126/ Science. 1258524.
33. Colluzzi M. Advances in the studies of Afro tropical malaria vectors. Parassitologia, 1992; 35:23-29.
34. Nigatu W, Curtis CF, Lulu M. Test for association of DDT resistance with inversion polymorphism in Anopheles arabiensis from Ehtiopia. Journal of American Mosquito Control Association. 1995; 11(2):238-240.
35. Krzywinski J, Wilkerson RC, Besansky NJ. Toward Understanding Anophelinae (Diptera, Culicidae) phylogeny: Insights from nuclear single-copy genes and the weight of evidence. Systemtic Biology. 2001; 50(4):540-556
36. Wilding CS, Weetman D, Rippon EJ, Steen K, Mawejje HD, Bursukov I, Donelly MJ. Parellel evolution or purifying selection, not introgression, explains similarity in the Pyrethroid detoxification linked GSTE4 of Anopheles gambiae and An. arabiensis. Molecular Genet Genomic Doi: 2014; 10.1007/500.