J Mol Evol (1996) 43:236–240

JOURNAL OF MOLECULAR EVOLUTION

Heterogeneity of the Glutathione Transferase Genes Encoding Enzymes Responsible for Insecticide Degradation in the Housefly

Michael Syvanen,[1] Zonghan Zhou,[1] Jonathan Wharton,[2] Claire Goldsbury,[2] Alan Clark[2]

[1] Medical Microbiology and Immunology, School of Medicine, University of California, Davis, CA 95616, USA
[2] School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand

Received: 22 November 1995 / Accepted: 23 February 1996

Abstract. One of the four glutathione-*S*-transferases (GST) that is overproduced in the insecticide-resistant Cornell-R strain of the housefly (*Musca domestica*) produces an activity that degrades the insecticide dimethyl parathion and conjugates glutathione to lindane. In earlier work, it was shown that the resistant Cornell-R carries an amplification, probably a duplication, of one or more of its GST loci and that this amplification is directly related to resistance. Using polymerase chain reaction (PCR) amplification with genomic DNA, multiple copies of the gene encoding the parathion-degrading activity (called MdGst-3) were subcloned from both the ancestral, insecticide-susceptible strain BPM and from the insecticide-resistant Cornell-R. In BPM, three different MdGst-3 genes were identified while in Cornell-R, 12 different MdGst-3 sequences were found that, though closely related to ancestral genes, had diverged by a few nucleotides. This diversity in MdGst-3 genomic sequences in Cornell-R is reflected in the expressed sequences, as sampled through a cDNA bank. Population heterozygosity cannot account for these multiple GST genes. We suggest that selection for resistance to insecticides has resulted in not only amplification of the MdGst-3 genes but also in the divergence of sequence between the amplified copies.

Key words: Housefly — Enzymes — Insecticide degradation — Glutathione transferase genes

Correspondence to: M. Syvanen

Introduction

Strains of the housefly, *Musca domestica*, have been naturally selected for resistance to insecticides such as lindane and the organophosphate triesters due to the intensive use of these chemicals in agriculture over the past 40 years. This provides us with the opportunity to observe, in a natural metazoan population, the genetic changes accompanying increased fitness, i.e., the opportunity to observe molecular evolution under positive selection over real time. In the resistant fly strain Cornell-R, changes in the glutathione transferases (GST) have been shown to be responsible for a large part of its organophosphate resistance (OP^R) and for a lower-level resistance to lindane (Oppenorth et al. 1972; Clark and Dauterman 1982; Clark and Shamaan 1984). Two genetic mechanisms for the emergence of resistance have been entertained. The first involves regulatory mutations causing overproduction of a wild-type GST, and the second involves the appearance of qualitatively new enzymes. In earlier work, we showed that resistance in Cornell-R was due to overproduction of certain GSTs (Wang et al. 1991) and that, furthermore, overproduction was caused, at least in part, by a gst gene amplification, probably a duplication, that was tightly correlated with the resistance phenotype. In the current study, we show that these same resistant flies also produce qualitatively new GSTs when compared to an ancestral susceptible strain.

In the housefly, four different theta-like GSTs (Pemble and Taylor 1992) have been isolated and sequenced in a cDNA gene bank prepared from the OP^R strain Cornell-R, and their activity has been expressed in

Table 1. Activity of the purified housefly theta-class glutathione transferases[a]

	GST activities with:		
	CDNB[b]	Methylparathion[c]	Lindane[c]
MdGST-1	3.51 ± 1.3	0.06	nd
MdGST-2	5.16 ± 0.66	nd	0.008
MdGST-3	6.6 ± 1.7	0.24 ± 0.02	0.045
MdGST-4	7.05 ± 1.7	nd	nd

[a] nd means no activity detected; for parathion this means an activity of less than 0.02–0.05 and for lindane an activity of less than 0.002–0.006.
[b] Activity is in μmoles product per min per mg protein.
[c] Activity is percent of activity observed with the CDNB substrate.

Escherichia coli (Syvanen et al. 1994). MdGst-1, -2, -3 and -4 constitute a homologous set that have diverged by 18–29% from one another at the amino acid sequence level. These current studies were initiated to identify which, if any, of these enzymes could detoxify the insecticides lindane and parathion and, further, to characterize the genomic loci responsible for resistance in Cornell-R.

Materials and Methods

Purification of Enzymes. The four housefly GSTs were expressed in *Escherichia coli* as described (Syvanen et al. 1994). After optimal enzyme induction, the GSTs were purified on a sulphobromophthalein-glutathione affinity column (Clark et al. 1990). Conjugation of glutathione to chlorodinitro benzene (CDNB) was assayed according to (Habig et al. 1974) and to methyl parathion according to (Dauterman 1986).

Polymerase Chain Reaction. The sequence of the two primers for MdGst-3 were CCC, then a *Bam*H1 site, followed by the MdGst-3 sequence TCAGTCATAATGACAGCCAAGG (forward primer) and CCC, then an EAG1 site followed by GTACTAACCGAAGCCAAGATGG (reverse primer). These flank 424 bp of the MdGst-3 gene that begins at 62 bp from the first AUG in its reading frame. After amplification the fragment was digested with *Bam*H1 and EAG1 and subcloned into pBluescript for sequencing using the dideoxy method provided in kit form from Stratagene.

Results

MdGst-1, MdGST-2, MdGST-3, and MdGST-4 were purified after expression in *E. coli* by affinity chromatography on a sulphobromophthalein-glutathione column. The ability of each enzyme to both degrade the insecticide dimethyl parathion and to conjugate glutathione to lindane was determined (Table 1). Of the four, the MdGst-3 isozyme has the greatest level of activity against these two substrates. In earlier experiments, we established a strong positive correlation between the level of transcription of MdGst-3 in a series of isogenic strains derived from Cornell-R and their level of OP resistance (Wang et al. 1991). In these experiments we

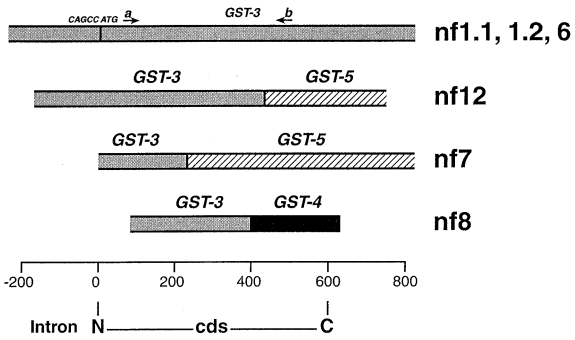

Fig. 1. Genomic loci carrying sequences highly similar to MdGst-3. Six different genes were found in the λEMB genomic DNA bank. The two nf1 alleles and nf6 contain the complete MdGst-3 coding sequence (>99% identical), while the other three are gene fusions between MdGst-3 and other Θ class GSTs found in the housefly. The four loci represented by the figure carry an intron in the 5′ noncoding region—the 3′ splice site and the beginning of the coding sequences all have the same CAGCCATG sequence, as shown. The sequence diverges substantially inside the intron. The finding of introns in the housefly GST genes is different from the situation observed in *Drosophila* (Toung et al. 1993). The *arrows* (labeled *a* and *b*) show the location of the PCR primers. The GenBank accession numbers are: *nf1.1,* X94279; *nf6,* X94280; *nf7,* X94281; *nf8,* X94284; and *nf12,* X94283. A complete description of these sequences is in preparation (Zhou, Z. and M. Syvanen, 1995).

identified either MdGst-3 or GST-4, but more likely MdGst-3, as the gene responsible for OP resistance; taken together with the results in Table 1, MdGst-3 is likely a major factor in contributing to OP^R in strain Cornell R, and probably offers a degree of protection against lindane. In addition, the MdGST-1 isozyme has detectable levels of activity against the substrate parathion. This is consistent with Fournier's interpretation (Fournier et al. 1992) that MdGST-1 is a major player in OP resistance. However, because we found that expression of MdGST-1 is negatively correlated with OP resistance in the series of isogenic strains (Wang et al. 1991), we believe it is unlikely that MdGST-1 is important for OP resistance.

In order to determine the genetic alterations that resulted in MdGst-3 overexpression or other changes contributing to OP^R, we have begun characterizing the genomic structure of the GST genes from Cornell-R. A phage lambda genomic bank containing 17–20 kb of chromosomal fragments was prepared and screened with a probe composed of a 650-bp fragment containing almost the entire MdGst-3 reading frame from the cDNA clone; nine independent phage clones were isolated. These nine clones carried housefly DNA from five different genomic regions. Each of these four regions carried multiple loci with MdGst-3 sequences. These multiple GST loci were subcloned and their nucleotide sequence was determined; a summary of these sequences is given in Fig. 1. As is shown, six different genes closely related to MdGst-3 were found, though not one contained exactly the same sequence as the cDNA copy of MdGst-3 that we used in Table 1. Three of the genes

Table 2. Variable sites in MdGst-3 gene family[a]

	78	117	126	129	144	147	156	177	183	186	208	252	300	312	333	337	342	357	360	384	412	421	426	433	456
3C	G	G	A	C	C	A	C	T	C	G	G	T	C	A	T	A	T	G	C	T	T	T	G	T	C
3A	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.
3f1.1	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.
3j	.	.	.	.	.	.	.	.	.	.	.	.	T	G	.	.	.	.	.	.	.	.	.	.	.
3k	.	.	G	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.
3f	.	.	.	.	.	.	T	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.
3h	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	C	.
3B	.	.	.	.	.	.	T	.	.	.	.	.	.	G	.	G	.	A	T	A	.	.	A	.	.
3f6	.	.	.	.	.	.	T	.	.	.	.	.	.	G	.	G	.	A	T	A	.	.	A	.	.
3c1	.	.	.	.	.	.	T	.	.	.	.	.	.	G	C	G	.	A	T	A	.	.	A	.	.
3g	.	.	.	.	.	.	T	.	.	.	.	C	.	G	.	G	.	A	T	A	.	.	A	.	.
3l	.	.	.	.	.	.	T	.	.	.	A	.	C	G	.	G	.	A	T	A	.	.	A	.	.
3e	A	A	T	T	A	T	C	.	.	A	.	.	.	G	.	.	C	.	.	A	.	.	A	.	A

[a] The numbers above the line designate the nucleotide position in base pairs from the ATG initiating codon. MdGst-3C is the MdGst-3 (accession #73575) published before (Syvanen et al. 1994). The sequence of the variants listed as 3A-h were determined from the products of the PCR reaction, while 3f6 and 3f1.1 were from the genomic clones diagrammed in Fig. 1a. 3A, 3B, and 3C were recovered from BPM. MdGst-3f1.1 and MdGst-3f1.2 have identical sequences between base pairs 78 and 456, though they differ outside of this region.

encode proteins predicted to have nearly the same amino acid sequence as MdGst-3 but none of the three has exactly the same nucleotide sequences as MdGst-3. In addition, there are three other genes that carry MdGst-3 sequences over their 5′ half, but they carry highly diverged GST-like sequences in their 3′ half. That is, they appear to be gene fusions that must have been formed by recombination events between the different but homologous GSTs. Finally, locus nf1 was sequenced from two independent clones—the two nf1 genes, now designated nf1.1 and nf1.2, were also found to have different nucleotide sequences.

The unexpected heterogeneity in the MdGst-3 genes immediately raises the question of whether this heterogeneity is related to insecticide resistance in Cornell-R. To address this, we examined genomic copies of MdGst-3 from both the sensitive strain, BPM, and the resistant Cornell-R. BPM is an insecticide sensitive strain isolated in Europe 40 or 50 years ago (Jesperson 1990); as such we may presume that it resembles the ancestral flies from which contemporary resistant flies evolved. First, we showed that BPM carried the same three gene fusions shown in Fig. 1 as did Cornell-R (data not shown), thereby indicating that these loci preexisted in the ancestral population and therefore arose independently of selection for OP^R.

To gain a clearer picture of the heterogeneity of the genomic MdGst-3 genes, we cloned and sequenced a 424-bp region inside MdGst-3, using PCR amplification from genomic DNA, obtained from 18 independent amplifications for BPM and 17 for Cornell-R. The location of the PCR primers a and b are shown in Fig. 1 and their sequence is given in Materials and Methods. In these amplifications, we used PCR primers whose sequence matched the four different copies of the complete MdGst-3s (but would exclude the fusion genes and would also amplify some of the variable regions). We recovered three different MdGst-3-like sequences from BPM, designated MdGst-3A (found ten times), MdGst-3B (five times), and MdGst-3C (twice). The recovery of three products with this distribution is consistent with one of two hypotheses; either that there are two loci, one of which is homozygous and the other heterozygous (chi-square test, $P = 0.75$), or that there are three different homozygous loci, (chi-square test, $P = 0.2$). The strain BPM having been maintained as a single breeding colony over 50 years should be homozygous, so possibly the three-locus hypothesis for MdGst-3 is the more reasonable.

With Cornell-R, on the other hand, we found 12 different 424-bp internal fragments of GST-3 among the 17 independent isolates. Of the 12 copies found in Cornell-R, exact copies of MdGst-3A, -3B, -3C are seen; the other nine are related to, but differ by a few nucleotides from either MdGst-3A or -3B or -3C. The sequence information is summarized in Table 2, where only the variable sites are shown. We can see that genomic clones nf1 and nf6 from the original genomic bank are identical to MdGst-3A and MdGst-3B, respectively. Most members belong to one of two clades—those related to MdGst-3A and MdGst-3C on the one hand or to MdGst-3B on the other. The relationships among these sequences are shown in the distance tree (Fig. 2). The minimal replacement tree of that same data defines the same two clades associated with 3A, and 3C or 3B. The total length of the tree is 29 substitutions. The original MdGst-3 (Syvanen et al. 1994) used in the experiment in Table 1 is the same as MdGst-3C.

Sequencing mistakes or taq polymerase-induced PCR errors cannot account for the variation in the genomic

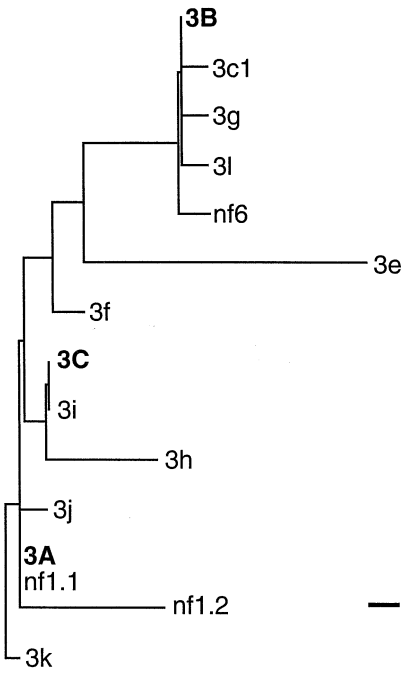

Fig. 2. Phylogram of the housefly MdGst-3 genes. The tree was constructed from an analysis of the 25 variable sites in the 424-bp internal region of MdGst-3 shown in Table 2 using the nearest-neighbor procedure of Saitou and Nei (1987) packaged in the GCG program. 3A, 3B, and 3C are the genes found only in strain BPM; all were found in Cornell R. The *space bar* represents the distance of one nucleotide substitution.

copies of MdGst-3 from Cornell-R. Consider the difference between BPM and Cornell-R. We know there are at least three different genes encoding MdGst-3 in BPM—we find these three, and only these three, among the 18 isolates from BPM. The extreme heterogeneity is only seen in the isolates from Cornell-R. We have also looked at MdGst-3 heterogeneity in a way not dependent on PCR amplification. First, after we discovered that MdGst-3 was heterogeneous at the genomic level, we determined the entire sequence in 12 independent MdGst-3 cDNA clones of which the original MdGst-3C in Table 1 was one. We found that these 12 clones now defined five different groups. But to add to the complexity, four of these are different from the genomic sequences defined above. The complete sequences for all 12 cDNA clones include the nontranslated 3′ ends which are much more heterogeneous than are the coding regions. Different members of the same group terminated with the same 3′ end. This means the sequence of the 3′ end correlated exactly with the internal differences. A random distribution of sequencing errors could not produce such a correlation. As another control for our sequencing fidelity, we sequenced a 300-bp region from 13 independent cDNA clones of the MdGST-1, a functionally constrained GST we described earlier that we believe is unrelated to insecticide resistance (Wang et al. 1991; Syvanen et al. 1994). We found identical sequences in 12 of the 13, and where one differed, at nine sites.

Discussion

In summary, we have found a total of 17 different MdGst-3 genes in the resistant strain. They were found among the λ-genomic bank, among the cDNA clones, and in the genomic copies generated by the PCR reaction. There is much greater heterogeneity among the genes from the insecticide-resistant strain Cornell-R compared to the ancestral strain BPM.

If selection for resistance to insecticides in the lineage leading to Cornell-R caused these variants of MdGst-3, then we could anticipate that these new genes code altered enzymes. If so, then we would predict that these changes code amino acid differences. We could not find strong support for this hypothesis. Of the 24 substitutions that distinguish the genes found in Cornell-R from the three genes found in BPM, nine are replacement changes while 15 are silent. For a comparison of the larger housefly GST gene family (i.e., comparing MdGst-3C to MdGST-1, -2, and -4) we have found that the replacement-changes-to-silent-changes ratio is in the range of 1:3 to 1:4. Therefore, we must consider the option that these differences among the different GST-3s are neutral.

If the strain Cornell-R carries many different kinds of MdGst-3 genes, a major question concerns their organization within the genome. We can make a few points: MdGst-3 in Cornell-R is found at a minimum of three distinct loci. In addition, one or more of these loci, as we showed previously (Wang et al., 1991), is amplified, possibly in a tandem duplication or unstable extrachromosomal element. The copy number of the tandem duplication varies from different Cornell-R populations, depending on the intensity of selection for OP^R. We would like to suggest that individual copies within this tandem duplication have diverged from one another, thus accounting for the heterogeneity seen with Cornell-R. Population heterozygosity could not account for the heterogeneity we have documented. There could be some heterozygosity since the line is not highly inbred, but it would have to be limited because the population descended from a single female and the consequent brother-sister matings on at least two occasions, and has been maintained as a single breeding colony for 22 years (i.e., approximately 260 generations).

Divergence of the sequence within this putative tandem duplication could occur by one of two means. First, many different alleles from the wild housefly population could have become concentrated into a single genome, perhaps by recombination and gene conversion events between different wild alleles and the individual copies within the tandem duplication. The presence of the different gene fusions (see Fig. 1) certainly shows that recombination events between the different GST loci do occur, so it seems possible that these genes may have evolved an enhanced recombination proficiency. Alternatively, the new copies may have evolved in the past 50 years. If this is the case, the sequence substitution rate

would have to be unusually high. For example, if these new alleles arose in the lineage leading to Cornell-R since the introduction of insecticides, then a substitution rate of 10^{-4} per nucleotide per year would be required. This is contrasted to a substitution rate of no greater than 4×10^{-9} per nucleotide per year (calculation not shown) that is based on a comparison of the housefly GSTs with those from *Drosophila* (Toung et al. 1993), where we assume 100-million-year divergence time between these two insects (Beverly and Wilson 1984). It is unlikely that Cornell-R preexisted in the ancestral population prior to insecticide selection, since the insecticide resistance in this strain is highly unstable; the fly reverts to susceptibility at a high rate.

An important mechanism for the evolution of new genes is thought to involve gene duplication followed by sequence divergence. It appears as if these two events may have occurred in the past 50 years, during which the housefly has been subject to intense pressure for insecticide resistance. Direct selection almost certainly caused the gene amplification though it is unclear if direct selection is responsible for the unusual divergence. Proof that any of the variants have enhanced insecticide-degrading activities will require isolating the separate MdGst-3 isozymes in Cornell-R that have diverged from the ancestral forms found in BPM and testing these enzymes for their ability to detoxify the insecticides dimethyl parathion and lindane. This work is in progress.

Acknowledgments. This work was supported by a grant from the US Army. The strains BPM and Cornell-R were kindly provided by Jørgen B. Jespersen and Frederick W. Plapp, Jr., respectively.

References

Beverly SM, Wilson AC (1984) Molecular evolution in *Drosophila* and higher *Diptera* II. A time scale for fly evolution. J Mol Evol 21: 1–13

Clark AG, Dauterman WC (1982) Characterization of affinity chromatography of glutathione S-transferases from different strains of housefly. Pest Biochem Physiol 17:307–314

Clark AG, Shamaan NA (1984) Evidence that DDT-dehydrochlorinase from the housefly is a glutathione S-transferase. Pest Biochem Physiol 22:249–261

Clark AG, Shamaan NA, Qureshi AR (1990) Synthesis and use of an isoform specific affinity matrix in the purification of glutathione S-transferases from the housefly *Musca domestica.* Protein Purif Exp 1:121–126

Dauterman WC (1986) Insecticide metabolism by multiple glutathione S-transferases in two strains of housefly, *Musca domestica* (L). Pest Biochem Physiol 25:169–175

Fournier D, Bride JM, Poirie M, Berge JB, Plapp FW Jr (1992) Insect glutathione S-transferases. Biochemical characteristics of the major forms from houseflies susceptible and resistant to insecticides. J Biol Chem 267:1840–1845

Habig WH, Pabst WJ, Jakoby WB (1974) Glutathione S-transferase—the first step in mercapturic acid formation. J Biol Chem 249:7130–7139

Jesperson JB (1990) Insecticide resistance. Danish pest infestation laboratory annual report. Lyngby, Denmark, pp 36–39

Oppenoorth F, Rupes V, El Bashir S, Houx N, Voerman S (1972) Glutathione-dependent degradation of parathion and its significance for resistance in the housefly. Pest Biochem Physiol 2:262–277

Pemble SE, Taylor JB (1992) An evolutionary perspective on glutathione transferases inferred from class-theta glutathione transferase cDNA sequences. Biochem J 287:957–963

Saitou N, Nei N (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Syvanen M, Zhou ZH, Wang JY (1994) Glutathione transferase gene family from the housefly *Musca domestica.* Mol Gen Genet 245: 25–31

Toung YP, Hsieh TS, Tu CP (1993) The glutathione S-transferase D genes. A divergently organized intronless gene family in *Drosophila melanogaster.* J Biol Chem 268:9737–9746

Wang JY, McCommas S, Syvanen M (1991) Molecular cloning of a glutathione S-transferase overproduced in an insecticide-resistant strain of the housefly (*Musca domestica*). Mol Gen Genet 227:260–266