Insect Molecular Biology (2008) **17**(4), 325–339

Transcriptomic profiles of *Drosophila melanogaster* third instar larval midgut and responses to oxidative stress

H.-M. Li*, G. Buczkowski*, O. Mittapalli†, J. Xie‡, J. Wu‡, R. Westerman§, B. J. Schemerhorn*¶, L. L. Murdock*†† and B. R. Pittendrigh*††

**Department of Entomology, Purdue University, West Lafayette, IN, USA; †Department of Molecular Ecology, Max-Planck Institute for Chemical Ecology, Jena, Germany; ‡Department of Statistics, Purdue University, West Lafayette, IN, USA; §Genomics Facility, Whistler Hall, Purdue University, West Lafayette, IN, USA; ¶The USDA-ARS, West Lafayette, IN, USA; and ††MPRINT – Molecular Plant Resistance and Nematode Team, Department of Entomology, Purdue University, West Lafayette, IN, USA*

Abstract

Oligoarray analysis was used to determine the number and nature of genes expressed in third instar *Drosophila melanogaster* larval midguts. The majority of transcripts were associated with protein synthesis and metabolism. Serine proteases were the main proteolytic enzymes detected. Some 40% of the cytochrome P450 genes and 74% of the glutathione S transferases (GSTs) in the genome of *D. melanogaster* were observed to be expressed in the midgut by oligoarray analysis. We also identified potential transcription factor binding motifs (TFBMs) of P450s, GSTs and carboxylesterases. Many of the midgut-expressed GST genes contained candidate TFBMs homologous to TFBMs in mammals that have been associated with responses to oxidative stress. We also investigated the response of GSTs in the midgut to dietary H_2O_2, which showed a dosage-based differential response.

Keywords: glutathione S-transferase, cytochrome P450, transcription factor binding motif, serine protease, midgut, insect.

Received 4 September 2007; accepted after revision 4 April 2008. Correspondence: Barry R. Pittendrigh, Department of Entomology, Purdue University, 901 West State Street, West Lafayette, IN 47907, USA. Tel.: + 1 765 494 7730; fax: +1 765 494 0535; e-mail: pittendr@purdue.edu

Introduction

Over the past several decades interest has grown in pesticides that target the insect digestive system. Particular attention has been given to proteins that affect insect-specific gut receptors because the genes encoding such proteins can be used in transgenic plants to serve as plant defences. For example, the soil bacterium *Bacillus thuringiensis* (*Bt*) produces a toxic protein that targets receptors in the insect midgut. *Bt* is considered to be a safe insect control agent from the perspective of both human health and the environment.

A significant threat to *Bt*-based insect control is the potential for resistance to evolve (Tabashnik *et al*., 1994; Bauer, 1995; Schnepf *et al*., 1998; Ferre & Van Rie, 2002; Chilcutt & Tabashnik, 2004). Although efforts have been made to discover other toxins useful in transgenic insect control, genes encoding *Bt* toxins are the only group of proteins that have been proven to be of commercial value thus far. Given the dearth of useful resistance gene classes, there is a need to identify novel potential target sites in the insect digestive system that may be exploited for the development of novel classes of insect control agents. For example, Baum *et al*. (2007) have recently demonstrated the concept of using transgenic plants expressing insect-specific dsRNAs as an insect control strategy. Discovery of novel insect digestive system target sites that may ultimately be useful in such a strategy will require a better understanding of the genes and corresponding proteins that are expressed in the insect midgut.

Better understanding of the genomics of the insect digestive system may also provide new insights into: (1) human digestion, nutrition, diseases of the digestive system and issues associated with weight gain, or loss, or both; (2) the evolution of the insect digestive system, which represents a crucial interface between the animal and its environment; and (3) a better understanding of plant–insect interactions. For example, it has been hypothesized that the divergence in total numbers of cytochrome P450s and glutathione S transferases (GSTs) associated with different insect genomes is a reflection of the diversity of the toxins they experience in their diet (Enayati *et al*., 2005). Thus, in the genomes of polyphagous insects we would expect a greater

number of genes encoding detoxification enzymes compared to monophagous insects. As a corollary to this hypothesis, if cytochrome P450s, GSTs and carboxylesterases play important roles in detoxification of compounds found in the diet, then we would expect to find a large number of P450s and GSTs associated with the midgut. To date, no comprehensive study has addressed the question of how many P450s, GSTs and carboxylesterases are associated with the midgut.

Currently, we have a limited understanding of the genomic profile of any insect digestive system. Recently, expressed sequence tag (EST) studies and cDNA spotted arrays have been used in developmental and nutritional studies as well as in an attempt to determine the complement (or subset) of genes specifically associated with insect digestive systems (Zinke *et al*., 2002; Li & White, 2003; Pedra *et al*., 2003; Arvestad *et al*., 2005; Mittapalli *et al*., 2005; Siegfried *et al*., 2005). The EST approach, however, may not accurately reflect the complement of genes being expressed (Wang *et al*., 2004).

Thanks to the availability of whole-genome oligoarrays in *Drosophila* (Adams *et al*., 2000), we now have the capacity to describe many of the genes expressed in a given tissue, i.e. those that are expressed at or above the level of detectability in the oligoarrays. Even so, we still do not know the complement of genes associated with the midgut of third instar *Drosophila* larvae (Li & White, 2003). The third instar *Drosophila* larva represents a developmental stage at which (1) high level of feeding occurs, (2) effective feeding bioassays can be performed, and (3) the midgut is large enough to be easily dissected.

In this study, we investigated the genes associated with the midgut of third instar *D. melanogaster* larvae. This allowed us to determine the number of cytochrome P450s and GSTs associated with the midgut. We also identified the potential DNA transcription factor binding motifs (TFBMs) of cytochrome P450s, GSTs and carboxylesterases expressed in the third instar larval midguts. As some of the candidate TFBMs for the midgut-associated GSTs were observed to be linked to the oxidative stress response, we also tested the effect of an oxidative stressor, hydrogen peroxide, on expression of GSTs.

Results and discussion

Functional annotations of Drosophila melanogaster *midgut transcriptome*

We observed 2943 probes ($P < 0.001$, ANOVA) representing 2923 different transcripts to be present on all three replicates of the oligoarrays (Supplementary Material Table S1). Sixty five per cent (1908/2923) of genes were functionally annotated using the gene ontology database. Based on the function annotations of the Affymetrix's NetAffx™ Analysis Center (http://www.affymetrix.com/analysis/index.affx), the transcripts were assigned to one of eight functional categories. A group of 810 of these genes (42.5%) was associated with protein synthesis and metabolism. It included a suite of genes involved in protein catabolism, biosynthesis, modifications, proteolysis and degradation. Two-hundred and forty-four genes were associated with transport (12.8%), 229 genes were associated with the cytoskeleton (12.0%) and 189 genes coded for proteins associated with DNA/RNA binding (9.9%). The number of genes involved in detoxification, transcription/translation, development and signalling were respectively 133 (7.0%), 121 (6.3%), 93 (4.9%) and 89 (4.7%) (Supplementary Material Table S1). Many genes in the midgut were associated with metabolism and protein synthesis; corroborating the results of Arbeitman *et al*. (2002).

EST midgut libraries vs. microarrays

Previous transcriptome studies of insect midguts involved sequencing and annotation of ESTs (Lehane *et al*., 2003; Pedra *et al*., 2003; Mittapalli *et al*., 2005; Siegfried *et al*., 2005). For example, in the tsetse fly (*Glossina morsitans*), Lehane *et al*. (2003) observed 8876 potential genes associated with the adult midgut, almost triple the number of expressed gene products that we observed in the *D. melanogaster* midgut. This difference could be because of: (1) differences among species; (2) different developmental stages or physiological states examined; (3) differences in techniques (e.g. selection criteria); or, (4) a combination of 1–3. The EST approach is known to overestimate the number of unique genes because of EST clustering errors, alternative splicing or both (Wang *et al*., 2004). Alternative splicing has previously been documented in insects (Hanrahan *et al*., 1999; Reenan, 2005). Thus, the oligoarray approach taken here may under-estimate the number of genes actually expressed in the *D. melanogaster* midgut, particularly those genes with low expression levels.

With the advantage of the availability of other insect oligoarrays we will be in a better position to make comparisons between species regarding (1) the nature of the genes associated with the insect midgut, and (2) the midgut-specific gene families that have gone through evolutionary expansion or contraction (Hahn *et al*., 2005). Such observations may give us greater insights into how food choices have shaped the evolution of the insect digestive system (e.g. generalists vs. specialists, legume feeders vs. cereal feeders or herbivores vs. carnivores).

Serine proteases and digestion

Digestive enzymes are critical for insects to process proteins, carbohydrates, lipids and other materials they imbibe with their diet. In the present study we identified 148 genes involved in proteolysis and peptidolysis based on their known molecular functions (Table 1). Serine-type endopeptidases (E.C.3.4.21), including trypsins (E.C.3.4.21.4)

Table 1. Genes related to digestion and detoxification expressed in third instar *Drosophila melanogaster* larval midgut. Classification was based on the functional annotation

Molecular functions	Number of genes
Proteolysis and peptidolysis	148
• Serine-type endopeptidase/trypsin /chymotrypsin	49
• Endopeptidases	22
• Carboxypeptidase	10
• Metallo-endopeptidase	11
• Ubiquitin-protein ligase	8
• Cysteine-like protease	4
• Aminopeptidase	2
• Others	42
Detoxification and P450s	133
• Glutathione S transferases	28
• Cytochrome P450s	36
• Carboxylesterases	9
• Others	60

and chymotrypsins (E.C.3.4.21.1), accounted for 33% (49/148) of all the proteolysis and peptidolysis enzymes present in the *D. melanogaster* midgut transcriptome (Table 1).

Our observations are consistent with results by Li & White (2003) showing that serine proteases are the most common proteolytic enzymes associated with the *Drosophila* midgut. Digestive serine proteases (trypsins and chymotrypsins), which function in the breakdown of ingested food, represent the major proteolytic enzymes in the midguts of many lepidopterans and dipterans (Chapman, 1998). Even so, some of the serine proteases observed in the midgut transcriptome may be involved in biological roles other than food digestion. For example, serine proteases are known to activate immune pathways that respond to pathogen detection (Gorman & Paskewitz, 2001).

Unlike *Drosophila*, the cowpea bruchid, *Callosobruchus maculatus* (F.), relies primarily on cysteine proteases to digest dietary proteins (Murdock *et al*., 1987). Pedra *et al*. (2003) observed mainly cysteine protease ESTs in the digestive system of cowpea bruchid larvae. Although Pedra *et al*. (2003) sequenced only 503 ESTs from the bruchid gut, six of the total ESTs (1.2%) coded for cysteine proteases while none coded for serine proteases. Our results do suggest that serine proteases are more common in the midguts of *D. melanogaster* compared to cowpea bruchids.

One hypothesis to explain the observed differences between *D. melanogaster* and cowpea bruchids is that seed-feeding insects (especially those that feed on legume seeds as do cowpea bruchids) encounter high levels of serine-specific protease inhibitors (e.g. Bowman–Birk type and Kunitz trypsin inhibitors) in their diet and this may have been the driving evolutionary force (selection pressure) behind the expansion of the cysteine proteases expressed in the midgut of *C. maculatus*. From the *Drosophila* genome database, there are 18 known serine proteases and one known cysteine protease (Flybase, http://flybase.bio.indiana.edu). However, there is no serine protease, and six known cysteine proteases in the ESTs of *C. maculatus* (Pedra *et al*., 2003). In addition, many coleopterans including *C. maculatus* have an acidic gut environment (Wolfson & Murdock, 1990). Such conditions may have favoured the expansion of cysteine protease families in the midgut. In keeping with this latter hypothesis, the western corn rootworm, *Diabrotica virgifera* LeConte, (1) has an acidic gut (Fabrick *et al*., 2002) and (2) an EST analysis of its midgut transcriptome revealed a large class of the cysteine proteases and only one serine protease (Siegfried *et al*., 2005). In contrast, the basic pH of the *D. melanogaster* midgut is very compatible with a large number of serine proteases.

Aminopeptidases and cadherins

We detected two aminopeptidases and no cadherins in the *D. melanogaster* midgut transcriptome. The aminopeptidases identified were granny smith (EC 3.4.11.1) and aminopeptidase P (EC 3.4.11.9). Both enzymes belong to different classes from that of aminopeptidase N (APN) (EC 3.4.11.2). Although an APN is encoded in the *D. melanogaster* genome, we did not detect its expression in our microarray analysis.

In Lepidoptera, cadherins and APN are thought to be receptors for *Bt* toxins in the midgut epithelial cells. Mutations in both cadherin and APN genes in lepidopterans have been associated with the resistance phenotype to *Bt* toxins (Ingle *et al*., 2001; Jenkins & Dean, 2001; Bel & Escriche, 2006). APN is considered as an essential gut enzyme associated with the binding of *Bt* in lepidopterans (Knight *et al*., 1994). *D. melanogaster* is not particularly susceptible to *Bt* Cry1 toxins (Gill & Ellar, 2002). However, when Gill & Ellar (2002) made transgenic *Drosophila* expressing an APN gene from *Manduca sexta* (L.) in the *Drosophila* digestive system, the transgenic *Drosophila* became susceptible to *Cry1Ac1*. Based on these observations, we suggest that wild-type *D. melanogaster* third instar larval midgut (1) do not express the APN or cadherin genes or (2) express the transcripts of these genes at levels below the sensitivity range of the oligoarrays.

Cytochrome P450s

Cytochrome P450 enzymes are involved in (1) biosynthesis and degradation of endogenous compounds and (2) metabolism of xenobiotics such as allelochemicals, drugs and insecticides (Feyereisen, 1999; 2005). Dietary factors can induce P450s (Snyder *et al*., 1995; Mittapalli *et al*., 2005; Krishnan & Kodrik, 2006). The *D. melanogaster* genome contains 90 cytochrome P450s (Tijet *et al*., 2001; Feyereisen, 2005) and in this study we observed 36 (40%, Table 1) expressed in the larval midgut. By comparison, Li & White (2003) only observed nine P450s (*Cyp4Ac3*, *Cyp4d2*, *Cyp6a14*, *Cyp6d4*, *Cyp9b2*, *Cyp9f2*, *Cyp12a4*, *Cyp12a5* and *Cyp311a1*) in the midgut of third instar

Table 2. Glutathione S transferases being observed in third instar *Drosophila melanogaster* larval midgut transcriptome by oligoarray analysis

GST class	Observed in the midgut transcriptome	Not observed in the midgut transcriptome
Delta	(GstD1, GstD2, GstD3, GstD7, GstD9, GstD10)*, (GstD4, GstD6, GstD8)†	GstD5†, CG17639
Epsilon	GstE1*, GstE3, GstE5, GstE6, GstE7, GstE8, GstE9, GstE10, CG5224, CG16936, CG11784	GstE2, GstE4, CG4688
Sigma	GstS1*	
Theta	CG1681, CG1702, CG30005	CG30000
Omega	CG6776, CG6673	CG6781, CG6662
Zeta	CG9362	CG9363
Gst-like	Mgstl	gfzf

*GSTs known to accept 4-hydroxynonenal (4-HNE) as a substrate (Sawicki *et al*., 2003).
†GSTs reported not to accept 4-HNE as a substrate (Sawicki *et al*., 2003).
All other GSTs were not tested by Sawicki *et al*. (2003) for 4-HNE substrate activities.

Drosophila larvae. The discrepancy in the number of P450s observed may be a result of the different types of arrays used or the methodology used to call the gene present. In contrast to the present study, Li & White (2003) used polylysine-coated glass slide arrays and an incomplete set of genes (of the total in the genome) from *Drosophila*.

However, our observations suggest that more P450s are transcribed in the midgut than previously thought (Li & White, 2003). It remains to be determined what role, if any, these 36 P450s, may play in xenobiotic metabolism of dietary materials.

Glutathione-S-transferases and oxidative stress

Glutathione-S-transferases (GSTs, EC 2.5.1.18) are a family of multi-functional enzymes involved in the detoxification of endogenous as well as xenobiotic compounds (Tu & Akgul, 2005). GSTs catalyse the conjugation of reduced glutathione (GSH) with xenobiotics such as plant allelochemicals and insecticides, thereby converting them to less toxic water-soluble forms (Grant & Matsumura, 1989; Zou *et al*., 2000; Singh *et al*., 2001). GSTs also play a role in oxidative stress, a condition cells experience when there is an increase in reactive oxygen species (ROS) such as superoxide, hydrogen peroxide and peroxynitrite (Fiers *et al*., 1999).

The *D. melanogaster* genome contains 38 GSTs (Tu & Akgul, 2005), 28 of which were observed in our microarray analysis (Table 1). We found five different GST classes in the third instar *D. melanogaster* midgut transcriptome (Table 2). Most of the GSTs in the midgut transcriptome were in the *delta* and *epsilon* classes. To validate the sensitivity of the oligoarray experiment, we carried out quantitative real-time PCRs (qRT-PCRs) to independently estimate gene expression levels (Fig. 1).

We measured the expression levels of GSTs absent in the microarrays (not midgut-associated) as well as GSTs observed in the microarrays (midgut-associated). Overall, 26 of the 28 GSTs we found associated with the midgut by microarray analysis were confirmed by qRT-PCRs to have higher expression levels than the genes defined as absent in the microarray experiments (Fig. 1 and Table 3). *GSTD8* and *MgstI* were observed to be midgut-associated in microarrays, but had very low expression levels as defined by qRT-PCR. The *Ct* values of all GST genes vs. the *Ct* values of the reference gene *RP49* are presented in Fig. 1.

Interestingly, Li & White (2003) did not observe any GSTs in their microarray experiments with *Drosophila* midgut. One difference between observations by Li & White (2003) and our results is that their arrays were made using a midgut-enriched EST library while our study employed Affymetrix microarrays. The microarrays performed by Li & White (2003) may have lacked some of the GSTs, have been less sensitive at detecting the GST transcripts as compared to our oligoarrays and/or their methodology of designating absent or present may have been different. However, our results suggest that a substantial number of GSTs in the *Drosophila* genome may be associated with the midgut. The functional role of these GSTs in the midgut remains to be elucidated.

The GSTs identified in this study could possibly be responsive to (1) midgut fermentation products (Ebert *et al*., 2003) or (2) oxidative stress in the midgut cells. For example, 4-hydroxynonenal (4-HNE) is a highly toxic aldehyde produced by lipid peroxidation in cells in response to oxidative stress (Ullrich *et al*., 1997). The major pathway of 4-HNE metabolism in mammals involves glutathione conjugation by GSTs (Alin *et al*., 1985; Videla *et al*., 2000). 4-HNE-protein adducts also occur in insects (Yan & Sohal, 1998; Singh *et al*., 2001). Thus, Sawicki *et al*. (2003) investigated the role of *Drosophila* GSTs in metabolizing 4-HNE. They found that six *delta* GSTs (*GSTD1*, *GSTD2*, *GSTD3*, *GSTD7*, *GSTD9* and *GSTD10*), one *epsilon* GST (*GSTE1*) and one *sigma* GST (GSTS1) metabolized 4-HNE. Interestingly, mRNA for each of these eight GSTs was found in our oligoarray analysis of the *D. melanogaster* larval midgut transcriptome (Table 2).

Identification of potential GST, cytochrome P450 and carboxylesterase regulatory regions

We compared the promoter regulatory regions corresponding to 800 bp upstream and 200 bp downstream of

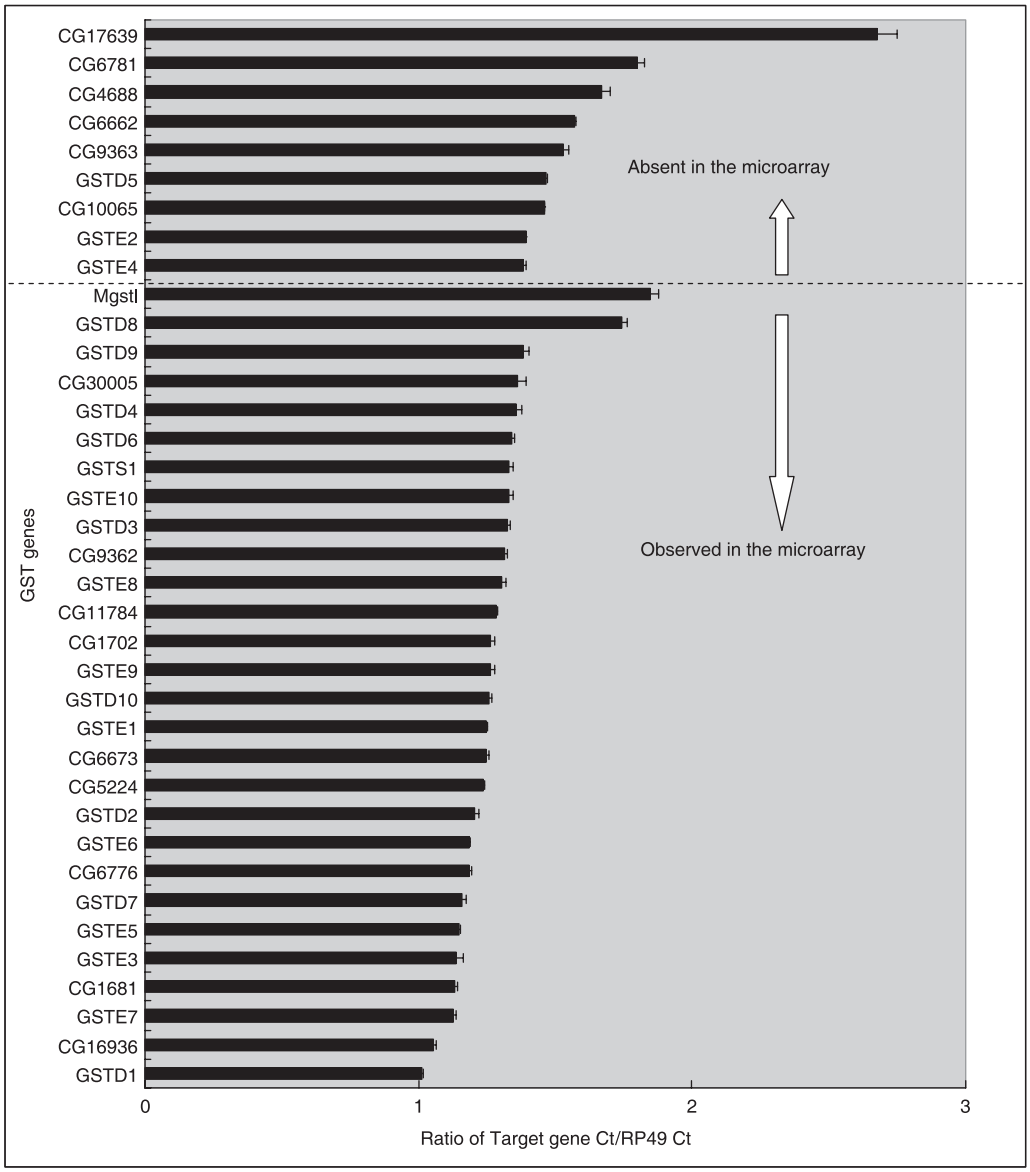

Figure 1. Fold changes and the standard errors of all glutathione S transferase (GST) transcripts in third instar *Drosophila melanogaster* larval midgut. RP49 is the reference gene.

the target gene transcription start site of cytochrome P450s and GSTs that were either associated or not associated with the midgut. Four potential transcription factor binding motifs (TFBMs) were observed for the midgut-associated GSTs: fork head box (FOX)-, glucocorticoid receptor element (GRE)-, E47- and NF-kappaB-like (Figs 2 and 4A), and four different potential TFBMs were observed for the midgut-associated P450s: TATA box-binding (TBP)-, hypoxia-inducible factor 1(HIF-1-), c-Myc/Max- and liver X receptor (LXR)-like (Figs 3, 4B and S1). Roles for each of these aforementioned TFBMs have been elucidated (Table 3) in mammals. However, it has hitherto not been known if the four similar motifs that we observed upstream of the midgut-associated GSTs have similar biological roles in *D. melanogaster*. For example, insects are not known to have glucocorticoids or glucocorticoid receptors (Yoshinaga & Yamamoto, 1991); however, one of the potential TFBMs was homologous to the mammalian glucocorticoid receptor element (GRE). If these potential GRE-like TFBMs in *D. melanogaster* are functional, their role in the life-history and cellular biology of this insect remains to be elucidated.

The four potential TFBMs observed for the midgut-associated P450s (Figs 3 and 4B) were not shared with the nonmidgut-associated P450s. These potential TFBMs (Table 3) have been reported in mammals with the initiation of transcription (van Roon-Mom *et al.*, 2005), gene expression

Table 3. Potential transcription factor binding sites of GSTs, cytochrome P450s, and carboxylesterases observed in third instar *Drosophila melanogaster* larval midgut transcriptome using oligoarray analysis

Gene group	TF motif name*	Known mammalian function	Distance†	Abundance rate§
GSTs	NF-kappaB-like	Nuclear factor kappa B is regulated by the intracellular reduction-oxidation state (Sen & Packer, 1996)	*0.09*	*0.58*
	glucocorticoid receptor element (GRE)-like	Suppressed under oxidative conditions and restored in the presence of reducing reagents (Okamoto *et al*., 1999)	*0.06*‡	*0.69*
	Fork head box (Fox)-like	Associated with signalling pathways and cell fate decisions in development (Kaestner *et al*., 2000)	*0.08*	*0.77*
	E-47-like	A helix-loop-helix protein (HLH), essential for both B- and T-lineage development (Ikawa *et al*., 2006)	*0.13*	*0.58*
P450s	TBP-like	TATA Box-binding protein is a general transcription factor that act on initiation of transcription (van Roon-Mom *et al*., 2005)	*0.06*‡	*0.78*
	HIF-1 like	Hypoxia-inducible factor-1, playing a role in formation of new blood vessels because of low oxygen levels (Semenza *et al*., 2006)	*0.11*	*0.47*
	c-Myc/Max-like	Plays a role in nitric oxide-mediated suppression of the small intestine-associated cytochrome P450 *Cyp3a4* (Watabe *et al*., 2003)	*0.05*‡	*0.50*
	LXR-like	Liver X receptors, associated with oxysterol inhibition of phenobarbital (PB)-mediated induction of cytochrome P450s in mammals (Gnerre *et al*., 2005)	*0.12*	*0.56*
Carboxyl esterase	PR-like	Progesterone receptor, an intracellular steroid receptor that specifically binds progesterone (Leonhardt *et al*., 2004)	*0.07*‡	*0.78*
	IRF-like	Interferon-regulating factors, cell growth inhibition (Kirchhoff *et al*., 1993)	*0.06*‡	*0.67*
	TFII-I-like	Gene regulation (Cheriyath & Roy, 2000)	*0.10*	*0.67*
	HSF2-like	Heat shock transcription factor (He *et al*., 2003)	*0.14*	*0.56*

*The transcription factor binding motif (TFBM) names were assigned based on the closest known motifs described in other organisms.
†The distance was defined as a dissimilarity measurement between any pair of position weight matrices (PWMs; Gordon *et al*., 2006). If the numerical values were smaller, the PWMs were considered to be closer to the mammalian TFBMs.
‡A distance of 0.07 or less indicated that the *Drosophila melanogaster* TFBM was relatively similar to the mammalian TFBM.
§Abundance rate was the frequency that a motif occurred in the given set of sequences.

at low oxygen levels (Semenza *et al*., 2006), nitric oxide-mediated suppression of the small intestine-associated cytochrome P450 (Watabe *et al*., 2003) and oxysterol inhibition of phenobarbital-mediated induction of P450s (Gnerre *et al*., 2005). However, to date we do not know what role, if any, these potential TFBMs play in *D. melanogaster*. Additionally, in those cases where such potential TFBMs are found to be functional, we currently do not know if their roles in *D. melanogaster* may be similar to those found in mammals.

Carboxylesterases are also involved in detoxification of xenobiotics (Russell *et al*., 1996) and in the metabolism of lipids and fats in *Drosophila* larvae (Campbell *et al*., 2003). We detected transcripts of nine carboxylesterases (Fig. 5) associated with the midgut of third instar *D. melanogaster* larvae. Among these transcripts, Esterase-3 and Esterase-7 have previously been reported to be expressed in *Drosophila* larvae (Campbell *et al*., 2003). Four potential TFBMs (progesterone receptor-like, interferon-regulating factors-like, TFII-I-like and heat shock factor 2-like) were observed among these nine midgut-associated carboxylesterases (Table 4 and Fig. 5). As insects are not known to have progesterone or interferon, the role of these potential TFBMs, if any, remains to be elucidated.

Further molecular studies are needed to determine if the detected TFBMs are functional and whether they have a role, if any, in the expression patterns of the given genes in *D. melanogaster*. We realise that some, or even all, of these motifs may not be functional. However, the analysis presented here (Figs 2–4) provides the basis for formulating testable hypotheses relating to the regulation of the digestive system-associated GSTs, P450s and carboxylesterases in the midgut of larval *D. melanogaster*.

As the strategy we used to observe the potential TFBMs was *de novo* motif discovery involving a combination of five multiple alignment approaches, it is likely that we failed to identify other potential TFBMs that exist in the 5′ region of the genes. Two motifs previously shown to be involved in the regulation of GSTs and P450s include antioxidant responsive element (ARE) and xenobiotic responsive element (XRE) (Rushmore *et al*., 1991; Zhang *et al*., 1998). We scanned for these two motifs in the 2 kb upstream

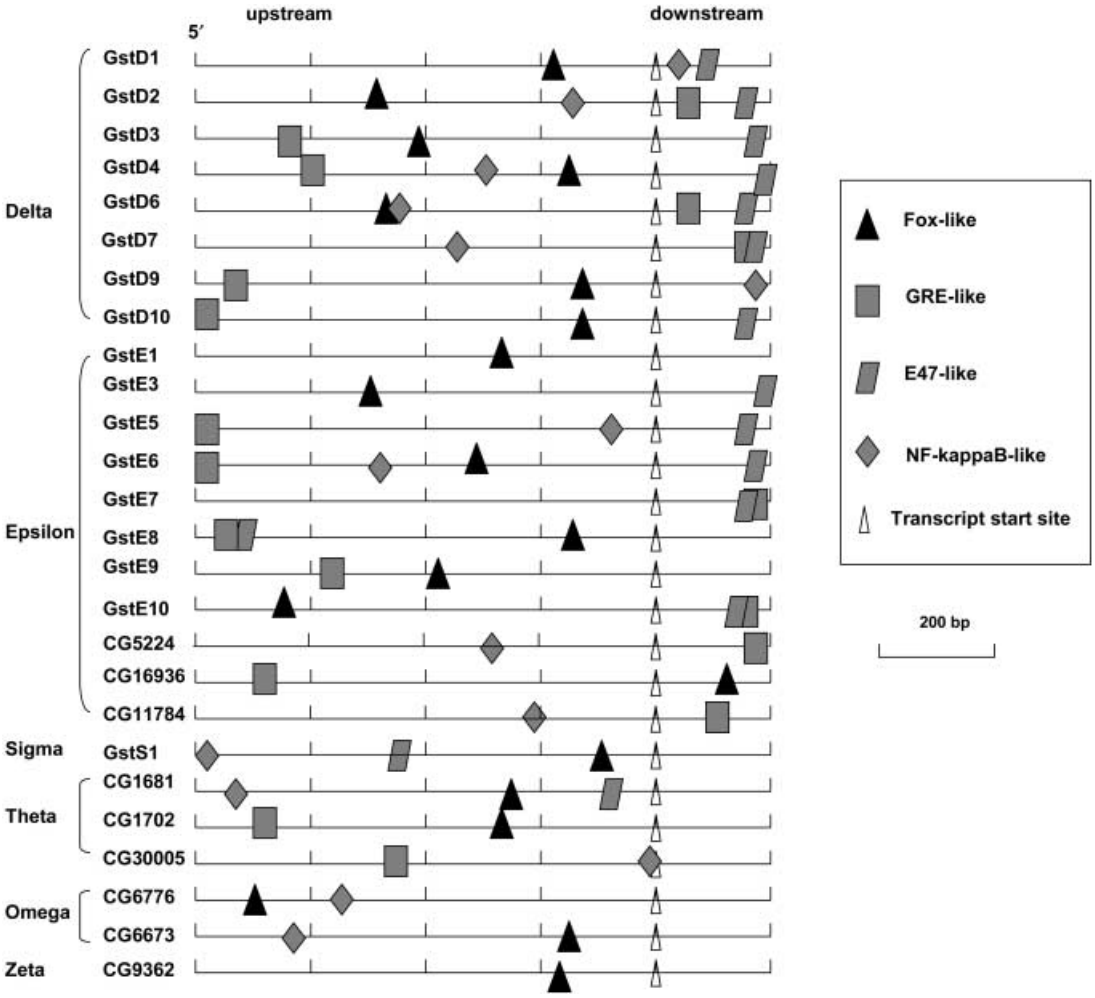

Figure 2. Glutathione S transferases (GSTs) associated with third instar *Drosophila melanogaster* larval midgut and the potential transcription factor binding motifs (TFBMs) observed 800 bp upstream and 200 bp downstream of the gene transcription start site (TSS). Different shapes represent different potential motifs. All GSTs are labelled by the gene name and categorized into their respective classes. The potential ARE and XRE TFBMs associated with the GSTs expressed in the midgut (based on microarray analysis) are given in Supplementary Material Fig. S2.

regulatory region of all the midgut associated GSTs, cytochrome P450s and carboxylesterases. The results are given in Supplementary Material Figs S2 and S3. Four ARE-like (sequence 5′-GTGACNNNG-3′) and 15 XRE (sequence 5′-GCGTG-3′) motifs were observed for the GSTs, while three ARE (sequence 5′-GTGACNNNGC-3′) and 23 XRE motifs were observed for the cytochrome P450s. We observed three carboxylesterases genes associated with the potential ARE-like motifs (sequence 5′-GTGACNNNG-3′), *CG3841*, *clt* and *Esterase-5*, and five carboxylesterase genes associated with the putative XRE motifs out of a total of nine midgut-associated carboxylesterases (Supplementary Material Fig. S4).

However, we also performed a similar screening on 200 randomly generated sequences and observed that 84.5 and 12.5% of these randomly generated sequences, respectively, contained matches to XRE and ARE. We further observed that 54, 64 and 56% of the respective midgut-associated GSTs, P450s and carboxylesterases contained the XRE motif. We also observed that 14, 8 and 33% of the respective midgut-associated GSTs, P450s and carboxylesterases contained the ARE motifs. Thus, the XRE and ARE motifs were typically more frequently observed associated with the randomly generated sequence than with the GSTs, P450s and carboxylesterases genes. It is possible that these potential XRE and ARE TFBMs were false positives; neither motifs were obtained from genes associated with the GSTs, P450s and carboxylesterases using the five motif discovery programs, which are *de novo* methods without prior information on the motif pattern. Alternatively, it is also possible that some of these AREs and XREs are functional; however, which (if any) of these motifs play a biological role in these aforementioned genes remains to be verified.

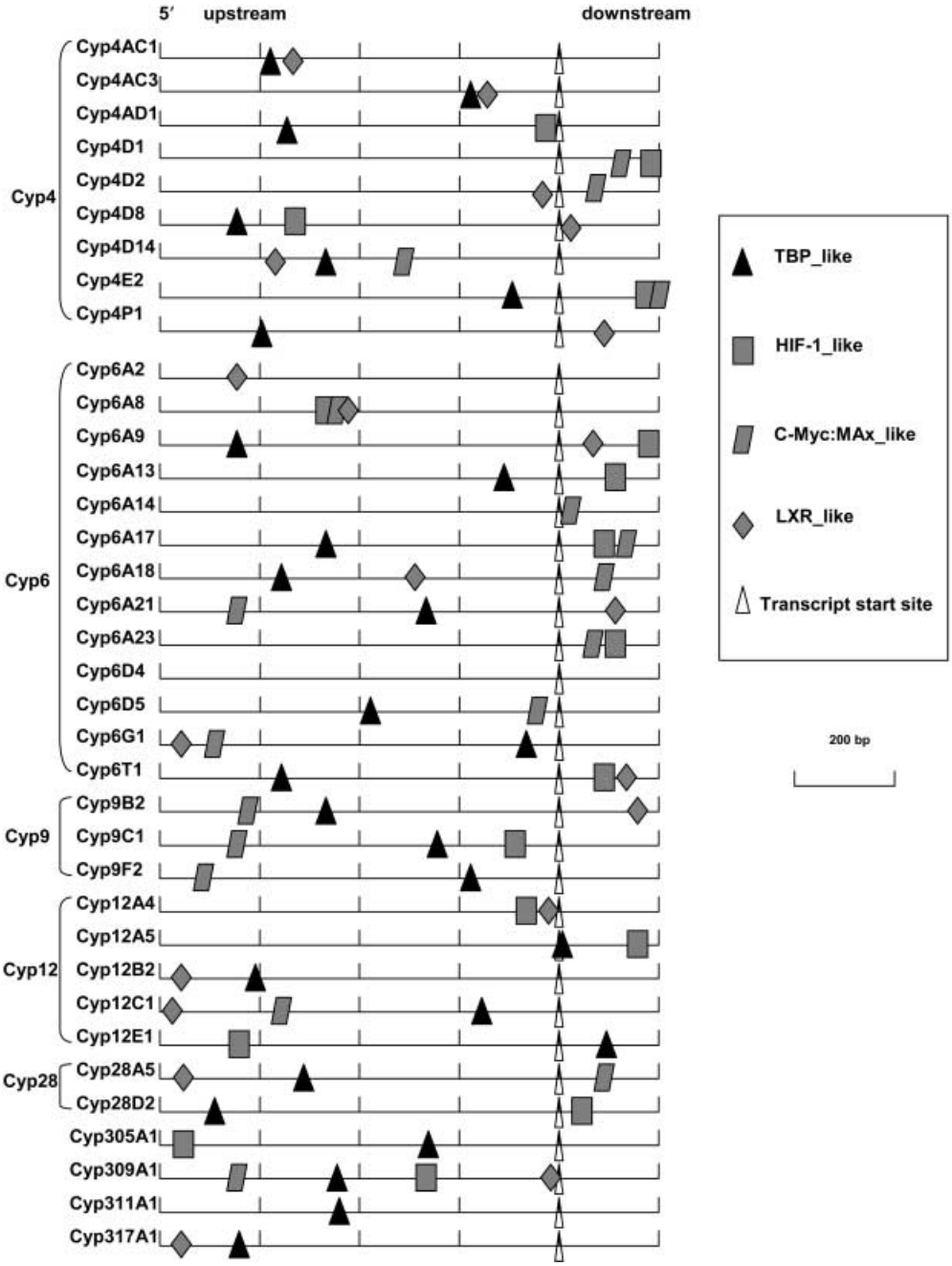

Figure 3. Cytochrome P450s associated with third instar in *Drosophila melanogaster* larvae midgut and the potential transcription factor binding motifs (TFBMs) observed 800 bases upstream and 200 bases downstream of the gene transcription start site (TSS). Different shapes present different potential motifs. All P450s are labelled by the gene name and listed in their respective classes. The potential ARE and XRE TFBMs associated with P450s expressed in the midgut (based on microarray analysis) are given in Supplementary Material Fig. S3.

GST response to different levels of dietary hydrogen peroxide

In order to determine whether the different midgut-associated GSTs responded to oxidative stressors in a similar manner, we fed hydrogen peroxide (H_2O_2) to *D. melanogaster* third instar larvae for 12 h at doses of 0.02, 0.1 and 0.5% (w/w) in artificial diets. Twelve of the genes with potential ARE and XRE motifs in their 5′ regions responded to at least one dose of dietary H_2O_2 (Table 3 and Supplementary Material Fig. S2) at the $P < 0.01$ level. Nine GST genes lacking both the potential XRE and ARE were differentially expressed in at least one dose of dietary H_2O_2 (Table 4 and Supplementary Material Fig. S2) at the $P < 0.01$ level. Further,

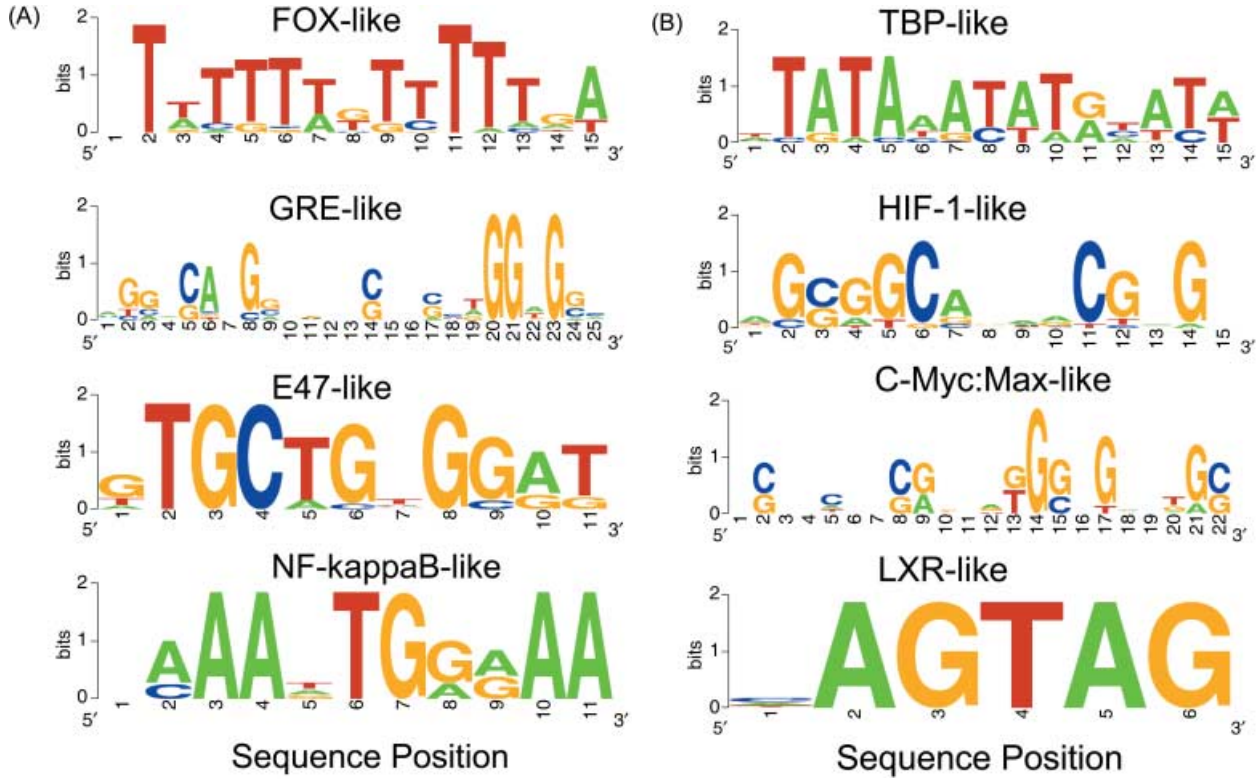

Figure 4. Sequence logo demonstration of all the transcription factor binding motifs (TFBMs) constituting the glutathione S transferases (GSTs) and P450s observed in the third instar *Drosophila melanogaster* larval midgut transcriptome using oligoarray analysis. (A) A list of potential TFBMs of the observed GSTs. (B) A list of potential transcription factor binding motifs of observed P450s.

three GST genes with potential XRE and/or ARE TFBMs were not differentially expressed in the presence of dietary H_2O_2 at the $P < 0.01$ level (Table 4 and Supplementary Material Fig. S2).

However, a trend did occur among some of the expression patterns of the GST genes and their potential TFBMs (including FOX-like, GRE-like, E-47-like and NK-kappB-like) (Fig. 6 and Supplementary Material Fig. S1). Based on the expression patterns of different GSTs that responded to dietary H_2O_2, we observed several types of FOX-like potential TFBMs, which we termed FOX-like-1, FOX-like-2 and FOX-like-3 (Supplementary Material Fig. S1). The potential FOX-like-1 TFBMs were associated with *GSTD3*, *GSTD9* and *GSTD10* and the potential FOX-like-2 TFBMs, which were different from the base consensus of FOX-like-1 were associated with *GSTS1*, *CG6776* and *CG1681* (Fig. 6 and Supplementary Material Fig. S1). The potential FOX-like-3 TFBM was associated with *GSTD1*, *GSTD2*, *GSTE1*, *CG16936* and *CG9362*, which also contained consensus sequences that differed from the two aforementioned groups *CG1681* (Fig. 6 and Supplementary Material Fig. S1). Two separate consensus sequences emerged among the potential GRE-like TFBMs associated with (1) *GSTD3*, *GSTD9* and *GSTD10* and (2) *GSTE5*, *GSTE6*, *GSTE9* and *CG1702* (Fig. 6 and Supplementary Material Fig. S1). Both *GSTE3* and *GSTE7* shared the common potential E-47-like TFBM and *GSTS1*, *CG6776* and *CG1681* also shared the potential NK-kappB-like TFBM (Fig. 6 and Supplementary Material Fig. S1). For the other nine differentially expressed GSTs (in the presence of dietary H_2O_2), no clear trend emerged in their association with the potential TFBMs.

We do not know if some or all of the potential TFBMs are functional, and in those cases where they are functional, what specific role, if any, they play in the response of the midgut-associated GSTs to the oxidative stressor, dietary H_2O_2. They may simply be associated with these genes, with limited or no role in response to this oxidative stressor. However, one thing is clear – there are likely to be other TFBMs, beyond ARE and XRE, involved in the response of the midgut-associated GSTs to dietary H_2O_2.

Experimental procedures

Fly culture maintenance and tissue dissection

The w^{1118} *D. melanogaster* strain was obtained from Dr Misha Ludwig at the University of Chicago. The population was reared in an environmental chamber at 24 °C, 60% humidity and 14 : 10 L/D cycle. Adult flies were transferred to fresh diet (Formula 4-24® Blue, Carolina Biological Supply Co., Burlington, NC, USA) every three weeks to maintain the strain. Eggs were allowed to develop to early third instar larvae over a 5-day period (108–110 h after the

Table 4. Quantitative real-time PCR results for 26 glutathione S transferase (GST) genes in response to three doses (0.02, 0.1 and 0.5% w/w) of dietary hydrogen peroxide. Fold expression is given for each transcript (gene name is given) at each dose, along with the SE and the *P*-value.

	0.02% H_2O_2 *			0.1% H_2O_2 *			0.5% H_2O_2 *		
Gene name	Fold Expression†	ΔCt + SE	P-value	Fold Expression†	ΔCt + SE	P-value	Fold Expression†	ΔCt + SE	P-value
GSTD1	0.82	−0.29 ± 0.16	0.0840	0.90	−0.15 ± 0.20	0.4692	1.13	0.18 ± 0.18	0.3370
GSTD2	1.05	0.07 ± 0.10	0.5040	0.77	−0.38 ± 0.13	0.0145	0.97	−0.05 ± 0.35	0.8903
GSTD3	1.48	0.57 ± 0.51	0.2826	1.23	0.30 ± 0.42	0.4888	**3.36**	**1.75 ± 0.36**	**< 0.0001**
GSTD4	**0.72**	**−0.47 ± 0.12**	**0.0006**	**0.40**	**−1.34 ± 0.17**	**< 0.0001**	**2.20**	**1.14 ± 0.14**	**< 0.0001**
GSTD6	**2.61**	**1.39 ± 0.40**	**0.0020**	0.15	−2.77 ± 1.39	0.0578	**0.30**	**−1.76 ± 0.56**	**0.0045**
GSTD7	0.89	−0.17 ± 0.10	0.1024	**0.41**	**−1.30 ± 0.25**	**0.0002**	0.98	−0.03 ± 0.15	0.8289
GSTD9	0.77	−0.38 ± 0.17	0.0368	0.83	−0.26 ± 0.12	0.0430	**1.81**	**0.86 ± 0.10**	**< 0.0001**
GSTD10	1.47	0.56 ± 0.26	0.0414	1.34	0.42 ± 0.47	0.3868	**6.83**	**2.77 ± 0.36**	**< 0.0001**
GSTE1	0.98	−0.02 ± 0.19	0.9100	0.72	−0.48 ± 0.17	0.0137	1.19	0.25 ± 0.23	0.3062
GSTE3	**1.67**	**0.74 ± 0.24**	**0.0048**	**0.52**	**−0.94 ± 0.20**	**< 0.0001**	**0.54**	**−0.88 ± 0.18**	**< 0.0001**
GSTE5	1.20	0.26 ± 0.26	0.3146	**0.41**	**−1.29 ± 0.16**	**< 0.0001**	**0.45**	**−1.16 ± 0.15**	**< 0.0001**
GSTE6	1.15	0.20 ± 0.25	0.4449	**0.21**	**−2.28 ± 0.24**	**< 0.0001**	**0.23**	**−2.12 ± 0.26**	**< 0.0001**
GSTE7	**1.69**	**0.76 ± 0.24**	**0.0056**	**0.66**	**−0.59 ± 0.16**	**0.0009**	**0.38**	**−1.41 ± 0.27**	**< 0.0001**
GSTE8	**1.58**	**0.66 ± 0.23**	**0.0084**	**0.39**	**−1.38 ± 0.16**	**< 0.0001**	0.36	−1.47 ± 0.60	0.0224
GSTE9	1.03	0.04 ± 0.51	0.9300	**0.34**	**−1.57 ± 0.36**	**0.0002**	**0.24**	**−2.05 ± 0.28**	**< 0.0001**
GSTE10	2.61	1.39 ± 0.71	0.0617	**3.68**	**1.88 ± 0.65**	**0.0079**	2.76	1.46 ± 0.64	0.0325
CG5224	**2.18**	**1.12 ± 0.28**	**0.0005**	1.02	0.03 ± 0.23	0.8886	1.44	0.53 ± 0.34	0.1292
CG16936	1.02	0.02 ± 0.09	0.8032	0.68	−0.56 ± 0.24	0.0392	0.97	−0.05 ± 0.21	0.8203
CG11784	2.02	1.01 ± 0.41	0.0200	**2.53**	**1.34 ± 0.36**	**0.0011**	**6.00**	**2.59 ± 0.40**	**< 0.0001**
GSTS1	**2.06**	**1.05 ± 0.37**	**0.0095**	**1.88**	**0.91 ± 0.30**	**0.0054**	**2.60**	**1.38 ± 0.26**	**< 0.0001**
CG1681	**2.06**	**1.05 ± 0.33**	**0.0041**	**5.31**	**2.41 ± 0.24**	**< 0.0001**	**10.06**	**3.33 ± 0.19**	**< 0.0001**
CG1702	0.88	−0.19 ± 0.11	0.0890	**0.41**	**−1.30 ± 0.21**	**< 0.0001**	**0.63**	**−0.68 ± 0.11**	**< 0.0001**
CG30005	1.55	0.63 ± 0.50	0.2187	**2.55**	**1.35 ± 0.39**	**0.0020**	1.85	0.89 ± 0.39	0.0308
CG6776	**1.74**	**0.80 ± 0.27**	**0.0075**	**1.99**	**0.99 ± 0.16**	**< 0.0001**	**2.35**	**1.23 ± 0.18**	**< 0.0001**
CG6673	**0.66**	**−0.59 ± 0.11**	**< 0.0001**	0.71	−0.50 ± 0.39	0.2250	0.59	−0.76 ± 0.46	0.1219
CG9362	1.73	0.79 ± 0.39	0.0538	1.49	0.57 ± 0.29	0.0614	1.29	0.36 ± 0.36	0.3200

*Highly significant changes ($P \leq 0.01$) are given in bold.

†Down-regulated transcripts, in the presence of dietary H_2O_2, are preceded with a negative value. All other transcripts were up-regulated in the presence of dietary H_2O_2.

egg hatched). The larvae were then removed from the diet and their midguts dissected out in 214 mM NaCl solution, and then transferred to RNAlater® (Ambion Inc., Austin, TX, USA). The hindgut and Malpighian tubes were excluded during the dissection. After dissecting, 50 midguts were pooled in the RNAlater and immediately frozen in liquid nitrogen, then transferred to −80 °C. Fifty midguts represented a single replicate for each microarray experiment. Three biological replicates were performed, with one biological replicate per oligoarray for a total of three oligoarrays.

Microarray hybridization, scanning and data analyses

Total RNA was isolated by using the TRIzol® Reagent (Invitrogen, Carlsbad, CA, USA) and RNeasy Mini Kit (Qiagen Inc., Valencia, CA, USA) following the manufacturers' protocols. The cDNA was generated from 1 μg total RNA using iScript cDNA Synthesis Kit (Bio-Rad Lab., Hercules, CA, USA) according to the manufacturer's instructions. We used the Affymetrix *Drosophila* Genome 1 arrays (Affymetrix, Inc., Santa Clara, CA, USA). Sample preparation, labelling, microarray hybridization and scanning followed Pedra *et al*. (2004). The affixed 25 oligonucleotide segments are referred to as 'probes' in the *D. melanogaster* genome array. One gene was represented by multiple probes in the array. The detection algorithm uses probe pair intensities to generate a detection *P*-value and assign a present, marginal or absent call by the Affymetrix Genechip® Expression Analysis. The data was normalized by Gene Chip Robust Multi-array Average (Supplementary Material Table S1). Overall, 2923 transcripts were considered present or observed in all three biological replicates, and were accordingly defined as being midgut-associated genes. The microarray data were deposited in the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and the corresponding accession numbers are GSM245144, GSM245146 and GSM245148.

Gene ontology analysis

We used the NetAffx™ Analysis Center (April 2005; http://www.affymetrix.com/analysis/index.affx) to determine the biological process, molecular function and cellular component information of the transcripts present in microarrays. We divided the transcripts into different categories based on their molecular functions or biological processes.

DNA transcription factor binding motif analysis

DNA-binding transcription regulators bind to specific genomic sequences to induce or repress gene expression. We first applied five different motif discovery programs to identify potential common regulatory sequences for a set of co-expressed genes, as each program had different strengths: AlignACE (Roth *et al*., 1998), MEME (Bailey & Elkan, 1995), MDscan (Liu *et al*., 2002), BioProspector (Liu *et al*., 2001) and Weeder (Pavesi *et al*., 2004). In the current analysis, we focussed on regions near the genes. Each promoter region analysed was 1000 bp long and contained 800 bp upstream and 200 bp downstream from the transcription start site of the gene.

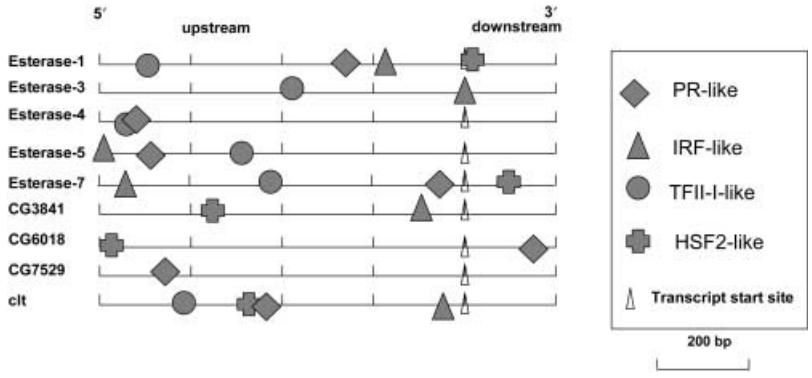

Figure 5. The locations of four candidate transcription factor binding motifs (TFBMs) in the carboxylesterases associated with third instar *Drosophila melanogaster* larval midgut observed 2 kb upstream of the gene transcription start site (TSS). The potential ARE and XRE TFBMs associated with the glutathione S transferases (GSTs) expressed in the midgut (based on microarray analysis) are given in Supplementary Material Fig. S4.

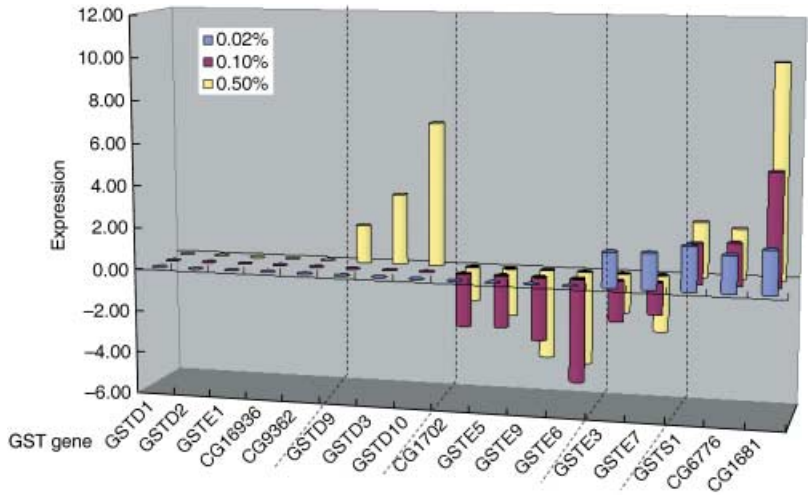

Figure 6. Responses of selected third instar *Drosophila melanogaster* larval midgut-associated glutathione S transferase (GST) genes to three doses of hydrogen peroxide (H_2O_2) at 0.02, 0.1 and 0.5% ($P \leq 0.01$). The common motifs observed from similar patterns of GSTs are given in Fig. S1. For genes without highly significant expression levels (ie $P < 0.01$), an expression ration of 0 is given in the figure. For details of expression levels and *P*-values please see Table 3.

In the second step, we combined the results of the different analytical programs into a comprehensive composite analysis. The results are represented in the form of position weight matrices (PWMs), whose columns define the probabilities of observing each nucleotide type (A, C, G, T) at each position of the motif (Gordon *et al*., 2006). Significant motifs from each of the five programs were pooled and clustered using either a k-medoids algorithm (Kaufman & Rousseeuw, 1990) or a hierarchical clustering algorithm (Hartigan, 1975). The functions were *pam()* and *hclust()* in R, which is a language and environment for statistical computing and graphics (http://www.r-project.org/). The centre of each cluster was selected to represent a potential transcription factor binding motif. To annotate the identified binding motifs, we compared the results with PWMs in the TRANSFAC database (Wingender *et al*., 2000).

The most similar transcription factors from TRANSFAC were used to index each of the candidate binding motifs we observed. We further verified the predictions of transcription factor binding sites by scanning the promoter sequences with the identified binding motifs (cluster centres). Significant matches were evaluated by *P*-values. A *P*-value was the probability that a motif with a similar matching score could be obtained in scanning randomly selected sequences. The abundant motifs with *P*-values less than 0.05 are considered significant and are presented in Figs 2 and 3. The sequence logo of a TFBM was built on a set of aligned transcription factor motif sequences. The height of a pile represents the information content at a certain position, whereas the height of a letter represents the individual weight of a base at a certain position. The sequence logos shown in Fig. 4 were generated using WebLogo (http://weblogo.berkeley.edu/). For the XRE and ARE motifs with given consensus, the 2 kb upstream regulatory regions of each genes were scanned for exact matches. Two hundred randomly generated sequences were also scanned for these motifs.

Hydrogen peroxide feeding experiment and qRT-PCR analyses

Eighty to 100 adult flies were placed in an 8.9 cm (diameter) Petri dish containing instant diet. They were allowed to lay eggs on the diet for 2 h and then removed. Eggs were allowed to develop to early third instar larvae over a 5-day period on the rearing diet. The larvae were then transferred from the rearing diets to the experimental vials, using a 3/0 NB-38 silverado brush (Yasutomo, South San Francisco, CA, USA). Each 2.5 cm (diameter) × 9.5 cm plastic experimental vial (Genesee Scientific Co., San Diego, CA, USA) was cut to a 3.5 cm height to allow for ease of transfer of the larvae. Twenty-five larvae were placed in each vial containing 364 mg Formula 4-24® Blue instant *Drosophila* medium and 1636 µl of solution with one of the following four treatments: (1) distilled water (control), (2) 0.02% (w/w) H_2O_2, (3) 0.1% H_2O_2, and (4) 0.5% H_2O_2. Hydrogen peroxide solution was dissolved in distilled water from a stock solution of 3% (w/v). Three biological replicates were performed for each treatment.

Total RNA was isolated from 50 larval midguts per treatment using TRIzol® Reagent (Invitrogen) and the RNeasy Mini Kit

Table 5. Primers for quantitative real-time PCR reactions

Gene name	Forward primer sequence	Reverse primer sequence
Rp49	CAGGCCCAAGATCGTGAAGAA	TGCGCTTGTTCGATCCGTAA
GSTD1	TGATCAATCAGCGCCTGTACT	GCAATGTCGGCTACGGTAAG
GSTD2	TCTGTACTTCGACATGGGAACTC	AGTGGACAGGATGGCAATGT
GSTD3	GGAACCCTGCATGACTCCTT	CGAAAGTGGAAACGCTGGA
GSTD4	GGAACCCTGCATGACTCCTT	CGAAAGTGGAAACGCTGGA
GSTD5	GAAAGCCCGGATCTGATGA	CATACCACCTGGCCACATTC
GSTD6	AAGCTGTGGGAGTCGAGTTCA	CAACAATGGCCCGAGTTTC
GSTD7	TGAAGCCAGAGTTCGTGAGG	CAGGGTGCCCATATCGAAGT
GSTD8	TTCGTCGAGGCCATCTATCC	AATGTCGGCAATGGTGAGG
GSTD9	GACGATGGTTTCGCCATCT	GGCTTCTTCACATCCTCGAA
GSTD10	AAGGCACTGGGTGTGGAGTT	TGCCGTACTTCTCCACCAGAT
CG17639	CAACGCCTCCAGTTCGATT	GCCCTCCAAGATTGTGTTTAGC
GSTE1	CTGAAGCTGCTGGAGACGTT	GGAGCTTCGTTGATCTCCTTG
GSTE2	TCGCAACTGACCATAGCTGA	CCGCAGATTGTCTTCCTCATAG
GSTE3	CGGCCGGAATGATTCTCTATAC	GCCTGTGGAATCTCCGTCTT
GSTE4	GGCGATCAGTTGACCATAGC	CCATTGGCCTCCTCGTAGTA
GSTE5	TGTATCCCAGAGATCTGCTCCA	TGGCATCGTAGCGTTCCTT
GSTE6	CTCTTCCAGGGACAGACGAAA	GCCACCGATGAAACGAGACT
GSTE7	GTAAACACTCGGGCCAAGGA	ACGCTGGAGGAGATCTTTCG
GSTE8	CGTGGTCTTTGTCAACGGACT	TCACCGGCAATGAAATCG
GSTE9	TACGAGGCCTATGACTTTCTGG	TAACTTGGGATAGCGCTTGG
GSTE10	GTTTGGCTGAGTTGAAGGATGC	CAACCAGGCGGATAATTTGG
CG5224	TGGTGGATGCCCGTTTGTA	GACAGTGCTCCAAGCCATCAT
CG16936	ATTGACTCGCACGCCATCT	TCTTGTCGATGGAGCAGTCC
CG11784	CTCATACCAGTGGAACGGGAAA	CTTTGTTCTCGGCCATGCAG
CG4688	CGACTTATCCATCGTGACCACA	ATCGTTTGGCGGAGCTTCT
GSTS1	CAATGCGGAGGGCTCAATA	AGGGAACTGGATGAGACGAAGA
CG1681	TGGCCACCACGGAGTATTT	AACTTGCGGGAGTTGAGGAA
CG1702	TCACCTGCGCCATGTACTTC	CACAGGCTGCAAAGATGTCC
CG30005	ACACCTGAACATTCGTCTTGC	CCACTCCCTCAATCAGTGCT
CG30000	GACAAACTTACCGTGGCAGACA	GGCCTCATCGTAGTAGGGATTG
CG6776	CGCCTCTCTGTAATCGAGTTGA	CCAGAACTCGTTGTGCTGCT
CG6673	TTCGAGCAGGAGATCACCAA	TTGACAGCCTCATCCTGAGC
CG6781	GACCACTCTATCCACGTGATCC	GCTCCCTTTCGTAGATGTCCA
CG6662	CCAGGATCGTGCTGTCAAGT	AAACGCCATCATCCCGTAG
CG9362	AAACTCGACCTCAGCCAGCA	CCCATTGCAGACTCTGATCCTT
CG9363	TCAATGCCCGAAGATTCCAC	GTCTTACGGCGCTTTGCAGT
Mgstl	ATCCAGCGTTTCACCTTCG	GATGTTCTCCAGGTCGTTGC

(Qiagen Inc.) following the manufacturers' protocols. Fifty larval midguts were used for each RNA extraction with three biological replicates per treatment. The cDNA was generated from 1 μg sample of total RNA using the iScript cDNA Synthesis Kit (Bio-Rad Lab.) according to the manufacturer's instructions. Primers were designed for the qRT-PCR analysis using the PCR Now program (http://pathogene.swmed.edu/rt_primer/). The primers of the ribosomal protein 49 (*RP49*) (accession number U92431) from *D. melanogaster* were used as an internal control to monitor equal loading (Table 5). Primers for GST genes were designed as shown in Table 3.

The transcript levels of all the genes were monitored by qRT-PCR using an iCycler Thermal Cycler system (Bio-Rad Lab.) with gene-specific primers and iQ SYBR Green Supermix (Bio-Rad Lab.). All qRT-PCR reactions contained 10 pmol of each primer and 2× of iQ SYBR Green Supermix. The program was 55 cycles of 95 °C for 90 s, 62 °C for 20 s, 72 °C for 20 s. The levels of *RP49* transcripts among all samples were normalized to be equal. We measured the cycle value of all GSTs from the *D. melanogaster* genome (total 38 as defined by Tu & Akgul, 2005). The fold change of each GST gene is presented in Fig. 1. Levels of all GST gene transcripts from the midgut of insects feeding on different doses of H_2O_2 were compared with insects feeding on the control diet (Table 3 and Fig. 6). The qRT-PCR data were analysed by the Relative Expression Software Tool-384 (REST 384 v1.0) (Pfaffl *et al*., 2002) and SAS (SAS Institute Inc., Cary, NC, USA).

Acknowledgements

This work was supported with funds provided to BRP by the Department of Entomology, Purdue University. HML was funded by the Ross fellowship of Purdue University and a grant from USDA-Agricultural Research Programs of Purdue University to BRP. We thank S. Balfe for technical assistance. We also thank the Affymetrix GeneChip® Laboratory of Purdue University for technical support.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G. *et al*. (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.

Alin, P., Danielson, U.H. and Mannervik, B. (1985) 4-Hydroxyalk-2-enals are substrates for glutathione transferase. *FEBS Lett* **179**: 267–270.

Arbeitman, M.N., Furlong, E.E.M., Imam, F., Johnson, E., Null, B.H., Baker, B.S. *et al*. (2002) Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**: 2270–2275.

Arvestad, L., Visa, N., Lundeberg, J., Wieslander, L. and Savolainen, P. (2005) Expressed sequence tags from the midgut and an epithelial cell line of *Chironomus tentans*: annotation, bioinformatic classification of unknown transcripts and analysis of expression levels. *Insect Mol Biol* **14**: 689–695.

Bailey, T.L. and Elkan, C. (1995) The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* **3**: 21–29.

Bauer, L.S. (1995) Resistance: a threat to the insecticidal crystal proteins of *Bacillus thuringiensis*. *Fla Entomol* **78**: 415–443.

Baum, J.A., Bogaert, T., Clinton, W., Heck, G.R., Feldmann, P., Ilagan, O. *et al*. (2007) Control of coleopteran insect pests through RNA interference. *Nat Biotechnol* **25**:1322–1326.

Bel, Y. and Escriche, B. (2006) Common genomic structure for the Lepidoptera cadherin-like genes. *Gene* **381**: 71–80.

Campbell, P.M. Robin, G.C.D.Q., Court, L.N. Dorrian, S.J., Russell, R.J. and Oakeshott, J.G. (2003) Developmental expression and gene/enzyme identifications in the alpha esterase gene cluster of *Drosophila melanogaster*. *Insect Mol Biol* **12**: 459–471.

Chapman, R.F. (1998) Alimentary canal, digestion and absorption. In *The Insects Structure and Function*, 4th edn, pp. 38–69. Cambridge University Press, Cambridge, UK.

Cheriyath, V. and Roy, A.L. (2000) Alternatively spliced isoforms of TFII-I. Complex formation, nuclear translocation, and differential gene regulation. *J Biol Chem* **275**: 26300–26308.

Chilcutt, C.F. and Tabashnik, B.E. (2004) Contamination of refuges by *Bacillus thuringiensis* toxin genes from transgenic maize, *Proc Natl Acad Sci USA* **101**: 7526–7529.

Ebert, M.N., Klinder, A., Peters, W.H., Schaferhenrich, A., Sendt, W., Scheele, J. *et al*. (2003) Expression of glutathione S-transferases (GSTs) in human colon cells and inducibility of GSTM2 by butyrate. *Carcinogenesis* **24**: 1637–1644.

Enayati, A.A., Ranson, H. and Hemingway, J. (2005) Insect glutathione transferases and insecticide resistance. *Insect Mol Biol* **14**: 3–8.

Fabrick, J., Behnke, C., Czapla, T., Bala, K., Rao, A.G., Kramer, K.J. *et al*. (2002) Effects of a potato cysteine proteinase inhibitor on midgut proteolytic enzyme activity and growth of the southern corn rootworm, *Diabrotica undecimpunctata howardi* (Coleoptera: Chrysomelidae). *Insect Biochem Mol Biol* **32**: 405–415.

Ferre, J. and Van Rie, J. (2002) Biochemistry and genetics of insect resistance to *Bacillus thuringiensis*. *Annu Rev Entomol* **47**: 501–533.

Feyereisen, R. (1999) Insect P 450 enzymes, *Annu Rev Entomol* **44**: 507–533.

Feyereisen, R. (2005) Insect cytochrome P450. In *Comprehensive Molecular Insect Science*, vol. 4, (Gilbert, L.I., Iatrou, K. and Gill, S.S., eds), pp. 1–77. Elsevier, Amsterdam.

Fiers, W., Beyaert, R., Declercq, W. and Vandenabeele, P. (1999) More than one way to die: apoptosis, necrosis and reactive oxygen damage. *Oncogene* **18**: 7719–7730.

Gill, M. and Ellar, D. (2002) Transgenic *Drosophila* reveals a functional *in vivo* receptor for the *Bacillus thuringiensis* toxin Cry1Ac1. *Insect Mol Biol* **11**: 619–625.

Gnerre, C., Schuster, G.U., Roth, A., Handschin, C., Johansson, L., Looser, R. *et al*. (2005) LXR deficiency and cholesterol feeding affect the expression and phenobarbital-mediated induction of cytochromes P450 in mouse liver. *J Lipid Res* **46**: 1633–1642.

Gordon, J.J., Towsey, M.W., Hogan, J.M., Mathews, S.A. and Timms, D. (2006) Improved prediction of baterial transcription start sites. *Bioinformatics* **22**: 142–148.

Gorman, M.J. and Paskewitz, S.M. (2001) Serine proteases as mediators of mosquito immune responses. *Insect Biochem Mol Biol* **31**: 257–262.

Grant, D.F. and Matsumura, F. (1989) Glutathione S-transferase 1 and 2 in susceptible and resistant insecticide resistant *Aedes aegypti*. *Pest Biochem Physiol* **33**: 132–143.

Ha, E.M., Oh, C.T., Bae, Y.S. and Lee, W.J. (2005) A direct role for dual oxidase in *Drosophila* gut immunity. *Science* **310**: 847–850.

Hahn, M.W., De Bie, T., Stajichm J.E., Nguyen, C. and Cristianini, N. (2005) Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res* **15**: 1153–1160.

Hanrahan, C.J., Palladino, M.J., Bonneau, L.J. and Reenan, R.A. (1999) RNA editing of a *Drosophila* sodium channel gene. *Ann N Y Acad Sci* **868**: 51–66.

Hartigan, J.A. (1975). *Clustering Algorithms*. Wiley, New York.

He, H., Soncin, F., Grammatikakis, N., Li, Y., Siganou, A., Gong, J. *et al*. (2003) Elevated expression of heat shock factor (HSF) 2A stimulates HSF1-induced transcription during stress. *J Biol Chem* **12278**: 35465–35475.

Ikawa, T., Kawamoto, H., Goldrath, A.W. and Murre, C. (2006). E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. *J Exp Med* **203**: 1329–1342.

Ingle, S.S., Trivedi, N., Prasad, R., Kuruvilla, J., Rao, K.K. and Chhatpar, H.S. (2001) Aminopeptidase-N from the *Helicoverpa armigera* (Hubner) brush border membrane vesicles as a receptor of *Bacillus thuringiensis* crylac delta-endotoxin. *Curr Microbiol* **43**: 255–259.

Jenkins, J.L. and Dean, D.H. (2001) Binding specificity of *Bacillus thuringiensis* Cry1Aa for purified, native *Bombyx mori* aminopeptidase N and cadherin-like receptors. *BMC Biochem* **2**: 12.

Kaestner, K.H., Knochel, W. and Martinez, D.E. (2000). Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev* **14**: 142–146.

Kaufman, L. and Rousseeuw, P.J. (1990) *Finding Groups in Data*. Wiley, New York.

Kirchhoff, S., Schaper, F. and Hauser, H. (1993) Interferon regulatory factor 1 (IRF-1) mediates cell growth inhibition by transactivation of downstream target genes. *Nucleic Acids Res* **21**: 2881–2889.

Knight, P.J.K., Crickmore, N. and Ellar, D.J. (1994) The receptor for *Bacillus thuringiensis* CryIA(c) delta-endotoxin in the brush border membrane of the lepidopteran *Manduca sexta* is aminopeptidase N. *Mol Microbiol* **11**: 429–436.

Krishnan, N. and Kodrik, D. (2006) Antioxidant enzymes in *Spodoptera littoralis* (Boisduval): Are they enhanced to protect gut tissues during oxidative stress? *J Insect Physiol* **52**: 11–20.

Lehane, M.J., Aksoy, S, Gibson, W, Kerhornou, A., Berriman, M., Hamilton, J. *et al*. (2003) Adult midgut expressed sequence tags from the tsetse fly *Glossina morsitans morsitans* and expression analysis of putative immune response genes. *Genome Biol* **4**: R63.

Leonhardt SA, Boonyaratanakornkit V, Edwards DP (2004) Progesterone receptor transcription and non-transcription signaling mechanisms. *Steroids* **68**: 761–770.

Li, T.-R. and White, K.P. (2003). Tissue-specific gene expression and ecdysone-regulated genomic networks in *Drosophila*. *Dev Cell*, **5**:59–72.

Liu, X., Brutlag, D.L. and Liu, J.S. (2001) BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput* **6**: 127–138.

Liu, X.S., Brutlag, D.L. and Liu, J.S. (2002) An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotechnol* **20**: 835–839.

Mittapalli, O., Neal, J.J. and Shukle, R.H. (2005) Differential expression of two cytochrome P450 genes in compatible and incompatible Hessian fly/wheat interactions. *Insect Biochem Mol Biol* **35**: 981–989.

Murdock, L.L., Brookhart, G., Dunn, P.E., Foard, D.E., Kelley, S., Kitch, L. *et al*. (1987) Cysteine digestive proteinase in Coleoptera. *Comp Biochem Physiol* **87B**: 783–787.

Okamoto, K., Tanaka, H., Ogawa, H., Makino, Y., Eguchi, H., Hayashi, S. *et al*. (1999). Redox-dependent Regulation of Nuclear Import of the Glucocorticoid Receptor. *J Biol Chem* **274**: 10363–10371.

Pavesi, G., Mereghetti, P., Mauri, G. and Pesole, G. (2004) Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res* **32**: W199–203 (web server issue).

Pedra, J.H., Brandt, A., Westerman, R., Lobo, N., Li, H.-M., Romero-Severson, J. *et al*. (2003) Transcriptome analysis of the cowpea weevil bruchid: identification of putative proteinases and alpha-amylases associated with food breakdown *Insect Mol Biol* **12**: 405–412.

Pedra, J.H.F., McIntyre, L.M., Scharf, M.E. and Pittendrigh, B.R. (2004) Genome-wide transcription profile of field- and laboratory-selected dichlorodiphenyltrichloroethane (DDT)-resistant *Drosophila*. *Proc Natl Acad Sci USA* **101**: 7034–7039.

Pfaffl, M.W., Horgan, G.W. and Dempfle, L. (2002) Relative Expression Software Tool (REST) for group-wise comparison and statistical analysis of relative expression results in Real-Time PCR. *Nucleic Acids Res* **30**: e36.

Reenan, R.A. (2005) Molecular determinants and guided evolution of species-specific RNA editing. *Nature* **434**: 409–413.

Roth, F.P., Hughes, J.D., Estep, P.W. and Church, G.M. (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat Biotechnol* **16**: 939–945.

van Roon-Mom, W.M., Reid, S.J., Faull, R.L. and Snell, R.G. (2005) TATA-binding protein in neurodegenerative disease. *Neuroscience* **133**: 863–872.

Rushmore, T.H., Morton, M.H and Pickett, C.B. (1991) The antioxidant responsive element. *J Biol Chem* **266**: 11632–11639.

Russell, R.J., Robin, G.C., Kostakos, P., Newcomb, R.D., Boyce, T.M., Medveczky, K.M. *et al*. (1996) Molecular cloning of an α-esterase gene cluster on chromosome 3R of *Drosophila melanogaster*. *Insect Biochem Mol Biol* **26**: 235–247.

Sawicki, R., Singh, S.P., Mondal, A.K., Benes, H. and Zimniak, P. (2003) 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* **370**: 661–669.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J. *et al*. (1998) *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol Mol Biol Rev* **62**: 775–806.

Semenza, G.L., Shimoda, L.A. and Prabhakar, N.R. (2006) Regulation of gene expression by HIF-1. *Novartis Found Symp* **272**: 2–8.

Sen, C.K. and Packer, L. (1996). Antioxidant and redox regulation of gene transcription. *FASEB J* **10**: 709–720.

Siegfried, B.D., Waterfield, N. and ffrench-Constant, R.H. (2005) Expressed sequence tags from *Diabrotica virgifera virgifera* midgut identify a coleopteran cadherin and a diversity of cathepsins. *Insect Mol Biol* **14**: 137–143.

Singh, S.P., Coronella, J.A., Benes, H., Cochrane, B.J. and Zimniak, P. (2001) Catalytic function of *Drosophila melanogaster* glutathione S-transferase DmGSTS1–1 (GST-2) in conjugation of lipid peroxidation end products. *Eur J Biochem* **268**: 2912–2923.

Snyder, M.J., Stevens, J.L., Andersen, J.F. and Feyereisen, R. (1995) Expression of cytochrome P450 genes of the CYP4 family in midgut and fat body of the tobacco hornworm, *Manduca sexta*. *Arch Biochem Biophys* **321**: 13–20.

Tabashnik, B.E., Finson, N., Groeters, F.R., Moar, W.J., Johnson, M.W., Luo, K. *et al*. (1994) Reversal of resistance to *Bacillus thuringiensis* in *Plutella xylostella*. *Proc Natl Acad Sci USA* **91**: 4120–4124.

Tijet, N., Helvig, C. and Feyereisen, R. (2001) The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. *Gene* **262**,189–198.

Tu, C.-P.D. and Akgul, B. (2005) *Drosophila* glutathione S-transferases. *Methods Enzymol* **401**: 204–226.

Ullrich, O., Huser, H.. Ehrlich, W. and Grune, T. (1997) Intracellular metabolism of 4-hydroxynonenal in primary cultures of rabbit synovial fibroblasts. *Free Radic Biol Med* **22**: 1153–1157.

Videla, L.A., Arisi, A.C., Fuzaro, A.P., Koch, O.R. and Junqueira, V.B. (2000) Prolonged phenobarbital pretreatment abolishes the early oxidative stress component induced in the liver by acute lindane intoxication. *Toxicol Lett* **115**: 45–51.

Wang, H., Yang, U., Lee, C. and Blume, J. (2004) Alternative splicing: session introduction. *Pac Symp Biocomput* **9**: 3–4.

Watabe, M., Isogai, Y., Numazawa, S. and Yoshida, T. (2003) Role of c-Myc in nitric oxide-mediated suppression of cytochrome P450 3A4. *Life Sci* **74**: 99–108.

Wingender, E., Chen, X. and Hehl, R. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* **28**: 316–319.

Wolfson, J.L. and Murdock, L.L. (1990) Diversity in digestive proteinase activity among insects. *J Chem Ecol* **16**: 1089–1102.

Yan, L.J. and Sohal, R.S. (1998) Mitochondrial adenine nucleotide translocase is modified oxidatively during aging. *Proc Natl Acad Sci USA* **95**: 12896–12901.

Yoshinaga, S.K. and Yamamoto, K.R. (1991) Signaling and regulation by a mammalian glucocorticoid receptor in *Drosophila* cells. *Mol Endocrinol* **5**:844–853.

Zhang, L., Savas, U., Alexander, D.L. and Jefcoate, C.R. (1998) Characterization of the mouse cyp1B1 gene. *J Biol Chem* **273**: 5174–5183.

Zinke, I., Schütz, C.S., Katzenberger, J.D., Bauer, M., Pankratz, M.J. (2002) Nutrient control of gene expression in *Drosophila*: microarray analysis of starvation and sugar-dependent response. *EMBO J* **21**:6162–6173.

Zou, S., Meadows, S., Sharp, L., Jan, L.Y. and Jan, Y.N. (2000) Genome-wide study of aging and oxidative stress response in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **97**: 13726–13731.

Supplementary material

The following supplementary material is available for this article:

Figure S1. Sequence logo demonstration of all the transcription factor binding motifs from glutathione S transferases (GSTs) based on their response to different doses of H_2O_2. All the factors were observed in the third instar *Drosophila melanogaster* larval midgut transcriptome.

Figure S2. Searches of antioxidant responsive element-like (ARE-like) (sequence 5′-GTGACNNNG-3′) and xenobiotic responsive element (XRE) (sequence 5′-GCGTG-3′) in the 2 kb upstream region of glutathione S transferases (GSTs) associated with third instar *Drosophila melanogaster* larval midgut.

Figure S3. Searches of antioxidant responsive element (ARE) (sequence 5′-GTGACNNNGC-3′) and xenobiotic responsive element (XRE) (sequence 5′-GCGTG-3′) in the 2 kb upstream region of cytochrome P450s associated with third instar *Drosophila melanogaster* larval midgut.

Figure S4. Searches of antioxidant responsive element-like (ARE-like) (sequence 5′-GTGACNNNG-3′) and xenobiotic responsive element (XRE) (sequence 5′-GCGTG-3′) in the 2 kb upstream region of carboxylesterases associated with third instar *Drosophila melanogaster* larval midgut.

Table S1. Normalized microarray data with probe IDs using Gene Chip Robust Multi-array Average (GCRMA) method.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1356-2583.2008.00808.x
(This link will take you to the article abstract).

Please note: Wiley–Blackwell are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.