BMC Genomics

BioMed Central

Research article



# The *Anopheles gambiae* glutathione transferase supergene family: annotation, phylogeny and expression profiles

Yunchuan Ding[†1], Federica Ortelli[†1,2], Louise C Rossiter[1,3], Janet Hemingway[1] and Hilary Ranson[*1]

Address: [1]Vector Research Group, Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK, [2]Present Address: Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Università degli Studi di Perugia, Perugia, 06122 Italy and [3]Present Address: Australian Cotton Research Institute, Narrabri NSW 2390 Australia

Email: Yunchuan Ding - yding@liv.ac.uk; Federica Ortelli - acpg@unipg.it; Louise C Rossiter - LouiseR@mv.pi.csiro.au; Janet Hemingway - hemingway@liv.ac.uk; Hilary Ranson* - hranson@liv.ac.uk

* Corresponding author †Equal contributors

Published: 13 August 2003

*BMC Genomics* 2003, **4**:35

Received: 11 June 2003
Accepted: 13 August 2003

This article is available from: http://www.biomedcentral.com/1471-2164/4/35



## Abstract

**Background:** Twenty-eight genes putatively encoding cytosolic glutathione transferases have been identified in the *Anopheles gambiae* genome. We manually annotated these genes and then confirmed the annotation by sequencing of *A. gambiae* cDNAs. Phylogenetic analysis with the 37 putative GST genes from *Drosophila* and representative GSTs from other taxa was undertaken to develop a nomenclature for insect GSTs. The epsilon class of insect GSTs has previously been implicated in conferring insecticide resistance in several insect species. We compared the expression level of all members of this GST class in two strains of *A. gambiae* to determine whether epsilon GST expression is correlated with insecticide resistance status.

**Results:** Two *A. gambiae* GSTs are alternatively spliced resulting in a maximum number of 32 transcripts encoding cytosolic GSTs. We detected cDNAs for 31 of these in adult mosquitoes. There are at least six different classes of GSTs in insects but 20 of the *A. gambiae* GSTs belong to the two insect specific classes, delta and epsilon. Members of these two GST classes are clustered on chromosome arms 2L and 3R respectively. Two members of the GST supergene family are intronless. Amongst the remainder, there are 13 unique introns positions but within the epsilon and delta class, there is considerable conservation of intron positions. Five of the eight epsilon GSTs are overexpressed in a DDT resistant strain of *A. gambiae*.

**Conclusions:** The GST supergene family in *A. gambiae* is extensive and regulation of transcription of these genes is complex. Expression profiling of the epsilon class supports earlier predictions that this class is important in conferring insecticide resistance.

## Background

Glutathione transferases (GSTs) are a diverse family of dimeric proteins found in almost all living organisms. Originally studied for their role in detoxification of endogenous and xenobiotic compounds, they have since been found to have additional important roles as transport proteins and in protection against oxidative stress [1]. Each GST subunit consists of two domains, each containing two binding sites, the G site and the H site. The highly conserved G site binds the tripeptide glutathione and is

largely composed of amino acid residues found in the N-terminal domain. The H-site or substrate binding site is more variable in structure and is largely formed from residues at the C-terminal [2].

Purification of independent homogenous GST preparations with differing substrate specificities indicated the presence of multiple forms of GSTs [3]. Subsequently, the availability of N-terminal sequence data led to the recognition of five classes of cytosolic GSTs in mammals, the alpha, mu, pi, theta and sigma classes [2,4,5] and an additional, structurally unrelated membrane bound microsomal class [6]. Recently the advent of large scale EST and full genome sequencing projects has led to a marked increase in the number of GST classes recognized. Some of these, such as the omega and zeta classes are represented in a wide range of species [7,8], whereas others, such as the mammalian kappa class [9], the insect epsilon class [10] and the plant tau and phi clases [11] have a more restricted distribution.

Most of these GST classes are encoded by multigene families. Alternative splicing [12,13] and the formation of heterodimers [14], can add a further level of heterogeneity to this enzyme family. With this level of diversity, assigning physiological functions to individual GSTs is a complex task, but progress towards this goal can be greatly facilitated by the process of cataloguing the number of genes within the supergene family. Armed with this information, details of expression profiles, induction mechanisms, tissue distribution etc. can be accurately obtained enabling biologically important questions to be addressed. Automatic annotation algorithms applied to assembled eukaryotic genomes provide projections of the sizes of gene families within a species. Using these tools the numbers of GST genes is estimated at 10 in *Saccharomyces cerevisiae*, 57 in *Caenorhabditis elegans*, 43 in *Drosophila melanogaster*, 37 in *Anopheles gambiae*, 46 in *Arabidopsis thaliana* and 40 in *Homo sapiens* [15]. Careful manual annotation is essential to confirm these predicted numbers. This process has led to revised sizes of the GST supergene family in *A. thaliana* and *A. gambiae* to 48 and 31 respectively [11,16]. (This gene count in *A. gambiae* includes three genes encoding putative microsomal GSTs but these will not be discussed further in this report).

To facilitate the functional characterization of insect GSTs, we have studied the annotation of each member of this supergene family in the mosquito *A. gambiae*. RT-PCR experiments demonstrate that all but one of the predicted GST genes are actively transcribed in adult mosquitoes and that alternative splicing of two GST genes contributes additional diversity. We compare this GST supergene family with that of a second Diptera, *D. melanogaster* and identify classes of GSTs that are conserved between the species and other classes that have undergone independent radiation.

The majority of studies on insect GSTs have focused on their role in conferring insecticide resistance (e.g. [17–19]) and, more recently, in protecting against cellular damage by oxidative stress [20,21]. Reactive oxygen species can be produced in response to infection by pathogens, and this phenomenon has been implicated in the defense mechanism of mosquitoes against malaria parasites [22]. *A. gambiae*, is therefore an ideal species in which to study the role of GSTs in both of these biological processes. This species is the major vector of malaria in Africa and as such is responsible for over 1 million deaths each year [23]. Efforts to control the disease by targeting the mosquito populations have relied on treatment of *Anopheles* breeding and resting sites with insecticides. The organochlorine DDT was the insecticide of choice for malaria control for much of the latter half of the 20th century and is still employed in public health campaigns today but, in many malarious regions, resistance has rendered DDT-based control programmes ineffectual. In *A. gambiae*, DDT resistance is associated with increased GST activity [24]. Genetic mapping using microsatellite markers has located two loci associated with DDT resistance in *A. gambiae* [25]. By aligning the cytogenetic position of these resistance loci with the *in situ* position of physically mapped GST genes we previously identified two candidate resistance-associated GSTs [10]. Analysis of the draft genome sequence of *A. gambiae* identified a further six GST genes within this region of the genome [16]. We now report that the expression of multiple members of this gene cluster is elevated in DDT resistant insects.

## *Results and Discussion*

### *Classification of* A. gambiae *GSTs*

Twenty eight genes putatively encoding cytosolic GSTs were identified in the *A. gambiae* genome (Table 1). Nomenclature guidelines originally proposed for vertebrate GSTs and later expanded to incorporate invertebrate classes [26], were employed when classifying the *A. gambiae* genes. This led to the renaming of nine genes previously described in the literature (changes are shown in Table 1).

The putative amino acid sequences of the *A. gambiae* GSTs were aligned using ClustalW [27] and GST-1 from *C. elegans* (Accession number CAA78471) as an outgroup (Figure 1). The alignment was manually truncated by the removal of twenty positions at both the N-terminal and C-terminal that contained excessive gaps and two small internal regions of poor alignment, one consisting of six residues in the linker region between the N and C domains and the second consisting of 21 residues in the C-terminal domain of which over 75% were gaps in the

**Table 1: Summary of the *A. gambiae* GST family. The length of the *GSTd6* gene and putative translation are not known (N.K.).**

| Gene name | Old name | No. of transcripts detected | Length of putative protein(s) | Length of gene (bp) | Genebank Accession number |
|---|---|---|---|---|---|
| *GSTd1* | GST1-1 | 4 | 219, 210, 217, 210 | 3780 | AF071160 |
| *GSTd2* | GST1-2 | 1 | 209 | 630 | Z71480 |
| *GSTd3* | | 1 | 218 | 701 | AF513638 |
| *GSTd4* | | 1 | 224 | 703 | AF513635 |
| *GSTd5* | | 1 | 216 | 720 | AF513634 |
| *GSTd6* | | 1 | N.K. | N.K. | AF513636 |
| *GSTd7* | GST1-7 | 1 | 218 | 2464 | AF071161 |
| *GSTd8* | GST1-8 | 1 | 211 | 749 | AF316637 |
| *GSTd9* | GST1-9 | 0 | 216 | 651 | AY255857 |
| *GSTd10* | | 1 | 211 | 731 | AF515527 |
| *GSTd11* | | 1 | 214 | 725 | AF513637 |
| *GSTd12* | GST1-10 | 1 | 211 | 715 | AF316638 |
| *GSTu1* | | 1 | 233 | 804 | AF515521 |
| *GSTu2* | | 1 | 222 | 832 | AF515523 |
| *GSTu3* | | 1 | 218 | 2329 | AF515524 |
| *GSTe1* | GST3-1 | 1 | 223 | 817 | AF316635 |
| *GSTe2* | GST3-2 | 1 | 226 | 830 | AF316636 |
| *GSTe3* | | 1 | 223 | 741 | AY070234 |
| *GSTe4* | | 1 | 225 | 743 | AY070254 |
| *GSTe5* | | 1 | 230 | 765 | AY070255 |
| *GSTe6* | | 1 | 226 | 774 | AY070256 |
| *GSTe7* | | 1 | 225 | 819 | AF491816 |
| *GSTe8* | | 1 | 217 | 789 | AY070257 |
| *GSTo1* | | 1 | 248 | 1473 | AY255856 |
| *GSTs1* | GST2-1 | 2 | 203, 203 | 3470 | L07880, AF513639 |
| *GSTt1* | | 1 | 229 | 862 | AF515526 |
| *GSTt2* | | 1 | 235 | 793 | AF515525 |
| *GSTz1* | | 1 | 223 | 14678 | AF515522 |

majority of sequences (Figure 1). This manually truncated alignment, consisting of 218 residues, was used to determine the phylogenetic relationship between the *A. gambiae* GSTs by both distance and parsimony methods. Two of the GST genes (*GSTd1* and *GSTs1*) are alternatively spliced to produce multiple transcripts with identical N-termini but differing C-termini (see below). To avoid distortion of the phylogeny that may have resulted from inclusion of conserved N-termini, a second alignment was generated, in which a further 85 residues from the N-terminal, were removed and this alignment was also subjected to parsimony and distance methods. Although several different optimal trees were generated from the two alignments and two phylogenetic methods, their topology was essentially the same, differing only in the placement of four GSTs, GSTu1, GSTu2, GSTu3 and GSTe8, discussed below. A representative distance tree, derived from the N-terminal truncated alignment is shown in Figure 2.

The two largest GST classes in *A. gambiae* are the insect specific delta and epsilon classes with 12 and 8 members respectively. Support for the monophyly of these two classes is low in the tree shown although, when the 218 residue alignment is used, the bootstrap values are more supportive (data not shown). Criteria for inclusion in a particular class is based primarily on amino acid sequence identity and phylogenetic relationship, but chromosomal location and immunological properties, where known, were taken into account. Thus *GSTe8*, although sharing less than 29% amino acid identity with other members of the epsilon class, is found immediately adjacent to the seven epsilon GSTs on chromosome 3R, is immunologically related [28] and, in the majority of the trees, formed a weakly supported monophyletic group with the seven *bona fide* epsilon GSTs. Thus this GST was classified as the eighth member of the epsilon class [28]. Three GSTs, *GSTu1*, *GSTu2* and *GSTu3*, are outliers from the major delta GST clade (Figure 2). The phylogenetic relationship of these GSTs to the remainder of the family was not consistent between the different trees. Furthermore, these GSTs share less than 37% amino acid identity with other members of the *A. gambiae* delta GST class (pairwise amino acid identities between the three tentative delta class GSTs and the remainder of the class range from 22.1% between GSTd4 and GSTu2 to 36.3% identity

**Figure 1**
Alignment of *A. gambiae* GSTs. ClustalW alignment of all putative subunits of *A. gambiae* cytosolic GSTs. Gaps introduced to maximise sequence similarity are shown by horizontal dashes. The regions of the alignment that were removed prior to phylogenetic analysis are shown in light grey. A vertical line demarks the putative end of the N-terminal domain. Residues to the left of this boundary were removed from the alignment in construction of the phylogenetic tree shown in Figure 2 (see text for further details).

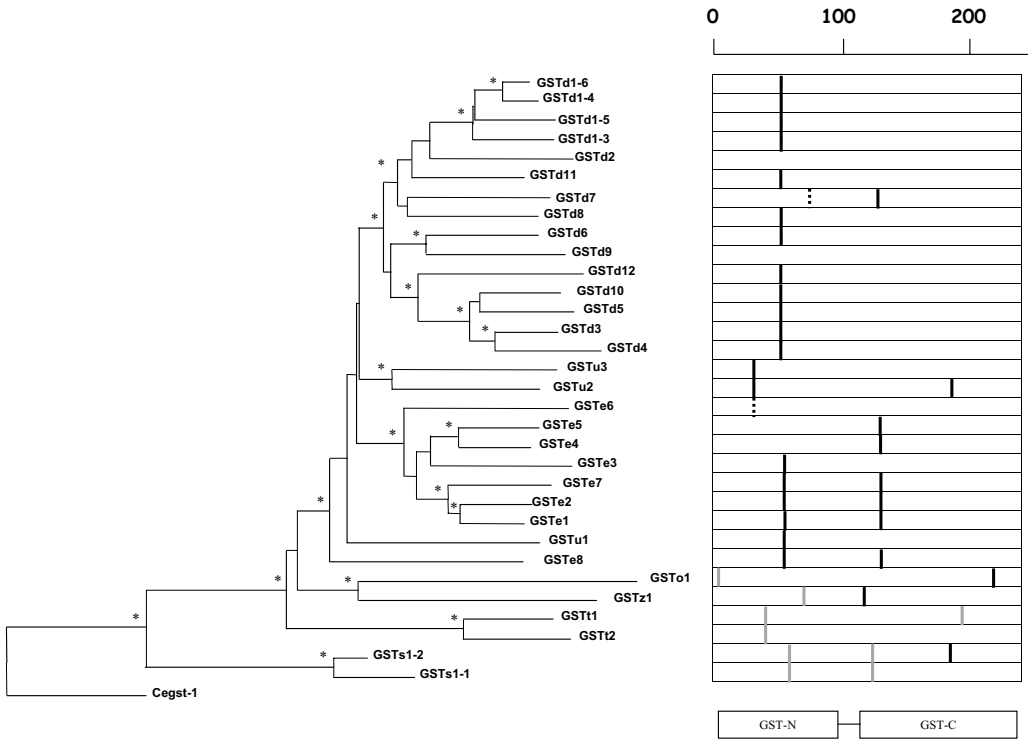


**Figure 2**
Neighbour-joining tree and introns positions of *A. gambiae* GSTs. The putative amino acid sequences of all 32 GST subunits in *A. gambiae* were aligned with GST-1 from *C. elegans* (Accession number CAA78471) using ClustalW (see text for details). The tree was constructed by the neighbour-joining method from a similarity matrix of pairwise comparisons made using the Jukes-Cantor algorithm. Nodes with distance bootstrap values (500 replicates) of > 70% are marked by *. The positions and phase of introns are shown on the right of the dendogram. Phase 0 introns are shown by a solid black line, phase 1 by dotted lines and phase 2 by grey solid lines.

between GSTd11 and GSTu3) and are not physically clustered with the majority of the delta class GSTs on chromosome 2L. Thus it is possible that these GSTs may belong to an as yet unrecognised GST class (or classes) and in the absence of clarifying immunological or biochemical data these GSTs have been designated as unclassified (denoted by a 'u').

Of the remaining, five *A. gambiae* GSTs, two belong to the theta class, originally thought of as the progenitor class of all GSTs [29]. The ubiquitous omega and zeta classes [7,8] are each represented by a single gene in *A. gambiae*, and a single sigma GST (alternatively spliced to produce two different transcripts (see below)) is also present. Support for the classification of these five GSTs is provided by a phylogenetic comparison with their orthologous classes from other organisms (Figure 3).

With the exception of the sigma class, all of the non-insect specific classes are expanded in *D. melanogaster* relative to *A. gambiae.* Neither species has any sequence related to the mitochondrial kappa class found in mammals [9]. The endogenous function of these mammalian GST classes in insects has not been clearly resolved and therefore the significance of the difference in size of these classes in the two Diptera examined, is, at present, unknown.

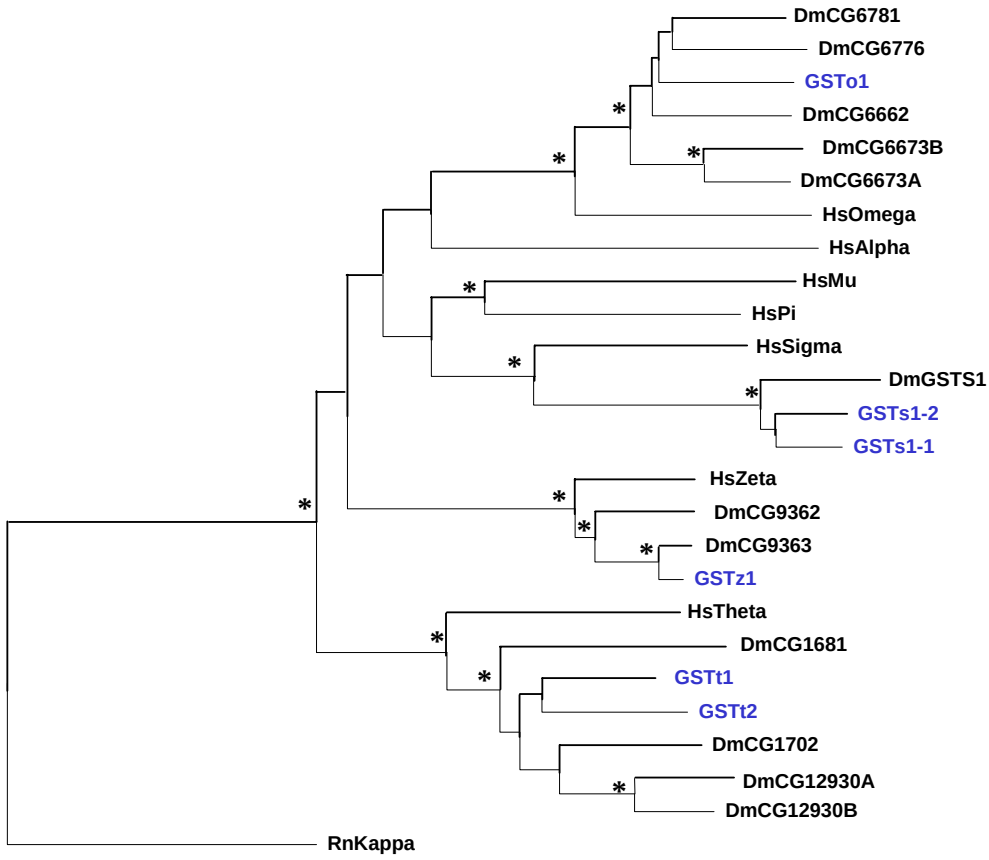


**Figure 3**
Neighbour-joining tree illustrating the relationship between representative insect and mammalian GST classes. GSTs from the non-insect specific classes from *A. gambiae*, *D. melanogaster* were aligned with representatives from mammalian classes using ClustalW. The tree was constructed by the neighbour-joining method from a similarity matrix of pairwise comparisons made using the Jukes-Cantor algorithm. Sequences shown in blue are from *A. gambiae*. Dm = *D. melanogaster*, Hs = *Homo sapiens*, Rn = *Rattus norvegicus*. Nodes with distance bootstrap values (500 replicates) of > 70% are marked by*.

### *Amplification of* A. gambiae *cDNAs*

Transcripts from 27 of the 28 *A. gambiae* GST genes were detected in fourth instar larvae or one-day-old adults by RT-PCR. Expression of *GSTd9* was not detected in any life stage and thus it is possible that this gene represents a silent pseudogene, as suggested earlier [10]. A second putative pseudogene is *GSTd6*. Utilisation of the first inframe stop codon for this gene (as present in the genome sequence database) would generate a transcript with 885 bp of coding sequence (> 130 bp longer than any of the other cytosolic GST genes in this species) and the putative translation would encompass a string of 14 glutamine residues. We therefore sequenced the intergenic region between *GSTd6* and the neighbouring gene, *GSTd11* to detect any possible frame shifts or sequencing errors that may have masked the stop codon but none were detected. A *GSTd6* transcript of 666 bp was amplified by RT-PCR but attempts at 3' RACE have so far failed to detect the 3' end of this transcript.

### *Intron positions and sizes*

The cDNA sequences of the *A. gambiae* GSTs were aligned with the genomic DNA sequences retrieved from the genome database and the position and sizes of the introns noted. Intron size ranged from 64 bp in *GSTe1* and *GSTd4* to 13,937 bp in *GSTz1* with the majority of introns ranging from 50 to 100 bp (Figure 4). Only two of the GSTs are intronless, *GSTd2* and the putative pseudogene, *GSTd9*.

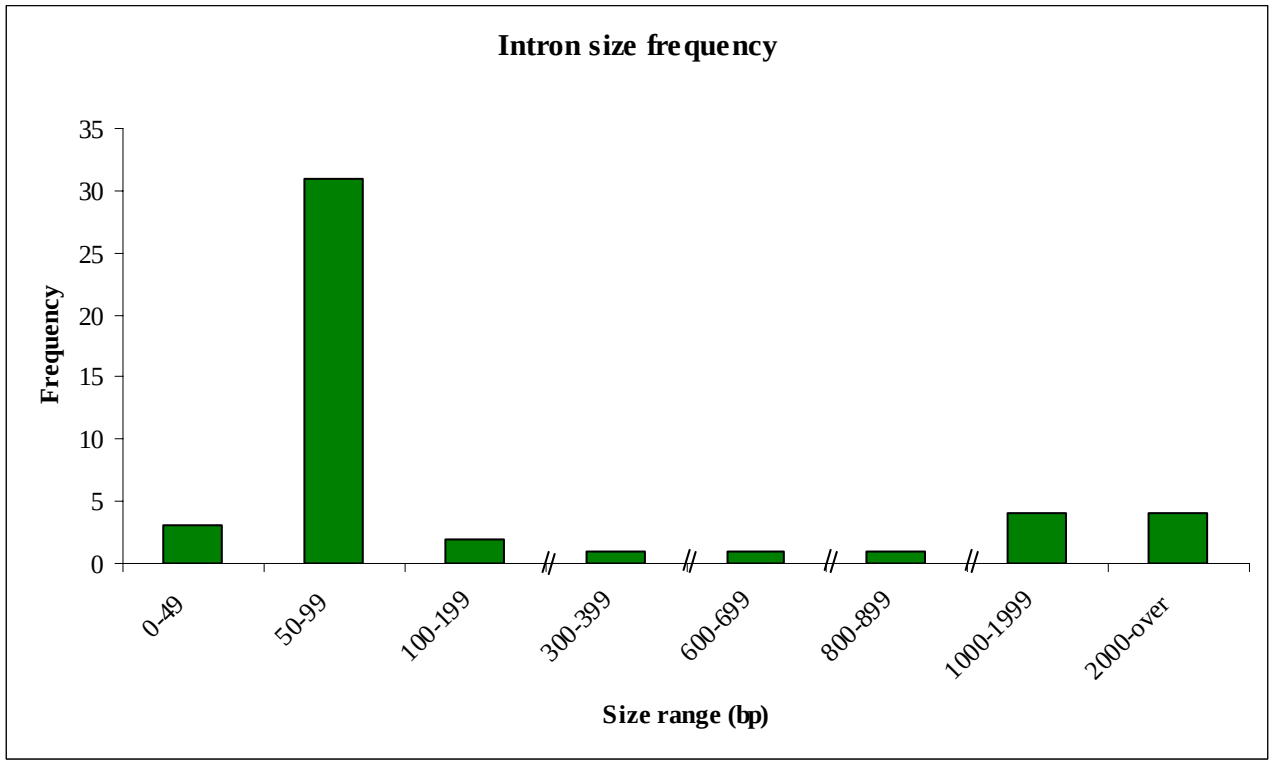


**Figure 4**
Histogram showing the frequency of intron sizes in the *A. gambiae* GST supergene family.

This contrasts with the situation in *D. melanogaster* where 20 of the 37 *D. melanogaster* GST genes are intronless [10,30].

There are thirteen unique intron sites and 42 introns within the *A. gambiae* GST supergene family. The majority of the introns (28) are phase 0 introns, i.e. the intron does not interrupt a codon. There is a considerable conservation of intron positions within the different GST classes (Figure 2) and one intron is found in 17 GST genes spanning three different classes. Interestingly this conserved intron, found approximately 50 amino acid residues from the N-terminal, is also the splice site for the alternative transcripts of *GSTs1* and *GSTd1* (see below). It has been proposed, by proponents of the 'introns-early' hypothesis, that different exons correspond to different domains in a protein [31]. The highly conserved GST intron however, splits the N-terminal domain (roughly residue 1–80). In addition, the phase of this intron is not conserved between the different classes, and thus the classification of this as an ancient intron is not well supported.

### *Clustering of GST genes in the genome*

The *A. gambiae* GST genes are located on all three of the mosquito's chromosomes but two large, gene-specific clusters exist (Figure 5). The eight members of the epsilon GST class are clustered on chromosome 3R division 33B. In *D. melanogaster*, a cluster of 10 epsilon GSTs are located on chromosome 2R, division 55C [30] (Figure 6B). A search of the *D. melanogaster* genome identified a further four members of this family also present on chromosome 2R but at a distant location. These four singletons belonging to the *D. melanogaster* epsilon GST class do not form a monophyletic clade with the cluster of 10 *DmGSTE* genes (Figure 6A). Instead, one of these *Drosophila* GSTs (CG4688) is a probable ortholog of *A. gambiae GSTe8*.

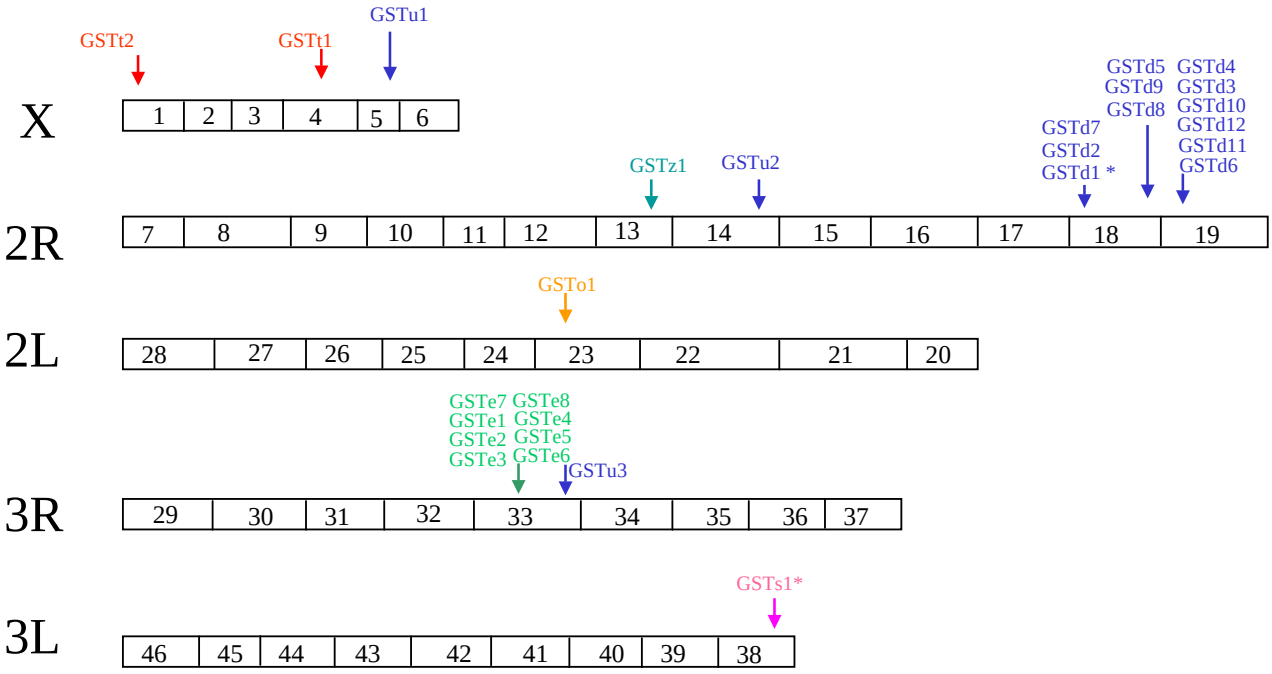


**Figure 5**
Schematic diagram indicating the organisation of GST genes in the *A. gambiae* genome. The numbers represent polytene chromosome divisions. Those genes marked with * are alternatively spliced to produce multiple transcripts.

Within the *A. gambiae* epsilon GST class there is evidence of recent internal duplications within this gene cluster. For example *GSTe4* and *GSTe5* have diverged recently and are located next to each other within the cluster. Similarly *GSTe7*, *GSTe1* and *GSTe2* are sequentially arranged and phylogeneticaly closely related (Figure 6A and 6B).

The *A. gambiae* delta class GSTs are located in two closely linked clusters of six genes on chromosome 2R divisions 18B and 19D (Figure 7B). Ten members of the *D. melanogaster* delta class are sequentially arranged on chromosome 3R, division 87B. We have identified an eleventh putative member of this class (CG17639) in *D. melanogaster* that is a clear ortholog of *A. gambiae GSTd7* [10] (Figure 7B). A detailed analysis of the phylogenetic relationship between members of the *A. gambiae* delta GST class, supports the suggestion by Holt et al [15] that the two clusters of paralogous delta GSTs are partly a result of segmental duplication. The clustering of *GSTd8* with *GSTd11*, *GSTd6* with *GSTd9* and *GSTd5* with *GSTd10* is consistent with a duplication of a cluster of three paralogous genes, with subsequent further local duplications in one block to give rise to *GSTd4* and *GSTd3* (Figure 7A and 7B).

In both *A. gambiae* and *D. melanogaster* local duplications in the epsilon and delta GST families have led to independent expansions of these gene classes. Subsequent diversification of these enzymes has presumably facilitated the adaptation of the two Diptera to their different ecological niches.

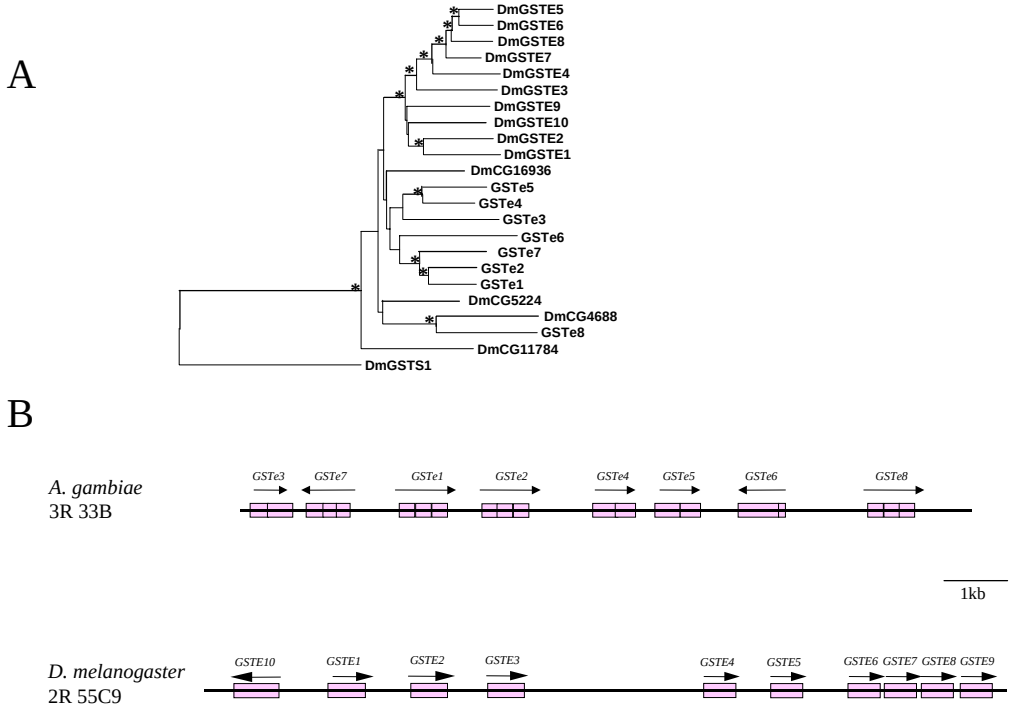


**Figure 6**
A: Neighbour-joining tree illustrating the relationship between *A. gambiae* and *D. melanogaster* epsilon class GSTs. ClustalW was used to align the putative amino acid sequences of the 8 epsilon class GST subunits in *A. gambiae* with the 14 putative epsilon class GSTs from *D. melanogaster* and with GSTS1 from *D. melanogaster* (as an outgroup). 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 marked by *. B, Epsilon class GST arrangement in *A. gambiae* and *D. melanogaster*. Arrows mark directions of transcription. The solid bars denote the genes and the vertical lines within these mark the approximate introns position.

### *Alternative Splicing*

An *A. gambiae* delta class GST, *GSTd1*, that is alternatively spliced to give four different mature transcripts each encoding subunits with differing biochemical properties, has already been described [12]. Here we report that a second member of this supergene family belonging to the sigma class, is also alternatively spliced in a similar manner producing two distinct transcripts both of which have been detected in adults by RT-PCR (Figure 8). As with *GSTd1* the two transcripts of *GSTs1* share a common 5' exon but differing 3' exons (64.3% identity at the amino acid level). As the carboxyl region of the protein contains the majority of the residues involved in substrate binding (the H-site), the two subunits encoded by the sigma GST gene in *A. gambiae* are likely to encode proteins with differing substrate specificities.

The sigma class in *D. melanogaster* is also represented by a single gene, *DmGSTs1*. The *Drosophila* GST has an N-terminal extension of 46 amino acid residues relative to the *Anopheles* ortholog. This N-terminal extension is not essential for catalytic activity and may play a role in attaching the *D. melanogaster* protein to indirect flight muscles [21]. The intron positions of *DmGSTS1* and *A gambiae GSTs1-1* are conserved indicating a common

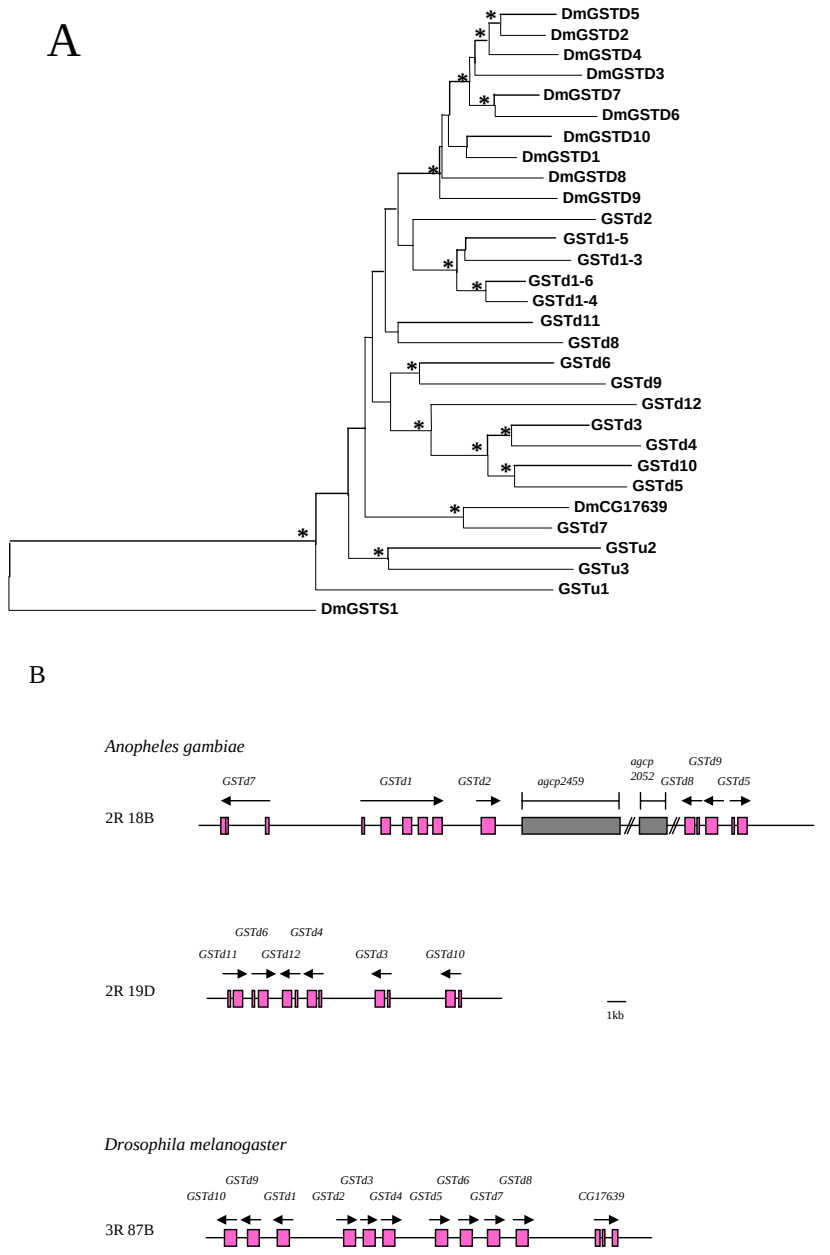


**Figure 7**
Neighbour-joining tree illustrating the relationship between *A. gambiae* and *D. melanogaster* delta class GSTs. ClustalW was used to align the putative amino acid sequences of the 18 delta GST subunits in *A. gambiae* with the 11 putative delta class GSTs from *D. melanogaster* and with GSTS1 from *D. melanogaster* (as an outgroup). 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 marked by *. B, Delta class GST arrangement in *A. gambiae* and *D. melanogaster*. Arrows mark directions of transcription. The solid rectangles denote the exons of the genes. *Agcp2459* and *agcp2052* are two putative genes, identified by the automatic analysis of the *A. gambiae* genome, that interrupt the delta GST cluster on chromosome 2R, division 18B.

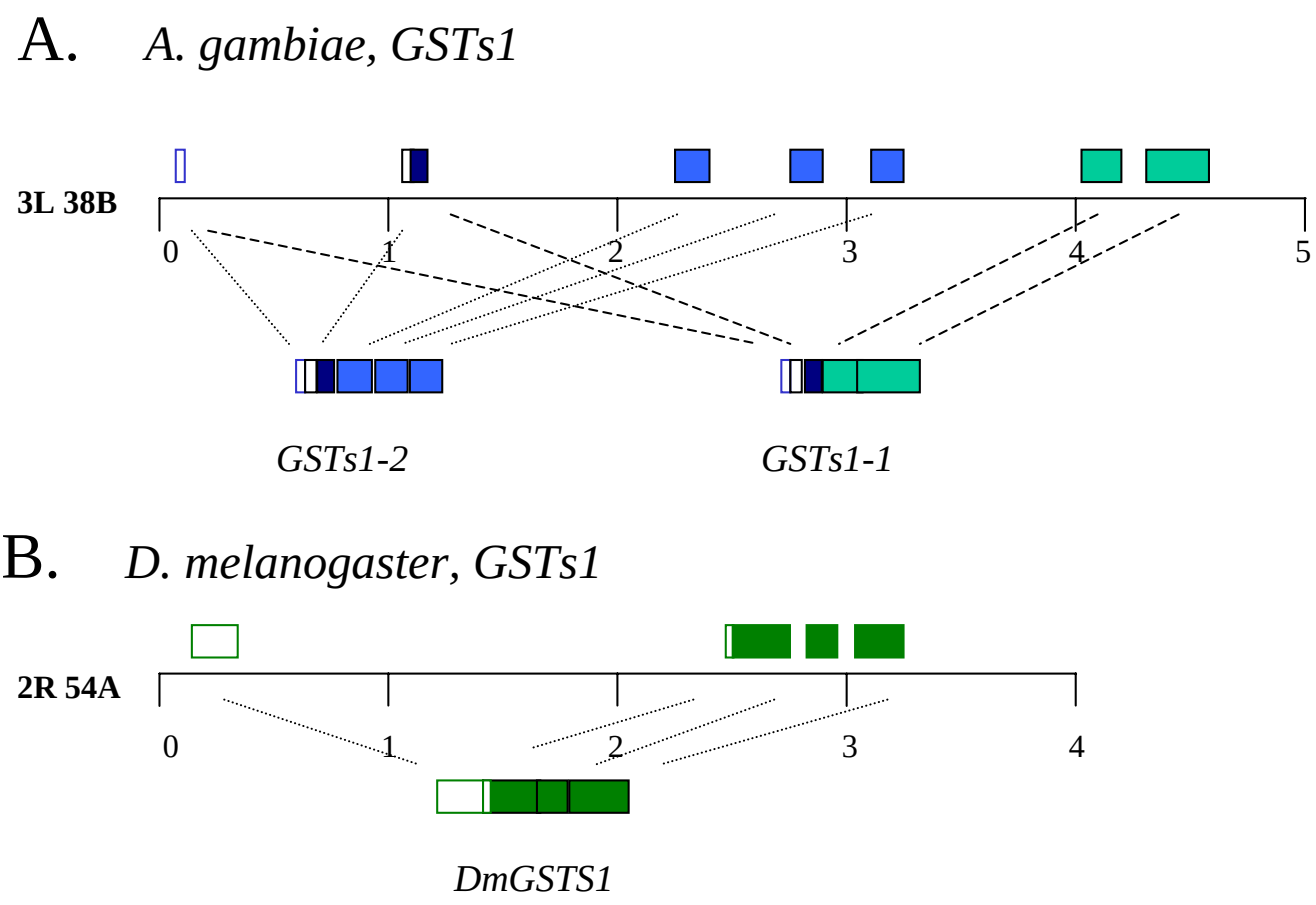


**Figure 8**
Schematic diagram showing the organisation of the sigma GST genes in *A. gambiae* and *D. melanogaster*. Each rectangle above the scale bar represents a different exon. The empty rectangles indicate 5'UTR regions whilst the rectangles in solid colours are coding sequence. The cDNAs produced by splicing of these genes are shown beneath the scale bar.

ancestor for these genes. Local duplication of the carboxyl exons of the *A. gambiae GSTs1* gene presumably occurred subsequent to the speciation event separating the two Diptera.

Alternative splicing as a means of increasing the level of heterogeneity of GST subunits appears to be a rare phenomenon outside of the genus *Anopheles*, having, as far as we are aware, only been reported in a GST from the nematode *Onchocerca volvulus* [13]. Two human cDNA clones encoding alternative transcripts of a mu class GST have been detected but both are incomplete and thus unlikely to encode functional GSTs [32].

### *Expression profiling of epsilon GSTs*

Our principal interest in insect GSTs stems from the role that these enzymes play in insecticide resistance. GSTs have been implicated in the conjugation, metabolism or remediation of harmful effects of all major classes of insecticides, but our studies have focused on the role of these enzymes in the detoxification of DDT. From genetic mapping studies [25] we know that a major DDT resistance locus co-localises with the epsilon GST cluster on chromosome 3R, division 33B. We have previously demonstrated that expression of one member of this GST class, *GSTe2*, is up-regulated in a DDT resistant strain [10] and that a recombinant homodimer of GSTe2 is very efficient at metabolising DDT [28]. In the present study we used quantitative PCR to compare the expression levels of all members of the epsilon class in DDT resistant and suscep-

tible strains. The results are shown in Figure 9A. Five of the eight GSTs are expressed at a significantly higher levels in the resistant strain. The difference in expression levels is greatest for *GSTe2* and *GSTe7* (Table 2). Of the five GSTs that are expressed at higher levels in the DDT resistant strain, only GSTe2 has confirmed DDT dehydrochlorinase activity. The ability of recombinant homodimers of GSTe3 and GSTe7 to metabolise DDT has not yet been determined. However, we have previously shown that neither GSTe1 or GSTe4 homodimers possess DDT dehydrochlorinase activity [28] and yet the expression of both of these genes is upregulated in the DDT resistant strain (Table 2). This result is most readily explained if it is assumed that the expression of multiple genes in the epsilon cluster is under the control of a common regulatory element and that an alteration in this element accounts for the increase in expression of the six epsilon GST genes. Nonetheless other explanations, such as modifications in the stability of the mRNA transcripts in the resistant strain can not be eliminated at this stage.

There was a large variation in the number of transcripts detected for each of the individual GSTs within both the susceptible and resistant strains. For example, the normalised copy number of *GSTe2* transcripts was over 26-fold greater than that of *GSTe3* in the resistant strain (Table 2). This indicates that the basal expression of individual GSTs within a cluster is independently regulated. To confirm this result, we repeated the qPCR for four of the GSTs using gDNA as a template. If the differences seen with cDNA reflect genuine differences in GST transcript copy number as we hypothesised, then the genomic copy number for each of the four GSTs would be approximately the same. Figure 9 and Table 2 support our hypothesis. Furthermore, as the genomic qPCR was carried out on DNA extracted from both susceptible and resistant mosquitoes, the results shown in Figure 9B support our unpublished data from Southern blots demonstrating that the increases in GST transcript levels seen in the resistant strain is not due to gene amplification (F. Ortelli, unpublished data).

As a final test of the reliability of our qPCR results we calculated the expected copy number of a single copy gene in 10.4 ng of *A. gambiae* genomic DNA (the starting amount of gDNA template in the qPCR reactions with Kisumu DNA). The theoretical copy number (33,120) was within an order of magnitude of the range of values obtained (11,993–17,205).

## Conclusion

The insect GST supergene family encodes a diverse set of proteins. The availability of the full genome sequence for two insect species has enabled the full extent of this protein family in insects to be realised. Multiple members of the epsilon class are upregulated in a DDT resistant strain of *A. gambiae* and it is proposed that this class plays a major role in the detoxification of xenobiotics. However, little is known about the endogenous substrates of insect GSTs. Functional genomics approaches will no doubt contribute to our understanding of the role of individual GSTs in insects and perhaps then the reason for the extensive diversity of this enzyme family will become clear.

## Methods

### *Mosquito Strains*

The DDT resistant ZAN/U strain of *Anopheles gambiae* s.s. originated from a field population collected from Zanzibar, Tanzania, in 1982. Adults from this strain have been maintained under regular selection pressure by exposure to Whatmans no.1 filter papers impregnated with 4% DDT according to standard WHO methods [33]. The Kisumu strain is susceptible to insecticides and originates from Kisumu in Western Kenya.

### *Annotation of* A. gambiae *GST genes*

Members of the GST supergene family were identified in the *A. gambiae* genome by BLAST searches [34] using multiple representative sequences from each GST class as query sequences. The sequences retrieved from the genome were manually annotated to predict transcription initiation and termination sites and intron/exon boundaries using BlastX comparisons of putative amino acid translations [35]. Primers pairs were designed to amplify the full length of the coding sequence of each gene from *A. gambiae* cDNA.

### *cDNA synthesis*

Total RNA was extracted from individual mosquitoes using the TRI reagent (SIGMA), according to the manufacturer's instructions. The RNA was treated with DNase to remove any contaminating genomic DNA and the mRNA was reverse transcribed into cDNA using superscript II (GIBCO BRL) and an oligo (dT) adapter primer (5'-GACTCGAGTCGACATCGA$(dT)_{17}$-3'). The PCR conditions for amplifying GST cDNAs were determined empirically for each GST. Products of the expected size were subcloned into pGEM T-easy vectors (Promega) and used as templates for sequencing. At least three independent clones were sequenced for each GST. Sequencing reactions were performed using Beckman chemistry and the resultant products analysed on a Beckman CEQ800 capillary sequencer.

### *Quantitative PCR*

Total RNA was extracted from pools of ten (five male and five female) one-day old adult mosquitoes and the mRNA was reverse transcribed into cDNA as described above. Plasmids containing the gene of interest were diluted to produce seven standard templates at concentrations

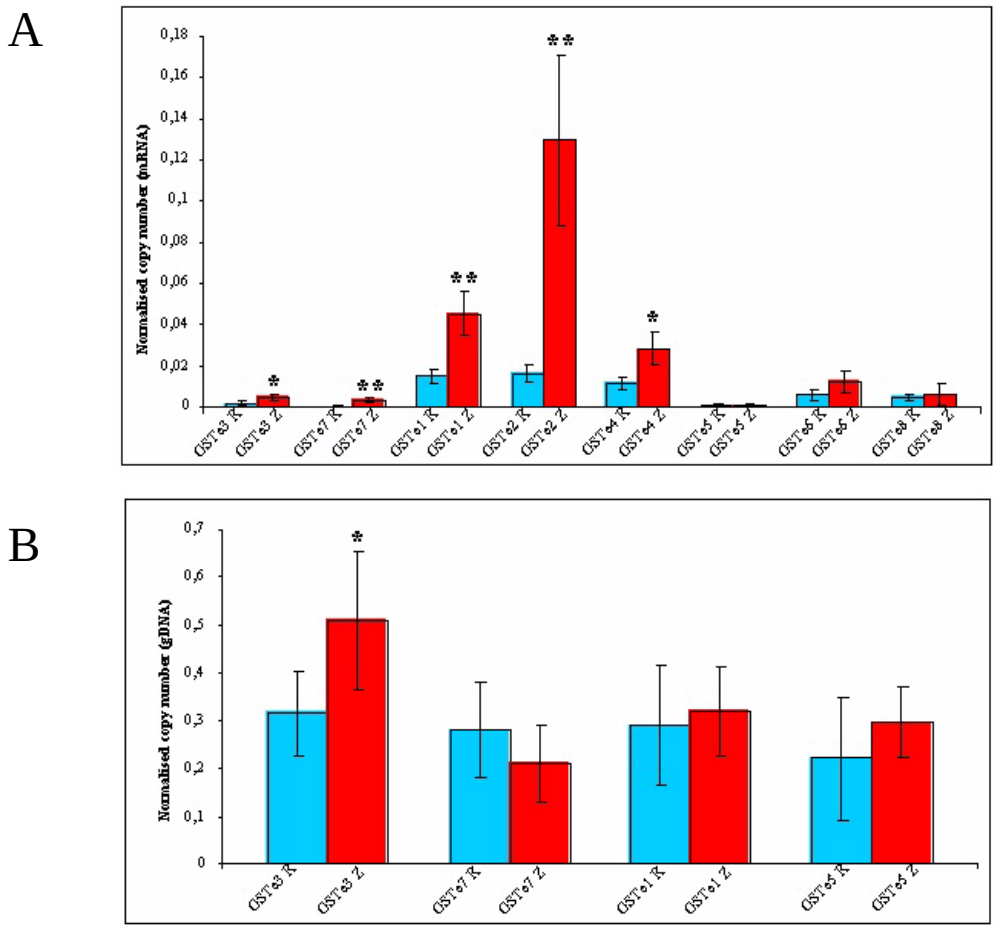


**Figure 9**
Expression levels of the epsilon class GSTs in two strains of *A. gambiae*. A Transcript copy number determined by qPCR using cDNA as a template. Results represent the average of three independent cDNA samples. B gene copy number, determined by using three independent aliquots of mass homogenates of gDNA as a template. In both cases, copy numbers were normalised against the copy number of the S7 genes and error bars represent standard deviations. Statistically significant differences (*$p < 0.05$, **$p < 0.01$) between ZAN/U and Kisumu are indicated. Kisumu is an insecticide susceptible strain, ZAN/U is resistant to DDT.

**Table 2: Quantitative PCR results of *A. gambiae* epsilon GST genes. The transcript copy number or gene copy number was determined using cDNA or gDNA respectively. The copy numbers were normalised for variations in initial template concentration by dividing each sample by the copy number of the ribosomal protein gene, S7. The final column shows the ratio of the transcript or gene copy number between the resistant ZAN/U and insecticide susceptible Kisumu strain. Statistically significant differences (*p < 0.05, **p < 0.01) between ZAN/U and Kisumu are indicated.**

| Gene | Kisumu Strain Normalised cDNA copy number × $10^2$(± S.D.) | ZAN/U Strain Normalised cDNA copy number × $10^2$(± S.D.) | Ratio of Copy Number (ZAN/U:KISUMU) |
|---|---|---|---|
| *GSTe1* | 1.50 ± 0.320 | 4.52 ± 1.054 | 3.0 ** |
| *GSTe2* | 1.66 ± 0.408 | 12.92 ± 4.125 | 7.8 ** |
| *GSTe3* | 0.15 ± 0.100 | 0.49 ± 0.133 | 3.3 * |
| *GSTe4* | 1.13 ± 0.299 | 2.82 ± 0.795 | 2.5 * |
| *GSTe5* | 0.08 ± 0.050 | 0.10 ± 0.065 | 1.3 |
| *GSTe6* | 0.58 ± 0.223 | 1.20 ± 0.551 | 2.1 |
| *GSTe7* | 0.03 ± 0.010 | 0.34 ± 0.152 | 11.0 ** |
| *GSTe8* | 0.48 ± 0.162 | 0.62 ± 0.541 | 1.3 |
| | **Kisumu strain Normalised gDNA copy number. (± S.D.)** | **ZAN/U strain Normalised gDNA copy number. (± S.D.** | **Ration of Copy Number (ZAN/U:Kisumu)** |
| *GSTe1* | 0.29 ± 0.125 | 0.32 ± 0.092 | 1.1 |
| *GSTe3* | 0.32 ± 0.088 | 0.51 ± 0.144 | 1.6 ** |
| *GSTe5* | 0.22 ± 0.129 | 0.29 ± 0.073 | 1.3 |
| *GSTe7* | 0.28 ± 0.098 | 0.21 ± 0.081 | 0.75 |

**Table 3: Primer sequences and PCR conditions for amplification of epsilon class GST genes by quantitative PCR.**

| Gene | Forward Primer (5' to 3') | Reverse Primer (5' to 3') | cDNA amplicon (bp) | gDNA amplicon (bp) | Annealing/Detection Temp (°C) |
|---|---|---|---|---|---|
| *GSTe1* | GTCAATGAGCCACTGCACTTC | GTGATCCGGCTACGTAATCG | 175 | 253 | 57/84 |
| *GSTe2* | ATCACCGAGAGCCACGCAATCAT | GCCACCGTTCGCTTCCTCGTAGT | 414 | 507 | 62/84 |
| *GSTe3* | GAGCTGACGGCAAAGATGATCG | CCTGCTTCACTAGATCCTTCGC | 229 | 298 | 61/83 |
| *GSTe4* | CGCCATTCAAACGACCATGCC | GATGGCGTGGCTGTCCCACACG | 229 | 229 | 62/85 |
| *GSTe5* | ATGGCAACGAACCCCATCATC | CACACCACCATCGTCAATCACC | 198 | 198 | 62/85 |
| *GSTe6* | GTACACGCACACGATTAGTCC | GCTGACCCTTGAAGACGTTC | 104 | 195 | 62/83 |
| *GSTe8* | GCCATGATTCTGTACTACGACG | GGTAAGCGTTAACTCACCGTG | 198 | 267 | 63/83 |
| *GSTe7* | GCAGATTGGTACTGTACACG | CTCGGATAGAGACCGTCGTC | 256 | 335 | 59/84 |
| *S7* | GCACGTCGTGTTCATTGCCG | GAACGTAACGTCACGGCCAGTCA | 292 | 441 | 60/86 |

ranging from 1 ng/μl to 1 fg/μl. The incorporation of the fluorescent dye SYBR Green during PCR amplification of these templates was detected using Roche LightCycler technology and Quantitect SYBR Green chemistry (Qiagen). For each experiment two replicates of each of the seven plasmid templates and two replicates of three cDNA samples from both strains, all synthesised from independent RNA extractions, were used. The binding sites of the primers used for template amplification were chosen to avoid regions of allelic variation within the gene sequences [28]. A control plasmid containing a partial fragment of the S7 ribosomal protein gene [36] was used to standardise the initial cDNA concentration in each sample. Details of the primer sequences and PCR conditions are given in Table 3. The samples were analysed using LightCycler Software V3 (Roche) and the number of GST copies in the cDNA sample was determined as described previously [37].

The experiment was repeated on genomic DNA samples extracted from approx 1 g of Kisumu or ZAN/U strains as described previously [38]. For the GST genes for which the initial primers spanned introns, new plasmids were constructed containing the genomic fragment of the gene.

### *Phylogenetic analysis*

Putative amino acid sequences of the GSTs were aligned using ClustalW [27]. The alignment was manually truncated as described in the Results and Discussion section. Evolutionary distances were calculated using the Jukes-Cantor algorithm [39] and phylogenetic trees were determined by the neighbor-joining method [40] with

TREECON [41] or by parsimony methods using MEGA2 [41].

## Abbreviations

DDT: 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane, GST: glutathione transferase

## Authors Contributions

YD and LR carried out the expression analysis. FO participated in the sequencing and performed the majority of the data analysis. JH conceived the study and participated in its design. HR contributed to the sequencing and analysis of data, drafted the manuscript and coordinated the project. All authors read and approved the final manuscript.

## Acknowledgements

Y. Ding was supported by a PhD fellowship award from the late Lord Leverhulme. F. Ortelli and L. Rossiter were supported by a Wellcome Trust project grant held by Ranson and Hemingway. H. Ranson is a Royal Society Dorothy Hodgkin Research Fellow. We are very grateful to A. Dana and F.H. Collins for sharing EST data from *A. gambiae* prior to publication.

## References

1. Hayes JD and Pulford DJ: **The glutathione S-transferase supergene family – Regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance.** *Crit Rev Biochem Mol Biol* 1995, **30:**445-600.
2. Mannervik B: **The isoenzymes of glutathione transferases.** *Adv Enzym Relat Areas Molec Biol* 1985, **57:**357-417.
3. Habig WH, Pabst MJ and Jakoby WB: **Glutathione S-transferases. The first enzymatic step in mercapturic acid formation.** *J Biol Chem* 1974, **249:**7130-7139.
4. Meyer DJ, Coles B, Pemble SE, Gilmore KS, Fraser GM and Ketterer B: **Theta, a new class of glutathione transferases purified from rat and man.** *Biochem J* 1991, **274:**409-414.
5. Meyer DJ and Thomas M: **Characterization of rat spleen prostaglandin H D-isomerase as a sigma-class GSH transferase.** *Biochem J* 1995, **311:**739-742.
6. Morgenstern R and DePierre JW: **Microsomal glutathione transferase. Purification in unactivated form and further characterization of the activation process, substrate specificity and amino acid composition.** *Eur J Biochem* 1983, **134:**591-597.
7. Board PG, Baker RT, Chelvanayagam G and Jermin LS: **Zeta, a novel class of glutathione transferases in a range of species from plants to humans.** *Biochem J* 1997, **328:**929-935.
8. Board PG, Coggan M, Chelvanayagam G, Easteal S, Jermiin LS, Schulte GK, Danley DE, Hoth LR, Griffor MC, Kamath AV, Rosner MH, Chrunyk BA, Perregaux DE, Gabel CA, Geoghegan KF and Pandit J: **Identification, characterisation, and crystal structure of the Omega class glutathione transferases.** *J Biol Chem* 2000, **275:**24798-24806.
9. Pemble SE, Wardle AF and Taylor JB: **Glutathione S-transferase class Kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue.** *Biochem J* 1996, **319:**749-754.
10. Ranson H, Rossiter L, Ortelli F, Jensen B, Wang X, Roth CW, Collins FH and Hemingway J: **Identification of a novel class of insect glutathione S-transferases involved in DDT resistance in the malaria vector, *Anopheles gambiae*.** *Biochem J* 2001, **359:**295-304.
11. Dixon DP, Lapthorn A and Edwards R: **Plant glutathione transferases.** *Genome Biology* 2002, **3:**3004.1-3004.10.
12. Ranson H, Collins FH and Hemingway J: **The role of alternative mRNA splicing in generating heterogeneity within the *Anopheles gambiae* class 1 GST family.** *Proc Natl Acad Sci U S A* 1998, **95:**14284-14289.
13. Kampkötter A, Volkmann TE, Hegi de Castro S, Leiers B, Klotz L, Johnson TE, Link CD and Henkle-Dührsen K: **Functional analysis of the Glutathione *S*-transferase 3 from *Onchocerca volvulus (Ov-GST-3)*: a parasite GST confers increased resistance to oxidative stress in *Caenorhabditis elegans*.** *J Mol Biol* 2003, **325:**25-37.
14. Dixon DP, Cole DJ and Edwards R: **Dimerisation of maize glutathione transferases in recombinant bacteria.** *Plant Mol Biol* 1999, **40:**997-1008.
15. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH and Hoffman SL: **The genome sequence of the malaria mosquito *Anopheles gambiae*.** *Science* 2002, **298:**129-148.
16. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, Unger MF, Collins FH and Feyereisen R: **Evolution of supergene families associated with insecticide resistance.** *Science* 2002, **298:**179-181.
17. Clark AG and Shamaan NA: **Evidence that DDT-dehydrochlorinase from the house fly is a glutathione *S*-transferase.** *Pest Biochem Physiol* 1984, **22:**249-261.
18. Ku SS, Chiang FM, Hsin CY, Yao YE and Sun CN: **Glutathione transferase isozymes involved in insecticide resistance of diamondback moth larvae.** *Pestic Biochem Physiol* 1994, **50:**191-197.
19. Vontas JG, Small GJ and Hemingway J: **Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in *Nilaparvata lugens*.** *Biochem J* 2001, **357:**65-72.
20. Zou S, Meadows S, Sharp L, Jan LY and Jan YN: **Genome-wide study of aging and oxidative stress response in *Drosophila melanogaster*.** *Proc Natl Acad Sci U S A* 2000, **97:**13726-13731.
21. Singh SP, Coronella JA, Benes H, Cochrane BJ and Zimniak P: **Catalytic function of *Drosophila melanogaster* glutathione S-transferase DmGSTS1-1 (GST-2) in conjugation of lipid peroxidation end products.** *Eur J Biochem* 2001, **268:**2912-2923.
22. Dimopoulos G, Christophides GK, Meister S, Schultz J, White KP, Barillas-Mury C and Kafatos FC: **Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge and malaria infection.** *Proc Natl Acad Sci* 2002, **99:**8814-8819.
23. Snow RW, Craig M, Deichmann U and Marsh K: **Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population.** *Bull WHO* 1999, **77:**624-640.
24. Prapanthadara L, Hemingway J and Ketterman AJ: **Partial purification and characterization of glutathione S-transferase involved in DDT resistance from the mosquito *Anopheles gambiae*.** *Pestic Biochem Physiol* 1993, **324:**97-102.
25. Ranson H, Jensen B, Wang X, Prapanthadara L, Hemingway J and Collins FH: **Genetic mapping of two loci affecting DDT resistance in the malaria vector *Anopheles gambiae*.** *Insect Mol Biol* 2000, **9:**499-507.
26. Chelvanayagam G, Parker MW and Board PG: **Fly fishing for GSTs: a unified nomenclature for mammalian and insect glutathione transferases.** *Chem-Biol Interact* 2001, **133:**256-260.
27. Thompson JD, Higgins DG and Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.
28. Ortelli F, Rossiter LC, Vontas J, Ranson H and Hemingway J: **Heterologous expression of four glutathione transferase genes**

**genetically linked to a major insecticide resistance locus, from the malaria vector, *Anopheles gambiae*.** *Biochem J* 2003, **373:**957-963.
29. Pemble SE and Taylor JB: **An evolutionary perspective on glutathione transferases inferred from class-theta glutathione transferase cDNA sequences.** *Biochem J* 1992, **287:**957-963.
30. Sawicki R, Singh SP, Mondal AK, Benes H and Zimniak P: **Cloning, expression and biochemical characterization of one Epsilon-class (GST-3) and ten Delta-class (GST-1) glutathione S-transferases from *Drosophila melanogaster*, and identification of additional nine members of the Epsilon class.** *Biochem J* 2003, **370:**661-669.
31. Gö M: **Correlation of DNA exonic regions with protein structural units in haemoglobin.** *Nature* 1981, **291:**90-96.
32. Ross VL and Board PG: **Molecular cloning and heterologous expression of an alternatively spliced human Mu class glutathione S-transferase transcript.** *Biochem J* 1993, **294:**373-380.
33. WHO: **Preparation, production and supply of test kits, impregnated papers and standard solutions for the evaluation of vector susceptibility to insecticides.** *WHO/VBC/EC 80.25* 1980.
34. ***A. gambiae* genome site** [http://www.ncbi.nlm.nih.gov/BLAST/Genome/FlyBlast.html]
35. Gish W and States DJ: **Identification of protein coding regions by database similarity search.** *Nature Genet* 1993, **3:**266-272.
36. Salazar CE, Mills-Hamm D, Kumar V and Collins FH: **Sequence of a cDNA from the mosquito *Anopheles gambiae* encoding a homologue of human ribosomal protein S7.** *Nucleic Acids Res* 1993, **21:**4147.
37. Paton MG, Karunaratne SHPP, Giakoumaki E, Roberts N and Hemingway J: **Quantitative analysis of gene amplification in insecticide resistant *Culex* mosquitoes.** *Biochem J* 2000, **346:**17-24.
38. Vaughan A, Rodriguez M and Hemingway J: **The independent gene amplification of electrophoretically indistinguishable B esterases from the insecticide-resistant mosquito *Culex quinquefasciatus*.** *Biochem J* 1995, **305:**651-658.
39. Jukes TH and Cantor CR: **Evolution of protein molecules.** In *Evolution of protein molecules* 1969:21-132.
40. Saitou N and Nei M: **The neighbour-joining method: a new method for constructing phylogenetic trees.** *Mol Biol Evol* 1987, **4:**406-425.
41. van de Peer Y and De Wachter R: **TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment.** *Comput Applic Biosci* 1994, **10:**569-570.
42. Kumar S, Tamura K, Jakobsen IB and Nei M: **MEGA2: Molecular Evolutionary Genetics Analysis Software.** *Bioinformatics* 2001, **17:**1244-1245.