

# Tissue-Specific Transcriptomics of the Exotic Invasive Insect Pest Emerald Ash Borer (*Agrius planipennis*)

Omprakash Mittapalli<sup>1\*</sup>, Xiaodong Bai<sup>1</sup>, Praveen Mamidala<sup>1,9</sup>, Swapna Priya Rajarapu<sup>1,9</sup>, Pierluigi Bonello<sup>2</sup>, Daniel A. Herms<sup>1</sup>

<sup>1</sup> Department of Entomology, Ohio Agricultural and Research Development Center, The Ohio State University, Wooster, Ohio, United States of America, <sup>2</sup> Department of Plant Pathology, The Ohio State University, Columbus, Ohio, United States of America

## Abstract

**Background:** The insect midgut and fat body represent major tissue interfaces that deal with several important physiological functions including digestion, detoxification and immune response. The emerald ash borer (*Agrius planipennis*), is an exotic invasive insect pest that has killed millions of ash trees (*Fraxinus* spp.) primarily in the Midwestern United States and Ontario, Canada. However, despite its high impact status little knowledge exists for *A. planipennis* at the molecular level.

**Methodology and Principal Findings:** Newer-generation Roche-454 pyrosequencing was used to obtain 126,185 reads for the midgut and 240,848 reads for the fat body, which were assembled into 25,173 and 37,661 high quality expressed sequence tags (ESTs) for the midgut and the fat body of *A. planipennis* larvae, respectively. Among these ESTs, 36% of the midgut and 38% of the fat body sequences showed similarity to proteins in the GenBank nr database. A high number of the midgut sequences contained chitin-binding peritrophin (248) and trypsin (98) domains; while the fat body sequences showed high occurrence of cytochrome P450s (85) and protein kinase (123) domains. Further, the midgut transcriptome of *A. planipennis* revealed putative microbial transcripts encoding for cell-wall degrading enzymes such as polygalacturonases and endoglucanases. A significant number of SNPs (137 in midgut and 347 in fat body) and microsatellite loci (317 in midgut and 571 in fat body) were predicted in the *A. planipennis* transcripts. An initial assessment of cytochrome P450s belonging to various CYP clades revealed distinct expression patterns at the tissue level.

**Conclusions and Significance:** To our knowledge this study is one of the first to illuminate tissue-specific gene expression in an invasive insect of high ecological and economic consequence. These findings will lay the foundation for future gene expression and functional studies in *A. planipennis*.

**Citation:** Mittapalli O, Bai X, Mamidala P, Rajarapu SP, Bonello P, et al. (2010) Tissue-Specific Transcriptomics of the Exotic Invasive Insect Pest Emerald Ash Borer (*Agrius planipennis*). PLoS ONE 5(10): e13708. doi:10.1371/journal.pone.0013708

**Editor:** Michael N. Nitabach, Yale School of Medicine, United States of America

**Received:** June 26, 2010; **Accepted:** October 4, 2010; **Published:** October 28, 2010

**Copyright:** © 2010 Mittapalli et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The research was supported by a grant from the USDA APHIS/USDA Forest Service Accelerated Emerald Ash Borer Research Program (GRT00011769/60016270), and by State and Federal funds appropriated to the Ohio Agricultural Research and Development Center, The Ohio State University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: mittapalli.1@osu.edu

<sup>9</sup> These authors contributed equally to this work.

## Introduction

The insect midgut and fat body are important metabolic tissues and represent key physiological interfaces during interactions with hosts. While both tissues are highly specialized and play major roles in the life of insects, they possess discrete physiological functions essential for growth and development. The midgut plays important roles in digestion, antioxidant defense and is thought to be one of the primary interfaces in countering the effects of dietary toxins [1–5], xenobiotics (ex. insecticides) and potential pathogens [6]. On the other hand, the insect fat body is a dynamic tissue, which plays vital roles in intermediary metabolism, energy storage/utilization, detoxification and immune response [7,8]. Remarkably, functions performed by the insect fat body are correlated to those of adipose tissue and liver of vertebrates. The combined action of the factors associated with the midgut and fat

body could determine the capacity of an insect to adapt to various ecological niches and thereby evolve.

Beetles are the most predominant group of insects constituting about 25% of all known life-forms [9]. Within Coleoptera, the Buprestidae, known commonly as metallic wood-borers, remains one of the most understudied groups of insects despite their ecological importance in forests and the status of some species as key pests of trees. For example, emerald ash borer (*Agrius planipennis* Fairmaire) which is endemic to Asia, has emerged as a devastating pest of ash trees in North America where it continues to kill millions of ash trees, primarily in the Midwestern United States and Ontario, Canada [10,11]. All North American ash that it has encountered, including green ash (*Fraxinus pennsylvanica* Marshall), white ash (*F. americana* L.) and black ash (*F. nigra* Marshall) are highly susceptible [11–13]. If *A. planipennis* sustains its current rate of invasion, it is considered to have the potential to

decimate ash on a continental scale [14,15] with economic and ecological impacts reminiscent of the invasions of the chestnut blight and Dutch elm disease pathogens [16,17].

Larvae of *A. planipennis* feed on the phloem and cambial tissues of ash trees, which usually results in the formation of S-shaped galleries that disrupt nutrient and water translocation, generally killing the trees in 3–5 years following attack [10]. *A. planipennis* is endemic to Far East Asia, including China, Korea, Japan and Siberia [18]. However, reports indicate that *A. planipennis* is rarely considered a pest in Asia [19,20], where Manchurian ash (*F. mandshurica* Rupr) is a primary host [21,22], and infestations appear restricted to stressed trees [23]. This implies that Manchurian ash may be generally resistant, with weakened trees preferentially colonized. A common garden experiment in Michigan confirmed that Manchurian ash has a much higher level of resistance to *A. planipennis* than do common North American ash species [24] and it has been reported that the resistance may be mediated by phenolic compounds in the phloem such as hydroxycoumarins and phenylethanoids, among other possible mechanisms [25].

Despite its status as a key pest, no functional genomics data exist for *A. planipennis*. In general, little is known at the tissue- level about the physiology-driven molecular strategies that invasive insects deploy to adapt to their new environment. This lack of fundamental molecular knowledge hinders our understanding of the insect's biology. However, the advent of second-generation sequencing such as 454 pyrosequencing [26] and Illumina/Solexa [27] offers a unique opportunity to study functional genomics [28] in non-model organisms. For example, the 454 pyrosequencing strategy has been successfully applied to unravel transcriptomic signatures in the tobacco hornworm, *Manduca sexta* L. [3], the soybean aphid, *Aphis glycines* [29] and the poplar leaf beetle, *Chrysomela tremulae* F. [30]. Here, we present data derived from 454 pyrosequencing of the midgut and fat body of *A. planipennis* larvae. Results obtained from this study will provide the basis for performing future gene expression and functional studies in *A.*

*planipennis*. Furthermore, the data will improve our understanding of the biology of this key invasive pest at the molecular level.

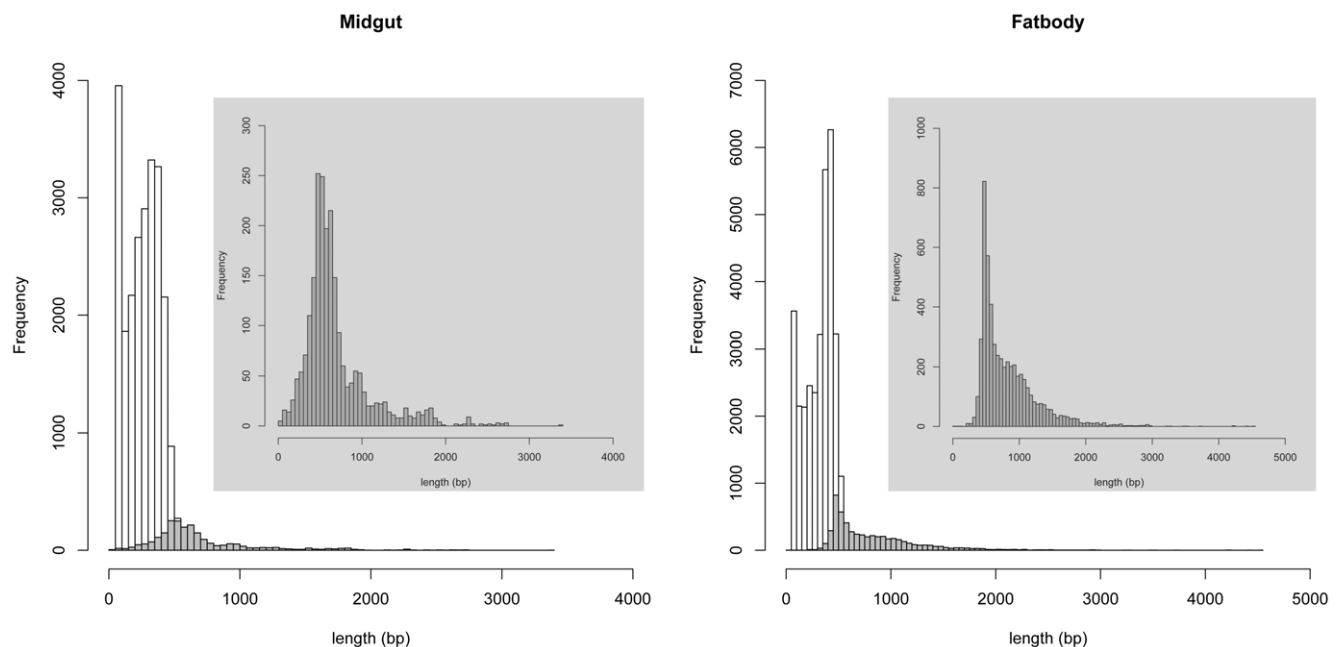
## Results and Discussion

### Transcriptome analysis

We obtained 126,185 transcriptomic reads for the midgut and 240,848 reads for the fat body of third-instar *A. planipennis* larvae. The reads for both tissues were assembled using Newbler program (Roche) after the removal of adapter sequences. To achieve better consistency, the contigs and singletons of the transcripts were renamed in the format “EABMG000001” with “EAB” for emerald ash borer, “MG” for midgut, and “000001” for an arbitrarily assigned number. Similar nomenclature was adopted to describe the fat body transcriptomic sequences with MG replaced by FB for fat body.

After the assembly of the sequences, 25,173 high quality ESTs (2,218 contigs and 23,495 singletons) were obtained for the midgut and 37,661 ESTs (5,376 contigs and 32,285 singletons) for the fat body (Table 1). The singleton sequences for the midgut ranged from 50 bp to 710 bp with an average length of 259 bp and total length of 6,091,300 bp; while the contig sequences ranged from 36 bp to 3,395 bp with an average length of 688 bp and total length of 1,526,781 bp. The singleton sequences for the fat body ranged from 50 bp to 606 bp with an average length of 314; while the contiguous sequences ranged from 38 bp to 4,501 bp with an average length of 825 bp and total length of 4,433,073 bp (Figure 1).

Among the midgut transcripts, 9,066 (36%) showed significant similarity ( $E$  value  $< 1e^{-5}$ ) to proteins in the GenBank nr database, while 14,253 (38%) sequences of the fat body ESTs showed significant hits (Table 1). The majority of the transcripts with similar sequences in the database (95%) matched to arthropodal proteins. The remaining midgut and fat body transcripts were similar to proteins of non-insect eukaryotes (5%) and bacteria (0.02%). A total of 14 sequences were similar to viral proteins and artificial sequences; and one midgut sequence matched to an *Archaea* protein (Table 1).



**Figure 1. Summary of transcriptomic sequences from *Agrilus planipennis* midgut and fat body.** The singleton sequences are represented by clear bars and the contig sequences by shaded bars (inserts). doi:10.1371/journal.pone.0013708.g001

**Table 1.** Summary of BLASTx search of the *Agrilus planipennis* sequences.

	MG*	FB
<b>Significant matches</b>	9,066	14,253
Artificial sequences	3	4
Archaea	1	0
Arthropoda	8,574	13,438
Bacteria	37	50
Other eukaryotes	448	757
Viruses	3	4
<b>Non- significant matches</b>	16,107	23,408
<b>Total</b>	25,173	37,661

\*MG = midgut; FB = fat body.  
doi:10.1371/journal.pone.0013708.t001

**Comparative analysis**

A comparative analysis of the derived *A. planipennis* midgut and fat body transcripts with protein sequences in the draft genomes of *Drosophila melanogaster* Meigen, *Anopheles gambiae* Giles, and *Tribolium castaneum* Hebst, revealed most sequence similarity (36.6%, 23,201 out of 62,834) to the *T. castaneum* genome [31] (Figure 2). Similar observations were reported by Pauchet et al. (2009) [30] for the midgut transcriptome of the chrysomelid *C. tremulae*. Nearly equal proportions of *A. planipennis* sequences showed hits to the non-beetle genomes including *A. gambiae* (29.4%) and *D. melanogaster* (29.1%; Figure 2). A total of 16,589 sequences were shared among all four insect species under comparison and about 62% of sequences (39,228 out of 62,834) did not show BLASTX similarity, implying either that they represented untranslated regions,

nonconserved coding regions, or genuinely novel proteins of *A. planipennis* (Figure 2 and Supplemental Table S1).

**Gene Ontology**

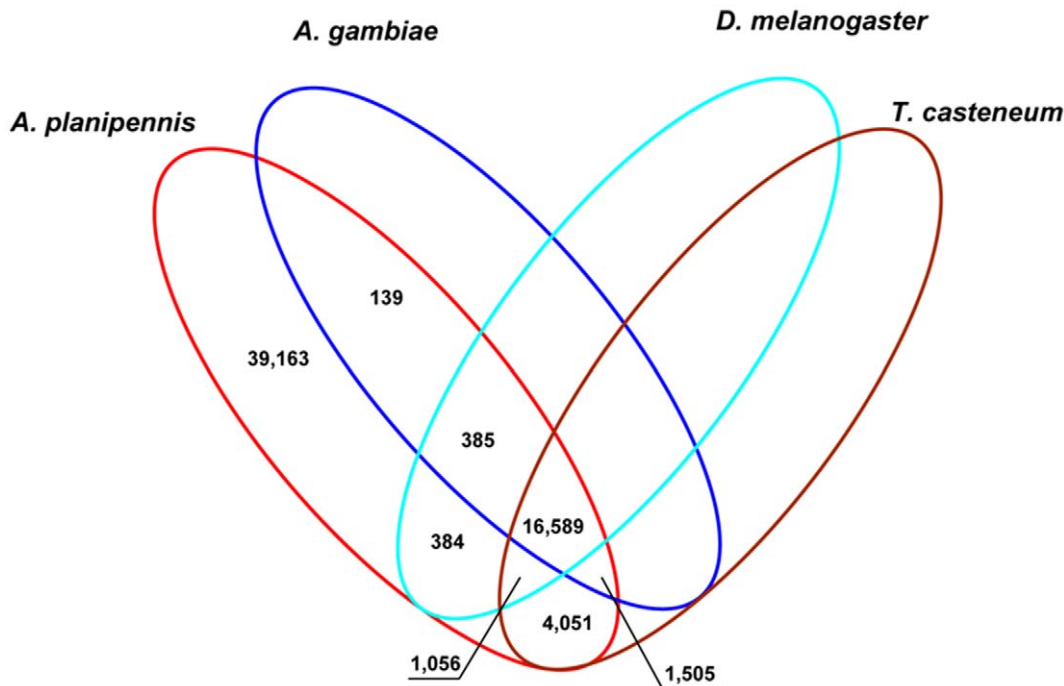
The obtained unigenes were assigned Gene Ontology (GO) based on significant homologies to proteins in GenBank. Since the two EST databases represent two different tissues, GO studies were performed separately to allow tissue- specific comparisons. GO terms were assigned to 7,304 midgut transcriptomic sequences (2,527 Biological Process, 585 Cellular Component and 1,169 Molecular Function; Figure 3 and Supplemental Table S2) and to 10,074 fat body transcriptomic sequences (2,827 Biological Process, 646 Cellular Component and 1,454 Molecular Function; Figure 4 and Supplemental Table S2).

**Metabolic pathways**

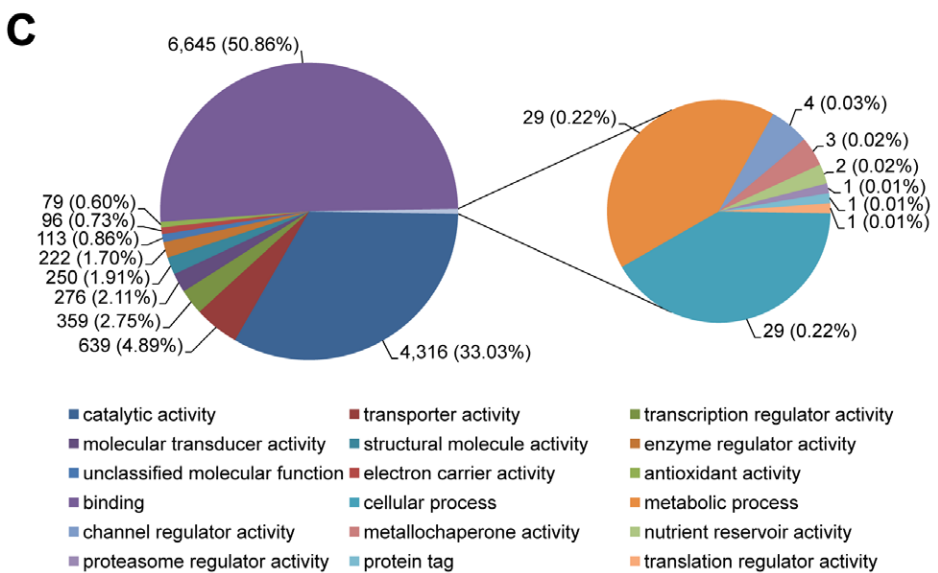
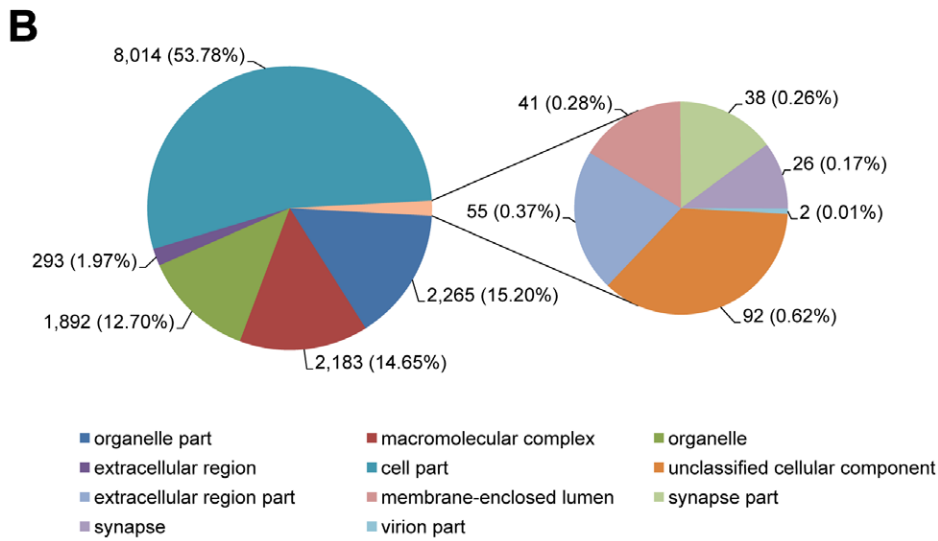
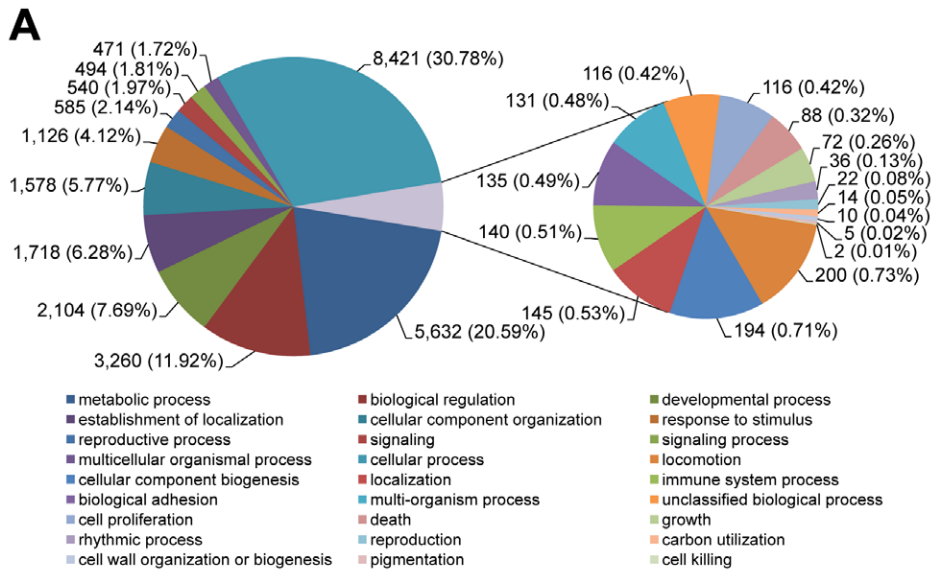
We identified 115 KEGG metabolic pathways for the midgut and 122 for the fat body (Supplemental Table S3 and S4). Of the pathways identified, it was interesting to note that the beta-lactam resistance pathway was midgut-specific. However, at this point it is unclear as to what the functional role of this pathway may have specifically in the midgut. The specific pathways identified in the fat body were related to steroid biosynthesis, tetracycline biosynthesis, D-arginine and D-ornithine metabolism, ubiquinone and other terpenoid backbone biosynthesis. Besides these pathways, both the midgut and fat body databases also revealed transcripts for biosynthesis of alkaloids (derived from ornithine), lysine, nicotinic acid, histidine and purine.

**Protein domains**

A Pfam domain search yielded 5,876 domains in 4,925 midgut transcripts and 8,852 domains in 7,205 fat body transcripts (Table 2). Among the identified Pfam domains, chitin binding Peritrophin-A (CBP-A) was found to be the highest in occurrence



**Figure 2.** Summary of the comparisons of *Agrilus planipennis* transcriptomic sequences with protein sequences from the draft genomes of *Anopheles gambiae*, *Drosophila melanogaster*, and *Tribolium castaneum*.  
doi:10.1371/journal.pone.0013708.g002



**Figure 3. Gene ontology (GO) terms for *Agrilus planipennis* larval midgut transcriptome.** (A) Biological Process, (B) Cellular Component and (C) Molecular Function.

doi:10.1371/journal.pone.0013708.g003

with 248 transcripts (4.2%) in the midgut and a mere 16 sequences (0.2%) in fat body. CBP-A domain is commonly found in proteins that constitute the peritrophic matrix (PM), a semipermeable membrane that lines the gut of most insects [6,32,33]. PM assists in digestion and protecting the midgut epithelium from abrasion as well as from pathogens including bacteria, viruses and parasites [34,35]. The presence of 0.2% fat body sequences with CBP-A domains is probably due to contamination during dissections and sample collection.

Several domains pertaining to detoxification enzymes were also identified in both the midgut and fat body transcripts of *A. planipennis*. First, compared to the fat body, the midgut showed a greater number of sequences matching domains of carboxylesterases (CEs). A similar high frequency of CEs was reported in the midgut of *C. tremulae* [30]. CEs are general detoxification enzymes involved in the resistance to carbamates, pyrethroids and organophosphate insecticides [36]. However, it is not clear what their role may be in the detoxification of plant allelochemicals [37]. Second, cytochrome P450 domains were found more frequently in fat body sequences than those of the midgut. These enzymes have been reported to detoxify a wide-range of xenobiotics including plant toxins and synthetic chemicals in an array of insect species [38].

WD-40 repeats were found in both midgut and fat body sequences. WD-repeat protein and its subunit domains were found significantly higher in fat body compared to the midgut (Table 2). WD-repeat proteins play an important role in RNA processing, signal transduction, cytoskeleton assembly, cell division and protein-protein interactions [29,39]. In the fat body we identified the katanin domain, which is part of a microtubule severing protein (heterodimer of 60 and 80 kDa subunits) that targets the centrosome using WD-40 containing subunit [40]. Other identified domains that are potentially involved in signal transduction and cytoskeleton related functions were GTPase Ras families [41].

P-element wimpy testis-induced (PIWI) - interacting RNAs (piRNA) play an important role in gene regulation in addition to silencing retrotransposons and repetitive sequences during germline development [42,43]. We identified some of the domains of proteins that are involved in the piRNA synthesis pathway. Interestingly, these domains were almost restricted to the fat body sequences of *A. planipennis* and included PIWI-like proteins and Tudor domain that are essential for piRNA production [42,44]. The midgut sequences revealed only Tudor domains. piRNAs differ from micro-RNAs (miRNAs) and small interfering RNAs (siRNAs) in several ways: 1) they interact only with PIWI and not with Argonaute; 2) piRNAs range from 24 to 31 nucleotide (nt) length compared to siRNAs and miRNAs, which are typically 21nt; and 3) piRNAs may positively regulate mRNA stability and translation whereas miRNAs and siRNAs have negative effect on transcriptional and translational process [45,46]. To-date, the function of piRNA is unclear and deciphering the functional role of molecules involved in piRNA synthesis would contribute novel insights into gene regulation in insects.

Apart from the above-described domains we found a significant number of other domains pertaining to trypsin, glycosyl hydrolase family, sugar transporters, protein kinase domain, mitochondrial protein domains, C<sub>2</sub>H<sub>2</sub> type zinc finger domains and RNA recognition motifs (Supplemental Table S5). The presence of a high number of trypsin and trypsin-like domains suggest that the gut of *A. planipennis* is predominantly composed of serine-proteases,

which seems to be also one of the major classes of digestive proteases in the midgut of the beetle *C. tremulae* [30].

### Genes of interest

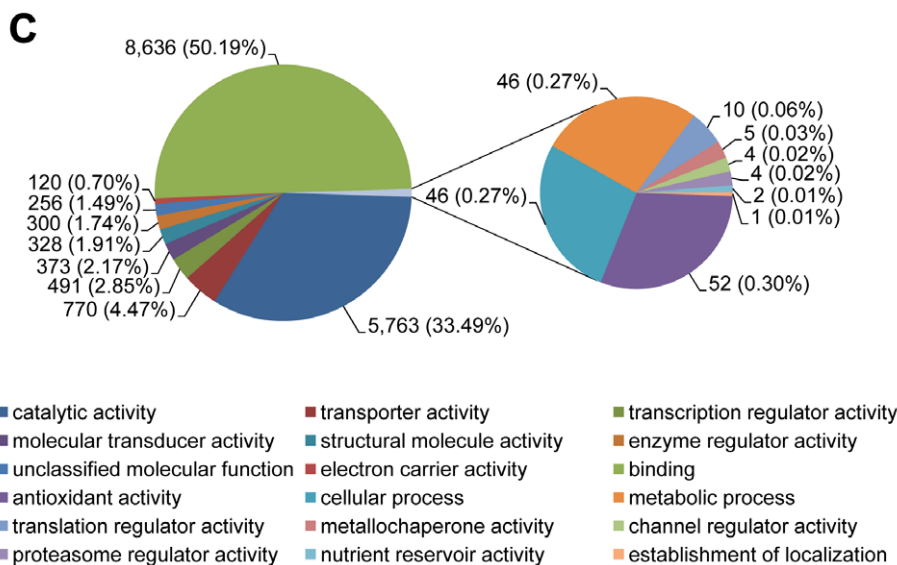
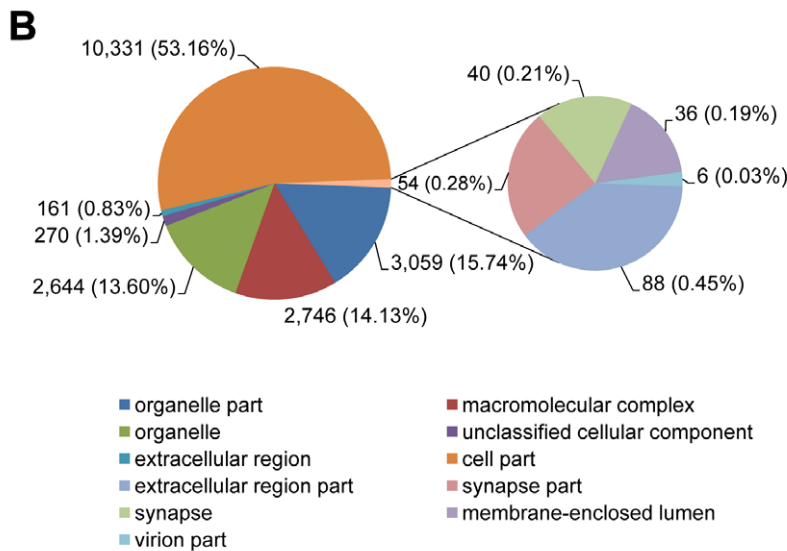
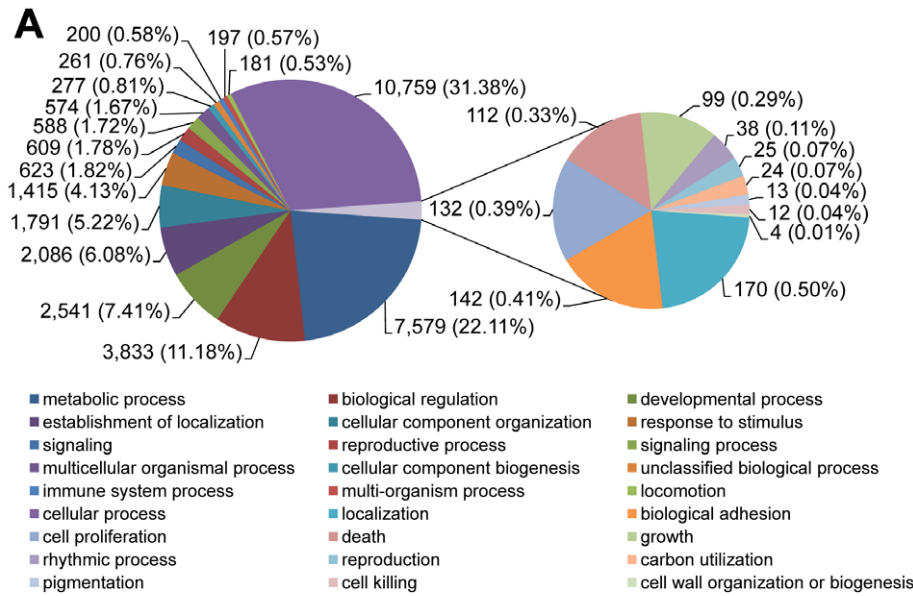
We are primarily interested in elucidating the functions of the genes involved in detoxification/antioxidant response (Table 3). Of particular interest are cytochrome P450s, which are known to be one of the most rapidly radiating groups of detoxifying enzymes in insects [47]. Cytochrome P450s metabolize toxic substrates including allelochemicals and insecticides to less harmful excretable forms via hydroxylation/oxidation [47,48]. Cytochrome P450s have been classified into four large clades: CYP2, CYP3, CYP4, and Mitochondrial clade [47]. In this study, the predominant occurrence of *A. planipennis* cytochrome P450s was members of the CYP3 clade, which is in agreement with those reported in various insect orders including Diptera, Lepidoptera and Coleoptera [47]. The genome of the red flour beetle *T. castaneum* encodes about 147 cytochrome P450s of which 70 belong to the CYP3 clade [31]. To begin ascertaining putative function, we assessed the mRNA levels of candidate *A. planipennis* cytochrome P450 genes belonging to each of the four clades in different larval tissues (midgut, fat body and cuticle).

Members of the CYP3 clade (mainly CYP6 and CYP9) revealed distinct expression patterns. Of the two *A. planipennis* CYP6 genes assayed, EABMG023904 showed highest mRNA levels in the midgut while contig EABMG001813 showed higher transcript level in the fat body (Figure 5A). Similar mRNA levels of tissue-specific CYP6 genes have been observed in other insect species [49–51]. It is thought that the high expression levels of CYP6 genes were associated with metabolism of plant allelochemicals and insecticides [52]. However, qPCR analysis for a CYP9 member revealed highest mRNA levels in larval cuticle of *A. planipennis* (Figure 5A). CYP9 in *A. planipennis* larvae could be involved in xenobiotics metabolism as reported in other insect taxa [52,53].

Although the CYP4 clade consists of numerous diverse genes, it seems to be the least studied group of cytochrome P450s in insects [48]. In this study, larval *A. planipennis* midgut showed the highest mRNA level for one of the CYP4 assayed (Figure 5B). Transcript levels in the fat body were significantly higher compared to levels in cuticle but significantly lower than the levels for the midgut. These observations are in agreement with CYP4 mode of action on dietary plant toxins reported in Lepidoptera (e.g. *Manduca sexta*; [54]) as well as against insecticides in Coleoptera (e.g. *Diabrotica virgifera virgifera* Leconte [55]). Apart from xenobiotic metabolism, members of CYP4 are also reported to have putative roles in ecdysteroid metabolism [56].

Of the CYP2 clade homologs identified in *A. planipennis* midgut, we assessed the expression of CYP306, which belongs to a set of cytochrome P450s called Halloween genes or otherwise known as the phantom gene, *phm* [57]. These genes are thought to be involved in ecdysone metabolism and therefore are important for insect growth and development. The CYP306 homolog of *A. planipennis* showed highest mRNA levels in the midgut and cuticle (Figure 5C). Similar tissue-specific profiles of CYP306 (*phm*) were observed in larva of *M. sexta* [57].

Mitochondrial P450s are very specific to animals and are not found in fungi and plants [48]. Insects have two types of mitochondrial P450, a conserved group involved in physiological functions and a diverse group known to participate in the detoxification of xenobiotics. Expression analysis of an *A.*





**Figure 4. Gene ontology (GO) terms for *Agrilus planipennis* larval fat body transcriptome.** (A) Biological Process, (B) Cellular Component and (C) Molecular Function.  
doi:10.1371/journal.pone.0013708.g004

*planipennis* CYP12 gene belonging to the mitochondrial clade revealed highest mRNA levels in the midgut compared to cuticle and fat body (Figure 5D), similarly to a study on resistance of *D. melanogaster* larvae to lufenuron, which was attributed to high expression of a CYP12 in the midgut and Malpighian tubules [58].

The above results suggest a potential involvement of cytochrome P450s in metabolism of host allelochemicals by *A. planipennis* larvae and perhaps in other physiological functions such as hormonal biosynthesis. However, additional experiments at the protein and transcript level (RNA interference) that elucidate gene function of these cytochrome P450s could provide insights into alternate control targets as in other insect systems [51].

**Molecular markers**

We predicted 317 simple sequence repeats (SSRs or microsatellites) in *A. planipennis* midgut sequences and 571 SSRs in fat body sequences (Table 4). The majority of SSRs were either di-nucleotide (58 for midgut and 94 for fat body) or trinucleotide (223 for midgut and 428 for fat body) repeats; while 26 and 46 sequences of the midgut and fat body, respectively, were predicted to be tetra-nucleotide repeats and 10 and 3 sequences of the midgut and fat body, respectively, were predicted to be penta-nucleotide repeats. The characteristics of these SSRs including length, start, stop motif units and foot print are summarized in Supplemental Table S6. Interestingly, we also identified 137 putative single nucleotide polymorphisms (SNPs) in 45 sequences of the midgut and 347 SNPs in 143 sequences of the fat body (Table 5) (Supplemental Table S7). The exploitation of markers which are linked to coding regions would facilitate the selection of polymorphic markers for *A. planipennis* [59]. However, prior to making any inferences of the predicted molecular markers in *A. planipennis*, validation has to be performed to account for false positives.

**Midgut microbiota**

A small subset of the sequences (79 of 14680) showed closer similarity to microbial, rather than animal genes. They might thus

represent low levels of microbial contamination in the dissection, or even horizontal transmission of genes from symbiotic microbes to the insect host (Supplemental Table S8). First, putative endo-glucanases were found that matched to *Streptomyces* species. Similar genera of microbiota in *A. planipennis* were identified by Vasanthakumar et al. [60]. In particular, they isolated *Streptomyces* sp. from midgut of larva; *Erwinia* sp. and *Burkholderia cepacia* from the guts of adult beetles. Second, putative endo-polygalacturonases were identified in the midgut of *A. planipennis*. Given that the diet of larval *A. planipennis* is primarily phloem-based, the recovered endo-glucanases and endo-polygalacturonases could facilitate in breakdown of plant cell walls [61] resulting in the release of nutrients required for larval growth and development.

A single bacterial sequence in the midgut transcriptome showed similarity with a *Wolbachia* sequence, an ubiquitous endosymbiont found across various insect taxa and nematodes [62–65]. *Wolbachia* is thought to interfere with insect reproduction through a process known as feminization, in which the sex ratios within a population become female-skewed [66,67].

**Conclusions**

This study represents the first report to elucidate tissue-specific gene expression in *A. planipennis*. Our analysis of *A. planipennis* midgut and fat body has revealed commonalities as well as potentially significant differences with other insect model systems and is beginning to provide insights into some of the fundamental physiological processes occurring in these tissues. Some noteworthy results of this study are 1) a significant number of cytochrome P450s that showed distinct tissue expression patterns; 2) a number of SNPs and microsatellite markers were predicted within the midgut and fat body databases, which upon validation could provide insights at functional polymorphism within and between tissues; and 3) some of the predicted microbial sequences encoding for enzymes involved in carbohydrate digestion and metabolism in the larval midgut of *A. planipennis* suggest the presence of putative endosymbionts. These traits along with the recovered sequences

**Table 2. Top Pfam domains identified in *Agrilus planipennis* sequences.**

Pfam accession	Pfam domain description	Number of occurrence in midgut	Number of occurrence in fat body
PF01607.17	Chitin binding Peritrophin-A domain	248	16
PF00135.21	Carboxylesterase	58	27
PF00232.11	Glycosyl hydrolase family 1	48	14
PF00083.17	Sugar (and other) transporter	43	63
PF00069.18	Protein kinase domain	77	123
PF00400.25	WD domain, G-beta repeat	54	111
PF07714.10	Protein tyrosine kinase	52	91
PF00076.15	RNA recognition motifs	43	86
PF00067.15	Cytochrome P450	37	85
PF00096.19	Zinc finger, C <sub>2</sub> H <sub>2</sub> -type	36	82
PF00071.15	Ras family	66	73
PF00089.19	Trypsin	98	55
PF00153.20	Mitochondrial carrier protein	28	51

doi:10.1371/journal.pone.0013708.t002

**Table 3.** Genes of interest in the midgut and fat body of *Agrilus planipennis* larvae.

Candidate genes	Number of occurrence in midgut	Number of occurrence in fat body
<b>Cytochrome P450</b>		
CYP2 clade	01	08
CYP2	—	01
CYP24	—	01
CYP49	—	01
CYP303	—	02
CYP306	01	03
CYP3 clade	37	45
CYP3	—	02
CYP6	26	37
CYP9	10	06
CYP28	01	—
CYP4 clade	03	16
CYP4	03	16
Mitochondrial CYP clade	04	06
CYP12	—	02
CYP301	03	01
CYP314	—	02
CYP315	01	01
<b>GST</b>	27	19
<i>Sigma</i>	08	01
<i>Omega</i>	01	02
<i>Theta</i>	17	15
<i>Microsomal</i>	01	01
Superoxide dismutase	05	07
Catalase	17	05
Carboxylesterases	16	05

doi:10.1371/journal.pone.0013708.t003

involved in piRNA synthesis provide substantial insights into the physiology-driven molecular mechanisms of *A. planipennis*.

## Materials and Methods

### Insect samples

*A. planipennis* larvae were collected from wood-lots consisting of green ash trees in Ottawa Lake, MI and Toledo, OH. Larvae were extracted by stripping off the bark of infested ash trees that were 10–25 cm in diameter at breast height. The ash trees sampled were alive but showed obvious symptoms of decline in response to *A. planipennis* colonization. The collected larvae were categorized according larval instars.

### Larval dissections and RNA isolation

About 50 late 3<sup>rd</sup>-instar larvae were dissected to extract midgut and fat body tissues. Dissections were performed as described by Vasanthakumar et al. [60]. The midgut tissue was dissected free of its contents and carefully rinsed in fresh PBS buffer prior to RNA isolation. Isolated larval midgut, cuticle and fatbody were placed in pre-chilled Trizol reagent in separate 1.5 ml eppendorf tubes. Following dissection, the tissues were stored at  $-80^{\circ}\text{C}$  until RNA was isolated. The dissected tissues were homogenized separately in TRIzol<sup>®</sup> and total RNA was isolated following the manufacturer's protocol and stored at  $-80^{\circ}\text{C}$  until use.

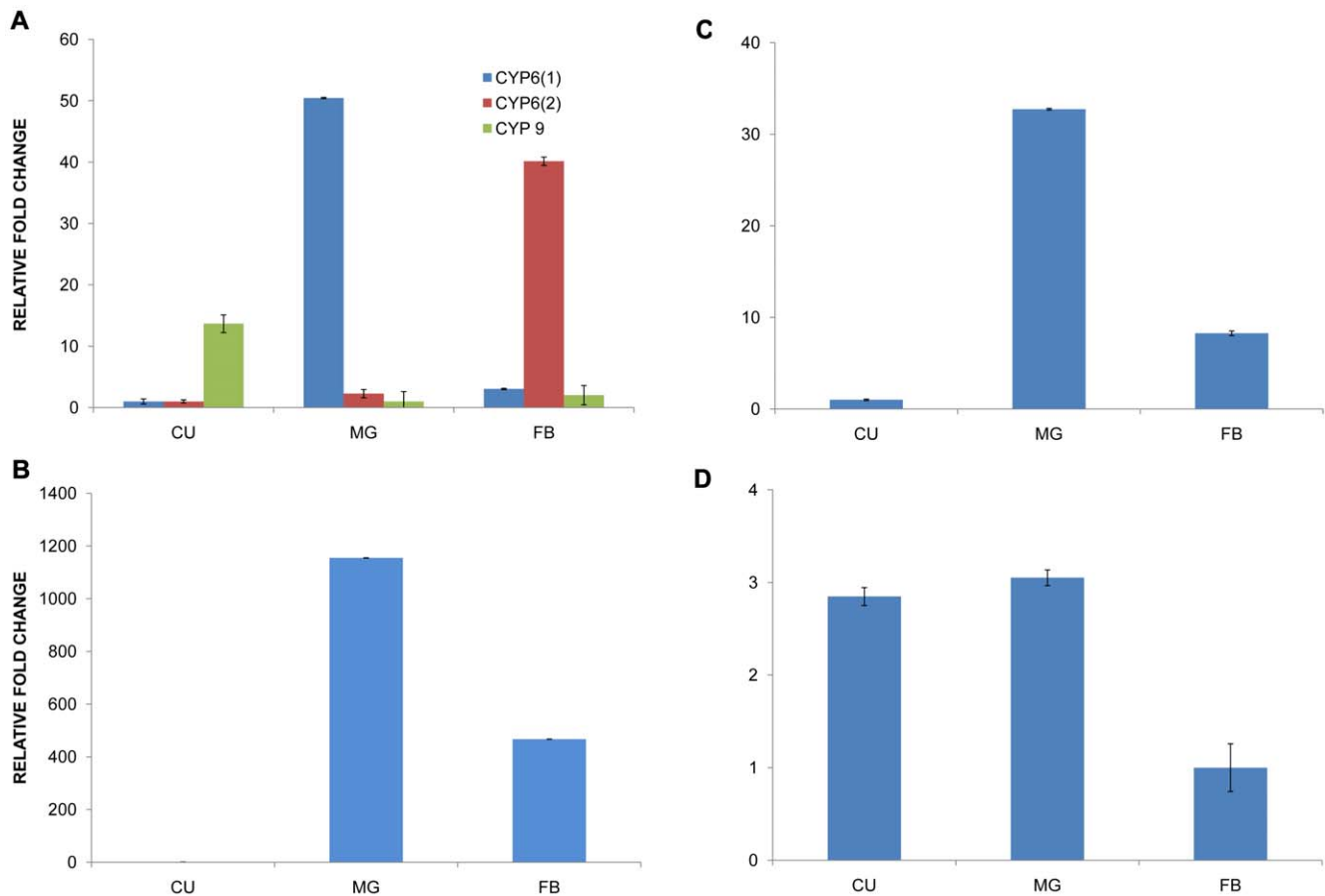
### cDNA library construction

The midgut and fat body RNA samples (10–20  $\mu\text{g}$ ) described above were used for constructing cDNA libraries. Prior to cDNA library construction, the total RNA was subjected to quality control using the Nanodrop 2000c (Thermo Scientific) and the RNA 6000 nanochip (Agilent Technologies) at the Purdue Genomics Core Facility (West Lafayette, IN). A SMART cDNA library construction kit (Clontech) was used to construct both libraries with modifications to suit the 454 GS FLX pyrosequencing. Instead of using the recommended MMLV reverse transcriptase, Super script II reverse transcriptase from Invitrogen was used for synthesizing the first-strand cDNA, and a modified PAGE-purified CDS II primer (5' - TAG AGG CCG AGG CGG CCG ACA TGT TTT GTT TTT TTT TCT TTT TTT TTT VN - 3') was used instead of the supplied one. The constructed cDNA libraries were shipped to the Purdue Genomics Core Facility for sequencing.

### 454 pyrosequencing

DNA bands of 500–800 bp were excised from the gel and purified. The isolated DNA was blunt ended, ligated to adapters and immobilized on Library Immobilization Beads. Blunt-end fragments were made with T4 polynucleotide kinase (PNK) and T4 DNA polymerase in a reaction that was held for 15 minutes at  $12^{\circ}\text{C}$  followed by 15 additional minutes at  $25^{\circ}\text{C}$ . The adaptors (Titanium A & B) used were provided with the 454 general library





**Figure 5. Relative mRNA levels of candidate cytochrome P450 genes in tissues of larval *Agrilus planipennis*.** (A) Expression levels of CYP3 clade members including CYP6 and CYP9. CYP6(1) represents contig EABMG023904 and CYP6(2) represents contig EABMG001813, (B) Expression levels of a CYP4, (C) Expression levels of mitochondrial CYP clade CYP12 and (D) Expression levels of CYP306. Tissues assayed include cuticle (CU), midgut (MG) and fat body (FB). Standard error of the mean for two biological replicates (nested with two technical replicates) is represented by the error bars.

doi:10.1371/journal.pone.0013708.g005

kit. The ligation reaction included ligase buffer, adaptors, and a ligase enzyme, which was incubated for 15 minutes at 25°C on a thermocycler. After the gaps were repaired, a single-stranded DNA library was isolated from the beads and quality controlled for the correct size using a LabChip 7500 machine. The concentration and the proper ligation of the adaptors were examined using qPCR. One-quarter of a pico-titer plate was sequenced for each tissue sample (midgut and fat body) following manufacturer's protocol. The pyrosequencing was performed as per the Genome Sequencer FLX Operation (Roche).

#### Bioinformatic data analysis

The transcriptome sequences of *A. planipennis* midgut and fat body were annotated by searching against the GenBank non-redundant database using BLASTx algorithms [68]. The *A. planipennis* transcriptome sequences were compared to proteins of the fruit fly *Drosophila melanogaster*, the African malaria mosquito *Anopheles gambiae*, and the red flour beetle *Tribolium castaneum* using BLASTx algorithm [68]. Domains within the sequences were identified by searching against the Pfam database [69] using the HMMER v3 program [70]. We used Blast2GO [71,72] to predict the functions of the sequences, assign Gene Ontology terms, and predict metabolic pathways in the KEGG database [73,74]. Microsatellites/SSR markers were identified using Msafinder

version 2.0.9 program [75] and single nucleotide polymorphisms (SNPs) were predicted using the gsMapper program (Roche) with an arbitrary criterion of at least 4 reads supporting the consensus or variant.

#### Quantitative real-time PCR

The concentration of total RNA samples was measured using NanoDrop 2000c Spectrophotometer from Thermo Scientific. The first strand cDNA was synthesized using Superscript IITM First Strand Synthesis kit for qRT-PCR from Invitrogen, following the manufacturer's protocol. For cDNA synthesis, 1 µl of oligo (dT)<sub>12-18</sub> primer (500 µg/µl) and 1 µl of dNTPs mix (10 mM each) were added to 8 µl of total RNA sample (~0.25 µg/µl). The samples were heated in a thermo cycler at 65°C for 4 min. Then, the following reagents were added on ice to each sample: 2 µl of 10×RT buffer, 4 µl of 25 mM MgCl<sub>2</sub>, 2 µl of 0.1 M DTT, 1 µl RNase Out, 1 µl of SuperScript II reverse transcriptase and 1 µl of RNase H. The cycling parameters were 42°C for 90 min., 70°C for 15 min. and 37°C for 20 min. After cDNA synthesis, the concentrations of the first-strand samples were brought to 20 ng/µl, which were used as the template in qRT-PCR reactions. Cycling parameters included 95°C for 3 min and then, 40 cycles of 95°C for 10 s and 60°C for 30 s.

**Table 4.** Putative microsatellite loci predicted in *Agrilus planipennis*.

Number of repeats	Di-nucleotide repeats		Tri-nucleotide repeats		Quad -nucleotide repeats		Penta -nucleotide repeats	
	MG*	FB	MG	FB	MG	FB	MG	FB
5			137	251	9	34	5	2
6			60	107	8	4	3	
7			18	45	3	5	2	1
8	19	38	1	16	4	1		
9	11	29	5	7	1			
10	9	12	1	2				
11	4	7			1			
12	6	4						
13	3							
14			1					
15	1	1				1		
16	1	1						
17	1							
18	1							
...								
21						1		
...								
23	1							
...								
27		2						
...								
46	1							
<b>Subtotal</b>	<b>58</b>	<b>94</b>	<b>223</b>	<b>428</b>	<b>26</b>	<b>46</b>	<b>10</b>	<b>3</b>

\*MG = midgut; FB = fat body.  
doi:10.1371/journal.pone.0013708.t004

Quantification of candidate gene expression in tissues (midgut, fat body and cuticle) was based on the Relative Standard Curve method [2] using the threshold cycle (Ct) values. For all qRT-PCR analyses the mRNA coding for a ribosomal protein (RP7) was assessed and included as the internal control. Assessment of the mRNA encoding for ribosomal proteins has been demonstrated to be an appropriate internal control in other studies [76].

**Table 5.** Putative single nucleotide polymorphisms identified in *Agrilus planipennis*.

SNP type	MG*	FB
Transition		
A-G	52	124
C-T	37	111
Transversion		
A-C	9	20
A-T	17	41
C-G	8	30
G-T	14	21
<b>Total</b>	<b>137</b>	<b>347</b>

\*MG = midgut; FB = fat body.  
doi:10.1371/journal.pone.0013708.t005

**Statistical analysis**

Relative expression value (REV) for each gene was determined by dividing the quantities for the target samples by the quantities obtained for ribosomal protein. The REVs obtained for each gene in the tissues was analyzed by Analysis of Variance (ANOVA) (p = 0.05) using the PROC MIXED procedure of SAS (SAS Institute Inc. SAS.STAT User's Guide, Version 9.1). Two biological replicates nested with two technical replicates were included for the entire analysis. Biological replicates were included as a random effect in the model. Relative fold changes in the tissues were calculated by taking the sample (REV) that showed the least level of expression as 1X sample (calibrator.) [77]. The standard error represented the variance in the two biological replicates (2 technical replicates within each) included for this analysis.

**Data Deposition**

The Roche 454 reads of *A. planipennis* midgut and fat body were submitted to NCBI Sequence Read Archive under the accession number of SRA012309.3.

**Supporting Information**

**Table S1** Genome comparisons of *A. planipennis* midgut and fat body sequences.  
Found at: doi:10.1371/journal.pone.0013708.s001 (6.16 MB XLS)

**Table S2** Gene Ontology of *A. planipennis* midgut and fat body sequences.

Found at: doi:10.1371/journal.pone.0013708.s002 (3.11 MB XLS)

**Table S3** KEGG summary of *A. planipennis* midgut sequences.

Found at: doi:10.1371/journal.pone.0013708.s003 (0.04 MB XLS)

**Table S4** KEGG summary of *A. planipennis* fat body sequences.

Found at: doi:10.1371/journal.pone.0013708.s004 (0.04 MB XLS)

**Table S5** Pfam domain search of *A. planipennis* midgut and fat body sequences.

Found at: doi:10.1371/journal.pone.0013708.s005 (2.59 MB XLS)

**Table S6** Predicted microsatellite loci in *A. planipennis* midgut and fat body sequences.

Found at: doi:10.1371/journal.pone.0013708.s006 (0.10 MB XLS)

**Table S7** Predicted SNPs in *A. planipennis* midgut and fat body sequences.

Found at: doi:10.1371/journal.pone.0013708.s007 (0.07 MB XLS)

**Table S8** Midgut-specific microbiota sequences of *A. planipennis*.

Found at: doi:10.1371/journal.pone.0013708.s008 (0.04 MB XLS)

## Acknowledgments

Help provided by the Purdue Genomics Core Facility (Phillip San Miguel and Rick Westerman) with regards to the pyrosequencing and initial analysis of the raw sequencing data is much appreciated. Amy Stone, The Ohio State University Extension, assisted with collection of *A. planipennis* larvae. Help provided by Loren Rivera-Vega (Department of Entomology, The Ohio State University/OARDC) with sample collection and cDNA library preparation is appreciated. We would also like to thank Dr. Andrew Michel and Dr. David Denlinger (Department of Entomology, The Ohio State University) for proving critical comments on the manuscript.

## Author Contributions

Conceived and designed the experiments: OM PB DAH. Performed the experiments: OM SPR. Analyzed the data: OM XB PM SPR. Contributed reagents/materials/analysis tools: XB. Wrote the paper: OM XB PM SPR PB DAH.

## References

- Yu Q, Lu C, Li W, Xiang Z, Zhang Z (2009) Annotation and expression of carboxylesterases in the silkworm *Bombyx mori*. BMC Genomics 10: 553–566.
- Mittapalli O, Neal J, Shukle RH (2007) Antioxidant defense response in a galling insect. Proc Natl Acad Sci U S A 104: 1889–1894.
- Pauchet Y, Wilkinson P, Vogel H, Nelson DR, Reynolds SE, et al. (2010) Pyrosequencing the *Manduca sexta* larval midgut transcriptome: messages for digestion, detoxification and defence. Insect Mol Biol 19: 61–75.
- Tittiger C, Keeling CI, Blomquist GJ (2005) Some insights into the remarkable metabolism of the bark beetle midgut. Recent Adv. Phytochem 39: 57–78.
- Felton GW, Duffey SS (1991) Protective action of midgut catalase in lepidopteran larvae against oxidative plant defenses. Journal of Chemical Ecology 17: 1715–1732.
- Mittapalli O, Sardesai N, Giovanini MP, Williams CE, Shukle RH (2006a) Expression patterns of antibacterial genes in the Hessian fly. J Insect Physiol 52: 1143–1152.
- Hou Y, Zhao P, Liu HL, Zou Y, Guan J, et al. (2007) Proteomics analysis of fat body from silkworm (*Bombyx mori*). Sheng Wu Gong Cheng Xue Bao 5: 867–72.
- Arrese EL, Soulages JL (2009) Insect fat body: energy, metabolism and regulation. Annu. Rev Entomol 55: 207–225.
- Liebherr JK, McHugh JV (2003) Coleoptera (beetles, weevils, fireflies). In: Resh VH, Carde RT, eds. Encyclopedia of insects. San Diego, CA: Academic Press. pp 209–203.
- Cappaert D, McCullough DG, Poland TM, Siegert NW (2005) Emerald ash borer in North America: a research and regulatory challenge. Am Entomol 51: 152–163.
- Poland T, McCullough D (2006) Emerald ash borer: invasion of the urban forest and the threat to North America's ash resource. J Forestry 104: 118–124.
- Anulewicz AC, McCullough DG, Cappaert DL (2007) Emerald ash borer (*Agrilus planipennis*) density and canopy dieback in three North American ash species. Arbor Urban For 33: 338–349.
- Smith A (2006) Effects of community structure on forest susceptibility and response to the emerald ash borer invasion of the Huron River Watershed in southeastern Michigan. Columbus: M.S. Thesis, The Ohio State University 122.
- Herns DA, Stone AK, Chatfield JA (2003) Emerald ash borer: the beginning of the end of Ash in North America? In: Chatfield JA, Draper EA, Mathers HM, Dyke DE, Bennett PJ, Boggs JF, eds. 2004. Ornamental Plants: Annual Reports and Research Reviews 2003. pp 62–71. Ohio Agriculture Research and Development Center, Ohio State University Extension Special Circular 193.
- Gandhi JKJ, Herns DA (2010) North American arthropods at risk due to widespread *Fraxinus* mortality caused by the alien emerald ash borer. Biol Invas 12: 1839–1846.
- Burns RM, Honkala BH (1990) Silvics of North America. SDA Forest Service, Timber Management Research Agriculture Handbook 654, Washington, DC. Agriculture Handbook . pp 654 p. 877.
- Liebold AM, MacDonald WL, Bergdahl D, Mastro VC (1995) Invasion by exotic forest pests: a threat to forest ecosystems. For Sci Mon 30: 1–49.
- Akiyama K, Ohmomo S (2000) The buprestid beetles of the world. Iconic Series of Insects 4. Gekkan-Mushi Company, Ltd: Tokyo, Japan.
- Schaefer PW (2005) Foreign exploration for emerald ash borer and its natural enemies. In: Mastro VC, Reardon R, eds. Proceedings of Emerald Ash Borer Research and Technology Development Meeting, USDA Forest Health Technology Enterprise Team Publication FHTET-2004-15. Morgantown, W.Va: USDA Forest Service. pp 67–68.
- Bauer LS, Liu HP, Haack RA, Gao RT, Zhao TH, et al. (2005) Update on emerald ash borer natural enemies in Michigan and China. In: Mastro V, Reardon R, eds. Proceedings of the Emerald Ash Borer Research and Technology Meeting, Romulus, MI. USDA FS FHTET-2004-15. pp 71–72.
- Yu CM (1992) *Agrilus marcopoli* Obenberger. In: Xiao G-R, ed. Forest Insects of China. 2nd Edition.. Beijing, China: China Forestry Publishing House. pp 400–401.
- Haack RA, Jendek E, Liu H, Marchant KR, Petrice TR, et al. (2002) The emerald ash borer: A new exotic pest in North America. Michigan Entomological Soc Newsletter 47: 1–5.
- Gould J, Tanner J, Winograd D, Lane S (2004) Initial studies on the laboratory rearing of emerald ash borer and foreign exploration for natural enemies. In: Mastro VC, Reardon R, eds. Proceedings of Emerald Ash Borer Research and Technology Development Meeting. USDA Forest Health Technology Enterprise Team Publication FHTET-2004-15. Morgantown W.Va.: USDA Forest Service, 2005. pp 73–74.
- Rebek EJ, Herns DA, Smitley DR (2008) Interspecific variation in resistance to emerald ash borer (Coleoptera: Buprestidae) among North American and Asian ash (*Fraxinus* spp.). Environ Entomol 37: 242–246.
- Eyles A, Jones W, Riedl K, Cipollini D, Schwartz S, et al. (2007) Comparative phloem chemistry of Manchurian (*Fraxinus mandshurica*) and two North American Ash species (*F. americana* and *F. pennsylvanica*). J Chem Ecol 33: 1430–1448.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–80.
- Bentley DR (2006) Whole-genome re-sequencing. Curr Opin Genet Dev 16: 545–552.
- Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92: 255–264.
- Bai X, Zhang W, Ornatas L, Jun T, Mittapalli O, et al. (2010) Combining next-generation sequencing strategies for rapid molecular resource development from an invasive aphid species, *Aphis glycines*. PLoS ONE 5: e11370.
- Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, et al. (2009) Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. Insect Biochem Mol Biol 39: 403–413.
- Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, et al. (2008) The genome of the model beetle and pest *Tribolium castaneum*. Nature 452: 949–955.
- Elvin CM, Vuocolo T, Pearson RD, East IJ, Riding GA, et al. (1996) Characterization of a major peritrophic membrane protein, peritrophin- 44, from the larvae of *Lucilia cuprina*. J Biol Chem 271: 8925–8935.
- Bhandary B, Rajarapu SP, Rivera-Vega L, Mittapalli O (2010) Analysis of gene expression in emerald ash borer (*Agrilus planipennis*) using quantitative real time-PCR. J Vis Exp;DOI 0.3791/1974.
- Lehane MJ (1997) Peritrophic matrix structure and function. Annu Rev Entomol 42: 525–550.

35. Hegedus D, Erlandson M, Gillott C, Toprak U (2009) New insights into peritrophic matrix synthesis, architecture and function. *Ann Rev Entomol* 54: 285–302.
36. Hemingway J, Karunaratne SH (1998) Mosquito carboxylesterases: a review of the molecular biology and biochemistry of a major insecticide resistance mechanism. *Med Vet Entomol* 12: 1–12.
37. Yu Q, Lu C, Li W, Xiang Z, Zhang Z (2009) Annotation and expression of carboxylesterases in the silkworm, *Bombyx mori*. *BMC Genomics* 10: 553.
38. Feyereisen R (1999) Insect P450 enzymes. *Annu Rev Entomol* 44: 507–5033.
39. Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* 24: 181–185.
40. Hartman JJ, Mahr J, McNally K, Okawa K, Iwamatsu A, et al. (1998) Katanin, a microtubule-severing protein, is a novel AAA ATPase that targets to the centrosome using a WD-40 containing subunit. *Cell* 93: 277–287.
41. Vlachou D, Schlegelmilch T, Christophides GK, Kafatos FC (2005) Functional genomic analysis of midgut epithelial responses in *Anopheles* during *Plasmodium* invasion. *Curr Biol* 15: 1185–95.
42. Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, et al. (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129: 69–82.
43. Wang G, Reinke V (2008) A *C. elegans* Piwi, PRG-1, Regulates 21U-RNAs during spermatogenesis. *Curr Biol* 18: 861–867.
44. Chuma S, Hosokawa M, Kitamura K, Kasai S, Fujioka M, et al. (2006) Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. *Proc Natl Acad Sci U S A* 103: 15894–15899.
45. Kim VN (2006) Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev* 20: 1993–1997.
46. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128: 1089–1103.
47. Werck-Reichhart, Feyereisen R (2000) Cytochromes P450: A success story. *Genome Biol* 1: 3003.1–3003.9.
48. Feyereisen R (2006) Evolution of insect P450. *Biochem Soc Trans* 34: 1252–1255.
49. Mittapalli O, Neal JJ, Shukle RH (2005) Differential expression of two cytochrome p450 genes in compatible and incompatible hessian fly/wheat interactions. *Insect Biochem Mol Biol* 35: 981–989.
50. Berenbaum MR, Cohen MB, Schuler MA (1992) Cytochrome P450 monooxygenase genes in oligophagous Lepidoptera. In: Mullin CA, Scott JG, eds. *ACS Symposium Series* 505. Washington, DC. pp 114–124.
51. Mao YB, Cai WJ, Wang JW, Hong GJ, Tao XY, et al. (2007) Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat Biotechnol* 25: 1307–1313.
52. Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, et al. (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Mol Biol* 15: 615–636.
53. Yang Y, Chen S, Wu S, Yue L, Wu Y (2006) Constitutive overexpression of multiple cytochrome P450 genes associated with pyrethroid resistance in *Helicoverpa armigera*. *J Econ Entomol* 99: 1784–1789.
54. Snyder MJ, Stevens JL, Andersen JF, 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.
55. Scharf ME, Parimi S, Meinke IJ, Chandler LD, Siegfried BD (2001) Expression and induction of three family 4 cytochrome P450 (CYP4) genes identified from insecticide-resistant and susceptible western corn rootworms, *Diabrotica virgifera virgifera*. *Insect Mol Biol* 10: 139–146.
56. Davies L, Williams DR, Aguiar-Santana IA, Pedersen J, Turner PC, et al. (2006) Expression and down-regulation of cytochrome P450 genes of the CYP4 family by ecdysteroid agonists in *Spodoptera littoralis* and *Drosophila melanogaster*. *Insect Biochem Mol Biol* 36: 801–807.
57. Rewitz KF, Rybczynski R, Warren JT, Gilbert LI (2006) Identification, characterization and developmental expression of Halloween genes encoding P450 enzymes mediating ecdysone biosynthesis in the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* 36: 188–199.
58. Bogwitz MR, Chung H, Magoc L, Rigby S, Wong W, et al. (2005) CYP12A4 confers lufenuron resistance in a natural population of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 102: 12807–12812.
59. Vasemagi A, Nilsson J, Primmer CR (2005) Expressed Sequence Tag-Linked Microsatellites as a Source of Gene-Associated Polymorphisms for Detecting Signatures of Divergent Selection in Atlantic Salmon (*Salmo salar* L.). *Mol Biol Evol* 22: 1067–1076.
60. Vasanthakumar A, Handelsman J, Schloss PD, Bauer LS, Raffa KF (2008) Gut microbiota of an invasive subcortical beetle, *Agrylus planipennis* Fairmaire, across various life stages. *Environ Entomol* 37: 1344–1353.
61. Bishop JG, Ripoll DR, Bashir S, Damasceno CMB, Seeds JD, et al. (2005) Selection on glycine  $\beta$ -1,3-endoglucanase genes differentially inhibited by a *Phytophthora glucanase* inhibitor protein. *Genetics* 169: 1009–1019.
62. Bourtzis K, Nirgianaki A, Markakis G, Savakis C (1996) *Wolbachia* infection and cytoplasmic incompatibility in *Drosophila* species. *Genetics* 144: 1063–1073.
63. Werren JH, Windsor DM (2000) *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proc R Soc Lond B Biol Sci* 267: 1277–1285.
64. Jayaprakash A, Hoy MA (2000) Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect Mol Biol* 9: 393–405.
65. Salzberg SL, Hotopp JC, Delcher AL, Pop M, Smith DR, et al. (2005) Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol* 6: R23.1–R23.8.
66. Heddi A, Grenier AM, Khatchadourian C, Charles H, Nardon P (1999) Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. *Proc Natl Acad Sci U S A* 96: 6814–6819.
67. Zimmer C (2001) *Wolbachia*: A Tale of Sex and Survival. *Science* 292: 1093–1095.
68. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
69. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: 281–288.
70. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14: 755–763.
71. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
72. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36: 3420–3435.
73. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36: D480–484.
74. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28: 27–30.
75. Thurston MI, Field D (2005) Msatfinder: detection and characterisation of microsatellites. Available: <http://www.genomics.ceh.ac.uk/msatfinder/>.
76. Nakabachi A, Shigenobu S, Sakazume N, Shiraki T, Hayashizaki Y, et al. (2005) Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. *Proc Natl Acad Sci U S A* 12: 5477–5482.
77. Pfaffl WM (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29: 2002–2007.